

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

- Tighter regulations
- Intricate data sources

Many processes still rely on manual coding

- Even with CDISC standards

Manual programming can be repetitive, inefficient, and prone to errors

- 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

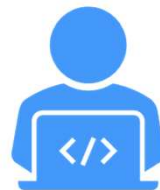


LLMs help bridge the gap between human expertise and machine automation in programming

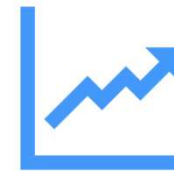
Goals of the Presentation



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

```
%put Project Type: .....&project_type.;
%put Cluster Size: .....
%put Min. Obs per Cluster: .....
%put Pseudo T Squared: .....
%put R-squared: .....

/* Reminder: i = project-t
j = time-series
m = minimum-c
```

Run Selected or All SAS Code	F3
Run All SAS Code	F8
Run Region	
Change All Occurrences	Ctrl+F2
Refactor...	Ctrl+Shift+R
Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
SAS Copilot: Add comments	
SAS Copilot: Explain	
SAS Copilot: Generate code	
Command Palette...	Ctrl+Shift+P

DEBUG CONSOLE TERMINAL

cf_row_2-1_ts_cluster.sas

clt_medic_v2.sas

cf_row_1_data

C: > Users > stsztu > OneDrive - SAS > Documents > git > sas > parallel-proc

Add comments | Explain

```
27 options autosignon=yes
28     →     noconnectwait →
29     →     noconnectpersist →
30     →     sascmd='!sascmd' →
31     →     ;
32
33 %macro parallel(data=, workers=);
34     → %let dsid = %sysfunc(open(&data.));
35     → %let n = %sysfunc(attrn(&dsid., nlobs));
36     → %let rc = %sysfunc(close(&dsid.));
```



```
57 run;
```

```
58  
59 %cleanTempFiles(kill=yes);  
60
```

Add comments | Explain | Generate code

```
61 /* Create a gradient boosting model that uses _CL  
62 ... Add the following features:  
63 ... 1. Partitioning with a 70/30 Train/Validation  
64 ... 2. Autotuning  
65 */
```


Add comments | Explain | Generate code

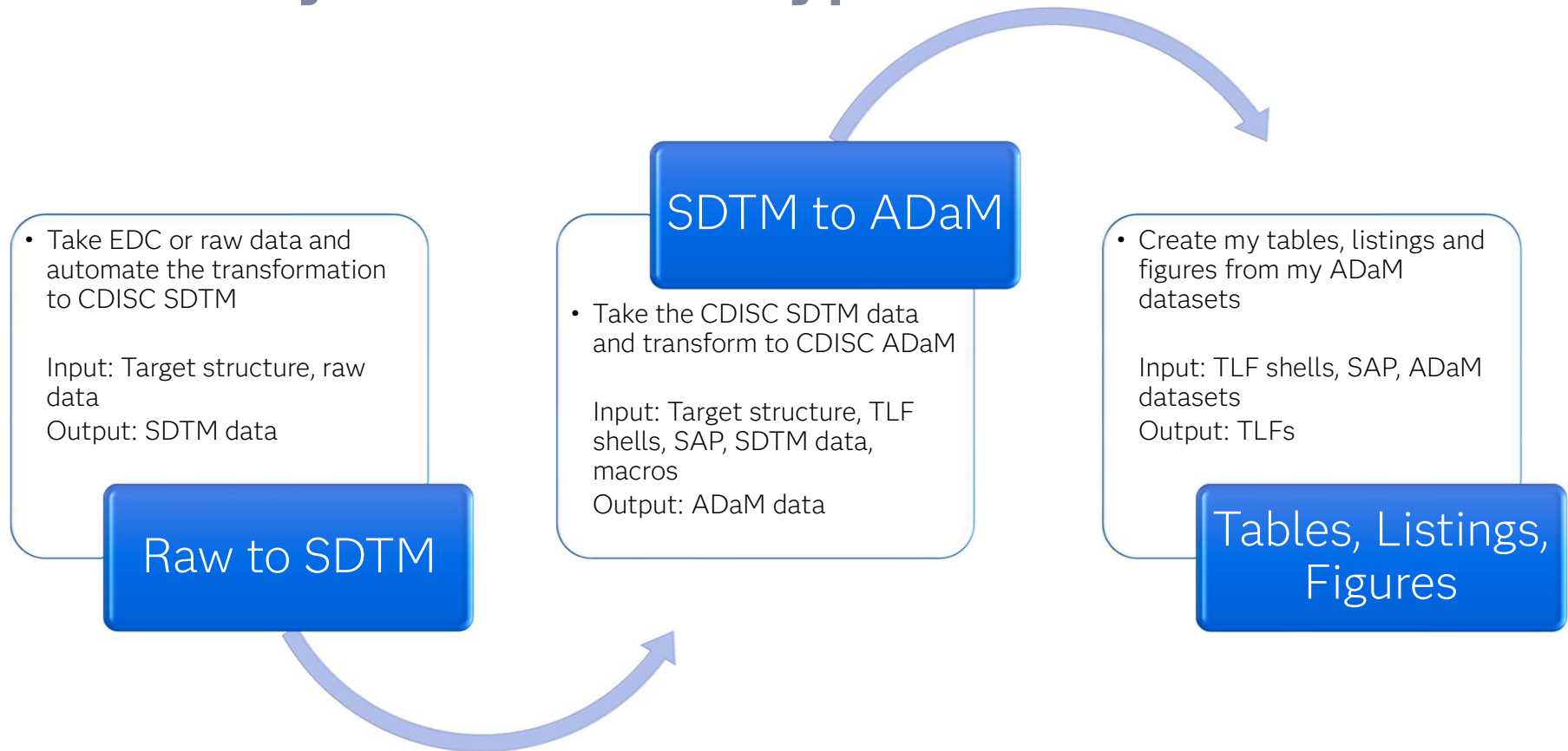
```
51 /* Create a gradient boosting model that uses _CLUSTER_ID_ as the target from whouse.cf_row_c  
52 ... Add the following features:  
53 ... 1. Partitioning with a 70/30 Train/Validation fit  
54 ... 2. Autotuning  
55 */  
56 proc gradboost data=whouse.cf_row_cluster_lookup;  
57   partition fraction(validate=0.3);  
58   target _CLUSTER_ID_ / level=nominal;  
59   autotune tuningparameters=(ntrees samplingrate vars_to_try learningrate lasso ridge);  
60 run;
```

I

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Agentic AI for Clinical Data Flow

Minimally Viable Prototype - Goals



Welcome, sinpan

Logout

Create New Clinical Study Project

CDISC_PILOT_Study_Project

Data Specifications

Base Directory

/nfsshare/sashls2/data/sinpan/Clinical_Data_Flow_At

Raw Data folder name

raw-data

SDTM datasets folder name

tabulations-sdtm

ADaM datasets folder name

analysis-adam

Programs folder name

programs

LLM Specifications

Gemini 2.0 Pro

SAS (Fine Tuned LLM)

gpt-4o

Set Temperature (0-1)

0.1

Start New Project

Retrieve Existing Clinical Study

No existing projects found.

Clinical Data Flow AI Assistant

Clinical Data Flow

General Study Information

SDTM IG version:SDTMIG v3.4

ADaM IG version:ADaMIG v1.3

SDTM Domains for Study:

☐ AE

☐ AG

☐ BE

☐ BS

☐ CE

☐ CM

☐ CO

☐ CP

☐ CV

☐ DA

☐ DD

☐ DM

☐ DS

☐ DV

☐ EC

☐ EG

☐ EX

☐ FA

Convert Raw Data to SDTM standards

Start here as the first step. Press Begin next to 'Identify' to get SDTM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce SDTM Datasets & SAS code.

Identify SDTM Datasets:

Begin

Generate SDTM Datasets:

Start

State:

In-Process

Extract TLF shells logic

After indicating valid SDTM and ADaM IG versions and SDTM domains upload the TLF shells document. After uploading, press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed.

TLF Shells:

Choose FileNo file chosen

Extract

State:

In-Process

Agent Outputs

SDTM Datasets

SDTM Code

TLF Logic

ADaM Datasets

ADaM Code

TLF Code

Dataset Name	Variable	Type	Description	Mapping Rule	Action
No data available					

Copyright © SAS Institute Inc. All rights reserved.

SDTM Domains for Study:

☐ SS

☒ SU

☐ SUPQUAL

☒ SV

☒ TA

☐ TD

☒ TE

☒ TI

☐ TM

☐ TR

☒ TS

☐ TU

☒ TV

☐ UR

☒ VS

Convert Raw Data to SDTM standards

Start here as the first step. Press Begin next to 'Identify' to get SDTM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce SDTM Datasets & SAS code.

Identify SDTM Datasets:

Begin

Generate SDTM Datasets:

Start

State:

In-Process

Extract TLF shells logic

After indicating valid SDTM and ADaM IG versions and SDTM domains upload the TLF shells document. After uploading, press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed.

TLF Shells:

Choose File No file chosen

Extract

State:

In-Process

Identify and Generate ADaM Datasets

Upload the SAP document and press Begin next to 'Identify' to get ADaM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce ADaM Datasets & SAS code.

SAP:

Choose File No file chosen

Clinical Data Flow AI Assistant

Agent Outputs

SDTM Datasets

SDTM Code

TLF Logic

ADaM Datasets

ADaM Code

TLF Code

Dataset Name

Variable

Type

Description

Mapping Rule

Action

AE

STUDYID

Required

Unique identifier for a study

Set to 'CDISCPLOT01', the study identifier for

AE

DOMAIN

Required

Two-character abbreviation

Set to 'AE'.

AE

USUBJID

Required

Identifier used to uniquely k

Map from ADVERSE_EVENTS.UNIQUE_SUBJ_ID.

AE

AESQ

Required

Sequence number given to

Generate a unique sequence number for each adverse event

AE

AESPID

Permissible

Sponsor-defined identifier. I

Map from ADVERSE_EVENTS.AE_ID.

AE

AETERM

Required

Verbatim name of the event

Map from ADVERSE_EVENTS.AE_TERM.

AE

AEDECOD

Required

Dictionary-derived text desc

Map from ADVERSE_EVENTS.AE_DECOD.

AE

AELLT

Expected

Dictionary-derived text desc

Map from ADVERSE_EVENTS.AE_LLT.

AE

AELLTCD

Expected

Dictionary-derived code for

Not available in raw data. A coding dictionary would be

AE

AEPTCD

Expected

Dictionary-derived code for

Not available in raw data. A coding dictionary would be

AE

AEHLT

Expected

Dictionary-derived text desc

Map from ADVERSE_EVENTS.AE_HLT.

AE

AEHLTCD

Expected

Dictionary-derived code for

Not available in raw data. A coding dictionary would be

AE

AEHLGT

Expected

Dictionary-derived text desc

Map from ADVERSE_EVENTS.AE_HLGT.

Not available in raw

Copyright © SAS Institute Inc. All rights reserved.

Clinical Data Flow AI Assistant

SDTM Domains
for Study:

- ☐ SS
 ☒ TE
 ☒ TV
- ☐ SU
 ☒ TI
 ☐ UR
- ☐ SUPPQUAL
 ☐ TM
 ☒ VS
- ☒ SV
 ☐ TR
- ☒ TA
 ☒ TS
- ☐ TD
 ☐ TU

Convert Raw Data to SDTM standards

Start here as the first step. Press Begin next to 'Identify' to get SDTM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce SDTM Datasets & SAS code.

Identify SDTM
Datasets:

Begin

Generate SDTM
Datasets:

Start

State: In-Process

Extract TLF shells logic

After indicating valid SDTM and ADaM IG versions and SDTM domains upload the TLF shells document. After uploading, press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed.

TLF Shells: Choose File No file chosen

Extract

State: In-Process

Identify and Generate ADaM Datasets

Upload the SAP document and press Begin next to 'Identify' to get ADaM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce ADaM Datasets & SAS code.

SAP: Choose File No file chosen

Agent Outputs

SDTM Datasets SDTM Code TLF Logic ADaM Datasets ADaM Code TLF Code

AE CM DM DS EX LB MH QS RELREC SC SE SV TA TI TS TV VS

Code

Last Run Log

Last Run Output

Last Run Code



Run Code

Save Code

Run Status: Success

```

/*=====*/
/* PROGRAM:      create_sdtm_vs.sas
/* DESCRIPTION:   Creates the SDTM Vital Signs (VS) domain from raw data.
/*
/* INPUT:        raw.vital_signs, raw.patient_information
/* OUTPUT:       sdtm.vs
/*
/* AUTHOR:       Clinical Data Programmer AI
/* DATE:         20 August 2025
/*
/* NOTES:        This program adheres to the provided SDTM specifications.
/*                Version 2: Corrected variable renaming in helper dataset and moved LENGTH
/*                statement to prevent warnings and errors.
/*=====*/

options nocenter ls=132 ps=64;

*-----*
* Define LIBNAMES for source Raw data and target SDTM datasets.
*-----*
libname RAW "/nfsshare/sashls2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCIPILOT_DEMO/raw-data";
libname SDTM "/nfsshare/sashls2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCIPILOT_DEMO/tabulations-sdtm";

*-----*
* 1. PRE-PROCESSING: Create a helper dataset with subject-level reference dates.
*-----*
/* Create a dataset with one record per subject containing the treatment start date.
/* This date will be used as the reference for calculating VSDY and VSLOBXFL.
*/
data dm_info;
  /* Use RENAME= dataset option to correctly rename the variable.
  */
  set RAW.PATIENT_INFORMATION(
    keep=UNIQUE_SUBJ_ID FIRST_TREATMENT_DATE
    rename=(UNIQUE_SUBJ_ID = USUBJID)
  );

  /* Create numeric reference date RFXSTDTC from FIRST_TREATMENT_DATE.
  */
  if FIRST_TREATMENT_DATE ne '' then RFXSTDTC = input(FIRST_TREATMENT_DATE, yymmdd10.);
  format RFXSTDTC yymmdd10.;

  keep USUBJID RFXSTDTC;
run;

/* Ensure one record per subject.
proc sort data=dm_info nodupkey;
  by USUBJID;
run;

```

Develop Code and Flows

View | Open | Save All

Content

shls2

data

sinpan

Clinical_Data_Flow_Automationwith...

CDISCILOT_DEMO

raw-data

tabulations-sdtm

ae.sas7bdat

cm.sas7bdat

dm.sas7bdat

ds.sas7bdat

ex.sas7bdat

lb.sas7bdat

mh.sas7bdat

qs.sas7bdat

relrec.sas7bdat

sc.sas7bdat

se.sas7bdat

sv.sas7bdat

ta.sas7bdat

ti.sas7bdat

ts.sas7bdat

tv.sas7bdat

vs.sas7bdat

CDISCILOT_Study

SampleData

Study01

RawDataPreprocessing.sas

CyTOF_all_subjects_100KCells

CyTOF_all_subjects_100KCells_Meta...

CyTOF_SDX112_Data

Start Page | ae.sas7bdat

_TEMP1.ae

Columns: 33 Rows: 1,191

Enter expression

#	STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AEDECOD	AELLT
1	CDISCILOT01	AE	01-701-1015	1	E07	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	APPLICAT
2	CDISCILOT01	AE	01-701-1015	2	E08	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
3	CDISCILOT01	AE	01-701-1015	3	E06	DIARRHOEA	DIARRHOEA	DIARRHEA
4	CDISCILOT01	AE	01-701-1023	1	E08	ERYTHEMA	ERYTHEMA	ERYTHEMA
5	CDISCILOT01	AE	01-701-1023	2	E09	ERYTHEMA	ERYTHEMA	LOCALIZE
6	CDISCILOT01	AE	01-701-1023	3	E08	ERYTHEMA	ERYTHEMA	ERYTHEMA
7	CDISCILOT01	AE	01-701-1023	4	E10	ATRIOVENTRICULAR BLOCK SECOND DEGREE	ATRIOVENTRICULAR BLOCK SECOND DEGREE	AV BLOCK
8	CDISCILOT01	AE	01-701-1028	1	E04	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	APPLICAT
9	CDISCILOT01	AE	01-701-1028	2	E05	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
10	CDISCILOT01	AE	01-701-1034	1	E08	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
11	CDISCILOT01	AE	01-701-1034	2	E07	FATIGUE	FATIGUE	FATIGUE
12	CDISCILOT01	AE	01-701-1047	1	E06	HIATUS HERNIA	HIATUS HERNIA	HERNIA H
13	CDISCILOT01	AE	01-701-1047	2	E06	HIATUS HERNIA	HIATUS HERNIA	HERNIA H
14	CDISCILOT01	AE	01-701-1047	3	E08	UPPER RESPIRATORY TRACT INFECTION	UPPER RESPIRATORY TRACT INFECTION	UPPER RE
15	CDISCILOT01	AE	01-701-1047	4	E09	BUNDLE BRANCH BLOCK LEFT	BUNDLE BRANCH BLOCK LEFT	LEFT BUN
16	CDISCILOT01	AE	01-701-1097	1	E04	ERYTHEMA	ERYTHEMA	ERYTHEMA
17	CDISCILOT01	AE	01-701-1097	2	E07	APPLICATION SITE VESICLES	APPLICATION SITE VESICLES	APPLICAT
18	CDISCILOT01	AE	01-701-1097	3	E05	PRURITUS GENERALISED	PRURITUS GENERALISED	GENERAL
19	CDISCILOT01	AE	01-701-1097	4	E06	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
20	CDISCILOT01	AE	01-701-1097	5	E06	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	APPLICAT
21	CDISCILOT01	AE	01-701-1097	6	E08	PRURITUS GENERALISED	PRURITUS GENERALISED	GENERAL
22	CDISCILOT01	AE	01-701-1097	7	E12	PRURITUS GENERALISED	PRURITUS GENERALISED	GENERAL
23	CDISCILOT01	AE	01-701-1097	8	E11	NASAL CONGESTION	NASAL CONGESTION	NASAL CO
24	CDISCILOT01	AE	01-701-1097	9	E10	PHARYNGOLARYNGEAL PAIN	PHARYNGOLARYNGEAL PAIN	SORE TH



Protocol: CDISCPIL01
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: CDISCPIL01
Population: All Subjects

Template 1
Summary of Populations

	Placebo (N=xxx)	Xanomeline Low Dose (N=xxx)	Xanomeline High Dose (N=xxx)	Total (N=xxx)
Population				
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.



Clinical Data Flow AI Assistant

Generate SDTM Datasets: Start

State: Approved

Extract TLF shells logic

After indicating valid SDTM and ADaM IG versions and SDTM domains upload the TLF shells document. After uploading, press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed.

TLF Shells: Choose File CDISCPLOT...elis_Small.pdf

Extract

State: In-Process

Identify and Generate ADaM Datasets

Upload the SAP document and press Begin next to 'Identify' to get ADaM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce ADaM Datasets & SAS code.

SAP: Choose File No file chosen

Identify ADaM Datasets: Begin

Macros File: Choose File No file chosen

Generate ADaM Datasets: Start

State: In-Process

Generate TLFs

Use this section to generate the TLF reports using specifications derived in Step 1 and 2.

Start: Start

State: In-Process

Agent Outputs

SDTM Datasets SDTM Code **TLF Logic** ADaM Datasets ADaM Code TLF Code

TLF 1

TLF 2

TLF 3



```

**Tables Layout:**
* **Title**: Template 1 Summary of Populations
* **column Headers**: Placebo (N=xxx), Xanomeline > Low Dose (N=xxx), Xanomeline > High Dose (N=xxx), Total (N=xxx)
* **Sub-column Headers (under each treatment arm)**:
* **Row Headers**:
  * Intent-To-Treat (ITT)
  * Safety
  * Efficacy
  * Completer Week 24
  * Complete Study

**Data Population Rules:**
* **Population**: All Subjects
* **N for columns**: N in column headers represents the number of subjects entered in study (i.e., signed informed consent) for each respective treatment arm or the total.
* **Intent-To-Treat (ITT)**:
  * Value: 'xxx (xxx)'
  * N: Count of subjects belonging to the Intent-To-Treat population within the respective column.
  * Percentage: Based on the "N" for the column.
  * Definition: The ITT population includes all subjects randomized.
* **Safety**:
  * Value: 'xxx (xxx)'
  * N: Count of subjects belonging to the Safety population within the respective column.
  * Percentage: Based on the "N" for the column.
  * Definition: The Safety population includes all randomized subjects known to have taken at least one dose of randomized study drug.
* **Efficacy**:
  * Value: 'xxx (xxx)'
  * N: Count of subjects belonging to the Efficacy population within the respective column.
  * Percentage: Based on the "N" for the column.
  * Definition: The Efficacy population includes all subjects in the safety population who also have at least one post-baseline ADAS-Cog and CIBIC+ assessment.
* **Completer Week 24**:
  * Value: 'xxx (xxx)'
  * N: Count of subjects who completed Week 24 within the respective column.
  * Percentage: Based on the "N" for the column.
* **Complete Study**:
  * Value: 'xxx (xxx)'
  * N: Count of subjects who completed the entire study within the respective column.
  * Percentage: Based on the "N" for the column.

**Footnotes/Notes:**
* 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.
  
```

patients were males or females or non-smoking potential, 55 years of age or older, 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, CDISCPIL01. 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, $\alpha=.05$) between the change in both the ADAS-Cog (11) and CIBIC+ scores, and drug dose (0, 50 cm² [54 mg], and 75 cm² [81 mg]).
- To document the safety profile of the xanomeline TTS.

3.1.2. Secondary

Datasets:

Begin

Generate SDTM Datasets:

Start

State:

Approved

Extract TLF shells logic

After indicating valid SDTM and ADaM IG versions and SDTM domains upload the TLF shells document. After uploading, press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed.

TLF Shells:

Choose File

CDISCILOT...els_Small.pdf

Extract

State:

Approved

Identify and Generate ADaM Datasets

Upload the SAP document and press Begin next to 'Identify' to get ADAM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce ADaM Datasets & SAS code.

SAP:

Choose File

CDISCILOT_SAP.pdf

Identify ADaM Datasets:

Begin

Macros File:

Choose File

No file chosen

Generate ADaM Datasets:

Start

State:

In-Process

Generate TLFs

Use this section to generate the TLF reports using specifications derived in Step 1 and 2.

Start:

Start

State:

In-Process

Clinical Data Flow AI Assistant

Agent Outputs

SDTM DatasetsSDTM CodeTLF LogicADaM DatasetsADaM CodeTLF Code

Dataset Name	Variable	Type	Description	Mapping Rule	Source	Action
ADSL	STUDYID	required	Study Identifier	Direct mapping from SDTM DM.STUDYID	ADaM IG	
ADSL	USUBJID	required	Unique Subject Identifier	Direct mapping from SDTM DM.USUBJID	ADaM IG	
ADSL	SUBJID	required	Subject Identifier for the St	Direct mapping from SDTM DM.SUBJID. This is required in ADSL	ADaM IG	
ADSL	SITEID	required	Study Site Identifier	Direct mapping from SDTM DM.SITEID. This is required in ADSL	ADaM IG	
ADSL	SITEGR1	permissible	Pooled Site Group 1	Derived by grouping SITEID values according to pooling	SAP/TFL: Q6	
ADSL	AGE	required	Age	Direct mapping from SDTM DM.AGE. If a different age is	ADaM IG	
ADSL	AGEU	required	Age Units	Direct mapping from SDTM DM.AGEU.	ADaM IG	
ADSL	AGEGR1	permissible	Pooled Age Group 1	Derived by grouping AGE values into categories (<65, 65-	SAP/TFL: TFL 3	
ADSL	AGEGR1N	permissible	Pooled Age Group 1 (N)	Numeric representation of AGEGR1. There must	SAP/TFL: TFL 3	
ADSL	SEX	required	Sex	Direct mapping from SDTM DM.SEX.	ADaM IG	
ADSL	RACE	required	Race	Direct mapping from SDTM DM.RACE.	ADaM IG	
ADSL	ITTFL	conditional	Intent-To-Treat Population F	Derived flag ('Y'/'N') indicating if the subject	SAP/TFL: TFL 1	
ADSL	SAFFL	conditional	Safety Population Flag	Derived flag ('Y'/'N') indicating if the subject	SAP/TFL: TFL 1	
				Derived flag		

Datasets:

Begin

Generate SDTM Datasets:

Start

State:

Approved

Extract TLF shells logic

After indicating valid SDTM and ADaM IG versions and SDTM domains upload the TLF shells document. After uploading, press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed.

TFL Shells:

Choose File

CDISCIPILOT...elis_Small.pdf

Extract

State:

Approved

Identify and Generate ADaM Datasets

Upload the SAP document and press Begin next to 'Identify' to get ADAM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce ADaM Datasets & SAS code.

SAP:

Choose File

CDISCIPILOT_SAP.pdf

Identify ADaM Datasets:

Begin

Macros File:

Choose File

Pre_DefinedMacros.sas

Generate ADaM Datasets:

Start

State:

In-Process

Generate TLFs

Use this section to generate the TLF reports using specifications derived in Step 1 and 2.

Start:

Start

State:

In-Process

Clinical Data Flow AI Assistant

Study Lineage

Agent Outputs

SDTM Datasets

SDTM Code

TLF Logic

ADaM Datasets

ADaM Code

TLF Code

ADSL

ADADAS

ADCBIC

ADNPI

ADLB

ADVS

ADTTE

ADAE

ADCM

ADEX

Code

Last Run Log

Last Run Output

Last Run Code

Run Code

Save Code

Run Status: Success

```
/* *****
/* PROGRAM:      ADADAS.sas
/* DESCRIPTION:  Creates the ADAS-Cog Analysis Dataset (ADADAS).
/*              This dataset contains one record per subject per analysis visit for the
/*              Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) total score.
/* *****
/* INPUT:        sdtm.qs, sdtm.dm
/* OUTPUT:       adam.adadas
/*
/* AUTHOR:       Clinical Data Programmer AI
/* DATE:         20 August 2025
/*
/* NOTES:        This program adheres to the provided ADAM specifications.
/*              It uses the predefined macro %mstudydy for study day calculation.
/* *****

options nocenter ls=132 ps=64;

*-----*
* Define LIBNAMES for source SDTM data and target ADAM datasets.
*-----*
libname SDTM "/nfsshare/sashls2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCIPILOT_DEMO/tabulations-sdtm";
libname ADAM "/nfsshare/sashls2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCIPILOT_DEMO/analysis-adam";

*-----*
* Define pre-defined macro(s) required for the program.
*-----*
%macro mstudydy (todate=, basedate=, studyday=studyday);
  %if &todate= |&basedate= %then %do;
    put 'missing parameters - aborting...';
  %end;
  %else %do;
    &studyday=&todate-&basedate+(&todate ge &basedate);
  %end;
%mend mstudydy;

*-----*
* 1. PRE-PROCESSING: Create helper datasets.
*-----*

* 1a. Create a subject-level dataset with Treatment Start Date (TRTSDT) from SDTM.DM. *;
* This information is analogous to what would be found in ADSL.
proc sort data=sdtm.dm out=dm_sorted (keep=STUDYID USUBJID RFXSTDTC ACTARMCD);
  by USUBJID;
run;

data adsl_vars;
  set dm_sorted;
```

SDTM dataset/variables the SDTM Datasets tab.

SDTM Datasets creation tab for details.

TFL logic extratction con Listings' tab.

ADaM dataset/variables the ADaM Datasets tab.

ADaM Datasets creation tab for details.

Type your message here.

Copyright © SAS Institute Inc. All rights reserved.

Code and Flows

OpenSave AllSAS

Start Pageadae.sas7bdat

_TEMP1.adae

Columns: 11Rows: 1,191

Enter expression

#	STUDYID	USUBJID	AESEQ	AETERM	AEDECOD	AEBODSYS
1	CDISCPILOT01	01-701-1015	1	APPLICATION SITE ERYTHEMA	APPLICATION SITE ERYTHEMA	GENERAL DISORDERS AND ADMINISTRATIO
2	CDISCPILOT01	01-701-1015	2	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATIO
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 ADMINISTRATIO
9	CDISCPILOT01	01-701-1028	2	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATIO
10	CDISCPILOT01	01-701-1034	1	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATIO
11	CDISCPILOT01	01-701-1034	2	FATIGUE	FATIGUE	GENERAL DISORDERS AND ADMINISTRATIO
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 ADMINISTRATIO
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 ADMINISTRATIO
20	CDISCPILOT01	01-701-1097	5	APPLICATION SITE PRURITUS	APPLICATION SITE PRURITUS	GENERAL DISORDERS AND ADMINISTRATIO
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

an

verse Events Chatbot

ayer

LI_TEST

LI_Test2

LI_Test3

linical_Data_Flow_Automationwith...

CDISCPILOT_DEMO

analysis-adam

adadas.sas7bdat

adae.sas7bdat

adcibic.sas7bdat

adcm.sas7bdat

adex.sas7bdat

adlb.sas7bdat

adnpi.sas7bdat

adsl.sas7bdat

adtte.sas7bdat

adv.sas7bdat

programs

raw-data

tabulations-sdtm

CDISCPILOT_Study

SampleData

Study01

RawDataPreprocessing.sas

yTOF_all_subjects_100KCells

yTOF_all_subjects_100KCells_Meta...

yTOF_SDY112_Data



biological event, which will be analyzed using

Source: derived from Survey 1 and 2.



Clinical Data Flow AI Assistant

Study Lineage

TLF Shells: Choose File CDISCPLOT...ells_Small.pdf

Extract

State: Approved

Identify and Generate ADaM Datasets

Upload the SAP document and press Begin next to 'Identify' to get ADaM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate' to produce ADaM Datasets & SAS code.

SAP: Choose File No file chosen

Identify ADaM Datasets: Begin

Macros File: Choose File No file chosen

Generate ADaM Datasets: Start

State: Approved

Generate TLFs

Use this section to generate the TLF reports using specifications derived in Step 1 and 2.

Start: Start

State: In-Process

Agent Outputs

SDTM Datasets

SDTM Code

TLF Logic

ADaM Datasets

ADaM Code

TLF Code

TLF 1

TLF 2

TLF 3

Code

Last Run Log

Last Run Output

Last Run Code



Run Code

Save Code

Run Status: Success

```

/*****
/* PROGRAM:      t_populations.sas
/* DESCRIPTION:  Creates the table that summarizes analysis populations.
/*
/* INPUT:        ADaM.ADSL, ADaM.ADADAS, ADaM.ADCIBIC
/* OUTPUT:       t_populations.rtf, t_populations.html
/*
/* AUTHOR:       Clinical Data Programmer AI
/* DATE:         26-OCT-2023
/* VERSION:      6.0 - Corrected errors from previous execution.
/*
/* NOTES:        This program adheres to the provided TLF layout specifications to
/*                create the summary of populations table.
/*                v6.0 fixes:
/*                1. Re-architected the summarization and formatting logic to be more robust.
/*                   Instead of transposing, the code now processes each treatment column
/*                   independently and then merges them horizontally. This resolves the
/*                   persistent "Variable not on file" error by ensuring the final dataset
/*                   structure is always created correctly.
/*
*****/

```

```

options nodate nonumber nonotes orientation=portrait leftmargin=1in rightmargin=1in;

* Define output paths and formats;
%let output_folder_path =
/nfsshare/sashls2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCPLOT_DEMO/outputs;
%let output_filename = t_populations;

ods rtf file="%output_folder_path/&output_filename..rtf";
ods html5 file="%output_folder_path/&output_filename..html";

```

TLF Log
Logic ta

TLF gen
TLF Co

TLF gen
TLF Co

Type you

SAS® Studio - Develop Code and Flows

NewOptionsViewOpenSave All

SAS Server

Name

NFScontent

sashls2

data

sinpan

Clinical_Data_Flow_Automationwith...

CDISCILOT_DEMO

analysis-adam

outputs

demographics_summary.html

demographics_summary.rtf

t_populations.html

t_populations.rtf

programs

raw-data

tabulations-sdtm

CDISCILOT_Study

SampleData

Study01

RawDataPreprocessing.sas

CyTOF_all_subjects_100KCells

CyTOF_all_subjects_100KCells_Meta...

CyTOF_SDY112_Data

Derived Variables - LLM

Forecasting

Exit images

Start Page

t_populations.html

demographics_summary.html

+

Display

☐

⋮

Chi-Square			
Likelihood Ratio Chi-Square	4	2.0929	0.7187
Mantel-Haenszel Chi-Square	1	0.3671	0.5446
Phi Coefficient		0.0761	
Contingency Coefficient		0.0759	
Cramer's V		0.0538	

WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 254

Template 3 Summary of Demographic and Baseline Characteristics

		Placebo (N=86)	Xanomeline Low Dose (N=84)	Xanomeline High Dose (N=84)	Total (N=254)	p-value[1]
Age (y)	<65 yrs	14(16)	8(10)	11(13)	33(13)	
	65-80 yrs	42(49)	47(56)	55(65)	144(57)	
	>80 yrs	30(35)	29(35)	18(21)	77(30)	
Sex	n	86	84	84	254	
	Female	53(62)	50(60)	40(48)	143(56)	
	Male	33(38)	34(40)	44(52)	111(44)	
Origin	n	86	84	84	254	
	Black	8(9)	6(7)	9(11)	23(9)	
	White	78(91)	78(93)	74(88)	230(91)	

Copyright © SAS Institute Inc. All rights reserved.

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



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.