

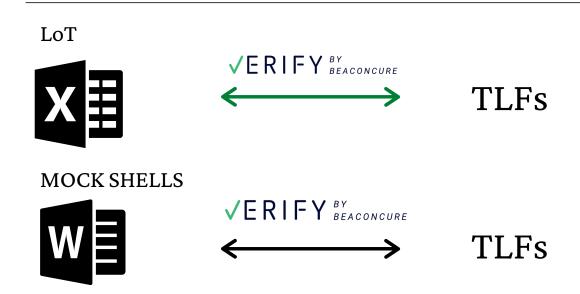
New Uses for Automation in the **Output Validation Process**

Ilan Carmeli, Hugh Donovan, Beaconcure

The tables that appear in the deliverable are the result of a four-step process:

- 1. Specification of the List of Tables (LoT) which defines the tables to be produced
- 2. Definition of the Tables, Listings and Figures Shells (TLFs) as a part of the Statistical Analysis Plan or a separate document
- 3. Generation of the tables
- 4. Generation of the in-text tables

Each step can be checked using automated checks incorporating machine learning and artificial intelligence. Similar techniques can be used to check for consistency of the deliverables from each step, e.g. are the in-text tables discrepant from the end-of-text tables, are the table shells the same as the tables, and have all tables in the LoT been produced with the same headings.



The current gold standard for validation in the industry is 100% double programming of all analysis datasets, tables, listings and figures. In a survey conducted by Beaconcure all 25 respondents use double programming for at least some of the outputs.

This approach, however, does not resolve cross-table discrepancies, and typically does not account for unanticipated data changes during clinical trials. Also, as it is a manual exercise, it is also subject to human error.

Today, with the support of ML-driven technology, the industry would do well to consider automating the output validation process.

An automation solution developed by Beaconcure, checks outputs in exactly the same way as figures within tables are commonly compared today. This technology can be used multiple times as data accumulates, identifying discrepancies in the output. One example of within table checks and discrepancies are shown below:

WITHIN TABLE VALIDATION

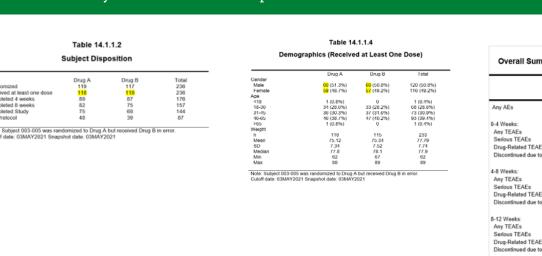
		Table 6 Drug Protocol Demo Shift Table of Lab Data - White Blood Cell Count at Last Visit Drug A (N=118)					
		Missing	Below Normal	Within Normal	Above Normal	Total	
Baseline	Missing	1 (0.8%)	0 (0.0%)	2 (1.7%)	0 (0.0%)	3 (2.5%)	
WBC	Below Normal Range	0 (0.0%)	12 (10.1%)	0 (0.0%)	0 (0.0%)	12 (10.1%)	
	Within Normal Range	0 (0.0%)	3 (2.5%)	63 (52.9%)	6 (5.0%)	72 (60.5%	
	Above Normal Range	0 (0.0%)	0 (0.0%)	12 (10.1%)	20 (16.8%)	32 (26.9%	
	Total	1 (0.8%)	15 (12.6%)	77 (64.7%)	26 (21.8%)	118 (99.2%)	

Despite its limitations, double programming is used for individual tables, but it does not compare the output across tables. Within an individual table, the two programs may produce the same answer, but it may not be correct. This is something we have seen in our analysis of tables. To overcome this limitation of double programming, statisticians perform a manual check for consistency across tables. While this will improve quality, there are shortcomings.

Verify, in addition to performing within table checks, can perform cross-table checks quickly and consistently for all deliverables. This is achieved by running a set of standard cross table checks.

CROSS TABLE CHECK

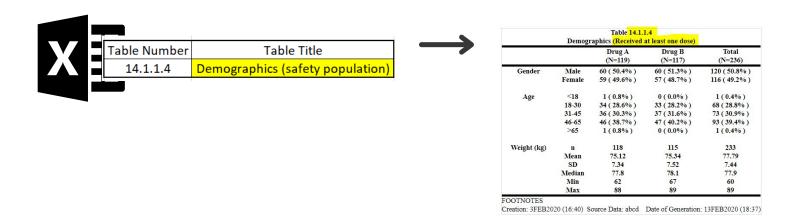
The discrepancies are highlighted in the output so that the reviewer can immediately see where the discrepancies are.



	Drug A (N-119)	Drug B (N-117)	Total (N=23
Any AEs	73 (61.3%)	88 (75.2%)	161 (68.
0-4 Weeks:			
Any TEAEs	75 (63.0%)	85 (72.6%)	156 (66.
Serious TEAEs	3 (2.5%)	17 (14.3%)	7 (5.99
Drug-Related TEAEs	17 (14.3%)	27 (23.1%)	44 (18.6
Discontinued due to TEAEs	7 (5.9%)	9 (7.7%)	16 (6.8
4-8 Weeks:			
Any TEAEs	48 (40.3%)	38 (32.5%)	86 (36.4
Serious TEAEs	1 (0.8%)	3 (2.6%)	4 (1.79
Drug-Related TEAEs	12 (10.1%)	20 (17.1%)	40 (16.9
Discontinued due to TEAEs	3 (2.5%)	4 (3.4%)	7 (3.09
8-12 Weeks:			
Any TEAEs	25 (21.0%)	20 (17.1%)	45 (19.1
Serious TEAEs	0 (0.0%)	1 (0.9%)	1 (0.49
Drug-Related TEAEs	7 (5.9%)	5 (4.3%)	12 (5.1
Discontinued due to TEAEs	1 (0.8%)	1 (0.9%)	2 (0.89
0-12 Weeks:			
Any TEAEs	83 (69.7%)	93 (79.5%)	176 (74)
Serious TEAEs	3 (2.5%)	7 (6.0%)	10 (4.2
Drug-Related TEAEs	28 (23.5%)	30 (25.6%)	58 (24.6
Discontinued due to TEAEs	11 (9.2%)	14 (12.0%)	25 (10.6

LoT

Specification of the List of Tables (LoT) which defines the tables to be produced is compared to identify discrepant titles as shown below:



CONCLUTION

The algorithms determine whether a certain check was relevant for particular tables, so there is no need to do study specific configuration. If a new check is required, it can be added to the standard set so it is available for all studies, not just one. This automated approach requires a statistician to resolve discrepancies raised by Verify, a task often requiring judgment and knowledge of the study. There would also still be a benefit to a statistician reviewing the main results of the study at a high level, for example are the results in the tables consistent with the p-values. Automation, while not replacing the statistician can:

- Increase the quality of statistical analysis
- Check the output comprehensively and consistently, both within and across tables
- Greatly reduce the time and effort to perform checks
- Improve the productivity of statisticians and improve their job satisfaction
- Eliminate all manual steps in the output validation process workflow