Paper ML08

Revolutionizing Clinical Trials with Intelligent Protocol Automation

Zach Martin, SAS Institute, Cary, NC, USA
Sundaresh Sankaran, SAS Institute, Cary, NC, USA

ABSTRACT

Designing clinical trial protocols and associated documents is time-consuming and labor-intensive, requiring ingest, aggregation and analysis of data from multiple sources. Conventional, manual approaches often result in inconsistency, errors, and delays. This hinders trial effectiveness and efficiency. We harness SAS's Natural Language Processing (NLP) capabilities along with Large Language Models (LLMs) to generate draft clinical protocols in a simple, governed and trusted manner. Low-code interfaces facilitate accessible prompt instructions and output specification using desired templates. Our unified AI platform, SAS Viya, enhances output quality by implementing a Retrieval Augmented Generation (RAG) approach which ensures right context, thus mitigating risk of hallucinations. This session helps you better appreciate how the power of AI, including Generative AI, accelerates and enhances clinical protocol development, leading to increased productivity and unlocking researchers' time and bandwidth for focus on high-value tasks.

INTRODUCTION

Any bravehearted soul who has ever attempted to read a Clinical Trial protocol document will share our sentiments. Clinical protocol documents tend to be complex and lengthy documents consolidating deep layers of sections. They are important because they are the foundation of clinical trials. A Clinical Trial Protocol provides a detailed plan outlining the background, objectives, design, methodology and statistical analysis of a clinical trial, and ensures consistency and standardization throughout the trial. Every clinical trial contains a written protocol that describes how a trial will be conducted to ensure patient safety, efficacy, and data integrity. Designed in a collaborative manner, it provides a roadmap for all resources and functions involved in clinical trials. The adherence to a protocol ensures compliance with regulatory standards such as those defined by the FDA and EMA.

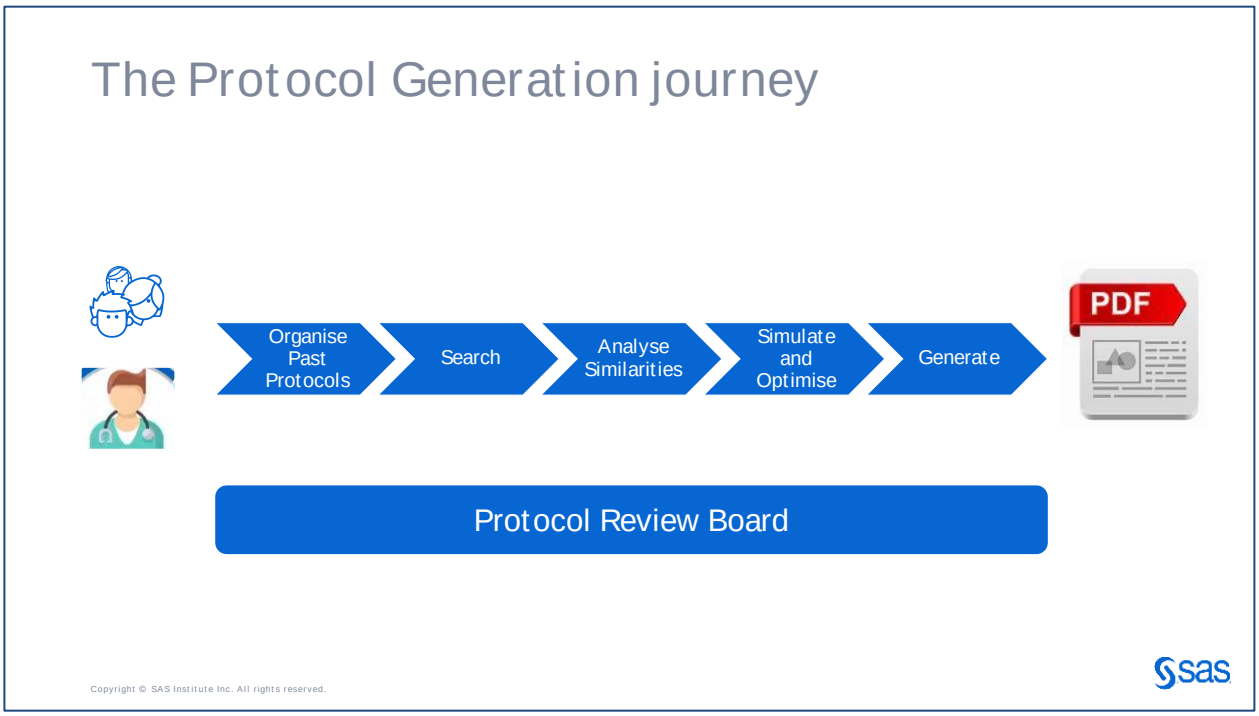
The ICH (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) has created the M11 template to help the industry standardize and digitize clinical trial protocols. Currently organizations have templates for different phases and of varying complexity. A move towards greater consolidation and standardization is expected in coming years.

In addition to different standards, the other major challenge is a need to increase productivity. Currently it takes anywhere between 6 to 18 months for a protocol to make its way through various stages of review until development. This is dictated by stringent industry guidelines which are administered by protocol review boards. A productive protocol development process helps to expedite bringing life changing treatments to market.

In this paper, we explore some options to enhance clinical trial protocols through Artificial Intelligence (AI) methods that make use of Natural Language Processing, Machine Learning, lookback analysis and scenario simulation.

THE PROTOCOL GENERATION JOURNEY

We present the protocol generation journey in the following illustration. It is characterized by multiple stakeholders who come from different backgrounds such as biostatisticians, programmers, operational and managerial profiles. They approach the process by searching past protocols, analyzing similar elements of past protocols, and matching them with specific requirements for the study under consideration. They then carry out simulation activities to help optimize for specific protocol criteria and all these exercises give rise to inputs that go into the final draft protocol, reviewed at various intervals by a protocol review board.



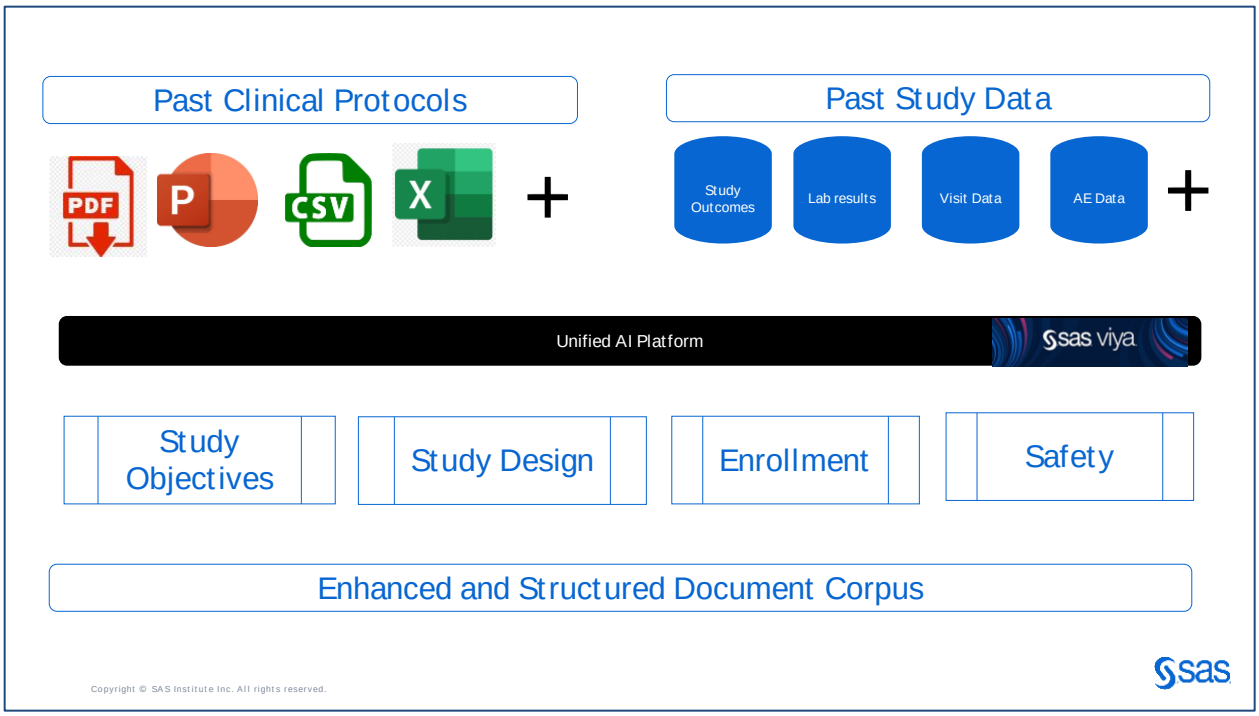
STRUCTURING OF PROTOCOL DOCUMENTS

Our first stage of applying AI to enhance and improve clinical protocol generation starts with data. In our goal to identify only the relevant sections of past protocols, we need to first figure out how to ingest the data into the system in the first place. Using SAS Viya for this exercise, we used available connectors to ingest unstructured text contained within past protocol documents of multiple formats, including PDFs, Word, PPT and other formats. The core metadata such as filename, study ID, date of creation etc. were also retained. At an early stage in the process, we used Natural Language Processing (NLP) to structure the content within these documents into logical segments, such as Objectives, Enrolment Criteria etc. Thus, we extracted specific sections of a protocol document which exhibit a common patterns into columns or tables of their own.

Some important sections which get persisted in a corpus of past protocols are:

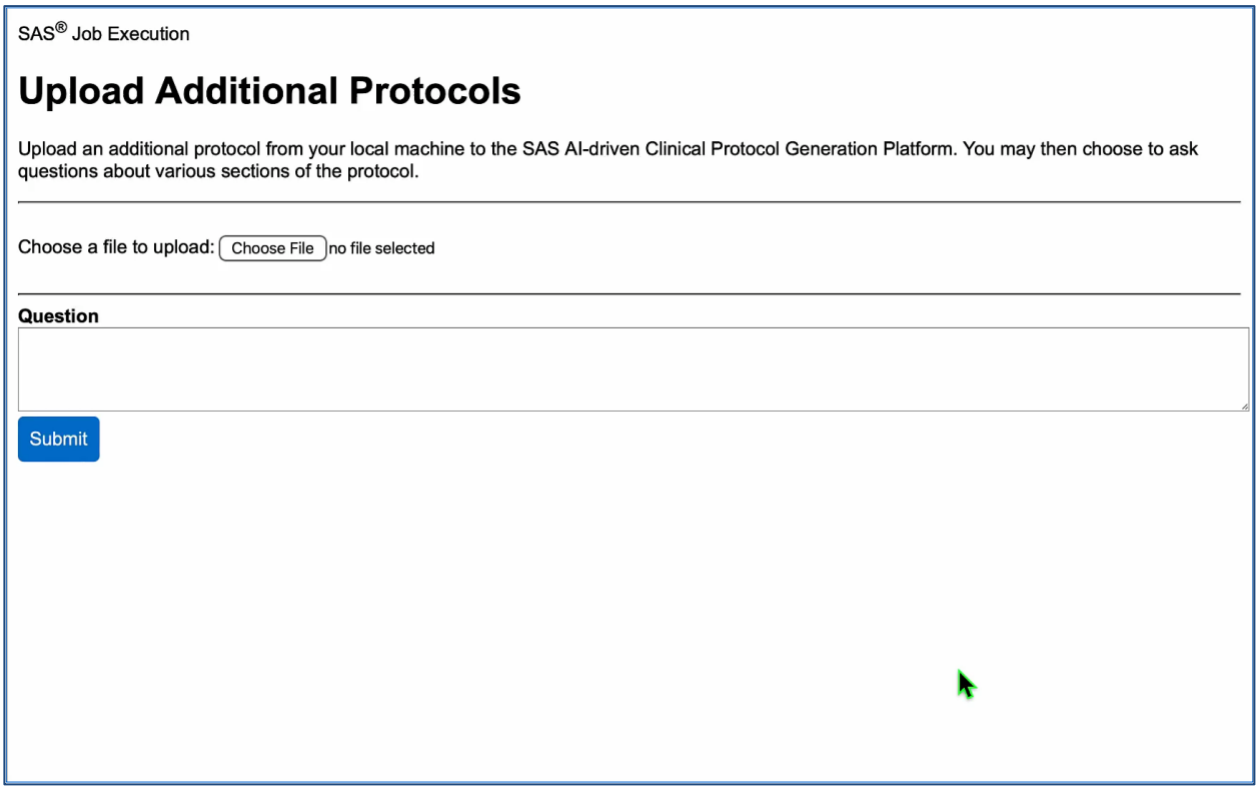
- Therapeutic area
- Study phase
- Design
- Study Objectives
- Enrolment / Eligibility Criteria (inclusion and exclusion)
- Study Endpoints

At this stage, a structured datamart of multiple sections has been built using past protocols. Past protocols have been ingested and structured in such a way that they are easily searchable when a new study on similar lines is undertaken.



ANALYSIS OF PAST PROTOCOLS

In addition to conventional search-based mechanisms to identify past protocols based on common shared characteristics and metadata, we harness Large Language Models (LLMs) and other AI methods to help us understand and summarize the contents of a protocol's section in an easy and coherent manner. For example, using the SAS Viya platform, we designed a convenient user interface through SAS Job Execution, which helped us point towards a protocol and make calls to an LLM to summarize key elements of the same such as the study phase, study criteria and a structured representation of inclusion and exclusion criteria. The LLM can even be hosted in-house (i.e. within internal company networks) in case life sciences organizations deal with non-public, or yet to be published protocols.



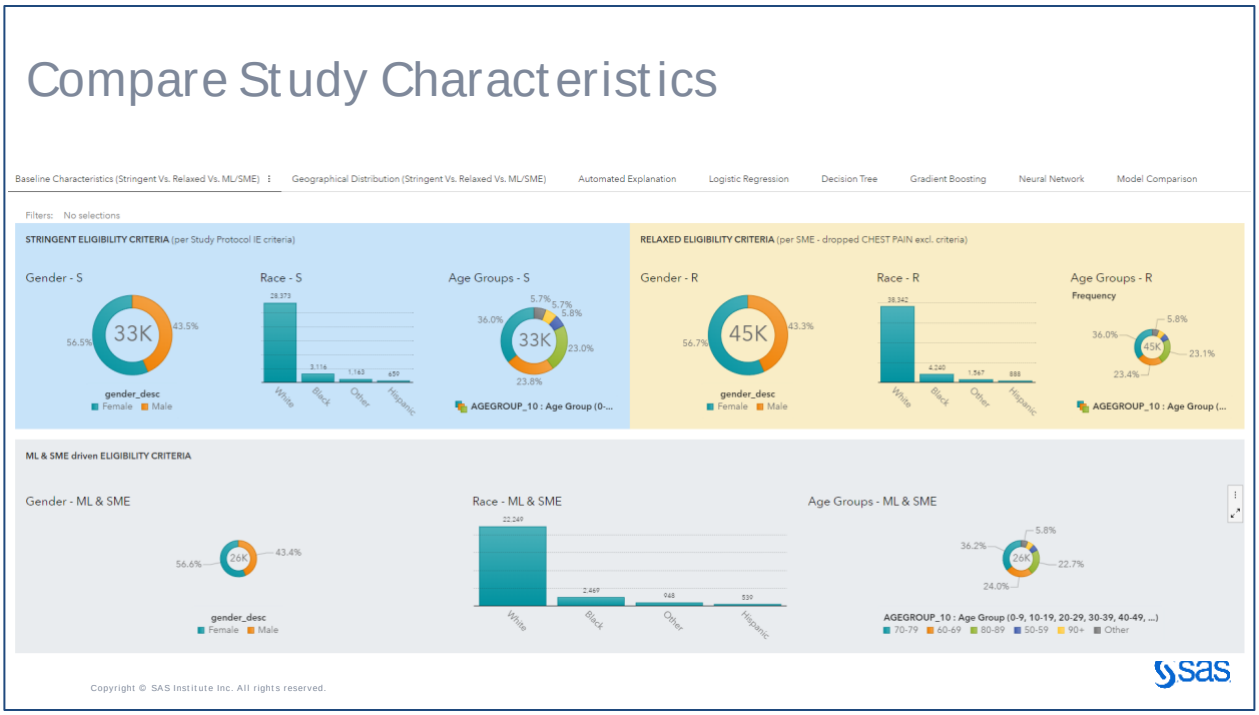
The benefit of harnessing LLMs and other AI techniques within a unified analytics platform such as SAS Viya is the seamless transfer of metadata as data goes from one transformation to another, reducing the need for redundancy or rework. This helps you obtain accurate results in an efficient manner. Furthermore, a common platform affords scope for multiple interfaces oriented towards multiple user personas, such as clinicians, data engineers, statisticians or business users, to access the same data through their applications of choice.

In addition to LLMs, conventional search interfaces and visualization offer different analytical techniques to narrow down search results and focus on protocols like your area of study. After filtering protocols, designers can also use LLMs to search for similar protocols around a target protocol and obtain a summary of those results instead of having to screen all similar protocols.

SIMULATION AND BENCHMARKING

Due to the extraction and structuring of different sections of a protocol into common elements, analysis and tweaking of multiple data elements can be carried out through simulation and benchmarking exercises. This helps the organization identify the right mix of controllable variables that go into the draft protocol.

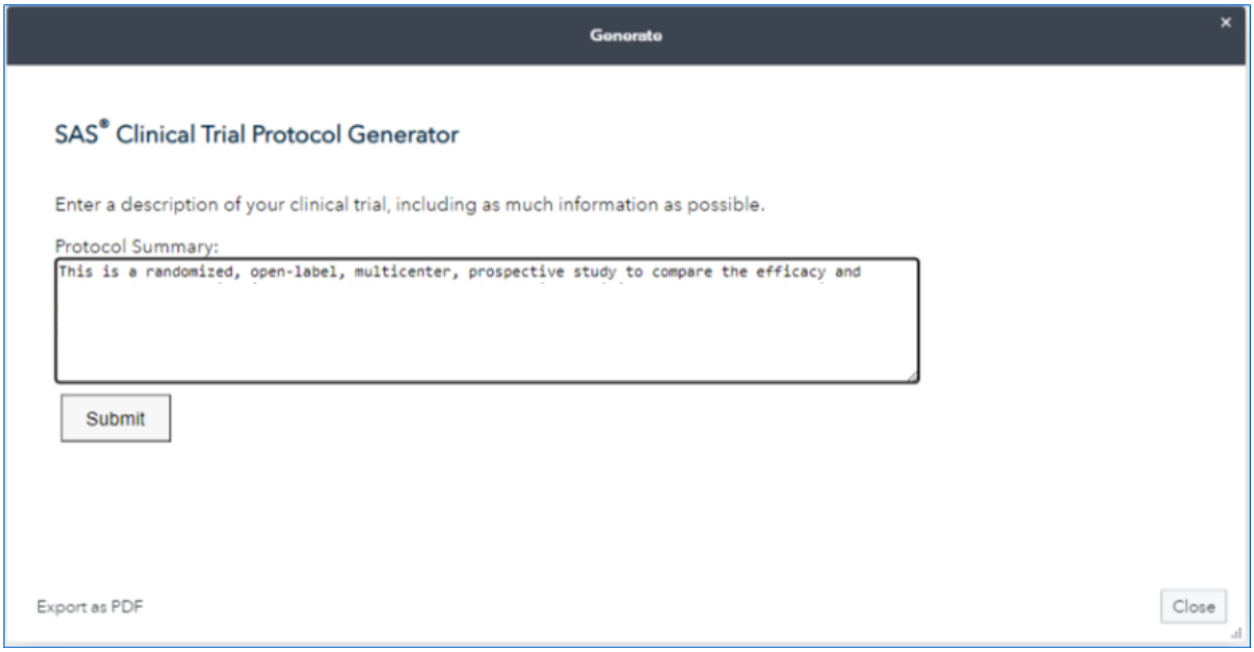
For example, you can simulate projected enrollment rates given different combinations of the site, inclusion and exclusion criteria. Projections can be discussed with the protocol review board prior to finalizing criteria.



DOCUMENT GENERATION

With the key upstream analysis taken care of, the actual generation of a draft protocol document is greatly eased. LLMs can be trained to use predefined templates (such as ICH M11 or an organization-specific template) through instructions built from the bottom up for each individual section separately. This enables better control over the final output that gets generated. The LLM output can be orchestrated from within an analytics platform on which the protocol documents reside, thus ensuring data security. Output can be saved within the platform and go through an approval workflow prior to release, ensuring built-in governance in the process. The draft protocol can also be collaboratively edited and versioned using specialized tools for this purpose.

An example of a custom-built interface to generate protocols is provided below. These can be further customized to include specific sections or parts of a protocol (provided purely for illustrative purposes) and even specify the source data to be used for generation.



CONCLUSION

While we believe that more nuanced analysis and customization is required as per every life sciences organization's unique data landscape, we identify building blocks which incorporate AI (including Generative AI) techniques to simplify and enhance the protocol generation process, and the critical upstream tasks of data management, visualization and analysis of past protocols for this purpose. We also highlight the crucial role played by a unified analytical platform in the entire process due to the many applications and transformation stages involved. In addition, open architecture which allows for easy integration and interoperability with external services such as Large Language Models, vector stores and web applications is useful.

REFERENCES

1. ICH M11 Template - https://www.ema.europa.eu/en/ich-m11-guideline-clinical-study-protocol-template-and-technical-specifications-scientific-guideline

ACKNOWLEDGMENTS

Acknowledgements to Sherrine Eid, Brittany Shriver & Soundarya Palanisamy, SAS Institute for their industry knowledge, assistance and inputs.

CONTACT INFORMATION

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

Author Name: Zach Martin
Company: SAS Institute
Address: 100 SAS Campus Dr

Author Name: Sundaresh Sankaran
Company: SAS Institute
Address: 100 SAS Campus Dr
Work Phone: +1 919 531 4921
Email: Sundaresh.sankaran@sas.com
Website: https://linkedin.com/in/sundareshsankaran