Beaconcure
TECHNOLOGIES

Revolutionizing Statistical Outputs Validation

Author: Ilan Carmeli, CPO

ilan@beaconcure.com
www.beaconcure.com

Beaconcure TECHNOLOGIES

Abstract

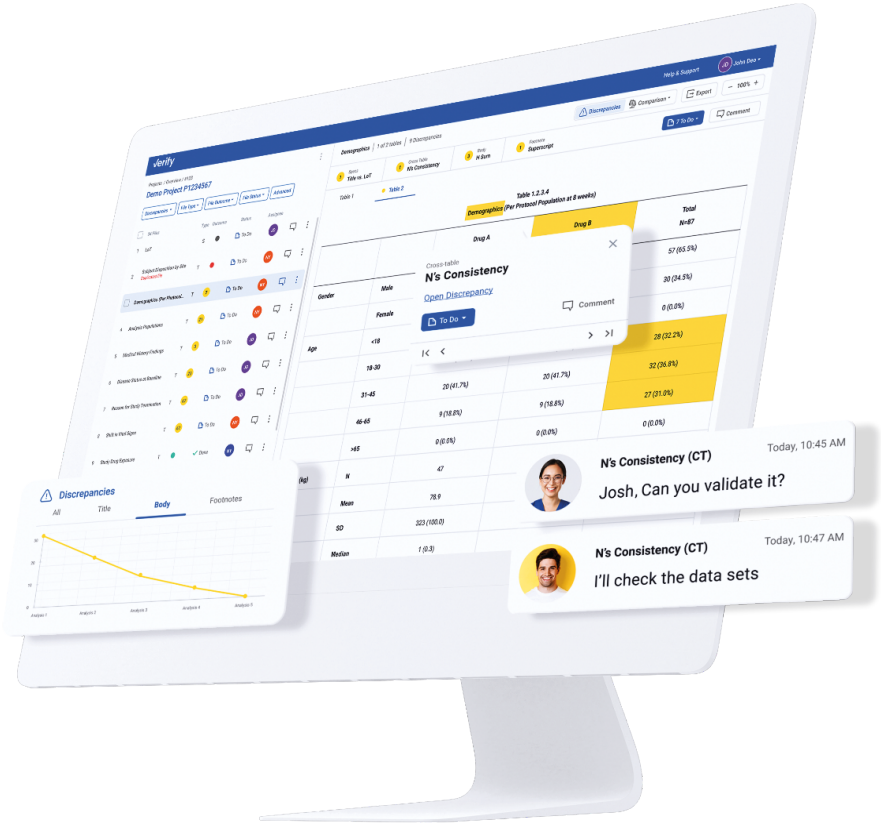
Automating the validation of statistical outputs from clinical trials is a time-consuming process, but the use of Machine Learning (ML) can provide a solution that is more efficient and reliable. ML has been successfully implemented in various industries, and it can have a significant impact on the drug development process. Beaconcure's automated output validation solution, 'Verify', is specifically designed for the pharmaceutical industry and utilizes ML algorithms to read and process clinical data in various formats. The system can identify errors and anomalies with 99.7% accuracy within hours, significantly reducing processing timelines while improving output quality. The system applies various algorithms to the processed data to identify groups and sub-groups and can validate single and cross-table content. The system flags discrepancies and directs the user to the relevant table for the resolution of identified discrepancies.

Beaconcure TECHNOLOGIES

Table of Contents

Introduction

The validation and quality control of the statistical output is an important aspect of the pharmaceutical industry, and it is typically performed manually by clinical programmers and biostatisticians. However, this approach is not only time-consuming but also prone to errors. In a survey conducted by Beaconcure in 2022, which received responses from 15 leading pharmaceutical companies, it was found that the manual review of a single table can take an average of 34 minutes and up to 28 hours for a small study with 50 outputs. These findings demonstrate the need for a more efficient and reliable approach to validate and ensure the quality of statistical output in the pharmaceutical industry. Additionally, the survey revealed that the exchange of study information primarily takes place via email, and tasks are assigned and shared through Excel spreadsheets without a standardized workflow. This further emphasizes the need for an automated and standardized solution to improve the validation and quality control of statistical output in the pharmaceutical industry.



Solution: Verify

Beaconcure collaborated with Pfizer's statistical programming team to develop Verify, an automated analytics platform that leverages machine learning tools to support the quality control (QC) process of clinical data trials. Verify includes features such as cross-table checks, in-table checks, specs checks, and date checks. The software is designed to be intuitive, saving time, reducing human error, and improving quality, ultimately accelerating regulatory approval and time-to-market.

Verify is used by customers worldwide as part of their validation processes to increase data quality and reduce submission time. Due to the power of automation, Verify is now applicable to all therapeutic areas and study phases (I, II, III).

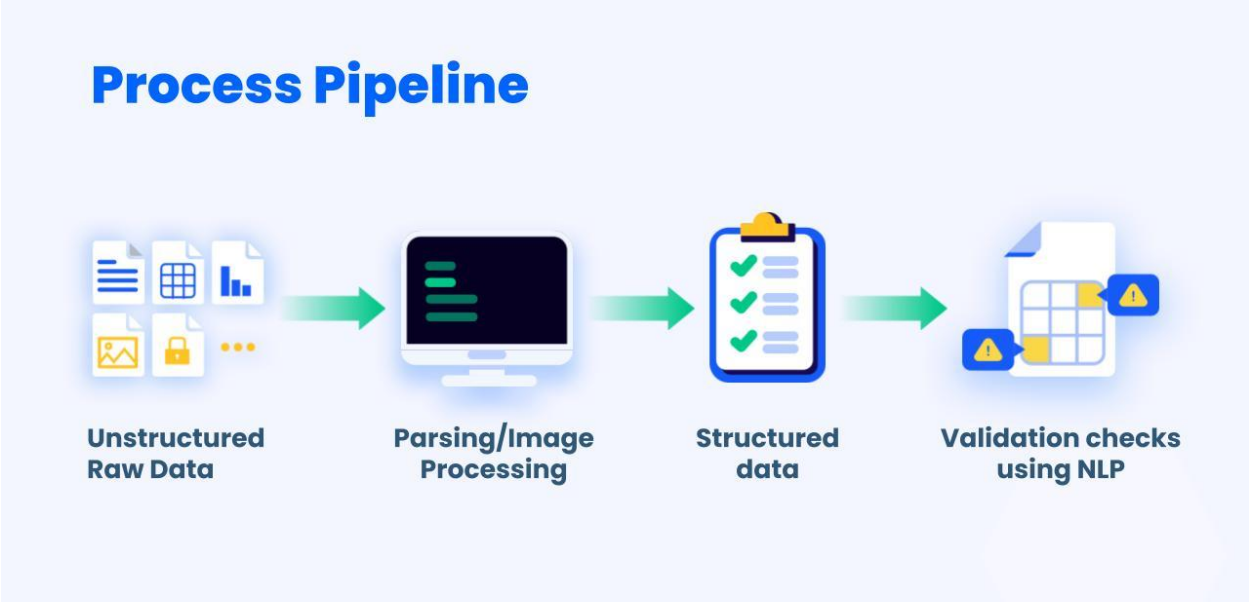


Fig. 1: High Level Validation Process Pipeline

Automated Checks

There are four categories of checks that are performed by Verify:

1. Cross-Table

By performing cross-table checks, numeric values representing the same data point throughout the study's outputs can be validated for consistency. An illustration of a discrepancy in a cross-table check is presented below, where the 'N=' in table headers is compared to a reference table that contains the corresponding population and treatment group.

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
8 weeks ITT	60.0	39.3	49.8
4 weeks (>=1 dose)	47.1	35.0	41.1
8 weeks (>=1 dose)	60.5	39.3	50.0
4 weeks per protocol	57.5	39.3	49.6
8 weeks per protocol	67.3	42.9	57.5

Subjects 001-005 did not receive any treatment and for the ITT analysis is counted as Worsening.
N = number of subjects in the specified group, or the total sample. This value is the denominator for the percentage calculations.
BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021

Table 1: The header cell is highlighted indicating a discrepancy in the 'N' value when compared to a reference table.

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

Table 2: Reference table with the cell reference highlighted, indicating a cross-reference to the 'randomized population' and 'drug A'.

2. Within-Table

Table checks are performed to validate the numeric logic within tables, such as the sum of numeric values and hierarchy checks. An instance of a discrepancy detected in the sum of a mutually exclusive sub-group in the table is demonstrated in the example presented below.

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	
Site 006	18	18	14	13	11	
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	1

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

Table 3: Example of a 'within table' check - discrepancy in a vertical summation is found.

3. Specs

Specs checks involve validating the outputs of source tables against specifications in the LoT and Mock Shell. An example of a discrepancy is presented below, where a difference in the titles was observed between the output table and the Mock table.

ae_over_5_safety.rtf Mock shell

Table 14.1.1.16
Treatment Emergent Adverse Events Occurring in >5% of Subjects in at least One Treatment Group in Study by System Organ Class and Preferred Term (Safety Population)

	Drug A (N=119)	Drug B (N=117)	Total (N=236)
Any TEAEs	61 (51.3%)	69 (59.0%)	130 (55.1%)
Gastrointestinal Disorders	39 (32.8%)	48 (41.0%)	87 (36.5%)
Diarrhoea	25 (21.0%)	35 (29.9%)	60 (25.4%)
Vomiting	20 (16.8%)	30 (25.6%)	50 (21.1%)
Infections & Infestations	10 (8.4%)	10 (8.5%)	20 (8.5%)
Respiratory, Thoracic and Mediastinal Disorders	11 (9.2%)	10 (8.5%)	21 (8.9%)
Cough	10 (8.4%)	8 (6.8%)	18 (7.6%)

N = number of subjects in the specified group, or the total sample. This value is the denominator for the percentage calculations.
Subjects are only counted once per event in each row.
BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021

ae_over_5_safety.rtf Mock shell

Table 14.1.1.16
Treatment Emergent Adverse Events Occurring in >5% of Subjects in at least One Treatment Group by System Organ Class and Preferred Term (Safety Population)
ae_over_5_safety
Table 14.1.1.16
Treatment Emergent Adverse Events Occurring in >5% of Subjects in at least One Treatment Group by System Organ Class and Preferred Term (Safety Population)

N = number of subjects in the specified group, or the total sample. This value is the denominator for the percentage calculations.
Subjects are only counted once per event in each row.
BEACONCURE CONFIDENTIAL SDTM Creation: DDMMMYYYY (HH:MM) Source Data: advs Table Generation: DDMMMYYYY

Table 4: difference in the title between the output table and the mock table results in a discrepancy and is highlighted in Verify. The words 'in study' do not appear in the mock shell title.

The Benefits of Verify for Clinical Study Deliverables Management

In today's industry, Excel spreadsheets are commonly used for deliverables management. While effective in presenting data discrepancies as a figure, it lacks the context of each anomaly, which is crucial for faster resolution. With the full context provided, programmers can understand the source of each discrepancy and eliminate it more efficiently.

By using Verify, the following challenges are addressed through its visualization capabilities:

- Project (clinical study) tracking
- Discrepancy tracking
- Concurrent work due to version misalignments

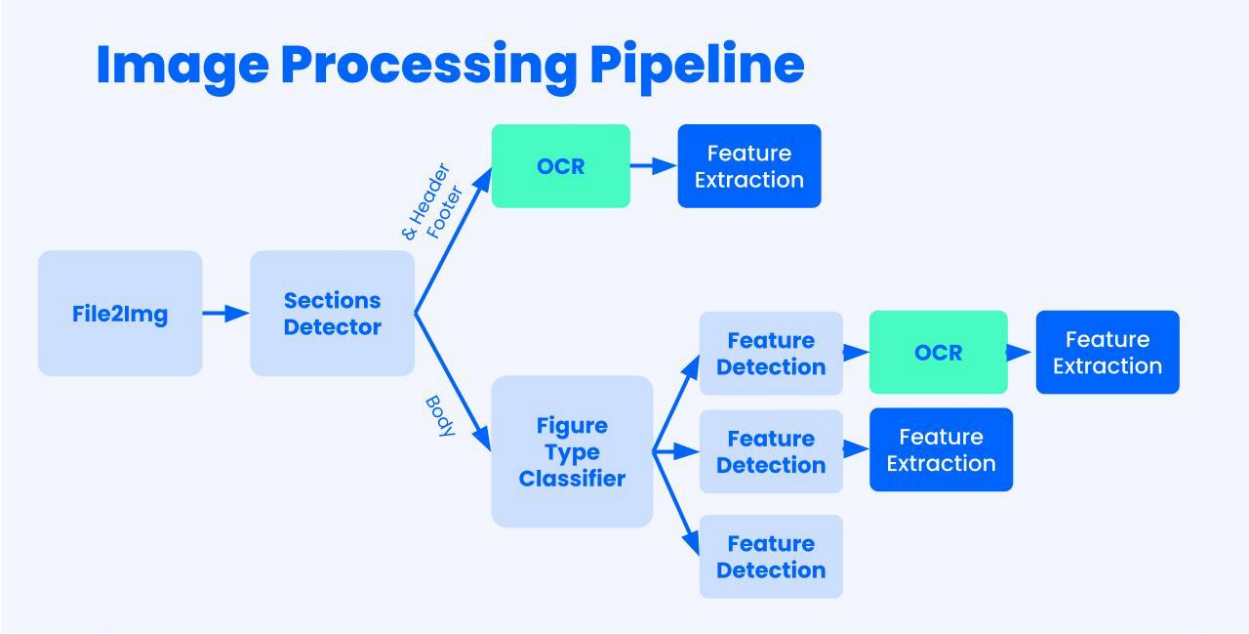


Fig. 2: Image Process Pipeline

Conclusion

The Verify platform has been extensively utilized in the processing of over 100 studies, which provides significant insight into the efficacy of the platform in supporting programmers and statisticians. Based on this experience, it is clear that the utilization of automation within Verify serves to de-risk the submission process through the provision of increased quality in data analysis and reporting.

Furthermore, Verify's ability to visualize discrepancies between the data being analyzed and the anticipated outcomes has proven to be a valuable tool in promoting team discussions and collaboration within the platform. The transparent presentation of results enables team members to identify any issues promptly and to work towards developing effective solutions to any challenges that may arise during the data analysis process.

Another notable benefit of Verify is the capacity to identify issues in content and format early in the process, which leads to accelerated delivery timelines. This early identification allows for prompt and efficient resolution of any issues, which contributes to the overall efficiency of the data analysis process.

Overall, the use of Verify-in data analysis and reporting has been a significant asset to research teams, contributing to increased quality, improved collaboration and communication, and more efficient timelines for project completion. These benefits highlight the value of incorporating automation into research processes to streamline workflows, reduce errors, and improve overall outcomes.

Contact Information

Beaconcure Technologies, Inc.

Israel
Israel Beck st 34
Tel Aviv

USA
117 Kendrick St, Needhm, MA

Email: hello@beaconcure.com
Web: www.beaconcure.com