Transforming Clinical Programming with Generative AI: Enhancing SDTM, ADaM, and TLF Processes



Matt Becker, Life Science Strategic Advisor, SAS
Pankaj Attri, Pre-Sales Solutions Architect, SAS

Introduction to GenAI in Clinical Trials

Life sciences industry requires faster, compliant clinical trial submissions

Generative AI (GenAI) can significantly automate SDTM, ADaM, and TLF processes

LLMs can reduce manual efforts while maintaining compliance and quality standards

Current State of Clinical Programming

Clinical trials are increasingly complex	Many processes still rely on manual coding	Manual programming can be repetitive, inefficient, and prone to errors
• Tighter regulations • Intricate data sources	• Even with CDISC standards	• Impacts submissions

The Need for Automation

Manual data mapping and coding

- Time-consuming
- Resource-intensive

Automation can improve accuracy and reduce burden

As trials expand globally, the demand for streamlined processes rises

Understanding GenAI and LLMs

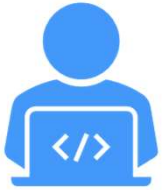
GenAI refers to algorithms capable of generating new content, such as code and summaries

Large Language Models (LLMs) are specialized GenAI tools trained on vast data sets

LLMs help bridge the gap between human expertise and machine automation in programming

Goals of the Presentation

Demonstrate practical applications of LLMs in clinical programming workflows



Explore automation in SDTM mapping, ADaM generation, and TLF programming

Showcase real-world case studies highlighting LLM benefits in efficiency and compliance

Common Challenges Faced

Mapping SDTM and ADaM datasets remains a manual, error-prone process often leading to inconsistencies

Common TLF coding is redundantly written, increasing the risk of errors and inefficiencies

Document management is often disjointed, complicating cross-referencing and updates

Business Impact of Automation

Delivers time savings in dataset mapping and TLF production, leading to quicker outputs

Cost efficiency emerges as reliance on external resources decreases

Increased submission readiness through consistent, compliant outputs across studies

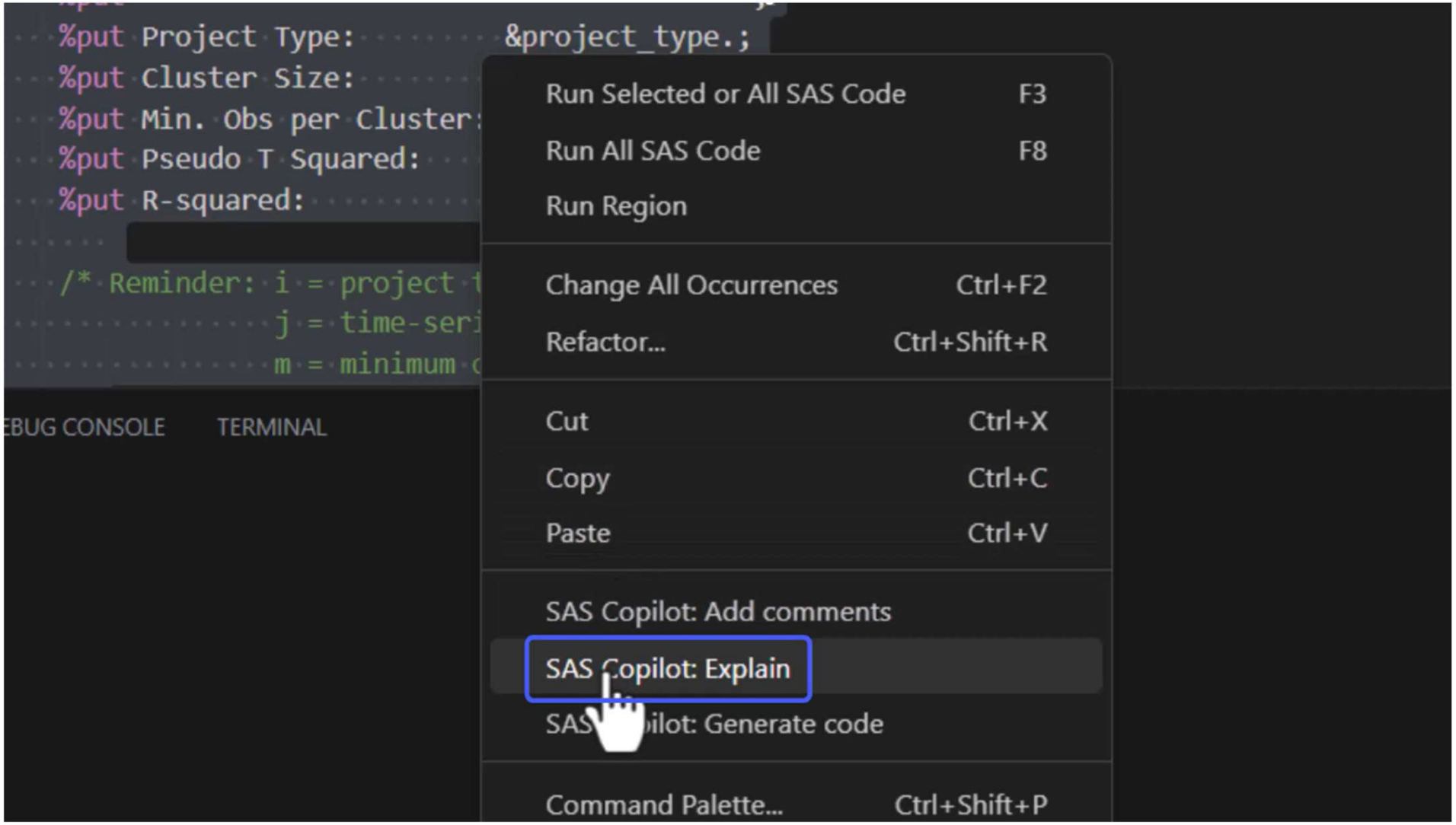
LLM Capabilities in Life Sciences

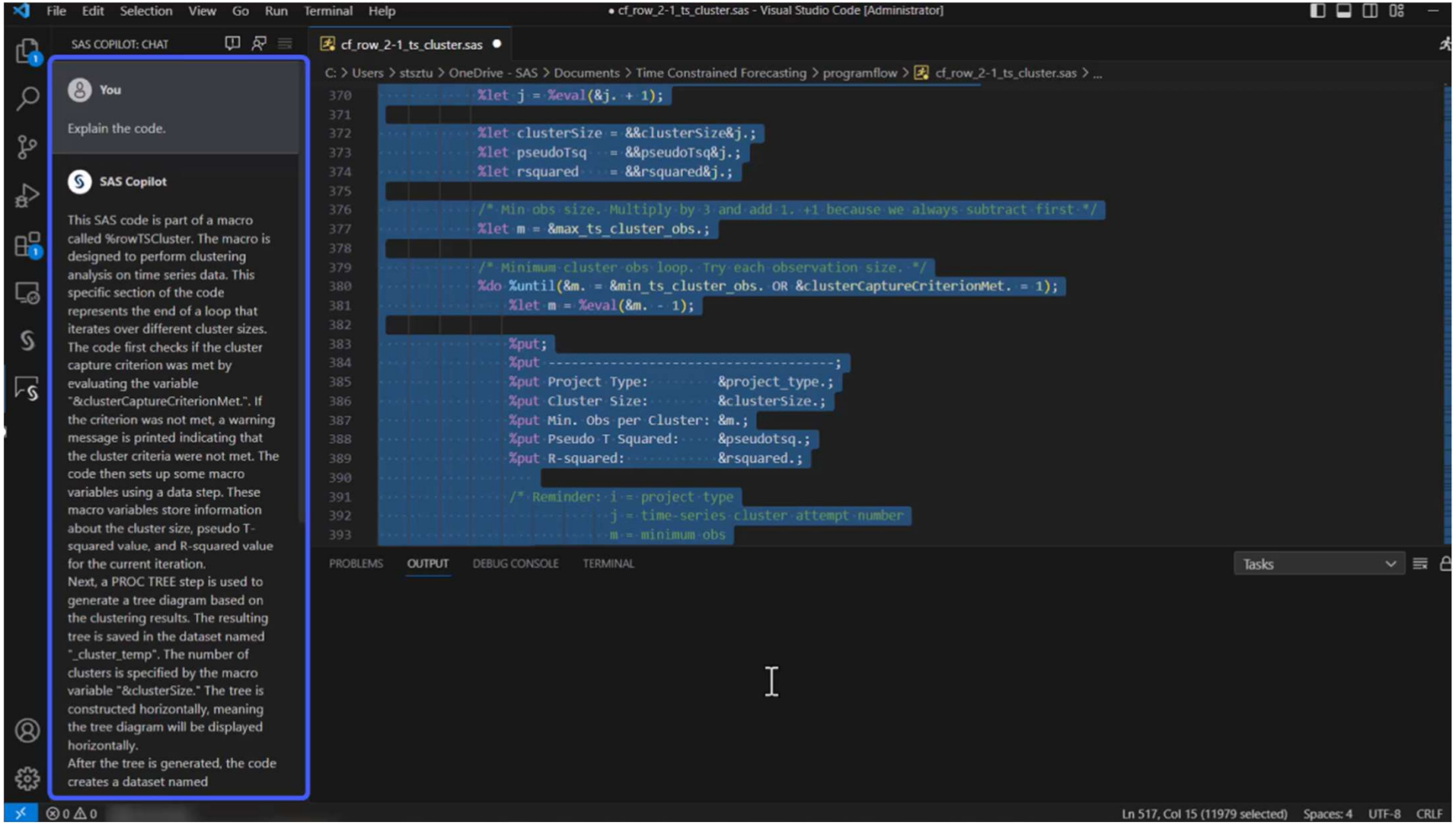
LLMs can translate natural language requirements into executable programming code

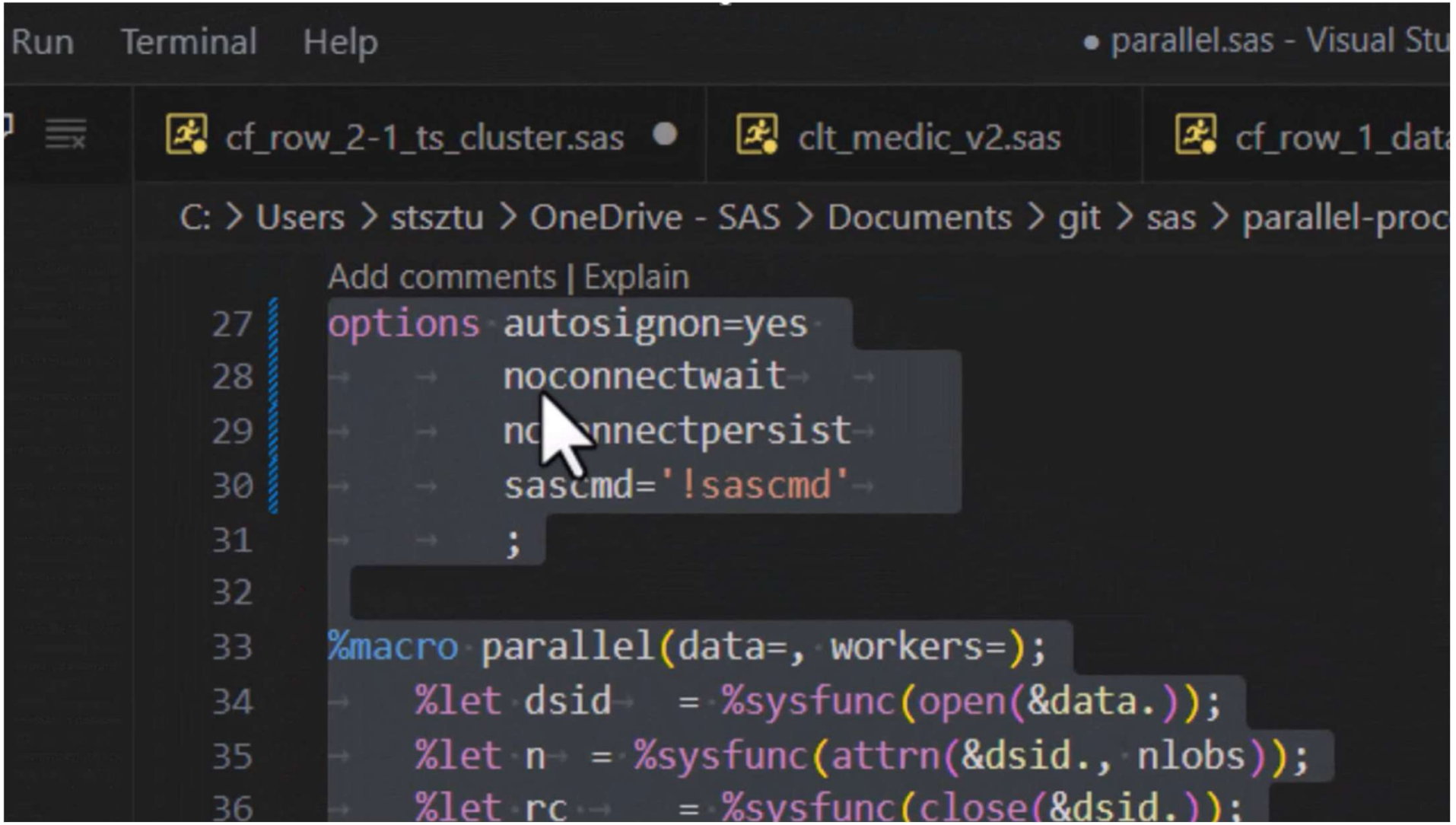
Can generate starter code for datasets by interpreting metadata specifications

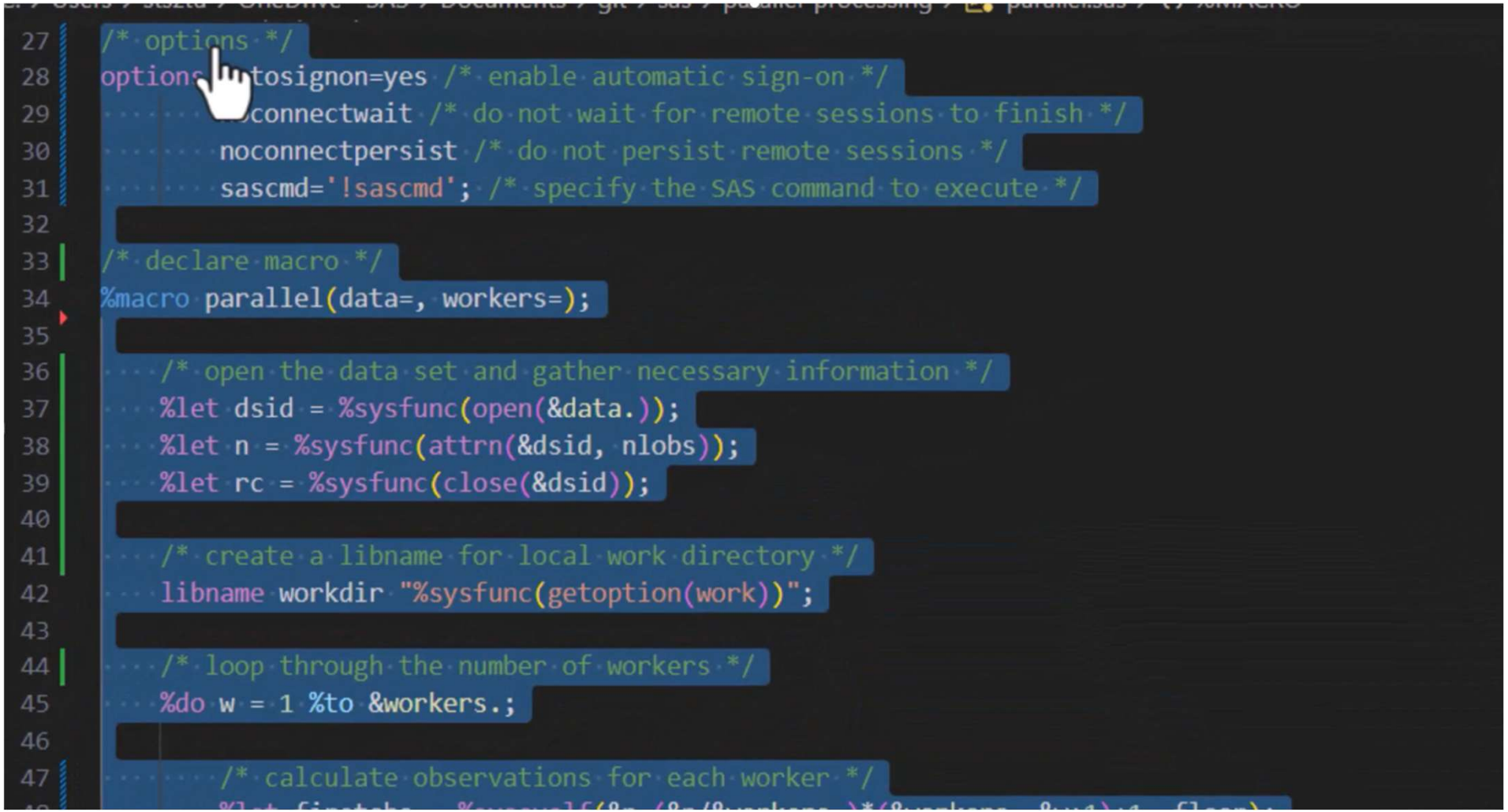
Assist in identifying inconsistencies in programming documentation and logic

SAS Co-Pilot

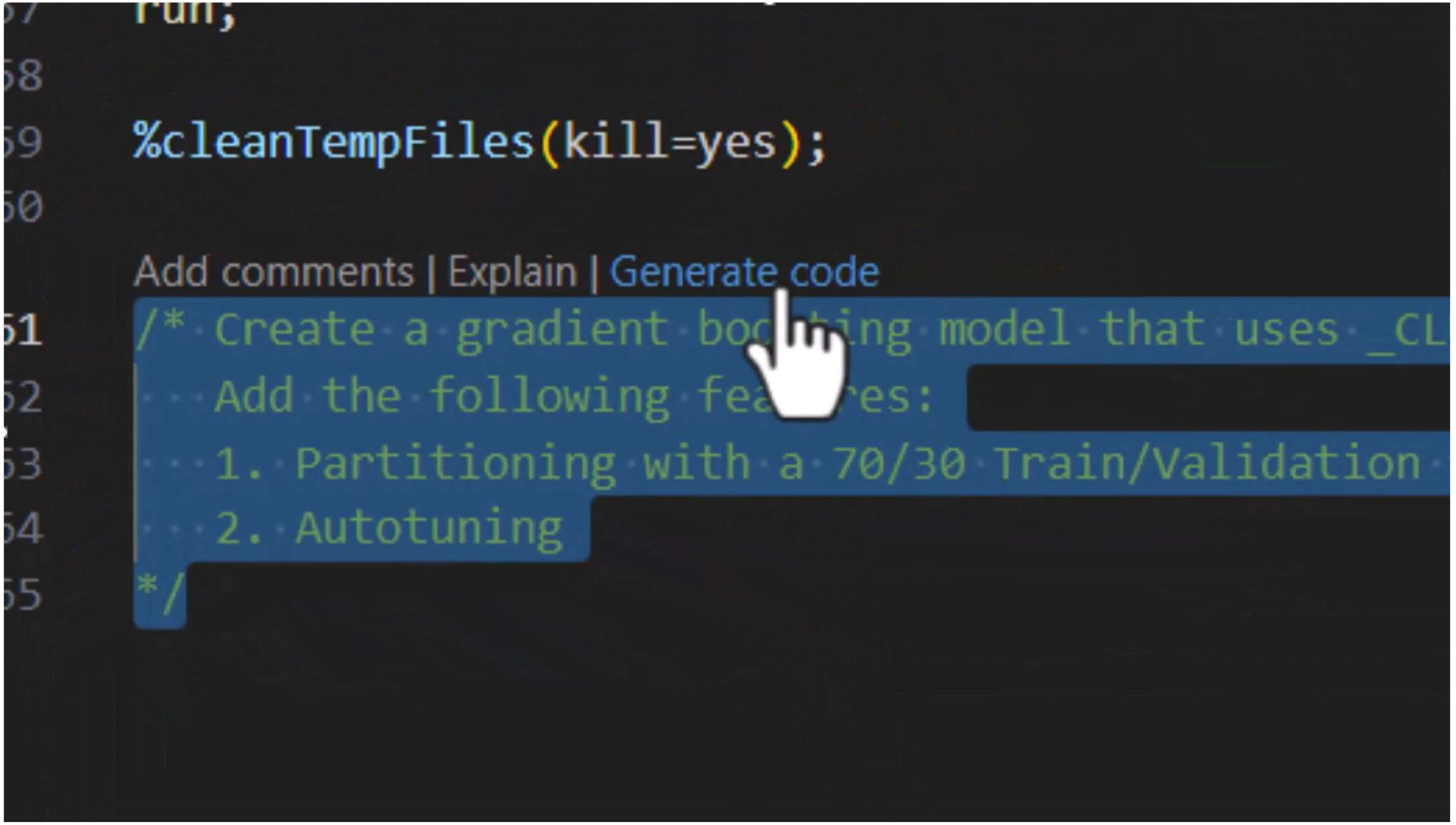


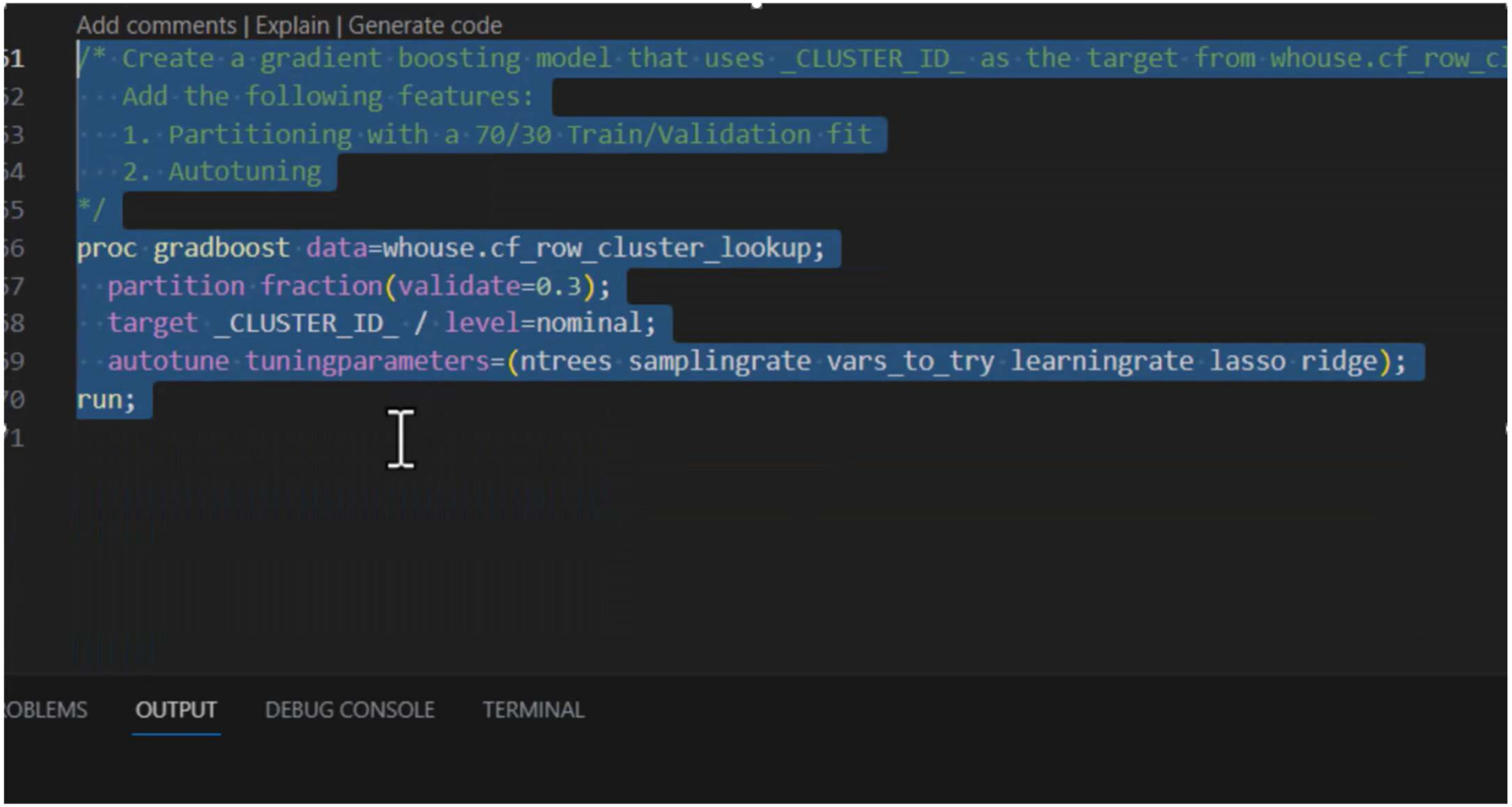






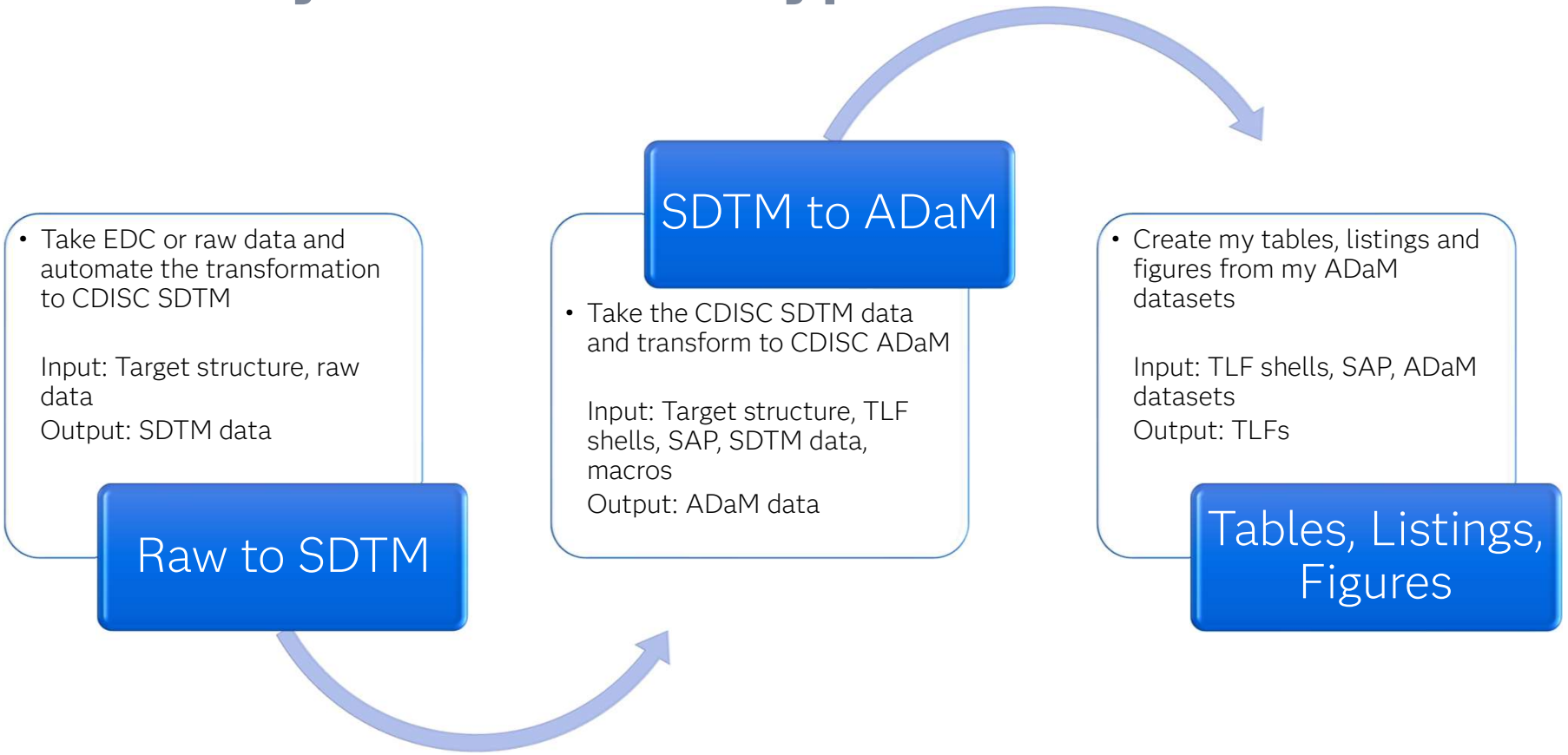
```
/* options */
options autosignon=yes /* enable automatic sign-on */
        noconnectwait /* do not wait for remote sessions to finish */
        noconnectpersist /* do not persist remote sessions */
        sascmd='!sascmd'; /* specify the SAS command to execute */

/* declare macro */
%macro parallel(data=, workers=);

    /* open the data set and gather necessary information */
    %let dsid = %sysfunc(open(&data.));
    %let n = %sysfunc(attrn(&dsid, nlobs));
    %let rc = %sysfunc(close(&dsid));

    /* create a libname for local work directory */
    libname workdir "%sysfunc(getoption(work))";

    /* loop through the number of workers */
    %do w = 1 %to &workers.;

        /* calculate observations for each worker */
```

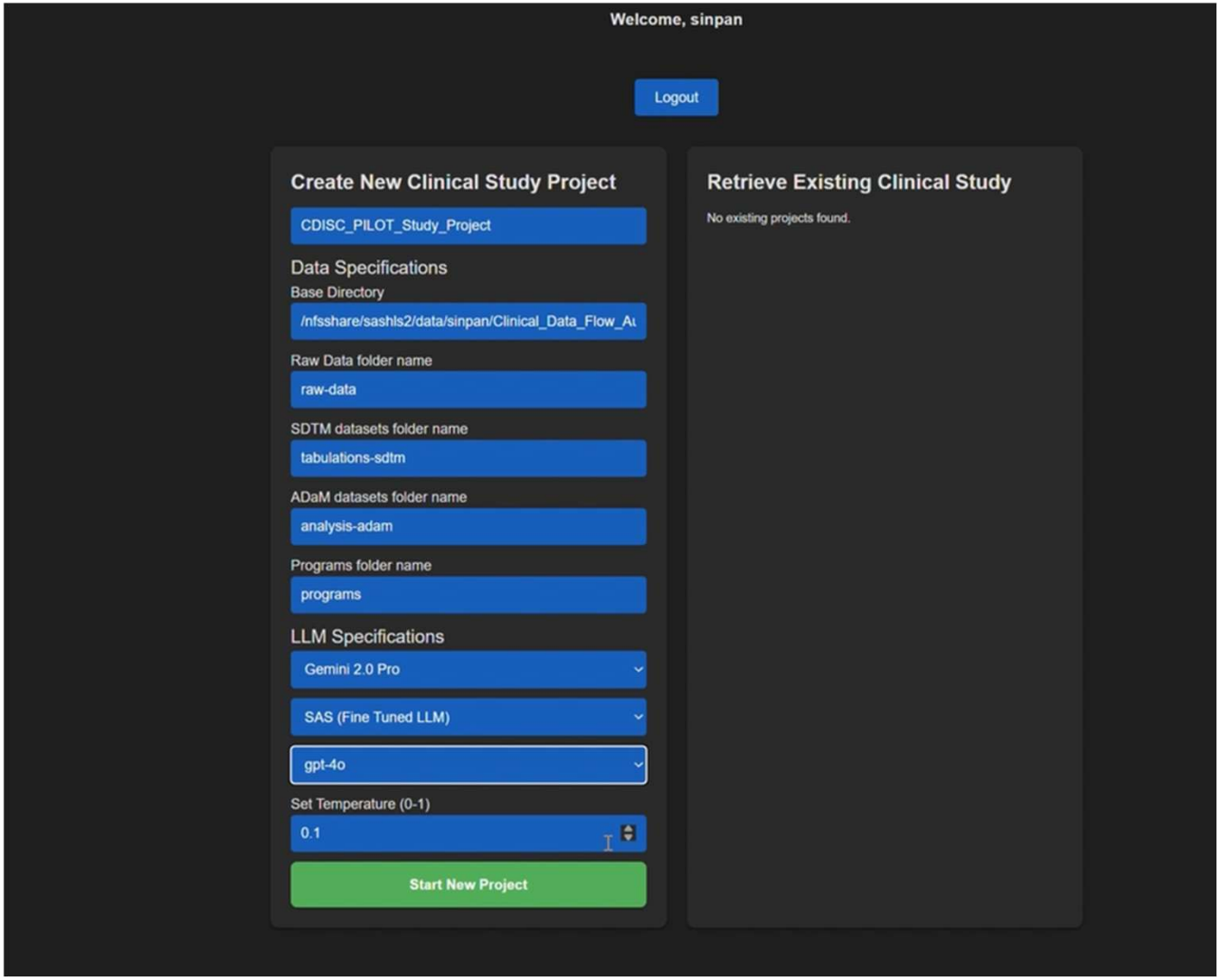


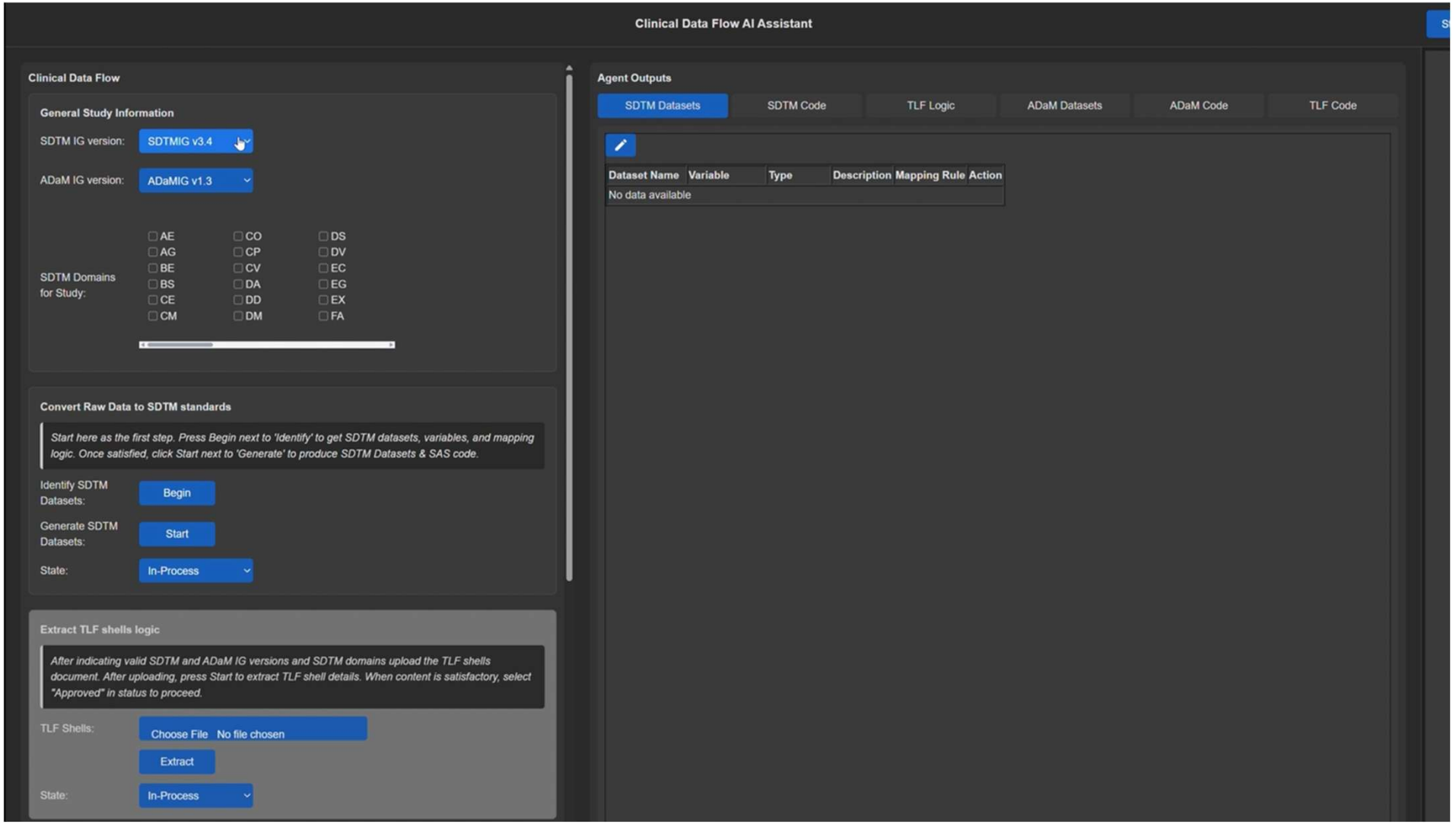


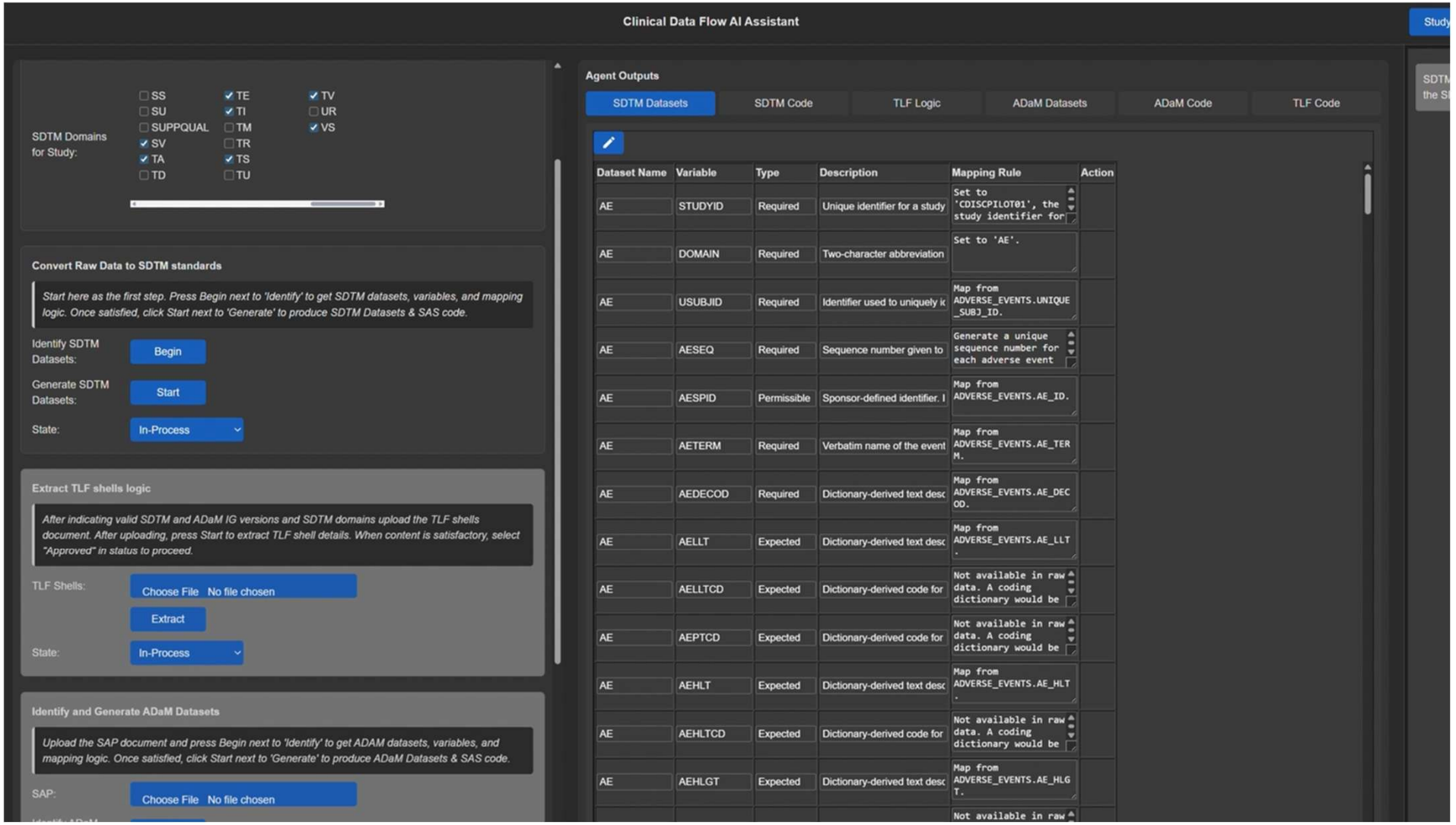
Agentic AI for Clinical Data Flow

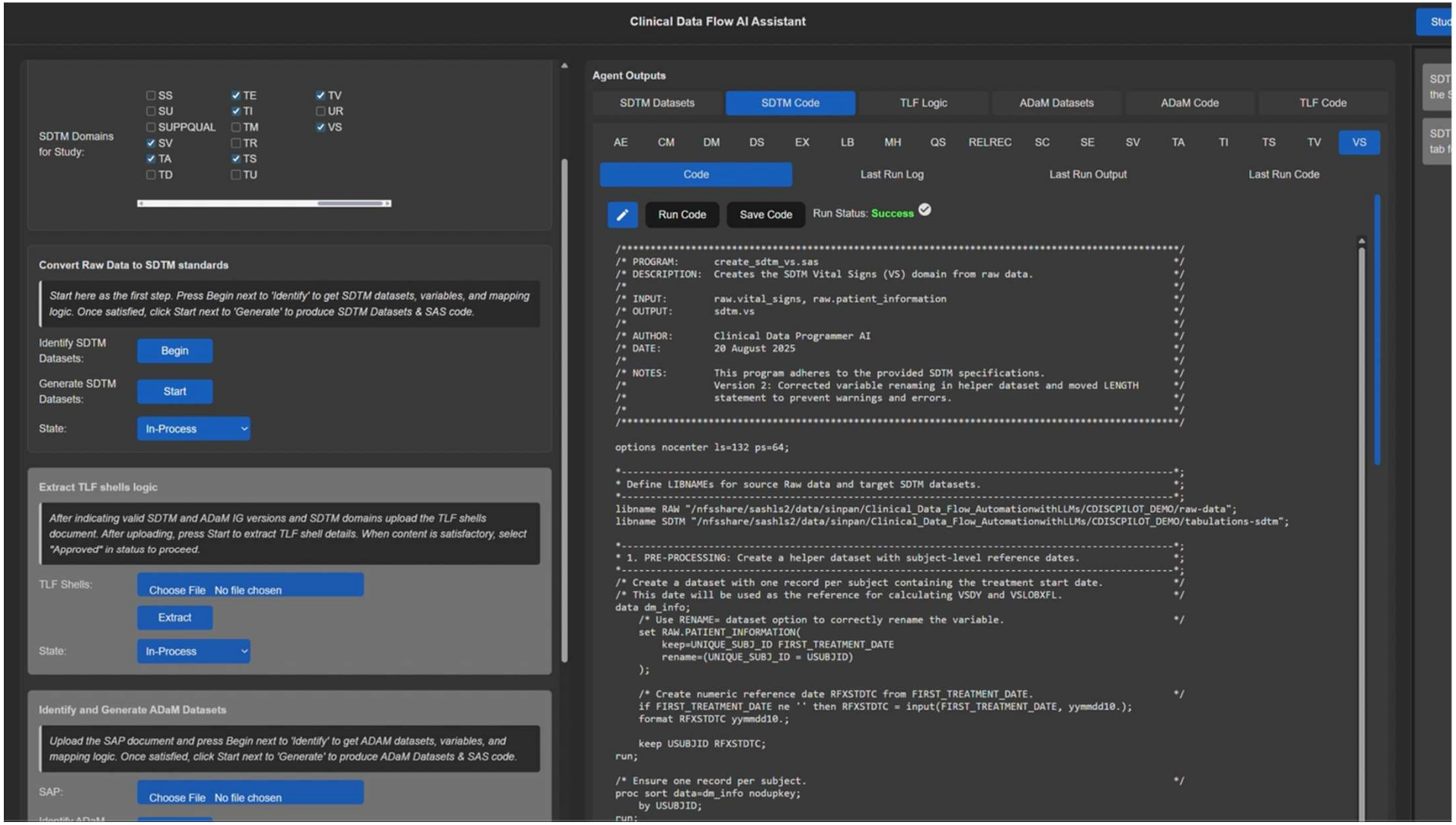
Minimally Viable Prototype - Goals

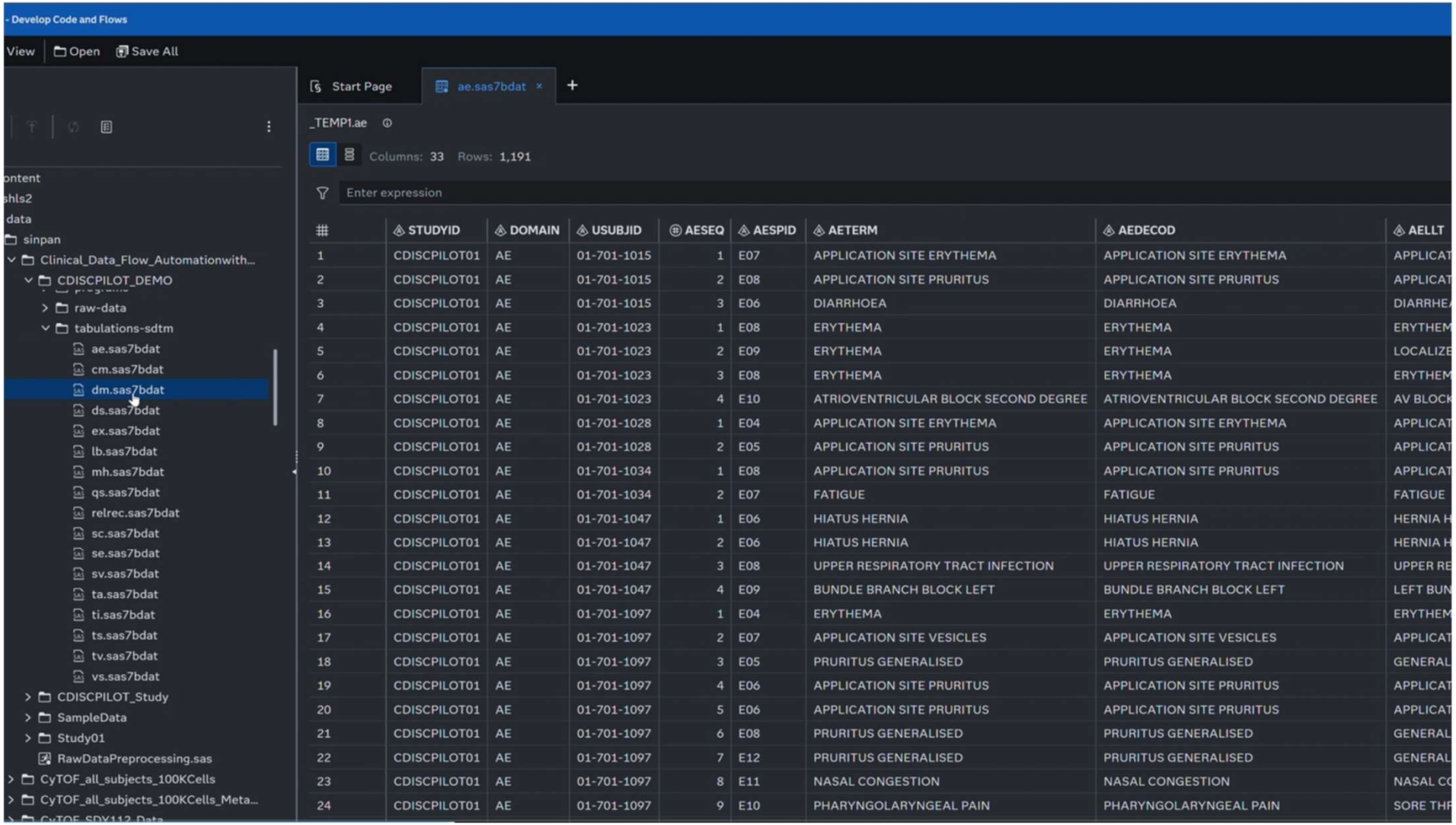












#	STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AEDECOD	AELLT
1	CDISCPILOT01	AE	01-701-1015	1	E07	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	APPLICAT
2	CDISCPILOT01	AE	01-701-1015	2	E08	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
3	CDISCPILOT01	AE	01-701-1015	3	E06	DIARRHOEA	DIARRHOEA	DIARRHEA
4	CDISCPILOT01	AE	01-701-1023	1	E08	ERYTHEMA	ERYTHEMA	ERYTHEM
5	CDISCPILOT01	AE	01-701-1023	2	E09	ERYTHEMA	ERYTHEMA	LOCALIZE
6	CDISCPILOT01	AE	01-701-1023	3	E08	ERYTHEMA	ERYTHEMA	ERYTHEM
7	CDISCPILOT01	AE	01-701-1023	4	E10	ATRIOVENTRICULAR BLOCK SECOND DEGREE	ATRIOVENTRICULAR BLOCK SECOND DEGREE	AV BLOCK
8	CDISCPILOT01	AE	01-701-1028	1	E04	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	APPLICAT
9	CDISCPILOT01	AE	01-701-1028	2	E05	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
10	CDISCPILOT01	AE	01-701-1034	1	E08	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
11	CDISCPILOT01	AE	01-701-1034	2	E07	FATIGUE	FATIGUE	FATIGUE
12	CDISCPILOT01	AE	01-701-1047	1	E06	HIATUS HERNIA	HIATUS HERNIA	HERNIA H
13	CDISCPILOT01	AE	01-701-1047	2	E06	HIATUS HERNIA	HIATUS HERNIA	HERNIA H
14	CDISCPILOT01	AE	01-701-1047	3	E08	UPPER RESPIRATORY TRACT INFECTION	UPPER RESPIRATORY TRACT INFECTION	UPPER RE
15	CDISCPILOT01	AE	01-701-1047	4	E09	BUNDLE BRANCH BLOCK LEFT	BUNDLE BRANCH BLOCK LEFT	LEFT BUN
16	CDISCPILOT01	AE	01-701-1097	1	E04	ERYTHEMA	ERYTHEMA	ERYTHEM
17	CDISCPILOT01	AE	01-701-1097	2	E07	APPLICATION SITE VESICLES	APPLICATION SITE VESICLES	APPLICAT
18	CDISCPILOT01	AE	01-701-1097	3	E05	PRURITUS GENERALISED	PRURITUS GENERALISED	GENERAL
19	CDISCPILOT01	AE	01-701-1097	4	E06	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
20	CDISCPILOT01	AE	01-701-1097	5	E06	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
21	CDISCPILOT01	AE	01-701-1097	6	E08	PRURITUS GENERALISED	PRURITUS GENERALISED	GENERAL
22	CDISCPILOT01	AE	01-701-1097	7	E12	PRURITUS GENERALISED	PRURITUS GENERALISED	GENERAL
23	CDISCPILOT01	AE	01-701-1097	8	E11	NASAL CONGESTION	NASAL CONGESTION	NASAL CO
24	CDISCPILOT01	AE	01-701-1097	9	E10	PHARYNGOLARYNGEAL PAIN	PHARYNGOLARYNGEAL PAIN	SORE THR

Protocol: CDISCPILOT01 Page 1 of n
Population: Intent-to-Treat

Template 3
Summary of Demographic and Baseline Characteristics

		Placebo (N=100)	Xanomeline Low Dose (N=100)	Xanomeline High Dose (N=100)	Total (N=300)	p-value [1]
Age (y)	n	xx	xx	xx	xx	
	Mean	xx.x	xx.x	xx.x	xx.x	0.xxx
	SD	x.xx	x.xx	x.xx	x.xx	
	Median	xx.x	xx.x	xx.x	xx.x	
	Min.	xx.x	xx.x	xx.x	xx.x	
	Max.	xx.x	xx.x	xx.x	xx.x	
	<65 yrs	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	0.xxx
	65-80 yrs	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	
	>80 yrs	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	
Sex	n	xxx	xxx	xxx	xxx	0.xxx
	Female	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	
	Male	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	
Origin	n	xxx	xxx	xxx	xxx	0.xxx
	Black	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	
	White	xx (xx%)	xx (xx%)	xx (xx%)	xx (xx%)	
	…					

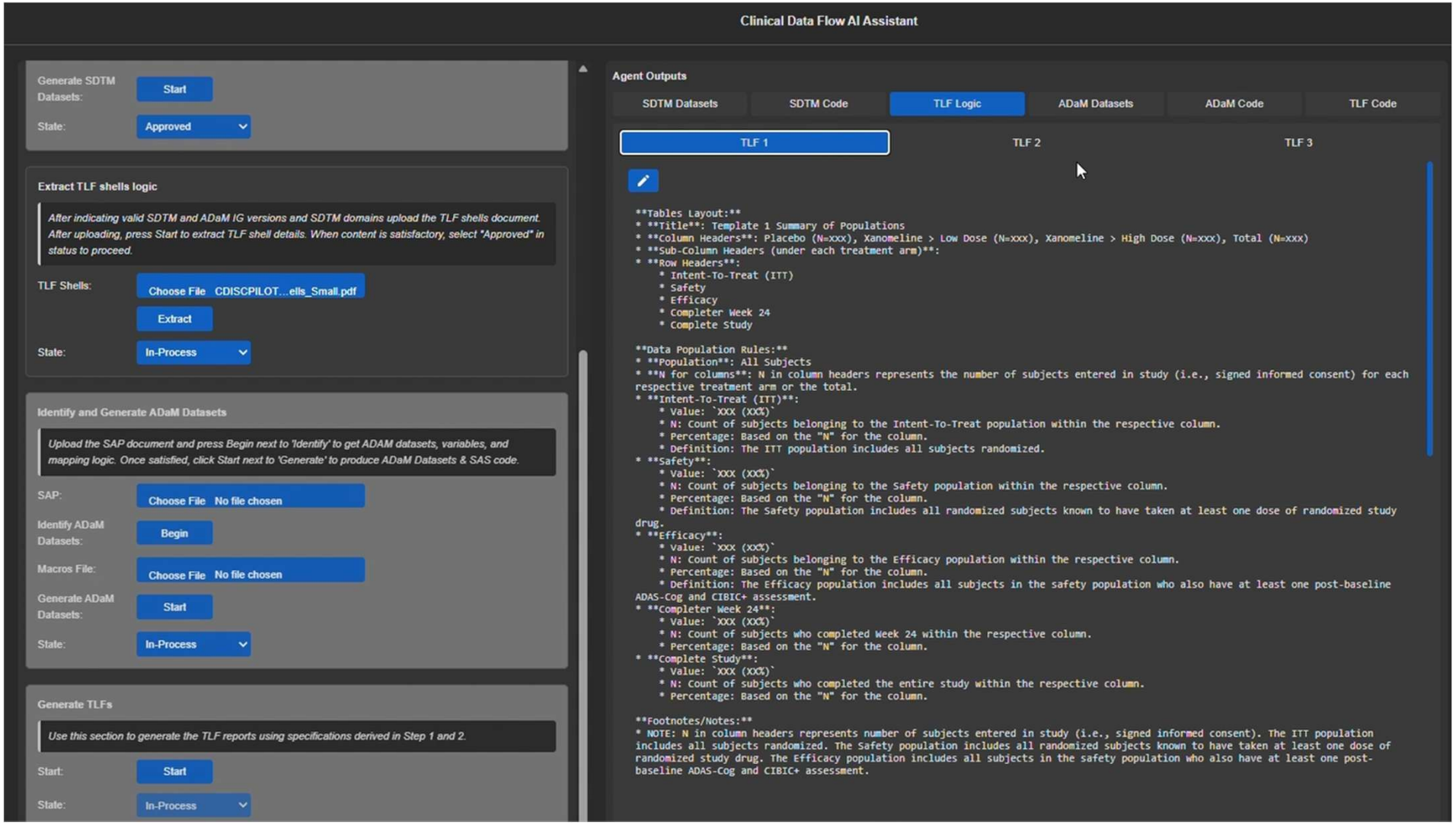
Also summarize: MMSE, Duration of disease (cont. and as <12 months, >=12 months), Years of education, Baseline Weight, Baseline Height, Baseline BMI (cont. and as normal(<25), overweight(25-<30), obese(>=30))
[1] P-values are results of ANOVA treatment group comparisons for continuous variables and Pearson's chi-square test for categorical variables.
NOTE: Duration of disease is computed as months between date of enrollment and date of onset of the first definite symptoms of Alzheimer's disease.

Protocol: CDISCPILOT01 Page 1 of n
Population: All Subjects

Template 1
Summary of Populations

Population	Placebo (N=xxx)	Xanomeline Low Dose (N=xxx)	Xanomeline High Dose (N=xxx)	Total (N=xxx)
Intent-To-Treat (ITT)	xxx (xx%)	xxx (xx%)	xxx (xx%)	xxx (xx%)
Safety	xxx (xx%)	xxx (xx%)	xxx (xx%)	xxx (xx%)
Efficacy	xxx (xx%)	xxx (xx%)	xxx (xx%)	xxx (xx%)
Completer Week 24	xxx (xx%)	xxx (xx%)	xxx (xx%)	xxx (xx%)
Complete Study	xxx (xx%)	xxx (xx%)	xxx (xx%)	xxx (xx%)

NOTE: N in column headers represents number of subjects entered in study (i.e., signed informed consent). The ITT population includes all subjects randomized. The Safety population includes all randomized subjects known to have taken at least one dose of randomized study drug. The Efficacy population includes all subjects in the safety population who also have at least one post-baseline ADAS-Cog and CIBIC+ assessment.



had probable Alzheimer's disease according to the NINCDS-ADRDA criteria, and an MMSE score of 10 to 23. The duration of treatment was 26 weeks, with 24 weeks of active treatment. A total of 295 patients were randomized into 1 of 3 treatment groups: xanomeline high dose, 97 patients; xanomeline low dose, 98 patients; and placebo, 100 patients; 166 were females and 129 were males.

2. PURPOSE OF THIS ANALYSIS PLAN

This analysis plan describes the analyses to be performed in the context of the first iteration of the CDISC SDTM/ADaM Pilot Submission, CDISCPILOT01. It should be noted that this document is not meant to represent all of the measures assessed or analyses performed in the original study.

3. STUDY OBJECTIVE(S) AND ENDPOINT(S)

3.1. Study Objective(s)

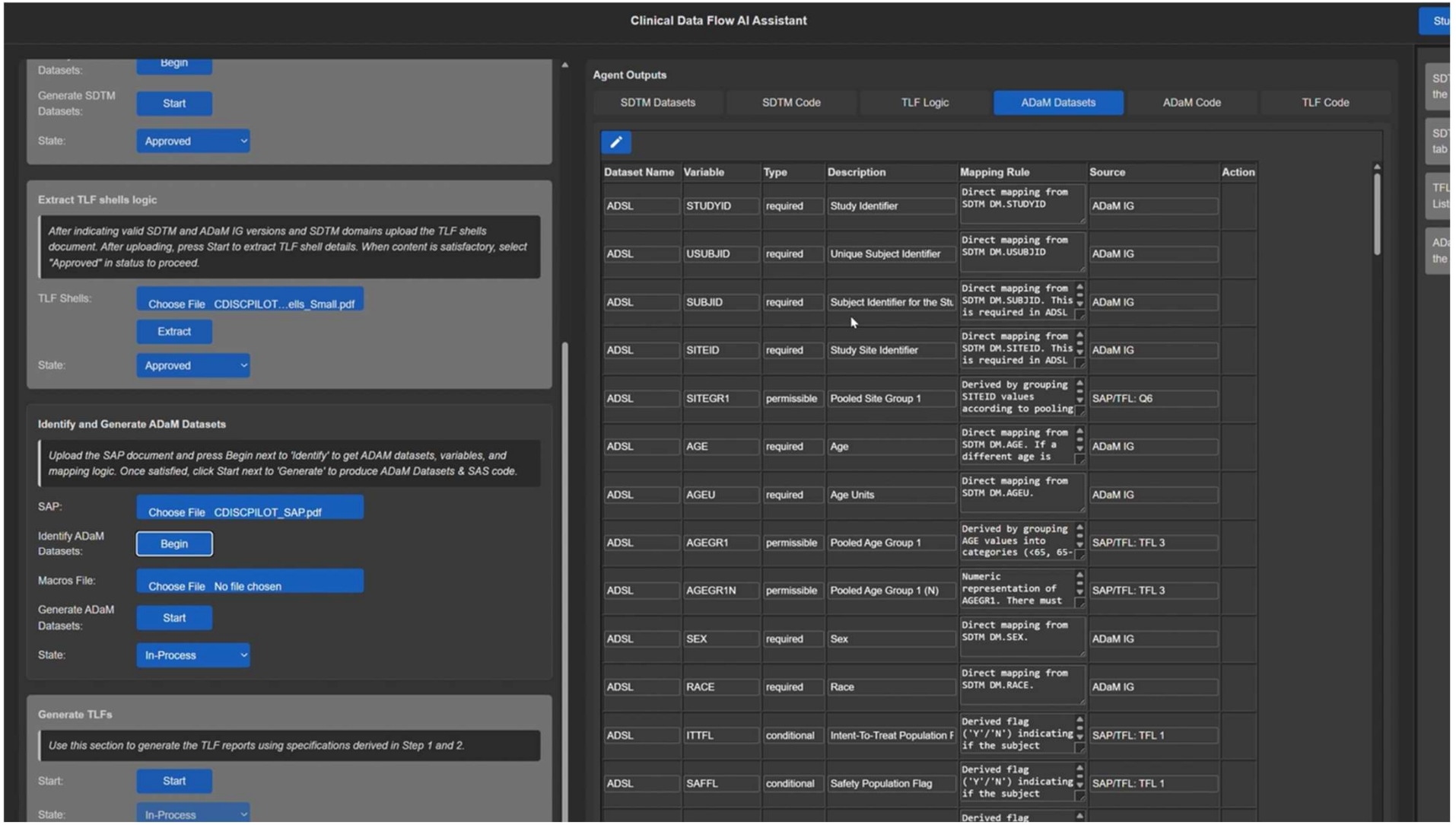
3.1.1. Primary

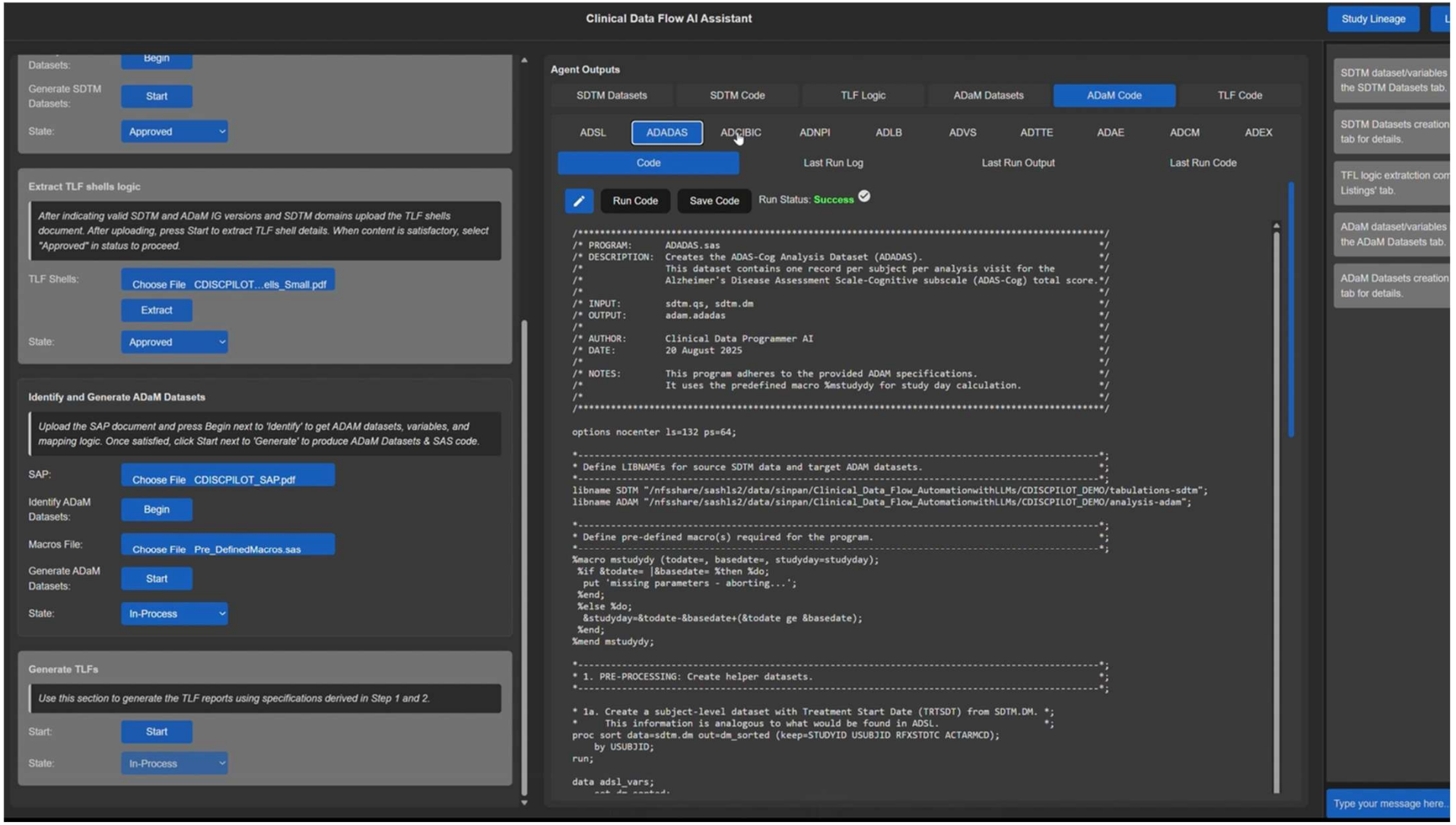
The primary objectives of this study are

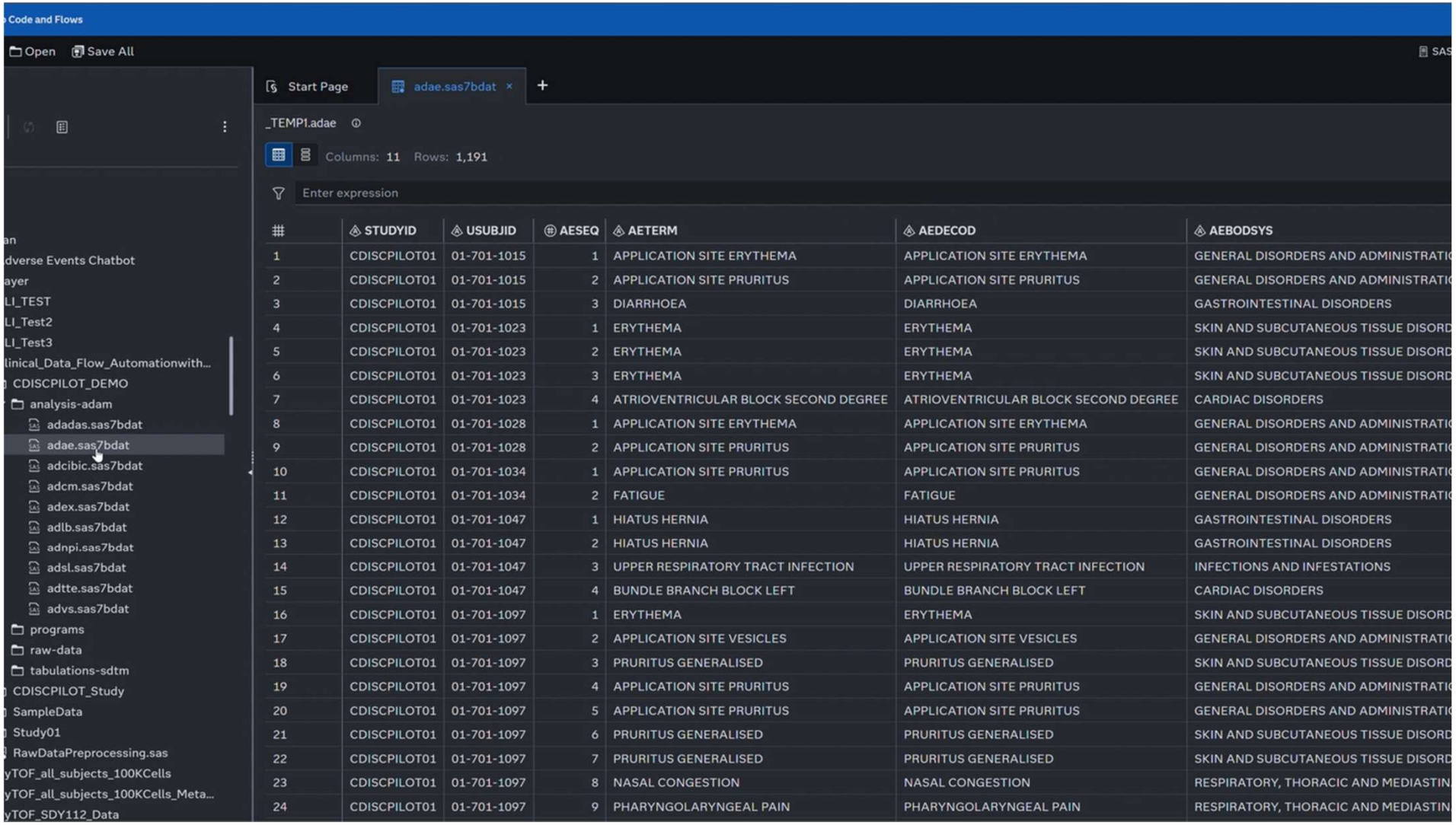
- To determine if there is a statistically significant relationship (overall Type 1 error rate, α=.05) between the change in both the ADAS-Cog (11) and CIBIC+ scores, and drug dose (0, 50 cm^2 [54 mg], and 75 cm^2 [81 mg]).

- To document the safety profile of the xanomeline TTS.

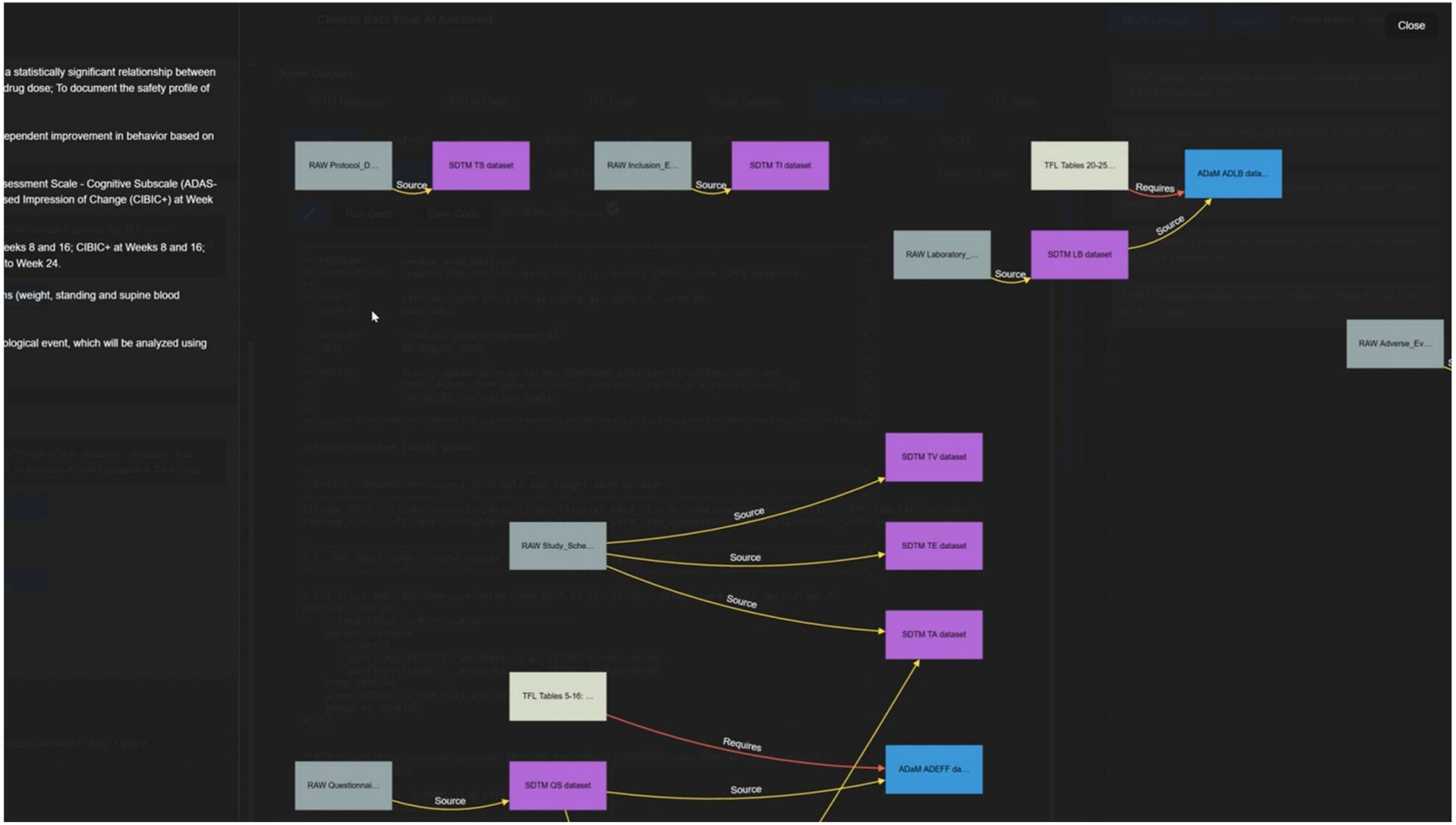
3.1.2. Secondary

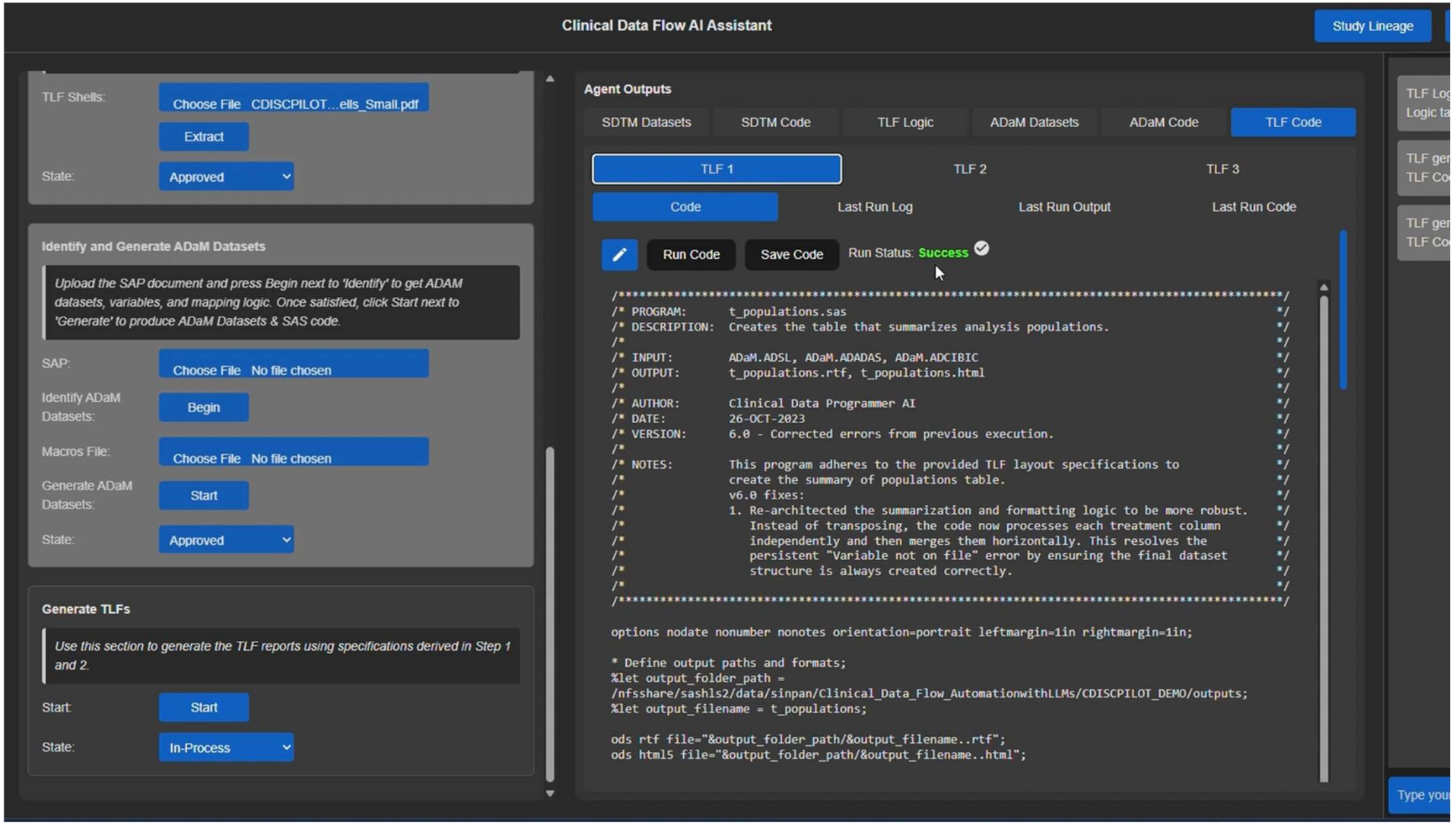


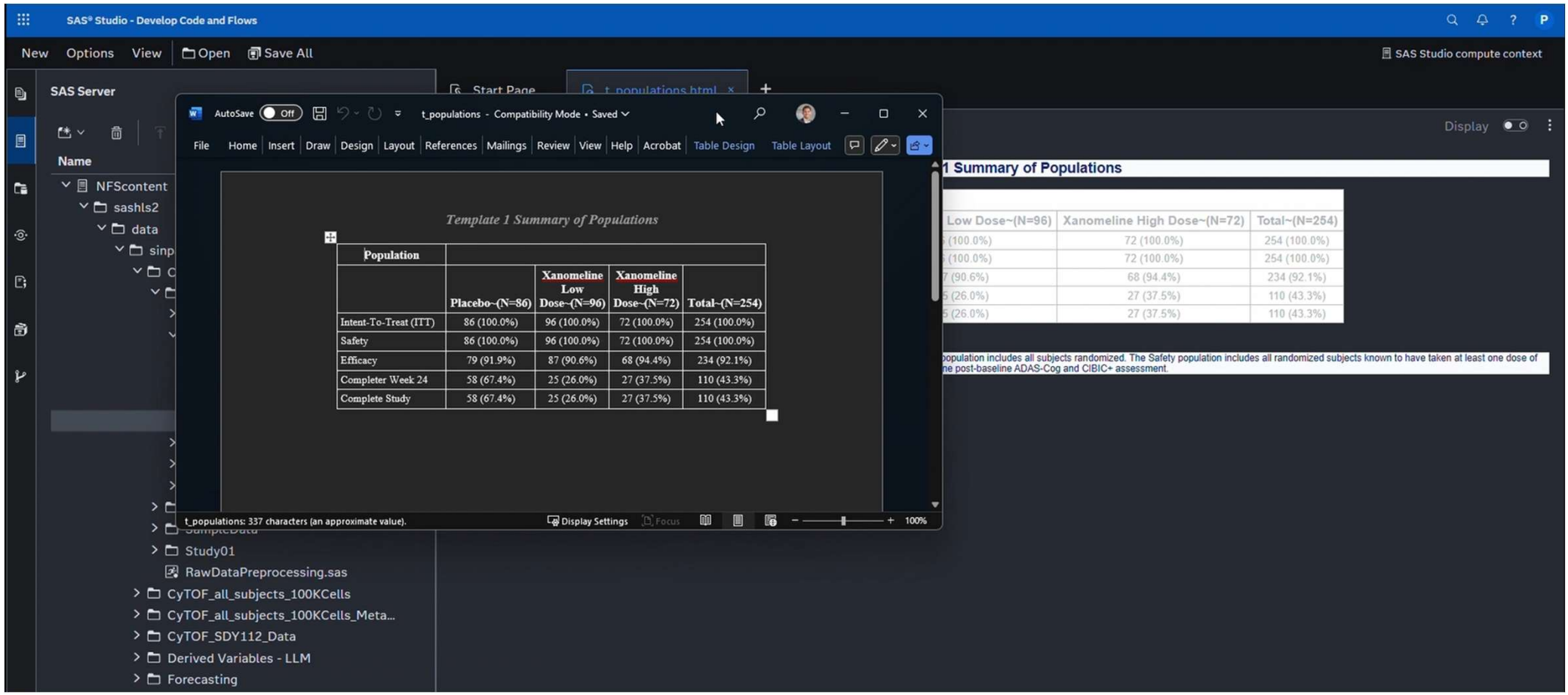


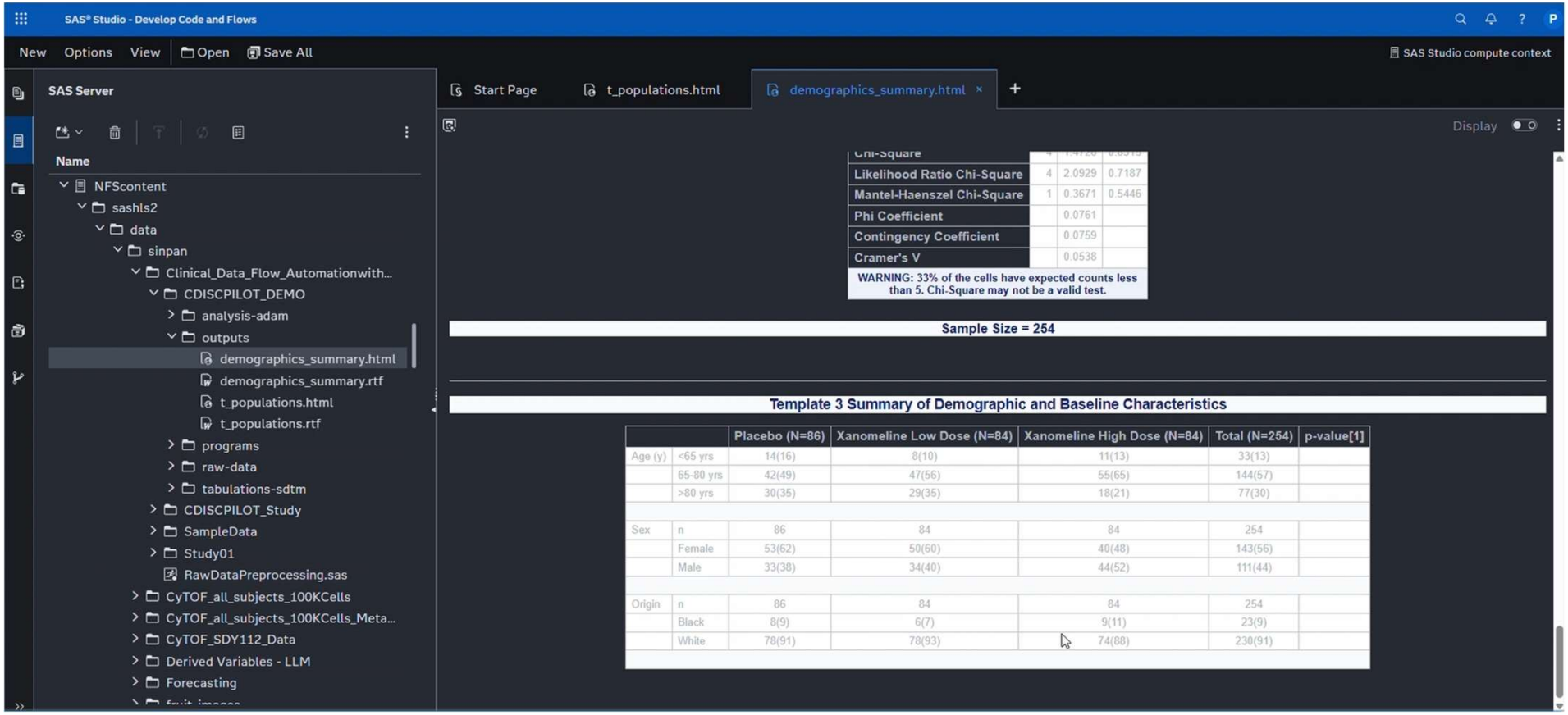


#	STUDYID	USUBJID	AESEQ	AETERM	AEDECOD	AEBODSYS
1	CDISCPILOT01	01-701-1015	1	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	GENERAL DISORDERS AND ADMINISTRATI
2	CDISCPILOT01	01-701-1015	2	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATI
3	CDISCPILOT01	01-701-1015	3	DIARRHOEA	DIARRHOEA	GASTROINTESTINAL DISORDERS
4	CDISCPILOT01	01-701-1023	1	ERYTHEMA	ERYTHEMA	SKIN AND SUBCUTANEOUS TISSUE DISORD
5	CDISCPILOT01	01-701-1023	2	ERYTHEMA	ERYTHEMA	SKIN AND SUBCUTANEOUS TISSUE DISORD
6	CDISCPILOT01	01-701-1023	3	ERYTHEMA	ERYTHEMA	SKIN AND SUBCUTANEOUS TISSUE DISORD
7	CDISCPILOT01	01-701-1023	4	ATRIOVENTRICULAR BLOCK SECOND DEGREE	ATRIOVENTRICULAR BLOCK SECOND DEGREE	CARDIAC DISORDERS
8	CDISCPILOT01	01-701-1028	1	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	GENERAL DISORDERS AND ADMINISTRATI
9	CDISCPILOT01	01-701-1028	2	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATI
10	CDISCPILOT01	01-701-1034	1	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATI
11	CDISCPILOT01	01-701-1034	2	FATIGUE	FATIGUE	GENERAL DISORDERS AND ADMINISTRATI
12	CDISCPILOT01	01-701-1047	1	HIATUS HERNIA	HIATUS HERNIA	GASTROINTESTINAL DISORDERS
13	CDISCPILOT01	01-701-1047	2	HIATUS HERNIA	HIATUS HERNIA	GASTROINTESTINAL DISORDERS
14	CDISCPILOT01	01-701-1047	3	UPPER RESPIRATORY TRACT INFECTION	UPPER RESPIRATORY TRACT INFECTION	INFECTIONS AND INFESTATIONS
15	CDISCPILOT01	01-701-1047	4	BUNDLE BRANCH BLOCK LEFT	BUNDLE BRANCH BLOCK LEFT	CARDIAC DISORDERS
16	CDISCPILOT01	01-701-1097	1	ERYTHEMA	ERYTHEMA	SKIN AND SUBCUTANEOUS TISSUE DISORD
17	CDISCPILOT01	01-701-1097	2	APPLICATION SITE VESICLES	APPLICATION SITE VESICLES	GENERAL DISORDERS AND ADMINISTRATI
18	CDISCPILOT01	01-701-1097	3	PRURITUS GENERALISED	PRURITUS GENERALISED	SKIN AND SUBCUTANEOUS TISSUE DISORD
19	CDISCPILOT01	01-701-1097	4	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATI
20	CDISCPILOT01	01-701-1097	5	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATI
21	CDISCPILOT01	01-701-1097	6	PRURITUS GENERALISED	PRURITUS GENERALISED	SKIN AND SUBCUTANEOUS TISSUE DISORD
22	CDISCPILOT01	01-701-1097	7	PRURITUS GENERALISED	PRURITUS GENERALISED	SKIN AND SUBCUTANEOUS TISSUE DISORD
23	CDISCPILOT01	01-701-1097	8	NASAL CONGESTION	NASAL CONGESTION	RESPIRATORY, THORACIC AND MEDIASTIN
24	CDISCPILOT01	01-701-1097	9	PHARYNGOLARYNGEAL PAIN	PHARYNGOLARYNGEAL PAIN	RESPIRATORY, THORACIC AND MEDIASTIN









AI as a Copilot for Programmers

LLMs enhance rather than replace human programmers, transforming routines into streamlined processes

The new workflow allows AI to draft code, which is then reviewed and validated by experts

Augmented intelligence fosters collaboration and preserves human oversight in programming tasks

Responsible Adoption of GenAI

Mitigate risks associated with LLMs through secure, private models and human oversight

Validation of AI-generated outputs is essential to ensure coding accuracy

Organizations must prepare for responsible, intelligent integration of AI in clinical operations

Thank you!

sas.com