



**Automate Scientific Document such as CSR
Generation using ML/NLP**

What are we trying to solve?

Clinical Study Report

- ❑ It is a document, typically very long, providing much detail about the methods and results. A CSR is a scientific document addressing efficacy and safety.
- ❑ Creating Clinical Study Report (CSR) is highly manual and time consuming.
- ❑ The content of CSR is constructed from Protocol, SAP, Safety narratives, in-text tables and medical writer's interpretation of study results.
- ❑ No automotive way to handle Manual Errors and multiple iterations

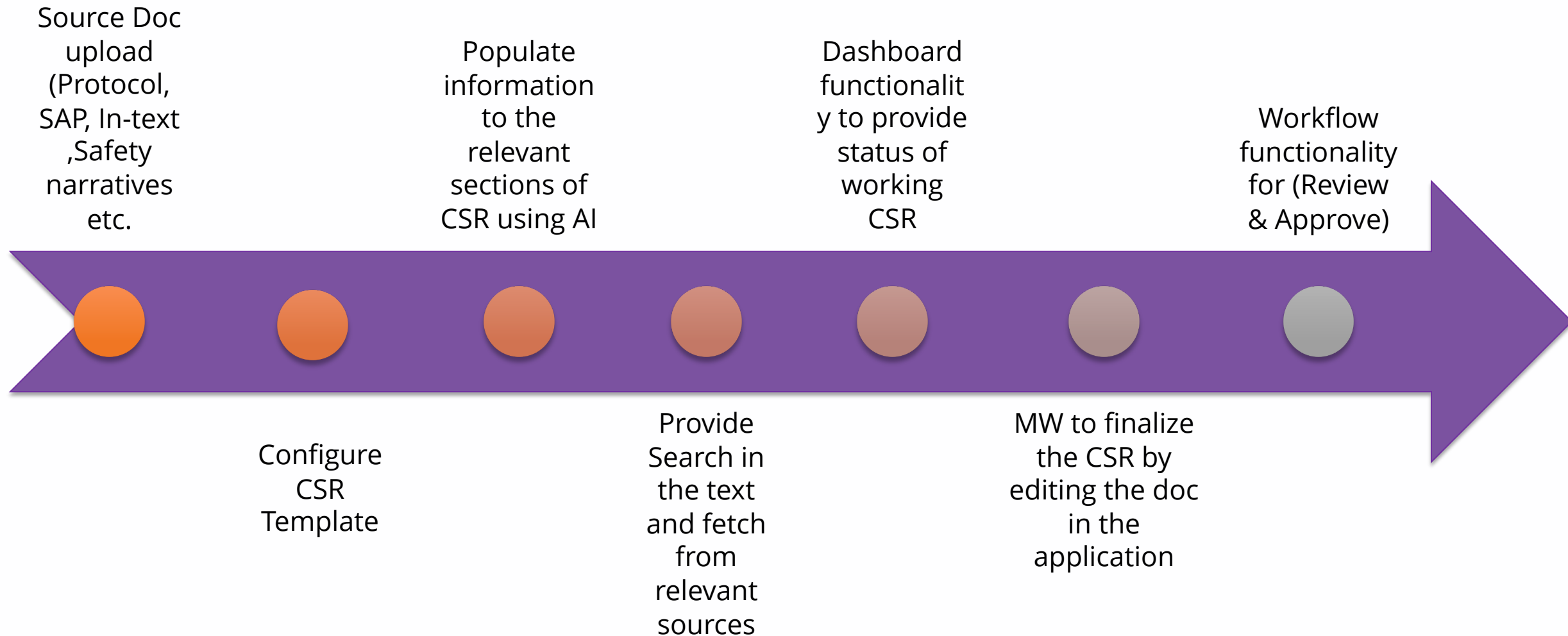
Why automation ?

- ❑ 70-80% of manual effort can be saved with automation, MW can focus on the Interpretation and discussion points for the remaining on 20-30%.
- ❑ Effective use and reuse of the template for an organization or therapeutic area or compound.
- ❑ The ML prediction algorithm along with NLP/NLG is used to identify/effectively used to read/write the texts from other documents such as Protocol, SAP, In-text Table & etc.,
- ❑ The CSR Template will be created based on ICH –E3 guidelines
- ❑ The medical writers can focus more on **discussion points and interpretation of study results.**

Blank CSR Template

1. TITLE PAGE
2. STUDY SYNOPSIS
3. TABLE OF CONTENTS
4. LIST OF ABBREVIATIONS & DEFINITION OF TERMS
5. ETHICS AND REGULATORY APPROVAL
6. INVESTIGATORS AND STUDY ADMINISTRATIVE STRUCTURE
7. INTRODUCTION
8. STUDY OBJECTIVES
9. INVESTIGATIONAL PLAN
10. STUDY POPULATION
11. RESULTS
12. SAFETY EVALUATION
13. DISCUSSION AND OVERALL CONCLUSIONS
14. TABLES, FIGURES AND GRAPHS
15. REFERENCES
16. APPENDICES

Steps



Use of ML/NLP

❑ **Prediction Accuracy :**

- ❖ The text from PDF document such as Protocol/SAP are extracted using ML/NLP and the texts are understood by the system engine.

❑ **Identifying Individual Sections:**

- ❖ ML model is used to predict the best matching content from various source documents(Protocol, SAP, In-text etc.,) for all the sections in a CSR.
- ❖ Named Entity Recognition (NER) which is a subprocess of NLP is used to identify the drug names, dosages, duration of drug, sponsor name and protocol number

❑ **Title Page & Synopsis:**

- ❖ Customized ML Algorithm used in title page and study synopsis.

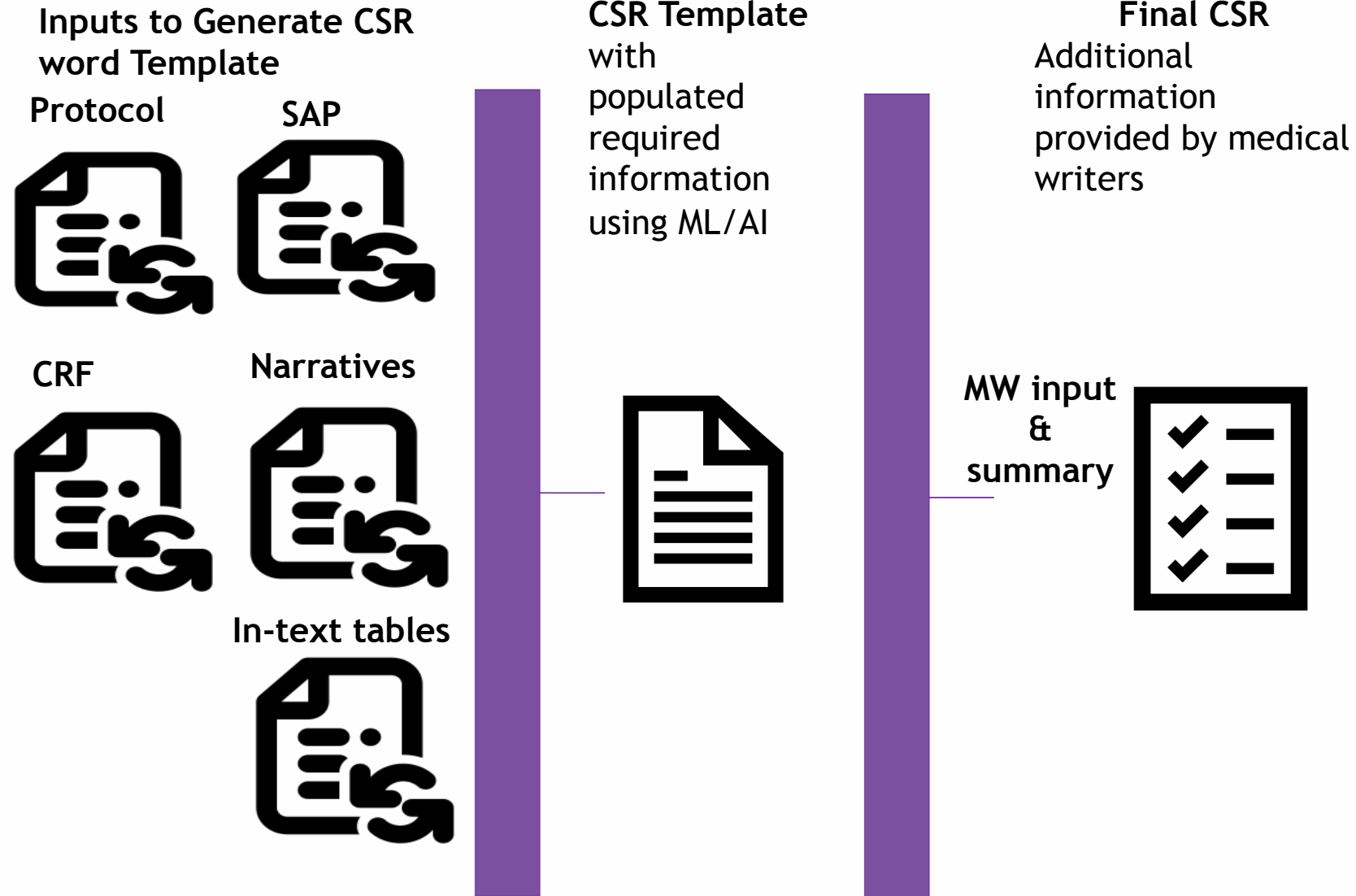
❑ **Deep Learning Model usage:**

- ❖ The section 10 (study population), section 11 (results) and section 12(safety evaluation) is more of In-text tables that utilizes Deep Learning Model to find the best matching In-text tables.

Use of Natural Language Generation (NLG)

- ❑ Using NLG to transform the content of CSRs from present tense to Past Tense and perform a grammar check of the content displayed.
- ❑ Use Latest Google's Transformer model T5 to generate the summary of any type of intext table along with the help of customized in-house algorithm
- ❑ Natural Language Generation to generate safety narrative text from the Serious adverse Events Listing present in the in
- ❑ Converting traditional TLF table to intext table using OpenCV and placing it in generated CSR. text table document.

Process



Conclusions

We developed an application to automate CSR generation with the following features,

- ❑ MW can add or modify the content by **section within the system**. The “**CSR EDIT**” feature in the user interface is better than the same in MS word.
- ❑ Medical writers Interpretation will be highlighted in yellow with a blank space by the system.
- ❑ Traceability Report(source file to final CSR mapping)
- ❑ Workflow integration(Author(s), Reviewer(s), Approver)
- ❑ Audit Log
- ❑ Narrative writing from Patient profiles/ Subject Listing .
- ❑ In-text table interpretation in simple English.
- ❑ Dashboard for the overall CSR, section vice and the workflow status
- ❑ Present tenses in source documents are converted to past tenses in the CSR using NLG.

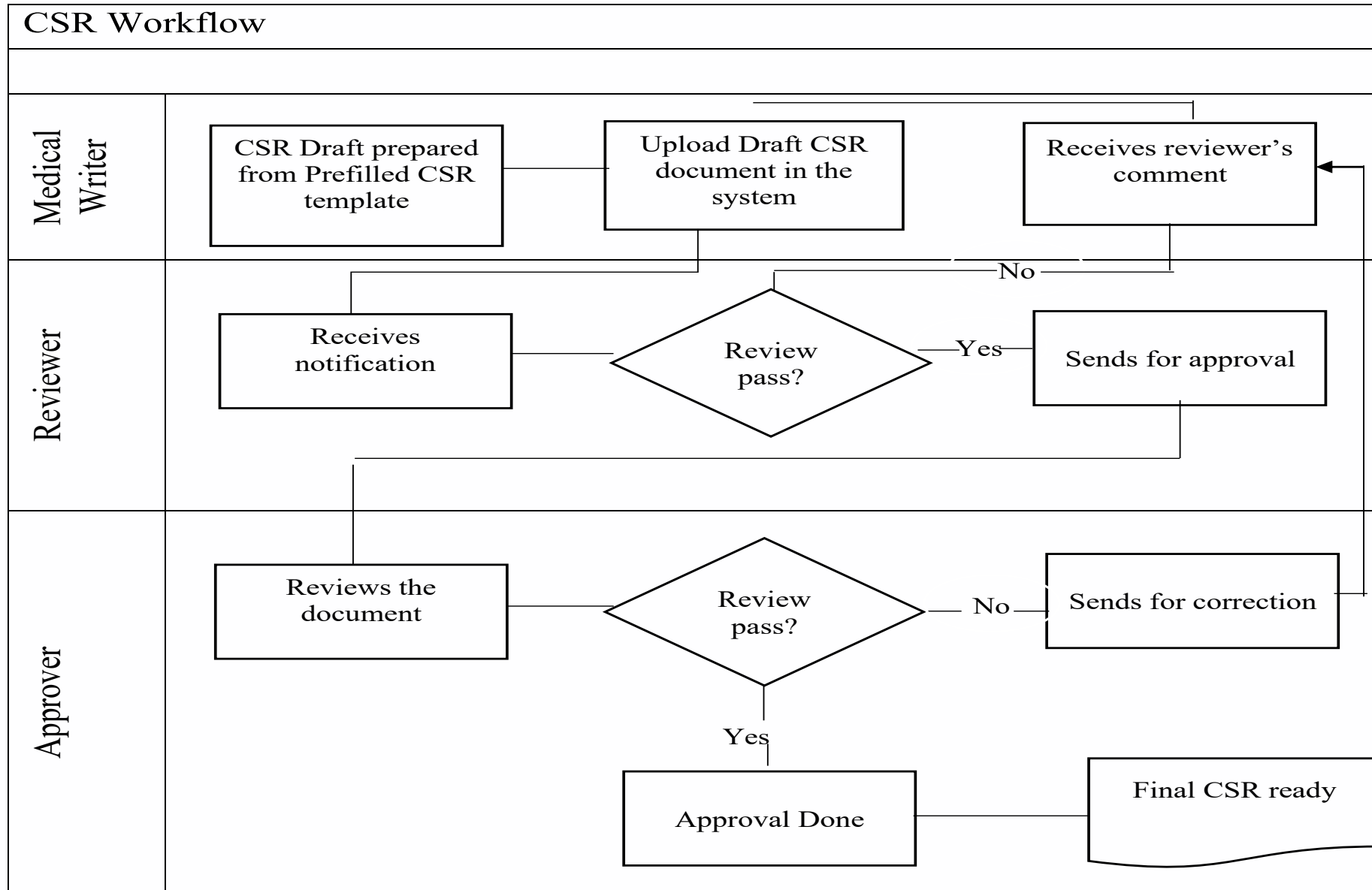
Application Edit Screen

The screenshot displays the ZYLIQ application interface for editing a document titled "Symbiance-017 - PKPD CSR". The interface includes a top header with the ZYLIQ logo and a user profile icon labeled "Writer Writer". A left sidebar contains a table of contents with sections 1 through 10, where section 5 is expanded. The main editing area shows the "1. TITLE PAGE" section with a toolbar containing icons for adding bookmarks, viewing bookmarks, saving changes, showing sections, undo, redo, undo all, redo all, image, table, find, link, track changes, and comments. The content of the title page is structured as follows:

1. TITLE PAGE	
STUDY TITLE:	Clinical Study Report
	A Randomized Crossover, Pilot Study in Healthy Subjects to Assess the Pharmacokinetics and Abuse-Deterrent Properties of Oxycodone Immediate-Release Formulations
INVESTIGATIONAL DRUG NAME:	xxxxxx (OC)
INDICATION:	
PROTOCOL NUMBER:	SYMBIANCE-017
DRUG DEVELOPMENT PHASE:	Phase 1
STUDY INITIATION DATE:	07-27-2020
STUDY COMPLETION DATE:	02-29-2020
GCP STATEMENT:	This study was conducted in compliance with the Declaration of Helsinki and Good Clinical Practice or Trials on Medicinal Products, including archiving of essential study documents.
DATE OF REPORT:	03-30-2021
Signature pages for clinical study report	
I have read this report and confirm that to the best of my knowledge it accurately describes the conduct and results of the study.	
Signed:	

At the bottom of the screen, the page number is indicated as "Page 1 of 1".

Workflow Integration



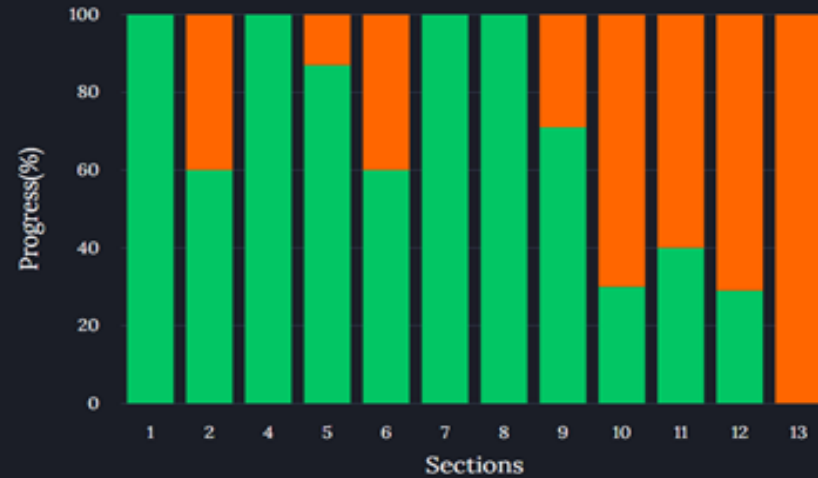
Dashboard Functionality

Symbiance-017 - PKPD CSR

Source Files

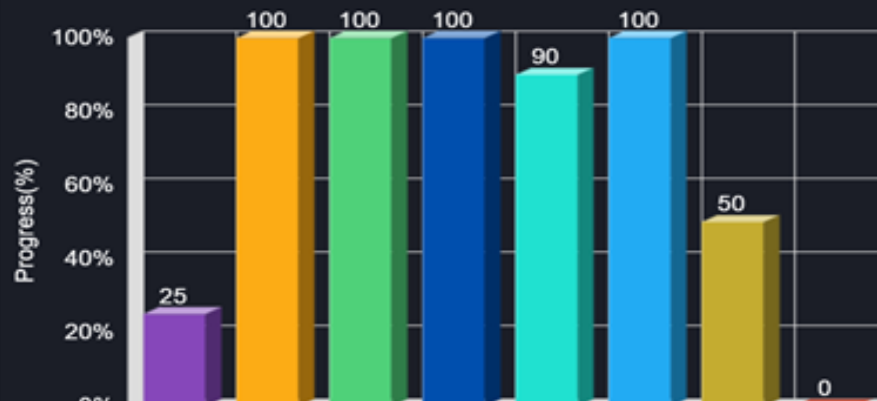
Clinical Study Statistics

This data is updated approximately every day.

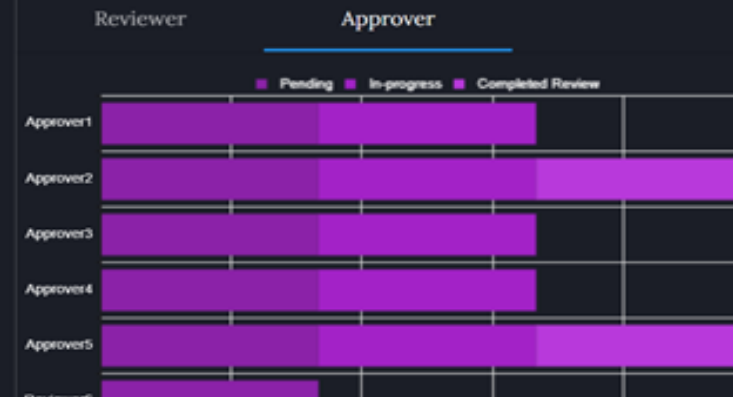


- Section1: 100%
- Section2: 60%
- Section3: 60%
- Section4: 100%
- Section5: 87%
- Section6: 60%
- Section7: 100%
- Section8: 100%
- Section9: 71%
- Section10: 30%
- Section11: 40%
- Section12: 29%
- Section13: 0%

9.INVESTIGATIONAL PLAN



Workflow





Thank you!!

***Contact:
Farha Fathima Feroze
Product Manager
Mobile: +1 (609) 945 7500***