Automating Clinical Study Report Development Using Generative AI

Presented by:

Aman Sandal

Niharikka Tyagi

About the Authors

Aman Sandal

Junior Full Stack Engineer
at Sycamore Informatics

Niharikka Tyagi

Software Developer
at Sycamore Informatics

AGENDA

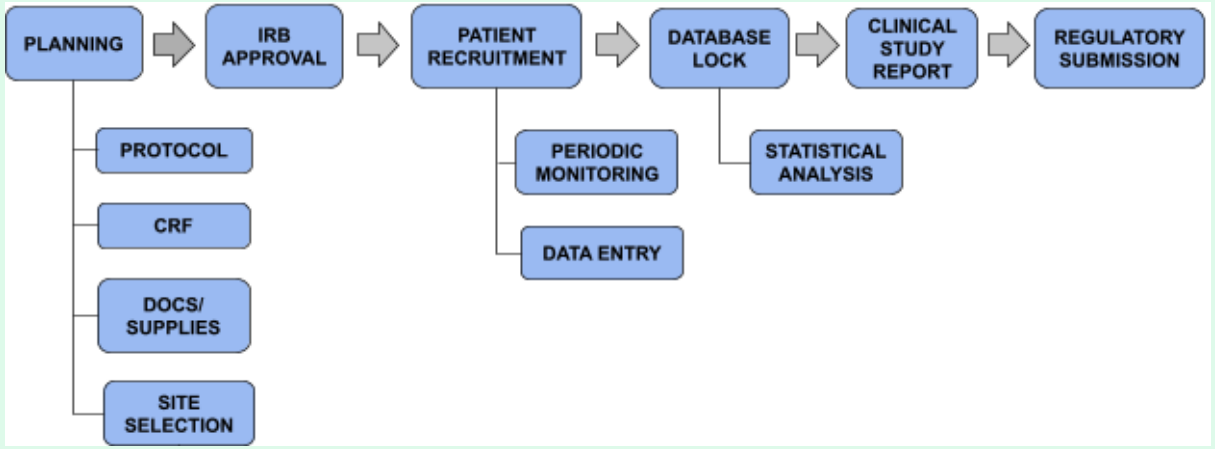
- Intro to Clinical Study Reports (CSRs)
- CSR Authoring Steps
- Challenges with Manual CSR Authoring
- Sycamore CSR Generator
- How the Tool works?
- Workflows
- Key Features of the Tool
- Conclusion
- Q&A

Intro to CSRs

- CSR is a comprehensive document detailing a clinical trial's **objectives, methods, results, and conclusions**
- Demonstrates a drug's **safety and efficacy** through structured evidence
- Includes **tables, listings, and figures** to support statistical findings
- Essential for **regulatory submissions** and scientific transparency
- Prior to the clinical trial, a **Protocol Study Document** is created to define the planned **objectives, study design, and methodologies**. This protocol serves as the foundation upon which the CSR is drafted

CSR Authoring Steps

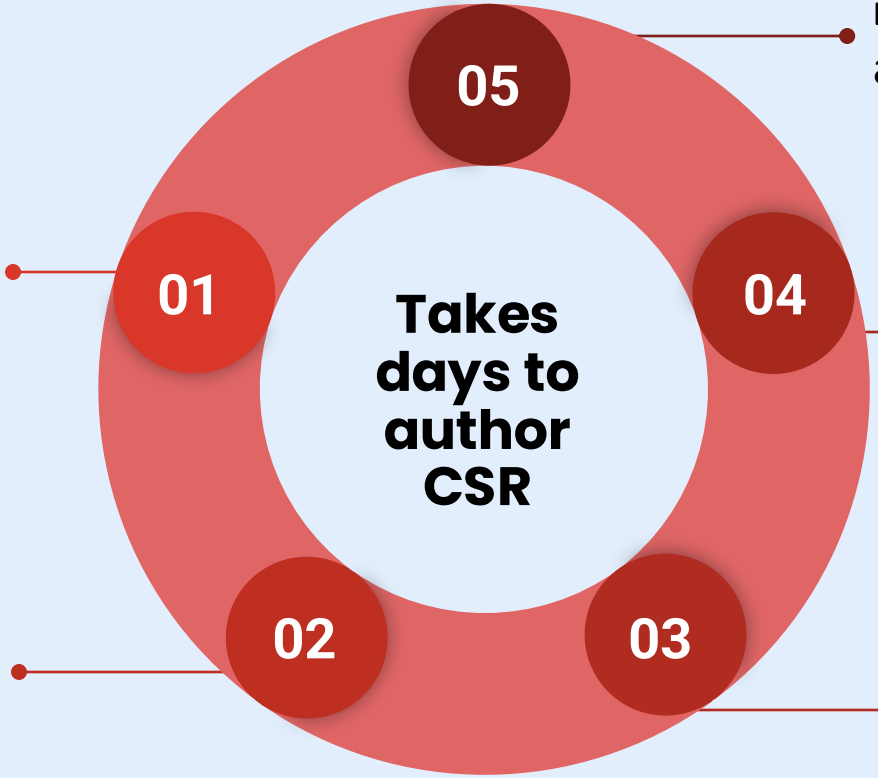
1. **Reusing content** from the Protocol Study Document
2. **Inserting TFLs** (Tables, Figures, Listings) generated by Biostatisticians, including multi-page tables
3. **Creating In-text summary tables** within the main CSR text
4. **Writing descriptive summaries** for in-text tables
5. **Integrating content** from the Statistical Analysis Plan (SAP)*



Challenges with Manual CSR Authoring

Incorrect summary due to **human interpretation of data** leads to additional cycle time and rework

Time Consuming process of manual **extraction, structuring and rewriting of content** using protocol documents and clinical trial data



Manual and error-prone calculations of In-text Summary Tables result in delays

Increased time for CSR review due to **inconsistent language** of protocol content **(future)** to CSR content **(past)** across studies even within same project

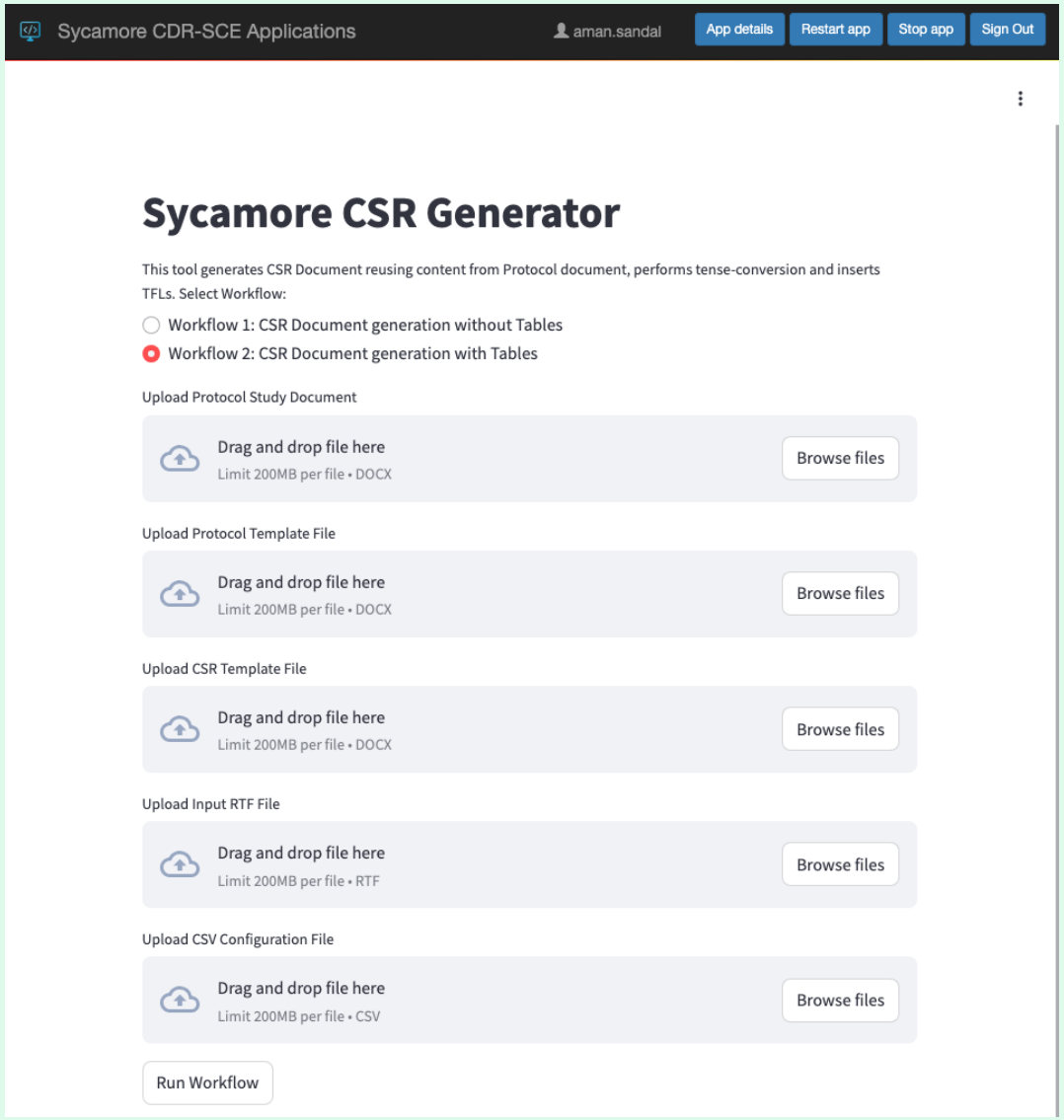
Inefficient use of authors time with laborious efforts to include Biostats TFLs (Mock, Dummy, Final) / **Multi-Page Tables in Section 14**

Sycamore CSR Generator

A Gen AI solution to reuse clinical content from Protocol, include and summarize the analysis results for the generation of a CSR.

Steps:

1. Prepare Templates
2. Upload Files
3. Tense Conversion using Gen AI
4. Multi-page Tables Placement in Section 14
5. In-text Summary Tables Generation (CSV config file)
6. Description Generation for In-text Summary Tables

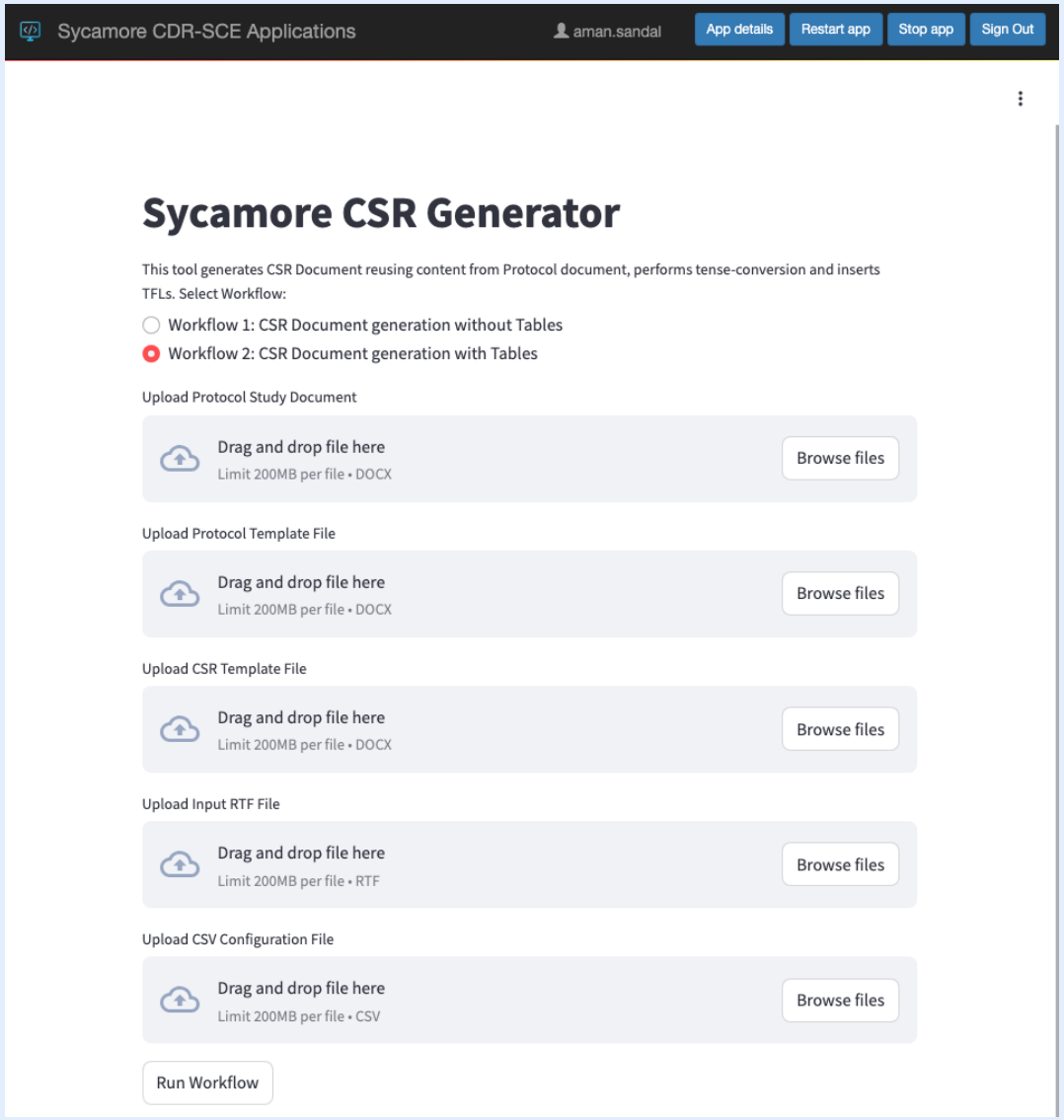


How the Tool works?

Step 1: Prepare Templates

Placeholder Mapping:
Place <sync_xyz> placeholders in Protocol and CSR Templates

Tense-conversion setup:
Use {{past:<sync_xyz>}} syntax in CSR template for tense-conversion



Protocol Template

2.2.2 Secondary Endpoints

Secondary Endpoint	Summary Measure
<sync_secondary_endpoint>	Number and percentage of subjects overall, by severity, by relatedness, by seriousness, and by action (ie, IP discontinuation)

2.3 Exploratory Objectives and Endpoints

2.3.1 Exploratory Objectives

Instructional Text (exactly same text present in both protocol template and protocol study document) gets omitted automatically from the generated CSR. The exploratory objectives of this study are:

<sync_exploratory_objective>

2.3.2 Exploratory Endpoints

The exploratory endpoints for this study are as follows:

<sync_exploratory_endpoint>

3 Study Design and Oversight

3.1 Overall Design

<sync_overall_design>

3.1.1 Screening Period

<sync_screening_period>

3.1.2 Lead-in Period

<sync_leadin_period>

CSR Template

8.2.2 Secondary Endpoints

{{past:<sync_secondary_endpoint>}}

8.3 Exploratory Objectives and Endpoints

8.3.1 Exploratory Objectives

{{past:<sync_exploratory_objective>}}

8.3.2 Exploratory Endpoints

{{past:<sync_exploratory_endpoint>}}

9 Investigational Plan

9.1 Overall Study Design and Plan: Description

{{past:<sync_overall_design>}}

9.1.1 Screening Period

{{past:<sync_screening_period>}}

9.1.2 Lead-in Period

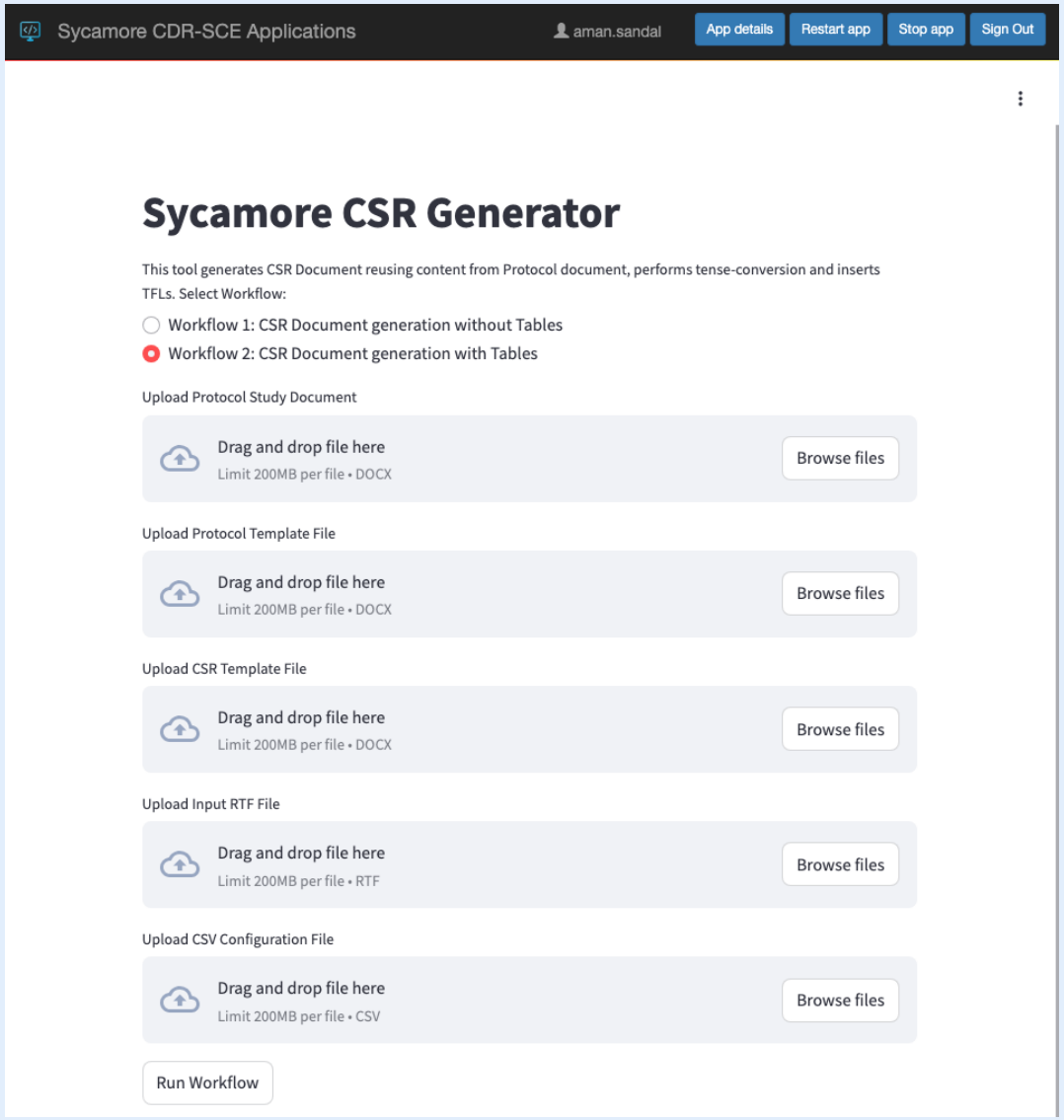
{{past:<sync_leadin_period>}}

How the Tool works?

Step 2: Upload Required Files

- Protocol Study Document
- Protocol Template
- CSR Template
- Multi-page Tables (RTF) File
- Summary Table Configuration File (CSV File) *

* Used by the tool for In-text Summary Tables Generation



Step 3: Tense-converted content pasted in CSR Template

3.1.2 Lead-in Period

During the Lead-in Period, subjects will have a study visit approximately 1 week before randomization (Visit 3, Day –7 ± 4 days). The subject's eligibility to continue into the Treatment Period will be assessed according to Randomization Criteria (Section 4.1.3). Assessments not satisfying the Randomization Criteria can be reassessed during the Lead-in Period if the Investigator has reason to believe the failed criteria(on) may have changed on repeat assessment. If all Randomization Criteria are satisfied, the subject will be eligible for randomization in the Treatment Period (Visit 4, Day 1).

3.1.3 Treatment Period

The Treatment Period will include 12 weeks of approximately monthly (every 4 weeks) administration of IP. Using an interactive response technology (IRT) system, eligible subjects will be randomized (1:1:2) to receive blinded IP (8 mg/kg CSL346, 16 mg/kg CSL346, or placebo).

Each subject's first dose of IP at Visit 4 (Day 1) will be an IV loading dose of 3 mg/kg CSL346 (subjects randomized to 8 mg/kg CSL346), 6 mg/kg CSL346 (subjects randomized to 16 mg/kg CSL346), or placebo (subjects randomized to placebo). All IV loading doses will be in a total volume of 10 mL and infused over the course of approximately 20 minutes. The subject will then be monitored for approximately 2 hours, during which AEs and vital signs will be recorded periodically as described in the Schedule of Assessments and Section 8.1.1. If no clinically relevant AEs are observed during this time, the subject's first SC dose of IP will be administered as an 8 mg/kg CSL346, 16 mg/kg CSL346, or placebo 20 mL SC infusion over the course of approximately 40 minutes.

Subjects will receive 3 subsequent SC infusions at Visit 7 (Week 4), Visit 8 (Week 8), and Visit 9 (Week 12) for a total of 4 SC doses.

Protocol Study Document: Present/Future Tense

9.1.2 Lead-in Period

During the Lead-in Period, subjects had a study visit approximately one week before randomization (Visit 3, Day –7 ± 4 days). The subject's eligibility to continue into the Treatment Period was assessed according to Randomization Criteria. If assessments did not satisfy these criteria, they could be reassessed during the Lead-in Period if the Investigator believed that the failed criteria might have changed on repeat assessment.

If all Randomization Criteria were satisfied, the subject would have been eligible for randomization in the Treatment Period (Visit 4, Day 1).

9.1.3 Treatment Period

The treatment period included 12 weeks of approximately monthly administration of IP. Eligible subjects were randomized (1:1:2) using an interactive response technology system to receive blinded IP, which consisted of 8 mg/kg CSL346, 16 mg/kg CSL346, or a placebo.

Each subject's first dose of IP at Visit 4 (Day 1) was an IV loading dose of 3 mg/kg CSL346 for those randomized to 8 mg/kg CSL346, 6 mg/kg CSL346 for those randomized to 16 mg/kg CSL346, or a placebo for those randomized to the placebo. The IV loading doses were administered in a total volume of 10 mL and infused over approximately 20 minutes.

Following this, subjects were monitored for about two hours during which adverse events (AEs) and vital signs were recorded periodically. If no clinically relevant AEs were observed, the subject's first SC dose of IP was administered as an 8 mg/kg CSL346, 16 mg/kg CSL346, or placebo 20 mL SC infusion over approximately 40 minutes.

Subjects received a total of four SC doses: one at Visit 7 (Week 4), another at Visit 8 (Week 8), and two at Visit 9 (Weeks 11-12). The subsequent SC infusions were administered every 4 weeks.

Generated CSR: Past Tense

Step 4: Multi-Page Tables

Table 14.1.1.1.1
Summary of Subject Disposition, Screening Period, Lead-in Period, Treatment Period, Follow-up Period
(Screened Analysis Set)

	CSL346 8 mg/kg n (%)	CSL346 16 mg/kg n (%)	Overall CSL346 n (%)	Placebo n (%)	Total n (%)
Study period: Lead-In Period					
Subjects entering lead-in period [1]					123 (37.6)
Lead-in failure [1]					11 (3.4)
Failure to meet randomization criteria [1]					4 (1.2)
Adverse event [1]					0
Pregnancy [1]					0
Death [1]					0
Lost to follow-up [1]					0
Protocol deviation [1]					0
Physician decision [1]					0
Withdrawal by subject [1]					4 (1.2)
Site terminated by sponsor [1]					0
Study terminated by sponsor [1]					0
Other [1]					3 (0.9)
Subjects randomized [1]	29 (8.9)	29 (8.9)	58 (17.7)	56 (17.1)	114 (34.9)
Subjects randomized but not treated with IP [2]	0	0	0	2 (3.6)	2 (1.8)
Subjects entering treatment period and treated with IP [2]	29 (100.0)	29 (100.0)	58 (100.0)	54 (96.4)	112 (98.2)

[1] Percentage based on number of subjects screened.
[2] Percentage based on number of subjects randomized to each treatment group.
[3] Data entered as free text allows identification of discontinuations due to COVID-19.

Source listing: 16.2.1.1, 16.2.1.2.1-2
Data snap shot date: 16DEC2022,/opt/zfs002/prd/csbeh251750/stats/interim/prog/tables/tdisp01.sas/09FEB2023/13:23

Table 14.1.1.1.1
Summary of Subject Disposition, Screening Period, Lead-in Period, Treatment Period, Follow-up Period
(Screened Analysis Set)

	CSL346 8 mg/kg n (%)	CSL346 16 mg/kg n (%)	Overall CSL346 n (%)	Placebo n (%)	Total n (%)
Study period: Treatment Period and Follow-Up					
Subjects entering treatment period and treated with IP [2]	29 (100.0)	29 (100.0)	58 (100.0)	54 (96.4)	112 (98.2)
Subjects discontinued the study medication during treatment period [2]	2 (6.9)	2 (6.9)	4 (6.9)	7 (12.5)	11 (9.6)
Primary reason for premature discontinuation of study medication [2]					
Adverse event [2]	1 (3.4)	2 (6.9)	3 (5.2)	0	3 (2.6)
COVID-19 associated adverse event [2][3]	0	0	0	0	0
Pregnancy [2]	0	0	0	0	0
Death [2]	0	0	0	0	0
COVID-19 associated death [2][3]	0	0	0	0	0
Lost to follow-up [2]	0	0	0	1 (1.8)	1 (0.9)
Physician decision [2]	0	0	0	0	0
COVID-19 associated [2][3]	0	0	0	0	0
Protocol deviation [2]	0	0	0	0	0
Withdrawal by subject [2]	0	0	0	3 (5.4)	3 (2.6)
COVID-19 associated [2][3]	0	0	0	0	0
Site terminated by sponsor [2]	0	0	0	0	0
Study terminated by sponsor [2]	0	0	0	0	0
Pandemic related [2]	0	0	0	1 (1.8)	1 (0.9)
Other [2]	1 (3.4)	0	1 (1.7)	2 (3.6)	3 (2.6)

[1] Percentage based on number of subjects screened.
[2] Percentage based on number of subjects randomized to each treatment group.
[3] Data entered as free text allows identification of discontinuations due to COVID-19.

Source listing: 16.2.1.1, 16.2.1.2.1-2
Data snap shot date: 16DEC2022,/opt/zfs002/prd/csbeh251750/stats/interim/prog/tables/tdisp01.sas/09FEB2023/13:23

These Multi-page Tables are pasted under appropriate subsections under Section 14

Step 5: Intext Summary Table Generation

Table 14.1.1.1.1 Summary of Subject Disposition, Screening Period, Lead-in Period, Treatment Period, Follow-up Period(Screened Analysis Set)

	CSL346 8 mg/kg n (%)	CSL346 16 mg/kg n (%)	Overall CSL346 n (%)	Placebo n (%)	Total n (%)
Subjects entering lead-in period [1]					123 (37.6)
Lead-in failure [1]					11 (3.4)
Failure to meet randomization criteria [1]					4 (1.2)
Withdrawal by subject [1]					4 (1.2)
Other [1]					3 (0.9)
Subjects randomized [1]	29 (8.9)	29 (8.9)	58 (17.7)	56 (17.1)	114 (34.9)
Subjects randomized but not treated with IP [2]	0	0	0	2 (3.6)	2 (1.8)
Subjects entering treatment period and treated with IP [2]	29 (100.0)	29 (100.0)	58 (100.0)	54 (96.4)	112 (98.2)
Subjects discontinued the study medication during treatment period [2]	2 (6.9)	2 (6.9)	4 (6.9)	7 (12.5)	11 (9.6)
Adverse event [2]	1 (3.4)	2 (6.9)	3 (5.2)	0	3 (2.6)
Lost to follow-up [2]	0	0	0	1 (1.8)	1 (0.9)
Withdrawal by subject [2]	0	0	0	3 (5.4)	3 (2.6)
Pandemic related [2]	0	0	0	1 (1.8)	1 (0.9)
Other [2]	1 (3.4)	0	1 (1.7)	2 (3.6)	3 (2.6)
Subjects discontinued the study during treatment or follow-up period [2]	2 (6.9)	3 (10.3)	5 (8.6)	7 (12.5)	12 (10.5)
Adverse event [2]	0	2 (6.9)	2 (3.4)	0	2 (1.8)
Death [2]	1 (3.4)	0	1 (1.7)	0	1 (0.9)
Lost to follow-up [2]	0	1 (3.4)	1 (1.7)	2 (3.6)	3 (2.6)
Withdrawal by subject [2]	0	0	0	4 (7.1)	4 (3.5)
Pandemic related [2]	0	0	0	1 (1.8)	1 (0.9)
Other [2]	1 (3.4)	0	1 (1.7)	0	1 (0.9)
Subjects completing the study [2]	27 (93.1)	26 (89.7)	53 (91.4)	47 (83.9)	100 (87.7)

[1] Percentage based on number of subjects screened.
[2] Percentage based on number of subjects randomized to each treatment group.
[3] Data entered as free text allows identification of discontinuations due to COVID-19.

The study, which aimed to evaluate the efficacy of an experimental treatment for a specific disease, was conducted during the COVID-19 pandemic. The study identifier and details on subject eligibility are not provided in the given snippet. However, it appears that the sample size may have undergone changes, as indicated by the varying numbers of subjects screened and randomized to each treatment group.

Critical findings from the study include:

* The most common reason for premature discontinuation of the study was adverse events, which affected 6.9% to 3.4% of the subjects depending on the treatment group.
* COVID-19 associated adverse events were reported in a negligible number of cases across both treatment groups.
* Pregnancy and disease relapse were not cited as reasons for premature discontinuation in any of the study groups.
* Death was reported in only one case, affecting 3.4% of subjects in one of the treatment groups, with COVID-19 associated deaths being a rare occurrence.
* Non-compliance with the study drug and lack of efficacy were not significant factors leading to premature discontinuation in either group.
* Lost to follow-up was a concern in both treatment groups, affecting 3.4% and 1.7% of subjects respectively, while withdrawal by subject affected a higher number of participants (7.1%) in one group.
* Withdrawal by parent/guardian or meeting protocol specified withdrawal criteria were not significant factors in either study group.
* Subjects completing the study showed a high completion rate across both groups, with percentages ranging from 93.1% to 83.9%, indicating a good adherence and engagement of subjects throughout the study.

The findings suggest that adverse events, particularly COVID-19 associated incidents, were not a significant concern in this study. However, the impact of COVID-19 on subject retention rates is notable, with higher percentages of lost to follow-up and withdrawal by subject observed in one treatment group compared to another. Despite these challenges, both groups demonstrated good adherence to the study protocol, as reflected in high completion rates across both cohorts. The low incidence of serious adverse events associated with the disease or study-related factors further supports the safe conduct of this trial amidst the pandemic context.

In-Text Summary Table
Generated from Multi-page Tables

Key Findings:
These assist medical writers to write descriptions for Intext Summary Tables

Step 5: Intext Summary Table Generation

Prepare a configuration CSV specifying:

- **table num:** List of tables to summarize (e.g., 14.1.1.1.1)
- **section:** Placeholder name in the CSR where the in-text summary and description will be inserted
- **description (Y/N):** Whether to generate an English description for the in-text summary table
- **table (Y/N):** Whether to generate a summarized version of the tables
- **drop 0 rows (Y/N):** Whether rows with zero counts should be included in the in-text summary table

	A	B	C	D	E
1	table num	section	description	table	drop 0 rows?
2	14.1.1.1.1	<table_disposition_subjects>	y	y	n
3	14.1.3	<table_protocol_deviations>	y	n	
4	14.1.1.3	<table_itt_analysis>	n	y	
5	14.1.1.3,14.4.1.11	<desc_2_tables>	y	n	
6	14.1.1.1.2	<table_analysis_set>	n	y	y
7	14.1.4.1.1	<table_itt_demog>	y	y	
8	14.1.5.1	<table_med_history>	y	n	
9					
10					

Workflows

Sycamore CSR Generator

This tool generates CSR Document reusing content from Protocol document, performs tense-conversion and inserts TFLs. Select Workflow:

- (●) Workflow 1: CSR Document generation without Tables
- ( ) Workflow 2: CSR Document generation with Tables

Upload Protocol Study Document

Drag and drop file here
Limit 200MB per file • DOCX
Browse files

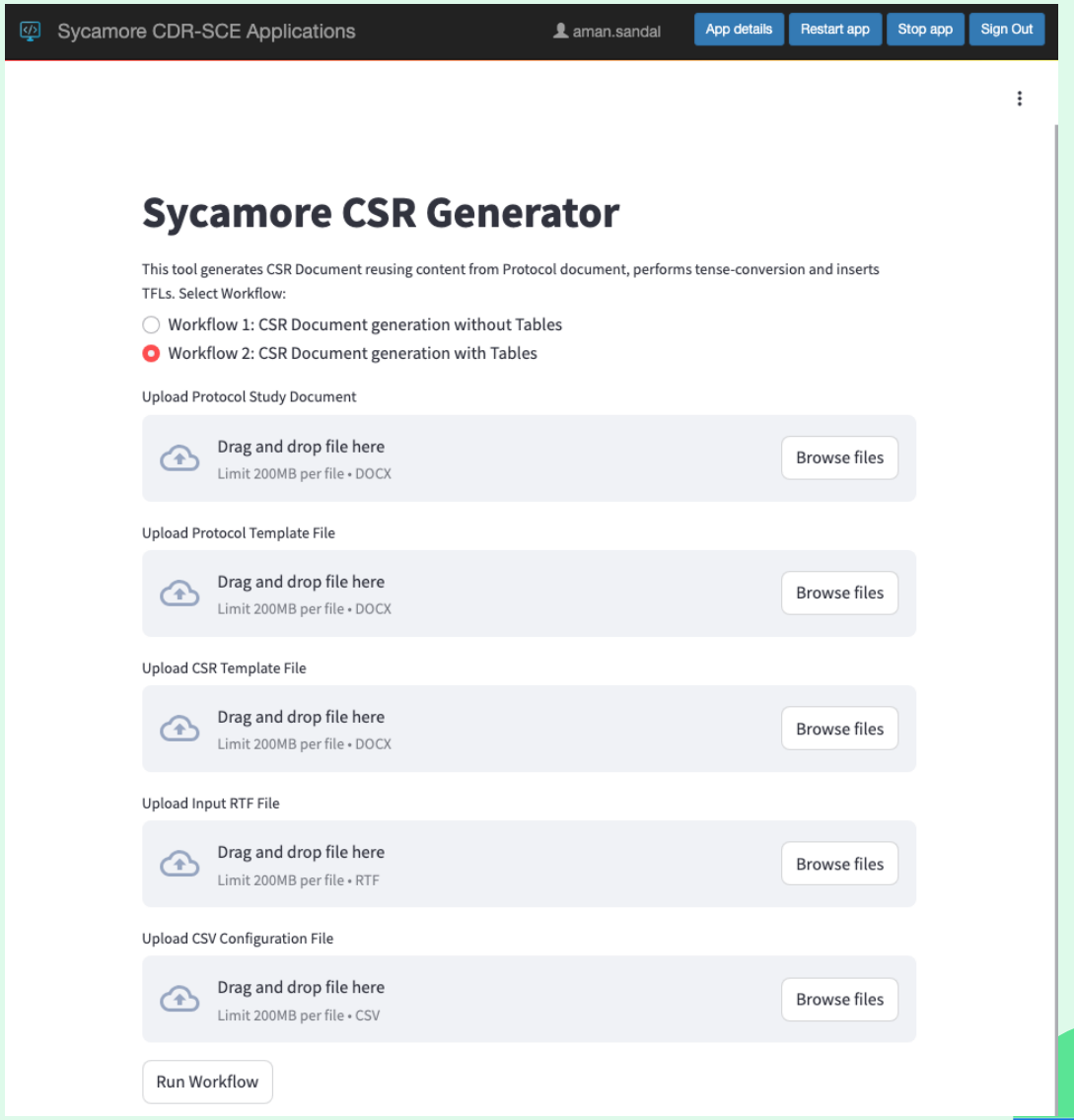
Upload Protocol Template File

Drag and drop file here
Limit 200MB per file • DOCX
Browse files

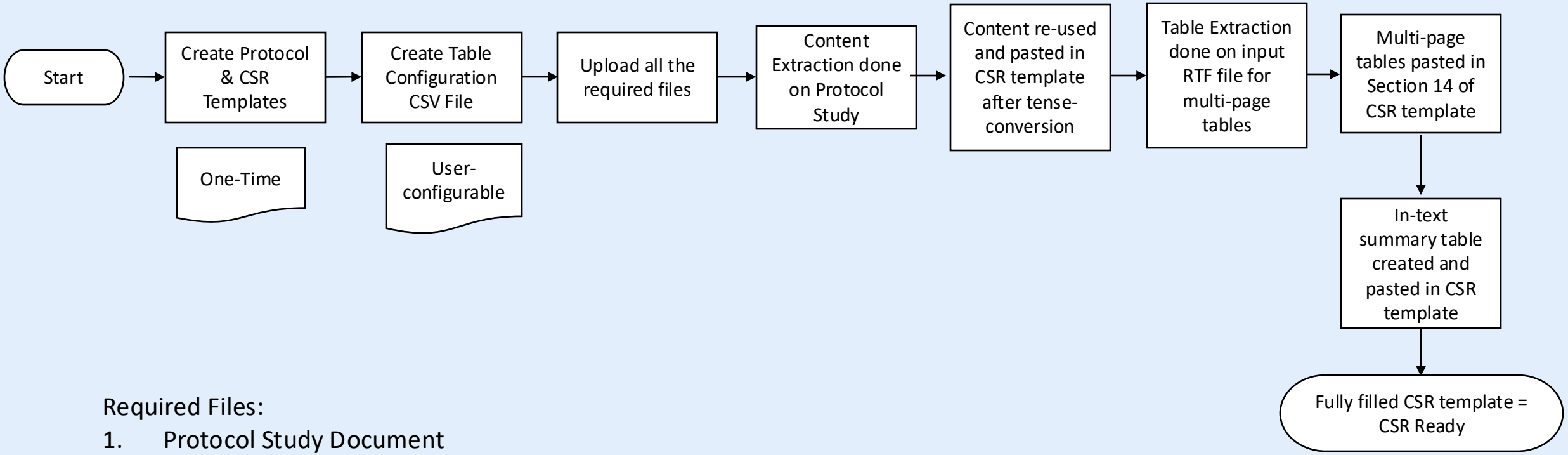
Upload CSR Template File

Drag and drop file here
Limit 200MB per file • DOCX
Browse files

Run Workflow



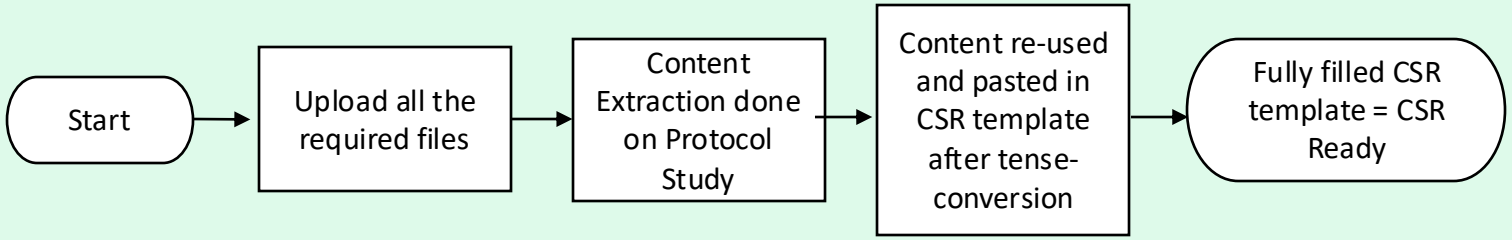
Workflow 2: Generate CSR using Protocol Study and TFLs



Required Files:

1. Protocol Study Document
2. Protocol Template
3. CSR Template
4. Multi-table RTF File
5. CSV Configuration File

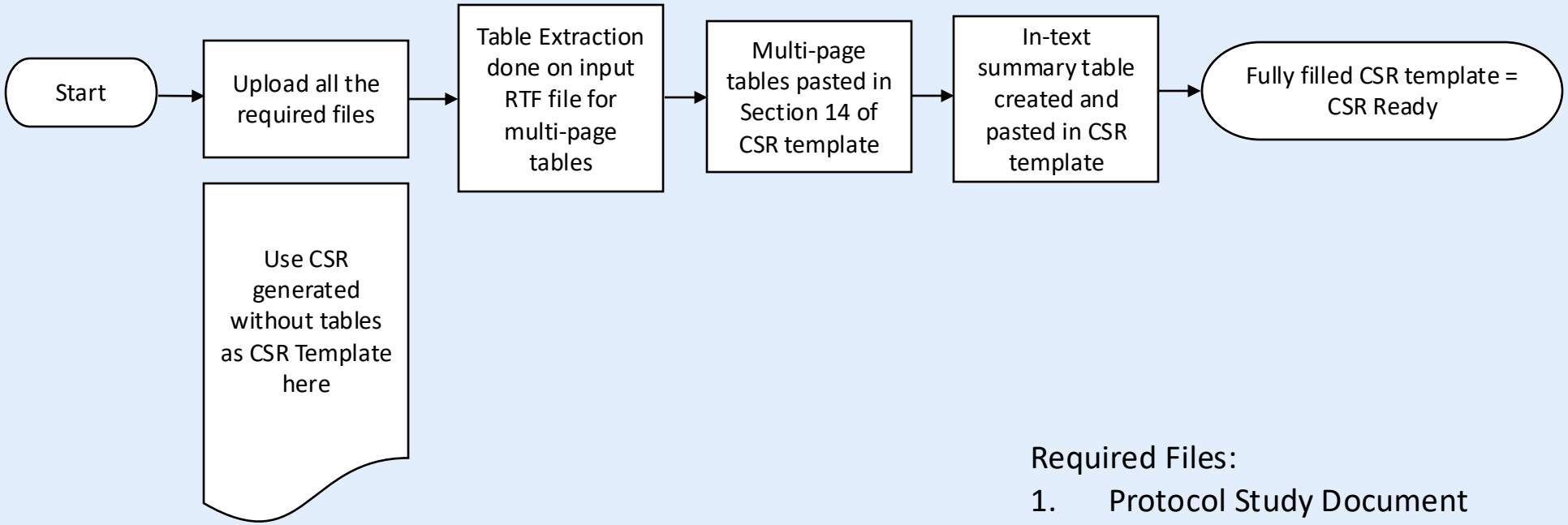
Workflow 1: Generate CSR using Protocol Study without Tables



Required Files:

1. Protocol Study Document
2. Protocol Template
3. CSR Template

Update CSR generated in Workflow 1 with TFLs



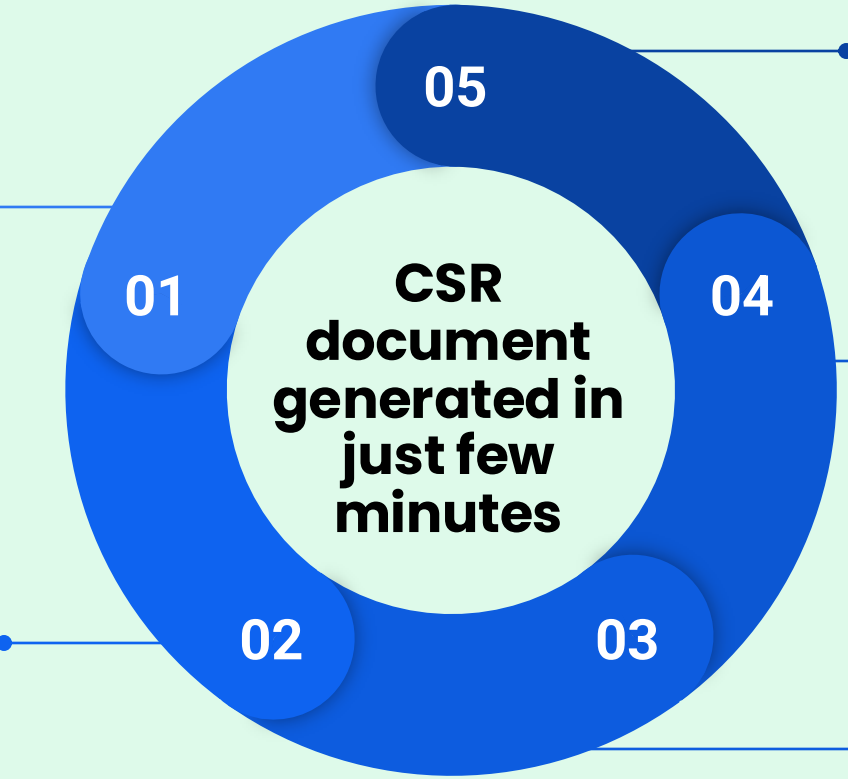
Required Files:

1. Protocol Study Document
2. Protocol Template
3. Generated CSR Document without tables
4. Multi-table RTF File
5. CSV Configuration File

Key Features of the Tool

Automated Content Extraction to extract protocol content and fill predefined CSR template

Automated Tense Conversion to convert future/present tense to past tense automatically using AI



Auto-Generated Key Findings provides key insights & trends of In-Text Summary tables using AI

In-Text Summary Table Generation using Multi-page tables **(RTF)** and placeholder-based insertion (configurable)

Seamless Multi-Page Table Insertion imports Tables, Listings, and Figures (TLFs) from external RTF files and inserts them into Section 14

Conclusion

- The AI-powered CSR generator holds the capability to transform clinical reporting by reducing weeks of manual work to minutes while ensuring accuracy and compliance.
- With Generative AI and structured data mapping, it enhances efficiency, accelerates insights, and simplifies the documentation.
- As it expands to include SAP integration*, it shall bring even greater transparency, reliability, and speed to clinical studies.

Thank You

Q&A