



Validation of Figures Using Computer Vision Approach

Mor Ram-On | VP of R&D

March 2023

Projects / Overview / #1961

PHUSE US 2023

Discrepancies

File type

File outcome

Advanced

1 file selected

	Type	Outcome	Status	Assignee	
1	LoT Demo.xlsx	S	●	To Do	SA
2	Fig14.1.2.3.pdf	F	2	To Do	SA
☒	Fig14.1.2.5.pdf	F	2	To Do	SA
4	Figure 14.1.2.4.pdf	F	2	To Do	SA

Fig14.1.2.5.pdf | 2 discrepancies

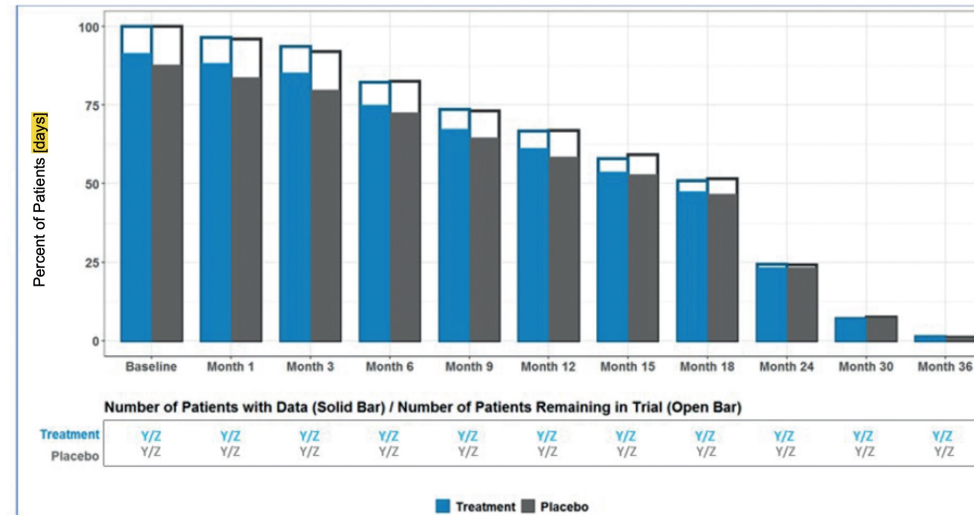
Export

100 %

2 To Do

Comments

Figure 14.1.2.5 Drug A
Proportion of Patients Remaining With Missing and Existing Laboratory Data Records, Safety Population, Phase 2



Note: The frequency of laboratory measurements presented here is based on actual data collected.
Note: The time frame (e.g., by day, week, month) that corresponds best with the prespecified visit is used as the study visit
BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs
Table Generation: 09OCT2021

Agenda

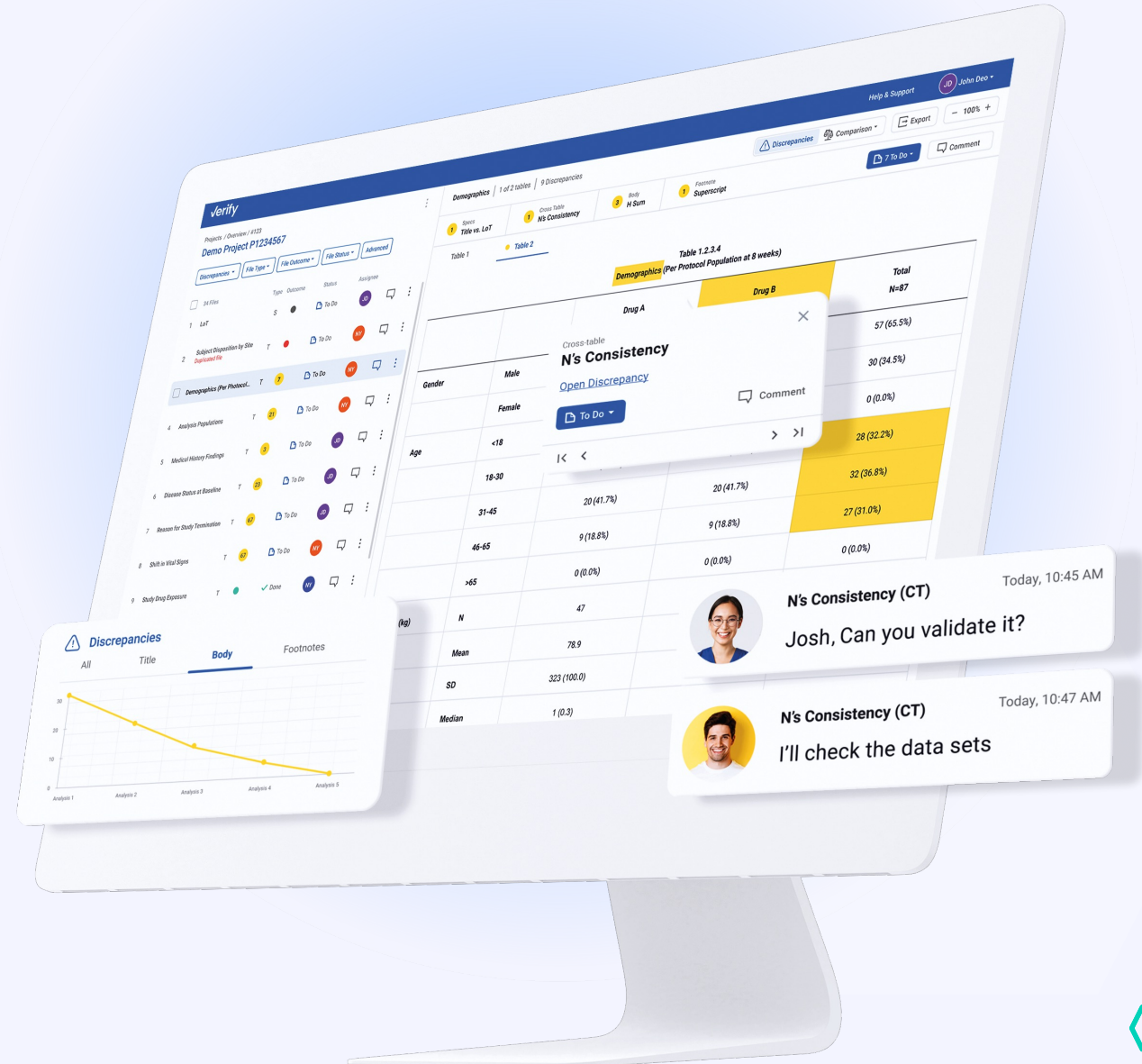
- **Introduction**
 - Verify Web Application
 - Validation process high-level pipeline
 - TLFs Type of Checks
- **Figure Analysis – What?**
 - Figure Structure
 - Figure Types
 - Figure Discrepancies
- **Figure Analysis – Why?**
 - Today's Alternatives
- **Figure Analysis – How?**
 - Pipeline building blocks



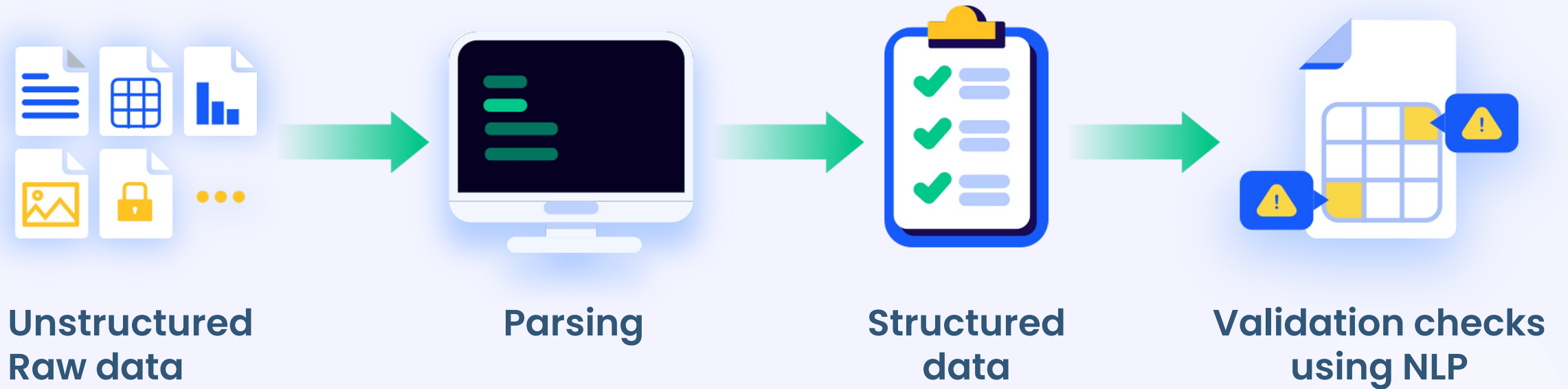
Verify Software

Developed by Beaconcure,
A Cloud Application for:

- ✓ Data Accuracy
- ✓ Accelerated Delivery
- ✓ Risk Reduction
- ✓ Process Improvement



High Level Pipeline



TLF of Checks Within TLF

Tables

- Vertical Sum
- Horizontal Sum
- Title Formats
- Footnotes Formats
- Superscript Reference
- ...

Listings

- Vertical Sum
- Horizontal Sum
- Title Formats
- ...

Figures

- Axes Units
- Title Formats
- Footnotes Format
- ...



Within TLF Check Example

Table 14.1.1.1
Subject Disposition by Site

	Randomized	Treated	Completed 4 weeks of treatment	Completed 8 weeks of treatment	Completed Study	Completed 4 weeks T
Site 001	25	24	24	23	20	1
Site 002	25	25	22	20	19	1
Site 003	24	24	22	22	21	1
Site 004	22	22	19	18	16	1
Site 005	0	0	0	0	0	0
Site 006	18	18	14	13	11	0
Site 007	25	25	22	20	17	1
Site 008	25	25	22	22	20	1
Site 009	24	23	20	20	18	1
Site 010	25	25	22	21	16	1
Site 011	25	25	22	21	17	1
Total	237	236	209	200	175	14

BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021



TLF of Checks Cross TLF

Tables & Listings

- Title vs. Mock shell
- Footnotes vs. LoT/TOC
- Protocol Number Consistency
- ...

Figures

- Legend Consistency
- Protocol Number Consistency
- ...



Cross TLF Check Example

Table 14.1.1.31
Summary of Complete Healing Rate (%) for each population by Treatment (Randomized Population)

Randomized	Drug A (N=118)	Drug B (N=117)	Total (N=235)
	(%)	(%)	(%)
4 weeks ITT	46.7	35.0	40.9

Table 14.1.1.2
Analysis Population

	Drug A	Drug B	Total
Randomized Population	120	117	237
Safety Population	119	117	236
Received at least one dose	119	117	236
Completed 4 weeks	89	87	176
Per Protocol Population at 4 weeks	80	61	141
Completed 8 weeks	82	75	157
Per Protocol Population at 8 weeks	48	39	87
Completed Study (12 weeks)	75	69	144

BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021



Figure Analysis – What?

- Figure Structure
- Figure Variety
- Figure Discrepancies

Figure Structure

We divide the figure into 3 sections:

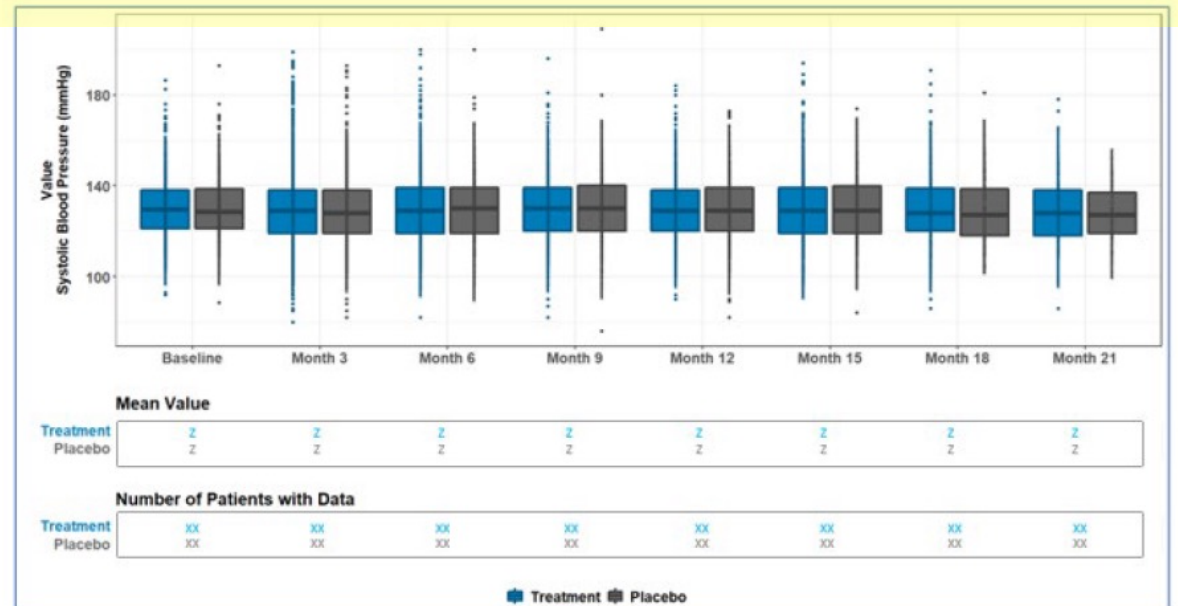
1. Header

Textual Data:

Title, Figure number,

Protocol Number, References

Figure 16. Median Interquartile Range of Systolic Blood Pressure Over Time by Treatment Arm,¹ Safety Population, Pooled Analysis



Source: [include Applicant source, datasets and/or software tools used].

¹Boxes span the interquartile range (25th to 75th percentile); horizontal line = median; whiskers = 1.5x the interquartile range; individual outliers are those beyond this range.

Figure Structure

We divide the figure into 3 sections:

1. Header

2. Footer

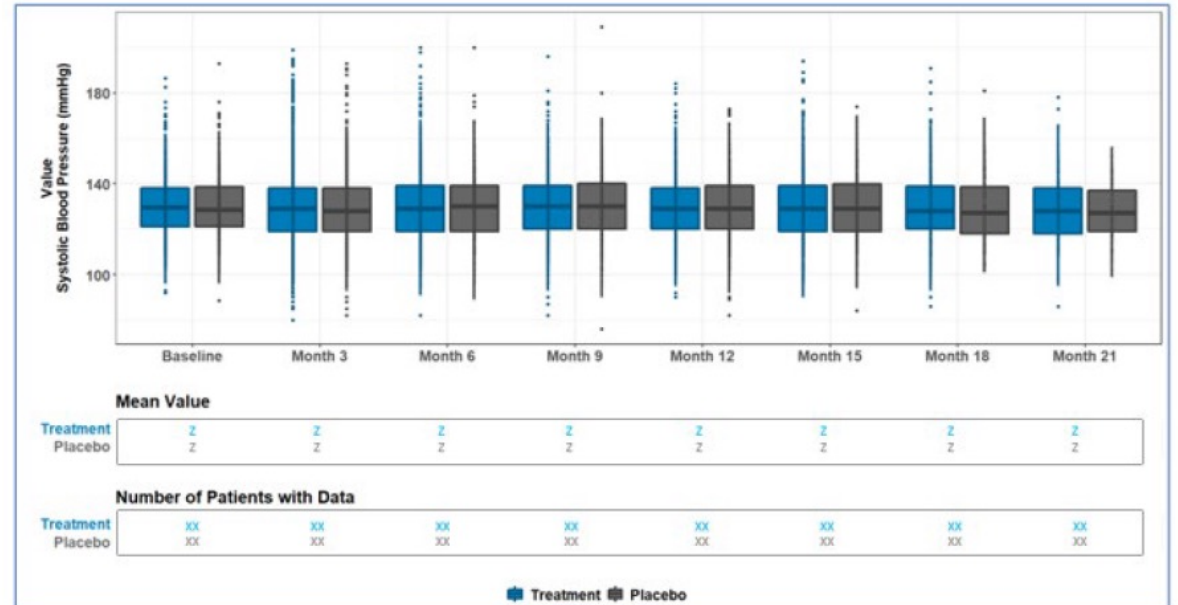
Textual Data:

Protocol Number,

References, Abbreviations, Source
data, Dates (Creation, Generation)

Programmer notes

Figure 16. Median Interquartile Range of Systolic Blood Pressure Over Time by Treatment Arm,¹ Safety Population, Pooled Analysis



Source: [include Applicant source, datasets and/or software tools used].

¹Boxes span the interquartile range (25th to 75th percentile); horizontal line = median; whiskers = 1.5x the interquartile range; individual outliers are those beyond this range.

Figure Structure

We divide the figure into 3 sections:

1. Header

2. Footer

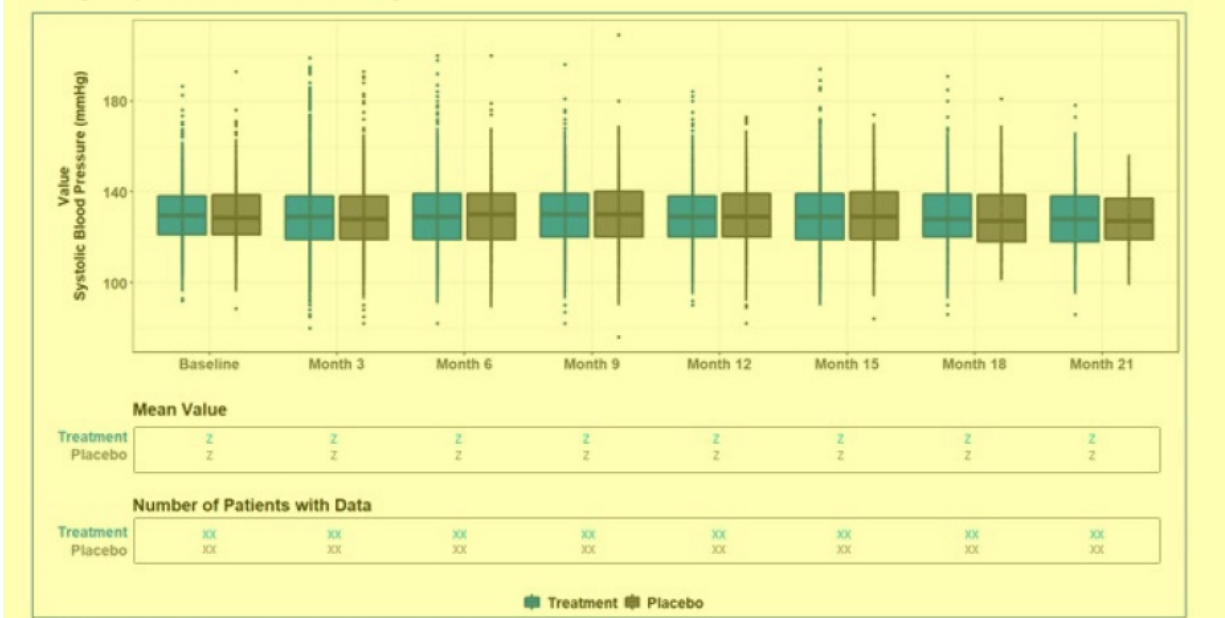
3. Body

Image & text:

Axes title, legend,

Axes ticks and data, data points, etc.

Figure 16. Median Interquartile Range of Systolic Blood Pressure Over Time by Treatment Arm,¹ Safety Population, Pooled Analysis



Source: [include Applicant source, datasets and/or software tools used].

¹Boxes span the interquartile range (25th to 75th percentile); horizontal line = median; whiskers = 1.5x the interquartile range; individual outliers are those beyond this range.

Figure Variety Complicates Analysis

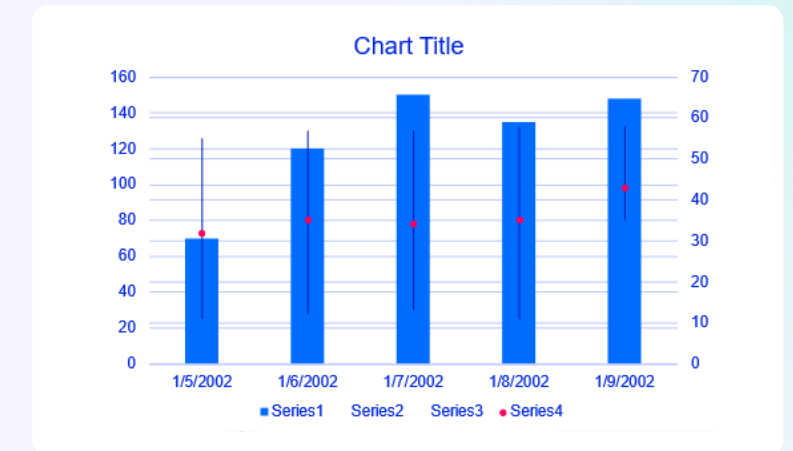
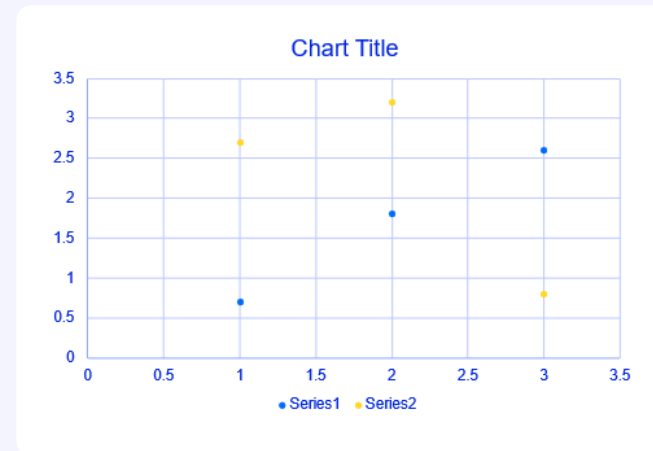
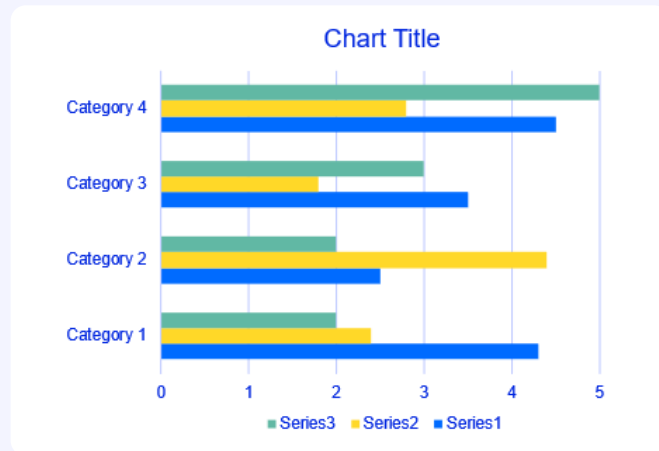
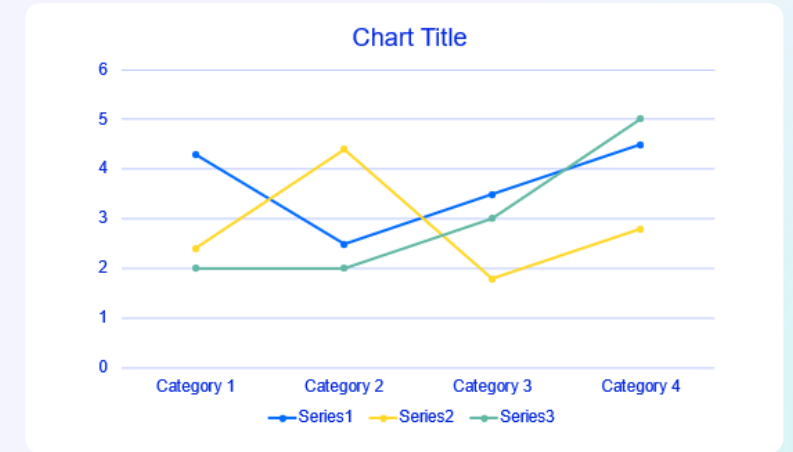
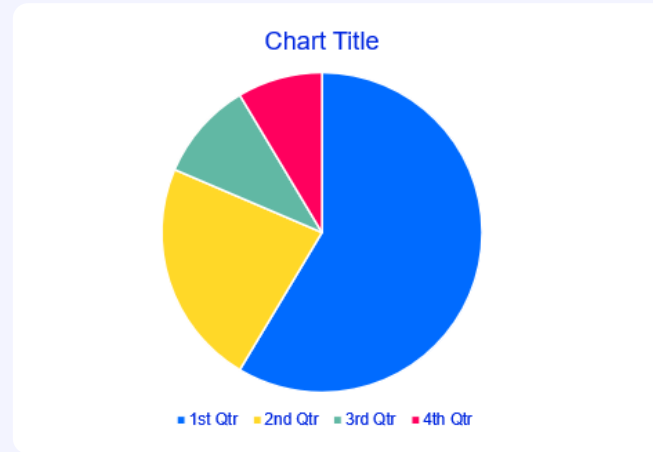
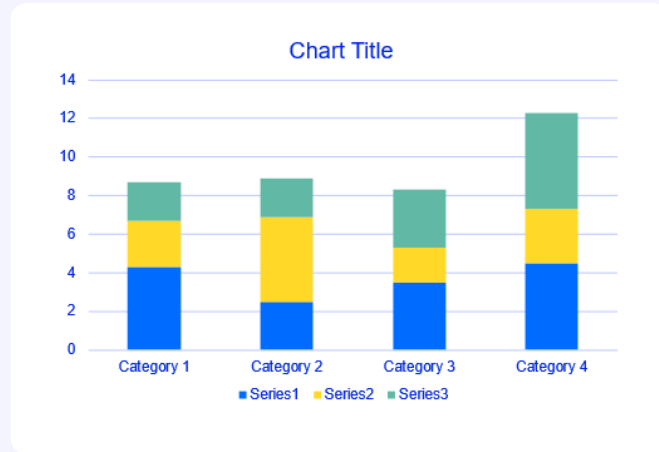


Figure Validation

Analysis involves all project resources

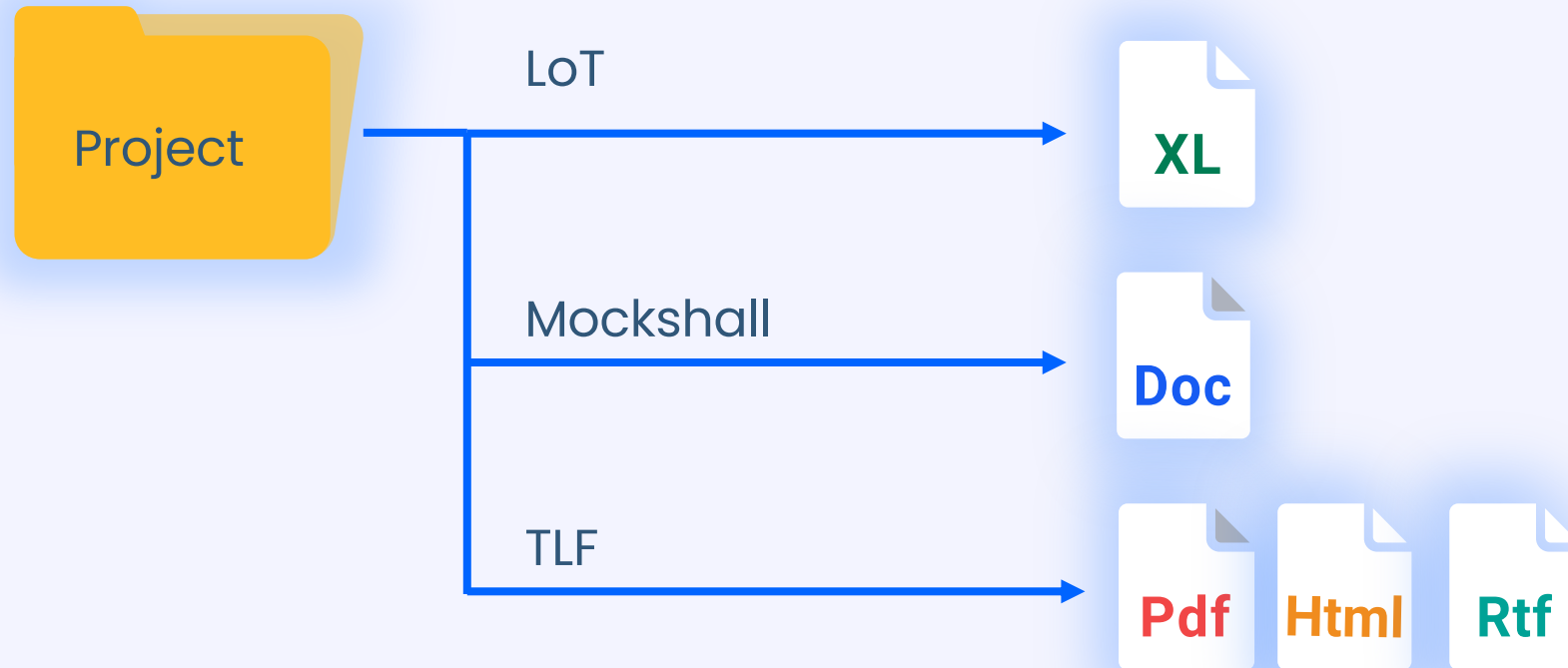


Figure Discrepancies

- **Figure vs. LoT**

Mismatch of figure number, figure title, figure source data

- **Figure vs. Mock shell**

Mismatch of figure number, figure title, figure source data

- **Axes Units**

Lack of axes units, where needed

- **Legend Consistency**

Inconsistency in series type (symbols), colors and names

Legend Consistency discrepancy

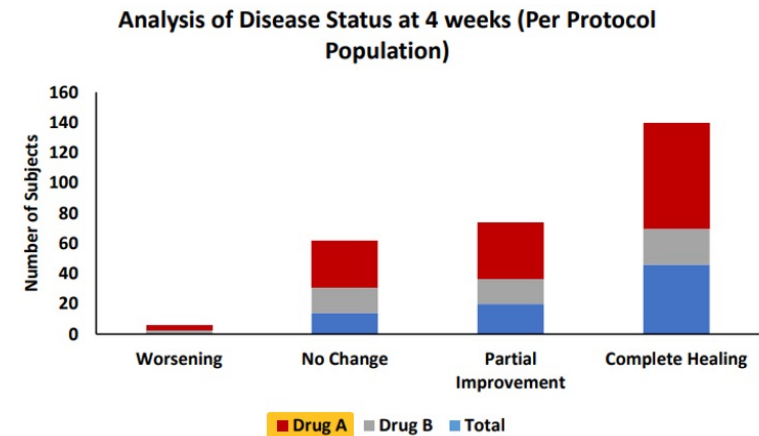
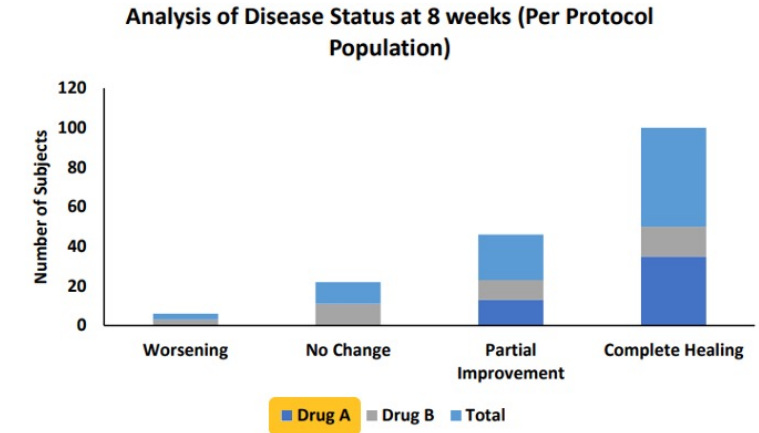


Figure Analysis – Why?

Current Solution – Manual

Table Vs. Figure



Table 2. Baseline Demographic and Clinical Characteristics, Safety Population, Pooled Analyses (or Trial X)

Characteristic	Drug Name Dosage X N = XXX n (%)	Drug Name Dosage Y N = XXX n (%)	Placebo N = XXX n (%)	Active Control N = XXX n (%)	Total Population N = XXX n (%)
Sex, n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Male	n (%)	n (%)	n (%)	n (%)	n (%)
Female	n (%)	n (%)	n (%)	n (%)	n (%)
Age, years	X.X (Y.Y)	X.X (Y.Y)	X.X (Y.Y)	X.X (Y.Y)	X.X (Y.Y)
Mean (SD)	X.X (Y.Y)	X.X (Y.Y)	X.X (Y.Y)	X.X (Y.Y)	X.X (Y.Y)
Median (min, max)	X.X (Y.Y, Z.Z)	X.X (Y.Y, Z.Z)	X.X (Y.Y, Z.Z)	X.X (Y.Y, Z.Z)	X.X (Y.Y, Z.Z)
Age groups (years), n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
≥17 to <65	n (%)	n (%)	n (%)	n (%)	n (%)
≥65	n (%)	n (%)	n (%)	n (%)	n (%)
≥65 to <75	n (%)	n (%)	n (%)	n (%)	n (%)
≥75	n (%)	n (%)	n (%)	n (%)	n (%)
Race, n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
American Indian or Alaska Native	n (%)	n (%)	n (%)	n (%)	n (%)
Black or African American	n (%)	n (%)	n (%)	n (%)	n (%)
Native Hawaiian or Other Pacific Islander	n (%)	n (%)	n (%)	n (%)	n (%)
White	n (%)	n (%)	n (%)	n (%)	n (%)
Other	n (%)	n (%)	n (%)	n (%)	n (%)
Ethnicity, n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Hispanic	n (%)	n (%)	n (%)	n (%)	n (%)
Not Hispanic or Latino	n (%)	n (%)	n (%)	n (%)	n (%)
Unknown	n (%)	n (%)	n (%)	n (%)	n (%)
Country of participation, n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
United States	n (%)	n (%)	n (%)	n (%)	n (%)
Country A	n (%)	n (%)	n (%)	n (%)	n (%)
Country B	n (%)	n (%)	n (%)	n (%)	n (%)
Country C	n (%)	n (%)	n (%)	n (%)	n (%)
Country D	n (%)	n (%)	n (%)	n (%)	n (%)
Other	n (%)	n (%)	n (%)	n (%)	n (%)

Table Vs. Figure

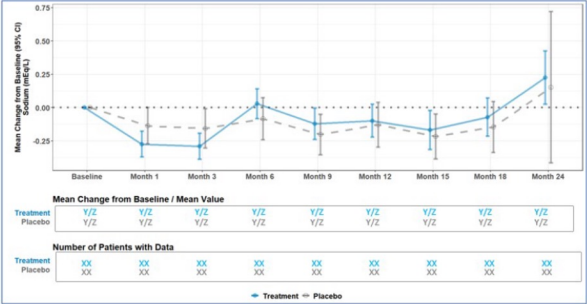
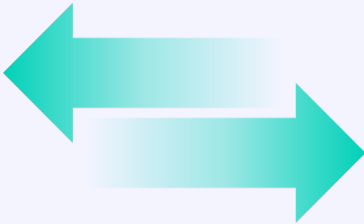
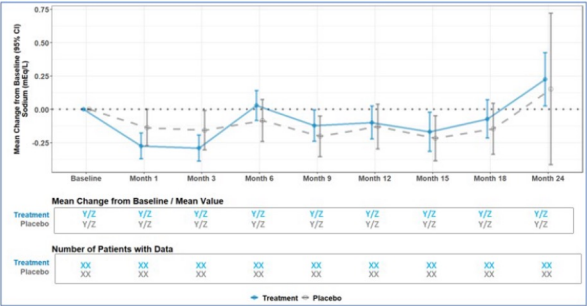
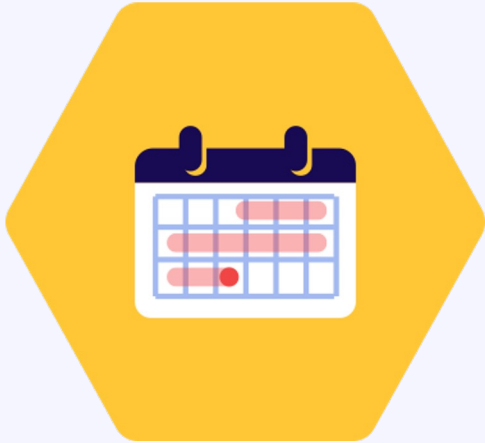


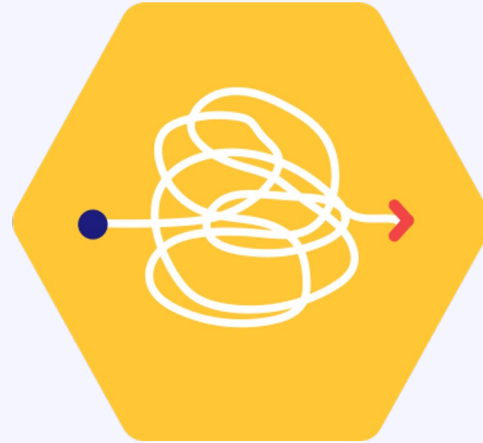
Figure Vs. Figure



Current Solution – Manual



Time Consuming



Complicated

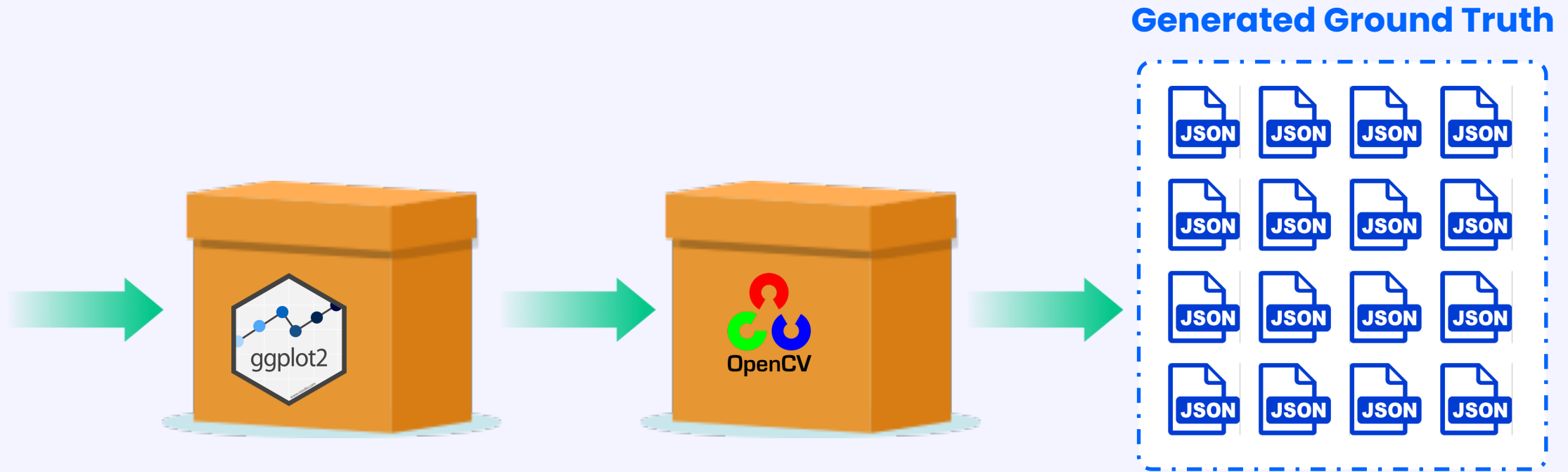


Error-prone

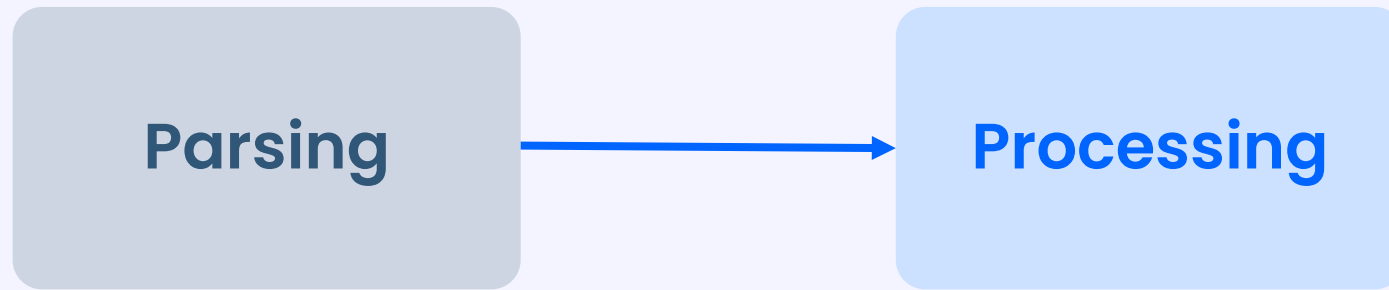


Figure Analysis – How?

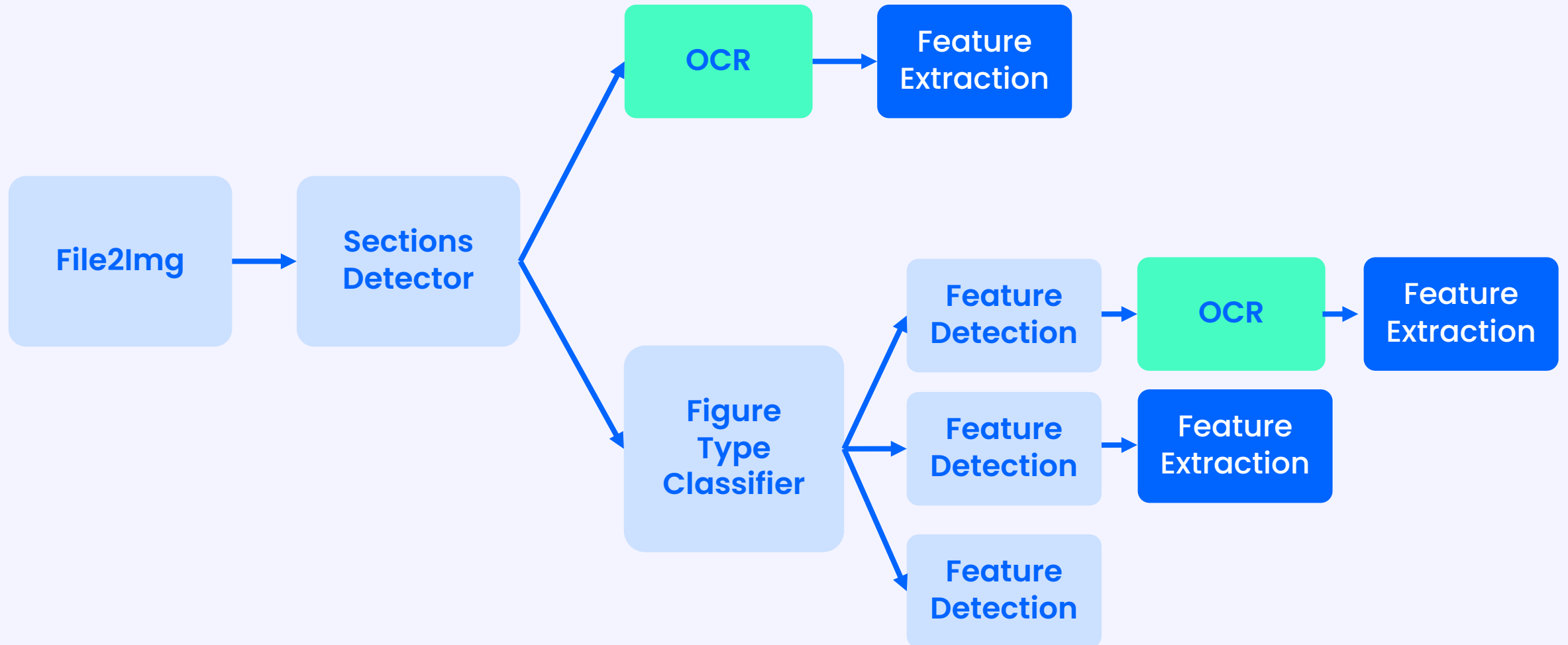
Synthetic Data Generation



High Level Pipeline



Parsing Pipeline





Thank You!

Visit us at the
Sponsor Zone

mor@beaconcure.com / www.Beaconcure.com