PharmaSUG 2026 - Paper DS-401

Why Standards Matter More Than Code in the Age of GenAI

Bhavin Busa, Clymb Clinical

ABSTRACT

The growing adoption of generative artificial intelligence (GenAI) in statistical programming and clinical reporting has introduced new opportunities for automation, alongside concerns related to reliability, reproducibility, and regulatory acceptability. While large language models (LLMs) are inherently probabilistic, their outputs can become significantly more consistent and predictable when guided by structured, machine-readable specifications.

CDISC Analysis Results Standard (ARS), together with emerging Analysis Concepts (AC) and Derivation Concepts (DC), provides a formalized framework for representing analysis intent, derivation logic, and output definitions. This paper demonstrates how ARS-based metadata can be used as controlled input to LLMs to enable rapid and reliable generation of Tables, Figures, and Listings (TFL) programs in SAS and R.

By combining ARS metadata, predefined programming standards, and reusable macro/function libraries, LLM-generated code can be produced in seconds with minimal refinement. This approach reduces manual programming effort, improves consistency, and strengthens traceability between specifications and outputs.

Importantly, this paradigm shifts the role of statistical programmers from manual code development to the design, validation, and governance of structured analysis specifications. The paper highlights how standards-driven GenAI workflows can support scalable, reproducible, and regulator-ready clinical reporting.

1. Introduction

Generative AI (GenAI) is rapidly transforming the landscape of statistical programming and clinical reporting. Large language models (LLMs) are increasingly being used to generate analysis code, automate repetitive programming tasks, and accelerate the production of Tables, Figures, and Listings (TFLs).

However, the adoption of GenAI in regulated environments introduces critical challenges. LLMs are inherently probabilistic, meaning that the same prompt can yield different outputs. This variability raises concerns about reproducibility, traceability, and regulatory compliance—key requirements in clinical trial reporting.

Traditional statistical programming workflows rely heavily on manual interpretation of Statistical Analysis Plans (SAPs), followed by custom code development in SAS or R. While GenAI has the potential to accelerate coding, using LLMs without structured guidance can lead to inconsistent outputs, ambiguity in implementation, and increased quality control burden.

A key question emerges:

How can GenAI be used reliably in clinical reporting?

This paper argues that the answer lies not in improving code generation alone, but in strengthening the standards and metadata that drive it. By leveraging CDISC standards such as ARS, along with emerging AC/DC frameworks, organizations can transform GenAI from a probabilistic tool into a controlled, standards-driven system [1, 2, 3].

2. The Limitation of Code-Centric Thinking

Historically, statistical programming has been code-centric. Programmers interpret SAPs, write custom code, and generate outputs, with code serving as the primary artifact of analysis implementation.

In this paradigm:

- Logic is embedded in code
- Variability arises from individual interpretation
- Reusability is limited
- Traceability is often retrospective

With the introduction of GenAI, this model becomes even more fragile. When LLMs are prompted with loosely defined instructions, they:

- Produce inconsistent implementations
- Vary in structure and style
- Require extensive review and correction

This highlights a fundamental limitation:

Code alone is not a reliable foundation for automation in the age of GenAI.

Instead, the reliability of outputs depends on the quality and structure of the inputs provided to the model.

3. Standards as the Control Layer for GenAI

CDISC Analysis Results Standard (ARS) provides a structured, machine-readable representation of analysis results, including:

- Analysis intent
- Population definitions
- Parameters and endpoints
- Statistical methods
- Display specifications

Complementing ARS, Analysis Concepts (AC) and Derivation Concepts (DC) define:

- What is being analyzed (AC)
- How it is derived (DC)

Together, these standards form a formal specification layer that can guide GenAI systems [1, 2, 3].

When ARS and AC/DC metadata are used as input to LLMs:

- Prompts become structured and unambiguous
- Outputs become more consistent
- Variability is significantly reduced

This effectively transforms GenAI from:

Prompt-driven generation → Specification-driven generation

4. ARS-Driven GenAI Workflow for TFL Generation

A standards-driven GenAI workflow can be structured as follows:

Structured Metadata (AC/DC, ARS)

⟶ Prompt Engineering Layer (templated prompts)

⟶ LLM Code Generation (SAS / R)

⟶ Execution on ADaM datasets

⟶ TFL Outputs + ARD

4.1 Inputs to the LLM

Instead of free-text prompts, the LLM receives:

- ARS metadata (analysis definition)
- AC/DC definitions (logic and derivations)
- Standardized templates (TFL layout and structure)
- Predefined macro libraries (SAS) or function libraries (R)

This creates a controlled input environment aligned with ARS-driven automation approaches [4, 5].

4.2 Example Prompt Structure

A structured prompt may include:

- Analysis objective
- Population definition
- Parameter and endpoint
- Statistical method
- Output format
- Required programming standards

Because these inputs are standardized, the LLM produces:

- Consistent code structure
- Aligned statistical logic
- Reusable programming patterns

4.3 Output Characteristics

The generated code typically demonstrates:

- Alignment with predefined macros/functions
- Standardized naming conventions
- Reduced variability across outputs
- Minimal need for manual correction

In many cases, production-ready code can be generated within seconds, requiring only targeted validation, as demonstrated in recent GenAI-based TFL automation work [6].

5. Improving Reliability and Reproducibility

A major concern with GenAI is reproducibility. Without structure, LLM outputs may vary across runs.

However, when driven by ARS metadata:

- Inputs are deterministic
- Prompts are templated

- Logic is predefined

This significantly reduces variability in outputs.

Key Improvements:

- Consistency: Standardized metadata ensures consistent interpretation across analyses.
- Traceability: Direct linkage between ARS definitions and generated code enables auditability.
- Reproducibility: Given the same metadata and prompt template, outputs are highly stable.
- Validation: Generated code can be validated against structured specifications rather than subjective interpretation.

6. Redefining the Role of Statistical Programmers

As GenAI automates code generation, the role of statistical programmers evolves.

Traditional Role

- Interpret SAP
- Write code
- Debug and refine
- Perform QC

Emerging Role in GenAI + Standards Environment

- Define structured analysis specifications (AC/DC, ARS)
- Design reusable templates and macros
- Validate metadata and generated outputs
- Govern standards and workflows

This represents a shift from:

Coding → Specification Design and Governance

7. Regulatory Considerations

Regulatory acceptance of GenAI in clinical reporting depends on:

- Traceability between specifications and outputs
- Reproducibility of results
- Transparency in analysis logic

Standards-driven GenAI workflows address these requirements by:

- Using ARS as a formal specification layer
- Maintaining linkage between metadata, code, and outputs
- Enabling consistent validation processes

Rather than viewing GenAI as a “black box,” this approach positions it as:

A controlled execution layer operating on validated standards

8. Practical Observations

Based on early implementations and pilot use cases, several observations emerge:

Area	Traditional Approach	GenAI without Standards	Standards-Driven GenAI
Code Consistency	Variable	Highly variable	Highly consistent
Programming Effort	High	Reduced but unstable	Significantly reduced
QC Effort	High	Very high (due to variability)	Reduced
Traceability	Limited	Weak	Strong
Reproducibility	Moderate	Low	High
Regulatory Readiness	Established	Uncertain	Strong potential

These observations highlight that:

GenAI alone does not solve problems - Standards make GenAI usable.

9. Future Directions

The integration of standards and GenAI opens new possibilities:

- Automated CSR narrative generation from ARD
- Real-time validation using structured metadata
- Expansion of AC/DC libraries for reusable analysis patterns
- Fully metadata-driven clinical reporting pipelines

As standards mature, GenAI systems will become increasingly reliable and scalable [6].

10. Conclusion

In the age of GenAI, the focus of clinical reporting must shift from code to standards.

While LLMs offer powerful capabilities for automating programming tasks, their reliability depends on structured, machine-readable inputs. CDISC standards such as ARS, along with emerging AC/DC frameworks, provide the foundation for transforming GenAI into a controlled, reproducible, and regulator-ready system.

By adopting a standards-driven approach, organizations can reduce manual effort, improve consistency, and enable scalable automation in clinical reporting. More importantly, this paradigm redefines the role of statistical programmers as designers and stewards of analysis specifications rather than manual coders.

Ultimately, in a GenAI-enabled world:

Standards -- not code -- become the primary drivers of quality, consistency, and trust.

12. References

1. CDISC ARS Model: https://cdisc-org.github.io/analysis-results-standard/
2. CDISC ARS User Guide version 1.0: https://wiki.cdisc.org/display/ARSP/Analysis+Results+Standard+User+Guide+v1.0

3. Busa, Bhavin; Harris, Brian. *Analysis Concepts Definition.* CDISC U.S. Interchange Oct 2025. Presentation Link
4. Bosman, Malan. '*Advancing TFL Automation: R-Based Meta-Programming Powered by ARS*' PHUSE US Connect, March 2025.
5. Busa, Bhavin; Marshall, Richard; LeRoy, Bess; 'All You Need to Know about the New CDISC Analysis Result Standards!' PharmaSUG, May 2023.
6. Dedhia Navin; Busa Bhavin. *'*AI Code Generator: Automating TFL Programming with ARS Metadata*'* PHUSE US Connect, March 2026.

Contact Information

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

Bhavin Busa
Principal & Co-founder,
Clymb Clinical
bhavin@clymbclinical.com