



PHUSE 2021 Live Presentation

How ML-driven Automation Technology Can Overcome
the Limitations of Double Programming

Ilan Carmeli - Co-Founder and Chief Product Officer, Beaconcure
Hugh Donovan - Former EVP clinical research services, Parexel



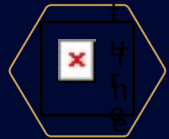
Topics

-  Key Findings From a 'Statistical Programming Output Validation Survey 2021' - Hugh
-  Double Programming Limitations - Hugh
-  New Approaches - The Power of Automation - Ilan

The Limitations of Double Programming



It is not truly independent programming



It typically involves a lot of manual checking



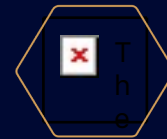
It is typically only applied if the data changes, not if the format changes



Although the numbers maybe the same, they may be wrong



Misinterpretation of the specifications



The double programming only works within a table

How to Apply ML Technology for Validating Outputs?

Main Technical Challenges

Clinical outputs are digital files designed for the **human eye**

Main challenges:

1. There is no standard for clinical tables in the life sciences industry
2. Different digital data formats (TXT, RPT, HTML, RTF, PDF, Docx)
3. Data includes free text (design for the human eye)
4. No machine-readable standard format
5. Unstructured tables
6. Metadata is missing

Define Your Goals

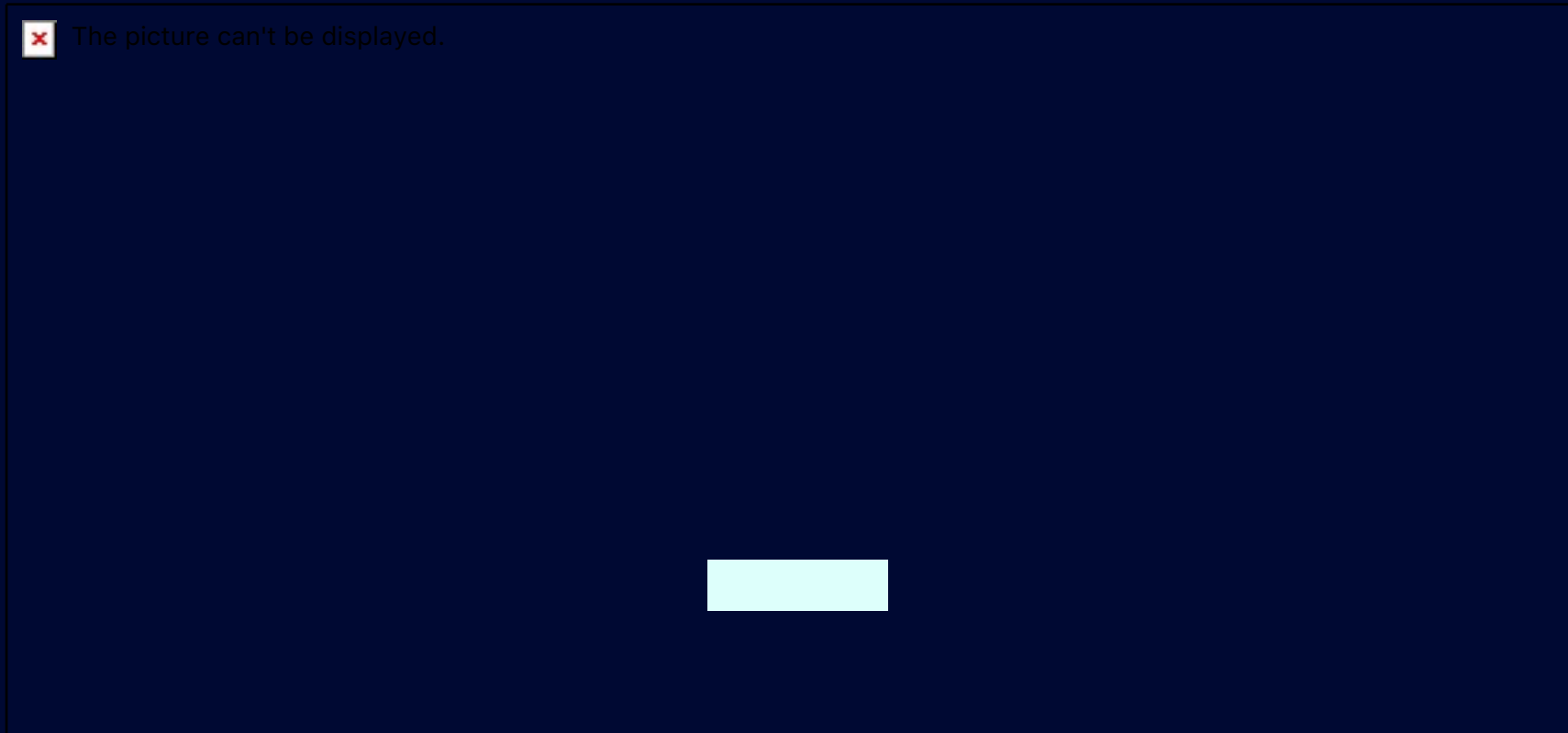
1. Accuracy sensitivity
2. What data type would you like to validate?
3. What kind of checks do you want to perform?
4. Get your ROI in place (Quality vs Cost vs Timelines)

Standardization

Develop algorithms to standardize your data that will support data type, outputs structure, clinical phase and TA.

Table Parsing

Parsing each format to convert it into a coherent, unified representation



Meta-Data Extraction

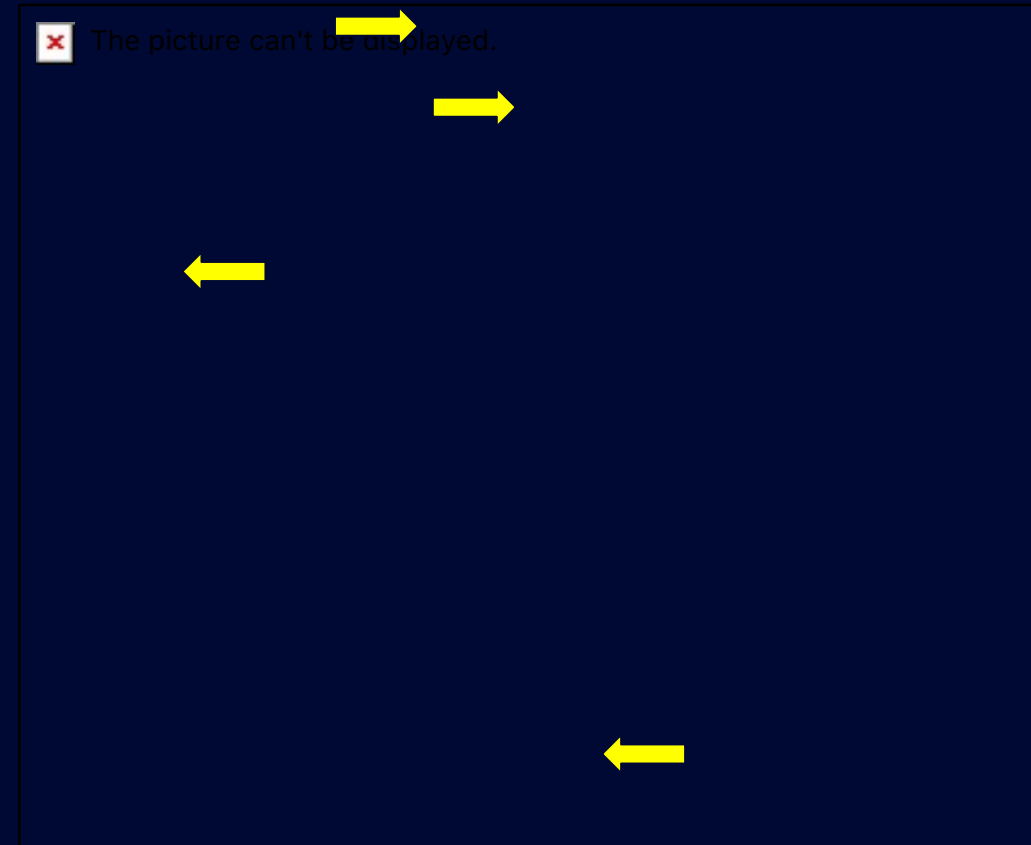
Extract meta-data automatically to enrich your data.
For example, data hierarchy, data type, etc.

Understanding the Content

Understand table data from:

- Headers that indicate cell content
- Table titles
- Table footnotes
- Table hierarchy

Example of a Post-Text table



Presentation in the Database

SAS File Output

The picture can't be displayed.

The picture can't be displayed.

Table number: 14.1.2.4
Table title: Overall Summary of Treatment Emergent Adverse Events (Safety Population)
Population: Safety
Table category: adverse events
Cell type: Data cell
Cell value: 3
Data type: number, decimal
Column Headers: [Drug B N=117]
Rows headers: [4-8 weeks, Serious TEAEs]
Arm treatment: Drug B
Subjects denominator (N): 117
Time: 4-8 weeks
AE category: Serious
Cutoff date: 5.3.2021
Snapshot date: 5.3.2021

Generalized Validation Checks

1. Review your list of checks for validating the outputs
2. Make sure that your data support such checks
3. Break complex and large checks to small pieces
4. Generalize your checks to support all studies and TA

Feedback from Clinical SME

Make sure to work closely with your SME.
The SMEs will be a very important piece in the automation efforts.

Example of Check Based ML

The 'Received at least one dose' population is not consistent between tables



The picture can't be displayed.

Table 14.1.1.4
Demographics (Received at Least One Dose)

	Drug A	Drug B	Total
Gender			
Male	60 (51.3%)	60 (50.8%)	120 (50.8%)
Female	59 (48.7%)	57 (49.2%)	116 (49.2%)
Age			
<18	1 (0.8%)	0	1 (0.4%)
18-30	34 (28.6%)	33 (28.2%)	68 (28.8%)
31-45	36 (30.3%)	37 (31.6%)	73 (30.9%)
46-65	46 (38.7%)	47 (40.2%)	93 (39.4%)
>65	1 (0.8%)	0	1 (0.4%)
Weight			
n	118	115	233
Mean	75.12	75.34	77.79
SD	7.34	7.52	7.74
Median	77.8	78.1	77.9
Min	62	67	62
Max	88	89	89

Note: Subject 003-005 was randomized to Drug A but received Drug B in error.
Cutoff date: 03MAY2021 Snapshot date: 03MAY2021

Table 14.1.2.4
Overall Summary of Treatment Emergent Adverse Events
(Safety Population)

	Drug A (N=119)	Drug B (N=117)	Total (N=236)
Any AEs	73 (61.3%)	88 (75.2%)	161 (68.2%)
0-4 Weeks:			
Any TEAEs	75 (63.0%)	85 (72.6%)	156 (66.1%)
Serious TEAEs	3 (2.5%)	17 (14.3%)	7 (5.9%)
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 (1.7%)
Drug-Related TEAEs	12 (10.1%)	20 (17.1%)	40 (16.9%)
Discontinued due to TEAEs	3 (2.5%)	4 (3.4%)	7 (3.0%)
8-12 Weeks:			
Any TEAEs	25 (21.0%)	20 (17.1%)	45 (19.1%)
Serious TEAEs	0 (0.0%)	1 (0.9%)	1 (0.4%)
Drug-Related TEAEs	7 (5.9%)	5 (4.3%)	12 (5.1%)
Discontinued due to TEAEs	1 (0.8%)	1 (0.9%)	2 (0.8%)
0-12 Weeks:			
Any TEAEs	83 (69.7%)	93 (79.5%)	176 (74.6%)
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%)

Note: Subject 003-005 was randomized to Drug A but received Drug B in error.
The Safety Population is all subject who received at least one dose of medication.
Any AE is any adverse event that began after the beginning of the screening period.
Any TEAE is any adverse event that began after the first dose.
Cutoff date: 03MAY2021 Snapshot date: 03MAY2021

The Benefits of Using a Software Based on ML

‘Automation of the validation process, supported by ML, reduces complexity, increases quality, streamlines business processes & workflows, and increases efficiency’



THANK YOU

Ilan@beaconcure.com

