Can Automation Improve Compliance, Consistency, & Efficiency for CGTP Data Analysis

Stephanie Pasas-Farmer, PhD[1]; Carrie Vyhlidal, PhD[2]
[1]BioData Solutions, LLC, [2]KCAS Bio

Introduction

With the promise of treating cancer, Parkinson's, and other rare diseases, cell and gene therapy products (CGTPs) are gaining traction in the biopharma industry. As of 2023, the FDA has approved 34 CGTPs and over 1500 clinical trials are currently ongoing globally. Despite the increasing number of CGTP drug candidates, the bioanalysis supporting these programs is still fraught with many hurdles. These challenges include limited or no well-defined guidance; lack of harmonization of validation approaches and acceptance criteria; and limited software tools to facilitate the management of the high volume of bioanalytical data being generated. Artificial intelligence (AI) offers a unique opportunity to alleviate these data analysis and throughput challenges by automating key processes that reduce time and increase efficiencies while maintaining data security, integrity, and compliance, such as 21 CFR part 11. More importantly, AI-enabled auditing data tools can facilitate consistency in scientific approaches and standardization of acceptance criteria.

Challenges

While molecular assays are not new to the world of drug development, their application in CGTPs is more recent. Consequently, the regulations are either not specific or have yet to be established. Validation approaches and acceptance criteria lack harmonization. This results in inconsistencies in data analysis across and within studies. Additionally, a lack of tools for tabulating the high volumes of raw data used for FDA submissions often leads to transcription errors, data analysis errors, and overburdened quality control staff.

Hypothesis and Workflow Solution

To overcome data analysis challenges facing molecular assays in highly regulated environments, an AI-enabled automation prototype was developed for evaluation. The tool would replace the common manual, repetitive, error-prone tasks and increase the efficiency of the laboratory scientists while improving the quality and compliance of the results and requiring minimal training. Firstly, a scientist would upload text/csv data file(s) to a validated and secure system with some minimal input fields. Due to the lack of harmonization around acceptance criteria, the system would permit study-specific customizable acceptance criteria to accommodate discovery data. Next, automated data analysis would substitute human processes of flagging any data outside of acceptance criteria per regulatory guidance and industry best practices; tabulating raw data to reportable tables; and interim quality control (QC) and quality assurance (QA) checks for data transcription. Finally, data output would be generated.

Results and Conclusion

The process of performing data calculation, creating reportable excel tables, and performing interim checks was completed in less than 1 minute. To make it more robust, R code simultaneously validated all calculations for additional validity and data integrity. It eliminated the need for interim QC and QA checks for transcription errors and data calculations as the data from the instrument is uploaded directly to the tool and converted to reportable output.

Based on data gathered from end users, the automation tool prototype yielded time savings of 96.4%, going from 5.8 FTE to 0.2 FTE (46.4 work hours to 1.6 work hours) for a project with 15 plates. The reported times include extracting reports from instruments, uploading files, performing calculations, completing interim QC and QA checks, and populating the tables. Additionally, cost savings from the time saved were 97.3% and despite customization costs, an ROI of 6X was observed. Some unmeasurable benefits of automated data analysis and tabulation include the time given back to the scientists to focus on more important tasks that require human discretion and decision-making capabilities.

97.3% COST SAVINGS

6X RETURN ON INVESTMENT

96.4% TIME SAVINGS

>99% REDUCTION IN TRANSCRIPTION & CALCULATION ERRORS

Other Benefits

Increased compliance and consistent approach to data analysis with each dataset

Figure 1. Results observed by developing a prototype of automation tool for CGTP data analysis.

Assay Date	**Run ID**	**LLOQ (10.0)**	**% RE**	**LQC (30.0)**	**% RE**	**MQC (2000.0)**	**% RE**	**HQC (800000.0)**	**% RE**	**ULOQ (1000000.0)**	**% RE**
11-Nov-2023	†DEMO_Run 001	9.621	-3.8	25.340	-15.5	2563.835	28.2	580820.806	-27.4	818802.492	-18.1
		7.729	-22.7	30.106	0.4	2082.506	4.1	604776.525	-24.4	828755.394	-17.1
12-Nov-2023	†DEMO_Run 002	10.846	8.5	37.271	24.2	▲1545.606	-22.7	834486.972	4.3	997907.835	-0.2
		10.876	8.8	31.931	6.4	2000.493	0.0	796732.662	-0.4	1006882.721	0.7
15-Nov-2023	DEMO_Run 005	13.725	37.2	31.188	4.0	1911.561	-4.4	753584.268	-5.8	928022.063	-7.2
		14.307	43.1	30.886	3.0	2031.282	1.6	802941.496	0.4	928050.509	-7.2
		14.289	42.9	26.989	-10.0	1820.932	-9.0	783090.560	-2.1	915585.283	-8.4
16-Nov-2023	†DEMO_Run 006	13.601	36.0	35.656	18.9	2207.174	10.4	750309.051	-6.2	972038.696	-2.8
		9.293	-7.1	40.441	34.8	2140.596	7.0	770427.264	-3.7	989099.962	-1.1
17-Nov-2023	DEMO_Run 007	16.704	#67.0	31.183	3.9	1911.005	-4.4	753232.349	-5.8	927582.986	-7.2
		14.305	43.1	30.881	2.9	2030.688	1.5	802565.024	0.3	927611.417	-7.2
18-Nov-2023	DEMO_Run 008	NR	NA	▲47.699	#59.0	1883.038	-5.8	832806.951	4.1	956174.396	-4.4
		13.552	35.5	35.848	19.5	NR	NA	▲810204.486	1.3	889542.770	-11.0
		101.574	#915.7	37.273	24.2	1931.844	-3.4	794449.719	-0.7	903147.931	-9.7
		13.588	35.9	32.238	7.5	2005.041	0.3	*Masked	NA	935218.013	-6.5
		12.608	26.1	36.045	20.2	2382.316	19.1	801436.196	0.2	926055.660	-7.4
		13.886	38.9	26.989	-10.0	1986.064	-0.7	902120.441	12.8	901677.633	-9.8
Mean		22.854		33.383		1989.377		803643.149		921697.151	
SD		27.679		5.818		153.951		42363.440		18321.015	
%CV		**121.1		17.4		7.7		5.3		2.0	
%RE		##128.5		11.3		-0.5		0.5		-7.8	
N		10		11		10		10		11	

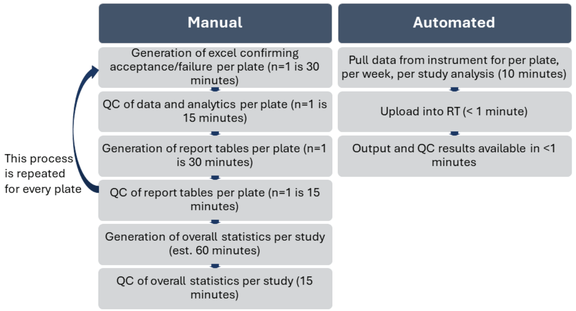
▲Mean of N = 2 values reported
NR: Not Reportable due to insufficient number of replicates.
*Masked, %CV between replicates not within acceptance criteria
#%RE outside of acceptance criteria.
** Overall %CV outside of acceptance criteria.
Overall %RE outside of acceptance criteria.
†Run DEMO_Run 001, DEMO_Run 002, DEMO_Run 006 does not meet all batch run acceptance criteria, thus the data is shown for transparency and not included in the statistical calculations.

Figure 3. QC Samples Table as generated by the qPCR prototype. The tool converted data from several run files into data tables for Regression Data, Negative Target Control Data, Calibration Standard Data, QC Samples, and True Unknown (Sample) Data. Using regulatory compliance and industry best practices, it performed statistical analysis to calculate Mean, S.D., %CV, %RE, and N values against the raw data measured by the instrument.

File Name	DEMO_Run 001_MM_20231111.eds										
Analysis Date/Time	2023-11-11 03:09:26 PM CDT										
Exported On	2023-11-11 03:10:15 PM CDT										
Well Position	**Sample**	**Quantity**	**Target**	**Task**	**Cq**	**Cq Mean**	**Cq Confide**	**Cq SD**	**Y-Interc**	**R2**	**Slope**
A01	Std 1	1000000	GENERIC_target	STANDARD	16.2	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A02	Std 1	1000000	GENERIC_target	STANDARD	16.2	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A03	Std 1	1000000	GENERIC_target	STANDARD	16.2	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A04	NTC		GENERIC_target	NTC	31.1	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A05	NTC		GENERIC_target	NTC	31.1	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A06	NTC		GENERIC_target	NTC	31.2	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A07	Sample 01	Redacted	GENERIC_target	UNKNOWN	18.8	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A08	Sample 01	Redacted	GENERIC_target	UNKNOWN	18.9	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A09	Sample 01	Redacted	GENERIC_target	UNKNOWN	18.6	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A10	Sample 02	Redacted	GENERIC_target	UNKNOWN	19	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A11	Sample 02	Redacted	GENERIC_target	UNKNOWN	19	Redacted	Redacted	Redacted	38.87	0.98	-3.8
A12	Sample 02	Redacted	GENERIC_target	UNKNOWN	18.9	Redacted	Redacted	Redacted	38.87	0.98	-3.8
B01	Std 2	100000	GENERIC_target	STANDARD	19.9	Redacted	Redacted	Redacted	38.87	0.98	-3.8
B02	Std 2	100000	GENERIC_target	STANDARD	19.9	Redacted	Redacted	Redacted	38.87	0.98	-3.8
B03	Std 2	100000	GENERIC_target	STANDARD	20	Redacted	Redacted	Redacted	38.87	0.98	-3.8
C04	ULOQ		GENERIC_target	POSITIVE_COI	16.4	Redacted	Redacted	Redacted	38.87	0.98	-3.8
C05	ULOQ		GENERIC_target	POSITIVE_COI	16.4	Redacted	Redacted	Redacted	38.87	0.98	-3.8
C06	ULOQ		GENERIC_target	POSITIVE_COI	16.4	Redacted	Redacted	Redacted	38.87	0.98	-3.8
C07	Sample 06	Redacted	GENERIC_target	UNKNOWN	18.1	Redacted	Redacted	Redacted	38.87	0.98	-3.8
C08	Sample 06	Redacted	GENERIC_target	UNKNOWN	18.3	Redacted	Redacted	Redacted	38.87	0.98	-3.8
C09	Sample 06	Redacted	GENERIC_target	UNKNOWN	18.2	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D04	HQC		GENERIC_target	POSITIVE_COI	17	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D05	HQC		GENERIC_target	POSITIVE_COI	17	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D06	HQC		GENERIC_target	POSITIVE_COI	17	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D07	Sample 08	Redacted	GENERIC_target	UNKNOWN	18.7	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D08	Sample 08	Redacted	GENERIC_target	UNKNOWN	18.6	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D09	Sample 08	Redacted	GENERIC_target	UNKNOWN	18.5	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D10	ULOQ (2)		GENERIC_target	POSITIVE_COI	16.4	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D11	ULOQ (2)		GENERIC_target	POSITIVE_COI	16.4	Redacted	Redacted	Redacted	38.87	0.98	-3.8
D12	ULOQ (2)		GENERIC_target	POSITIVE_COI	16.4	Redacted	Redacted	Redacted	38.87	0.98	-3.8

Figure 2. A highly redacted sample of a qPCR input file with limited attributes.

Workflow Comparison — Manual v. Automated Process



*Estimates based on 96-well plate

Figure 4. The workflow above highlights time savings achieved by automating the manual process of data compilation by the scientists and review by Quality Control team. The time savings can instead be used toward critical activities, such as experimental design, troubleshooting, or other critical decision-making tasks.

BIODATA SOLUTIONS Powered by Red Thread

CONTACT
Stephanie Pasas-Farmer, PhD
BioData Solutions, LLC
info@bdatasolutions.com
www.bdatasolutions.com
785-679-1600

REFERENCES
1. FDA. (n.d.). Approved Cellular and Gene Therapy Products. 2. Lohr, A. (2023). 2023's Market Outlook For Cell And Gene Therapies. Cell & Gene.
3. Hays, A., et al (2024). Recommendations for Method Development and Validation of qPCR and dPCR Assays in Support of Cell and Gene Therapy Drug Development. The AAPS Journal, 26(24)

CONTACT
Carrie Vyhlidal, PhD
KCAS Bio
Carrie.Vyhlidal@kcasbio.com
www.kcasbio.com
913-248-3038

kcas bio SCIENCE ACCELERATED