A Case Study Conducted With

arrowhead pharmaceuticals

BIODATA® SOLUTIONS
Powered by Red Thread

Simplifying Bioanalytical Review Using Advanced Computational Techniques

Fast-Growing Pharma Automates Bioanalysis Reviews for Outsourced Studies

bdatasolutions.com

SUCCESS HIGHLIGHTS

Pharmacokinetic Small Molecule Method Validation Module

Faster reviews: 50% decrease in time to review reports

Reduction in potentially required rework: 2,000% more red flags and 3,500% more yellow flags than manual processes

Increased compliance: Red Thread® checks for 92% FDA guidelines established for Accuracy & Precision, Sensitivity, and Stability tests

ARROWHEAD PHARMACEUTICALS AT-A-GLANCE

Arrowhead Pharmaceuticals develops medicines that treat intractable diseases by silencing the genes that cause them. Arrowhead therapies trigger the RNA interference mechanism to induce rapid, deep, and durable knockdown of target genes.

- Headquartered in Pasadena, CA
- Proprietary technology plat form called Targeted RNAi Molecules (TRiM™)
- Broad pipeline of early, mid, and later-stage candidates targeting rare and high prevalence diseases

BIODATA® SOLUTIONS
Powered by Red Thread

BUDGET AND RESOURCE LIMITATIONS DELAY REVIEWS

Arrowhead Pharmaceuticals develops RNAi-based therapies that treat intractable diseases by silencing the genes that cause them. As an emerging pharma with finite resources, Arrowhead outsources key components of the R&D process such as bioanalysis to contract research organizations (CROs).

To comply with regulatory oversight requirements, Txheng Yang, Associate Director, Bioanalytical Development at Arrowhead, needed to review bioanalytical study reports from CROs and perform Sponsor quality control (QC) reviews. For an organization in early development stage at the time, Txheng sought an alternative solution in order to maximize efficiencies and minimize costs.

"Finding a fast, efficient way to accurately review validation and sample analysis study data and reports was a top priority due to our growing backlog. Red Thread® was the solution that met our needs and enabled our scientists to spend time on more valuable activities," says Txheng.

WHAT IS RED THREAD?

Red Thread optimizes bioanalytical reviews by leveraging expert systems and computer vision components of artificial intelligence (AI) to read and audit data tables, facilitate remote and thorough data reviews, and proactively highlight issues to reduce downstream errors, while ensuring the highest levels of regulatory compliance. Using AI, it brings objectivity and rigor into the process of data review and auditing for peer reviews, QC checks, and QA audits.

After hearing about Red Thread by Ariadne Software® at an industry conference, Arrowhead decided to implement the innovative technology. "Red Thread helps us perform objective scientific reviews with the staff we already have," says Txheng.

Red Thread uses expert systems, a form of artificial intelligence that mimics the decision-making capabilities and expertise of a human, to perform data review by developing if-then rules to flag data green, red, or yellow based primarily on regulatory guidance. In cases where regulatory guidance is ambiguous, industry best practices dictate the if-then rules. The user then reviews the findings within the context of experimental design and makes the ultimate decision on whether to accept the findings or perform additional experiments or rework. The current version of Red Thread reads tabular data, where most of the critical information is presented within a report.

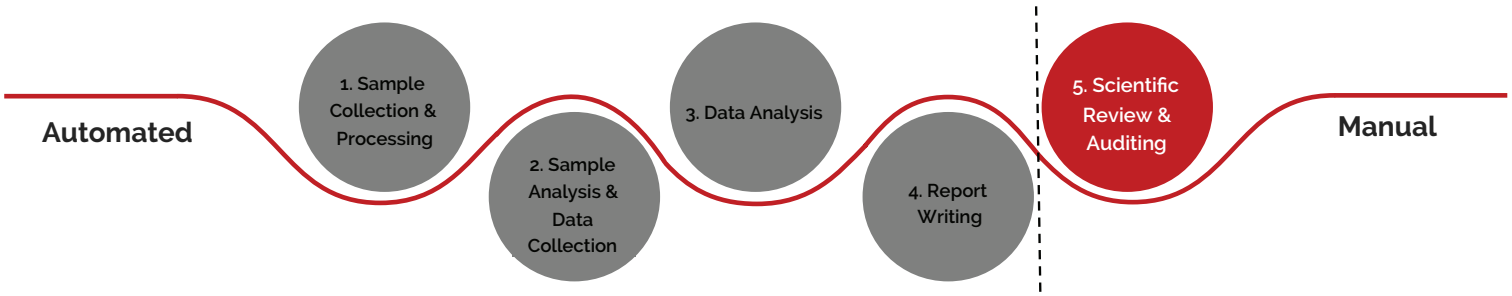
Red Thread's Alerting System

- **GREEN** Data is compliant with regulatory guidelines and/or best practices
- **YELLOW** Areas of risk
  - Data with no specific guidance or gaps in guidance, and therefore the data is subjectively interpreted
  - Recurring failure trends (data is near failure for the same test multiple times during a study or validation indicating a possible weakness in the assay or study conduct issues by the scientific staff)
  - Near-failure patterns (data that is near failure but still meets criteria, indicating possible risk of future failure)
  - Non-compliance with best practices (data that might require additional context before making a decision)
- **RED** Critical findings, such as missing data or values outside of acceptance criteria, that cause non-compliance with industry guidelines or best practices

RED THREAD SOLUTION OVERVIEW AT ARROWHEAD

Arrowhead uses Red Thread's small molecule method validation and sample analysis modules to verify toxicokinetic preclinical data reviews. "It's a very quick process for us to upload the data we receive from our CROs to the cloud. The Red Thread report generates in as little as five minutes, provides helpful flags that guide us deeper in our reviews, and reduces review time by more than half, with very little training," Txheng says. Arrowhead also takes advantage of Ariadne's service to assist with results interpretation when resources are unable to allocate time to this activity. "Ariadne is easy to work with and a truly collaborative partner, supporting us when we need it," says Txheng.

Red Thread: Serving an Unmet Industry Need



BIODATA® SOLUTIONS
Powered by Red Thread

Critical Data Review, Simplified

Process chart

Critical Data Review, Simplified

Steps in grey represent User actions and steps in red represent Red Thread's processing of the uploaded data set.

METHODOLOGY

Alexander Behling, Chemist II, and Jed Dallas, Associate Chemist II, of Arrowhead validated four bioanalytical methods in two species, rat and monkey, across both high and low linear dynamic ranges. Of the four method validation reports, two were audited manually by Bioanalyst A, and two were audited with Red Thread and reviewed by Bioanalyst B. The reports were reviewed for compliance with acceptance criteria established for all applicable bioanalytical parameters outlined in the FDA/EMA guidance (or industry best practices when guidance may be open to interpretation).

To ensure a blinded study, the reviewer auditing the output report generated by Red Thread (Bioanalyst B) received a brief demonstration of the capabilities of Red Thread and was asked to evaluate the results for two of the four reports to assess if Red Thread could meet their needs as a possible bioanalytical review tool. Similarly, the reviewer performing manual review (Bioanalyst A) was instructed to highlight all areas of concern (yellow flags) within the document. To encourage a fair comparison, Bioanalyst A was provided a very brief overview of Red Thread yellow flags. Neither bioanalyst saw each other's findings to ensure objectivity.

Arrowhead measured Red Thread's performance on three metrics:

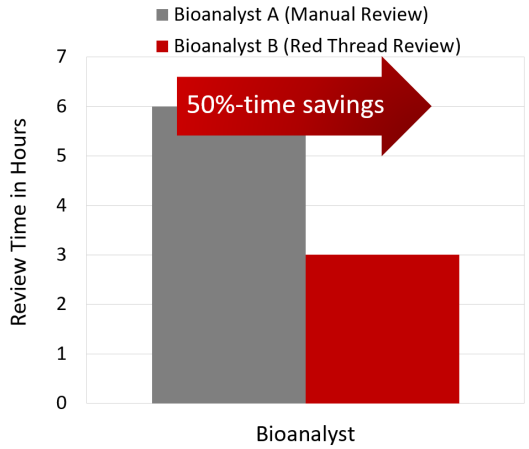
1. Improved efficiencies: Red Thread delivers time savings by enabling available resources to better manage workload and increase throughput. To measure time savings, the average time to perform a manual review was compared to the average time to run and review an output report generated by Red Thread.
2. Reduction in potentially required rework: To measure the improvement in data quality and reduction in missed findings/errors/trends that may lead to potential rework in sample analysis or clinical stages, Arrowhead compared manual and Red Thread findings by counting red and yellow flags reported by Bioanalyst A in manual analysis to those highlighted in the Red Thread output report.
3. Increased compliance: Arrowhead selected three bioanalytical parameters - accuracy and precision (A&P), sensitivity, and stability - to assess compliance against the FDA guidelines outlined in the 2018 guidance document. Where the guidance is open to interpretation, Red Thread is designed to audit data with the most conservative interpretation of the guidance, which may differ from internal standard operating procedures (SOPs). An exception to this described procedure would be when Red Thread is customized per the client's needs or SOPs.

BENEFITS AND RESULTS

Red Thread has enabled Arrowhead to reduce resource commitment, improve transparency with CROs, and provide data with higher compliance, quality, and integrity.

IMPROVED EFFICIENCIES

Due to the blinded nature of the study, Bioanalyst B had minimal training on how to analyze Red Thread outputs. Despite minimal training and lack of familiarity, Arrowhead experienced a 50% decrease in review time for the low and high range rat method validation reports for a preclinical toxicokinetic study. With additional training and repeated use of Red Thread, increased time efficiency is expected.



REDUCTION IN POTENTIALLY REQUIRED REWORK

On average using Red Thread, Arrowhead uncovered 35 additional yellow flags (warnings/trends) and 20 more red flags (critical findings) per report than manual reviewers in half the time. This enables remediation much earlier and provides more time for other critical tasks, such as designing and running experiments.

	Average Yellow Flags Per Report	Average Red Flags Per Report
Manual Findings	0	9
Red Thread Findings	35	29
Reduction in Missed Errors Using Red Thread	3,500%	2,000%

INCREASED COMPLIANCE

Arrowhead focused on three bioanalytical parameters to confirm that Red Thread flags data appropriately according to FDA and EMA guidelines.

ACCURACY AND PRECISION

Rationale: These tests tend to surface more errors because there is ambiguity in the acceptance or rejection of data, due to the subjective interpretation of the regulatory guidance recommendations.

Results: Red Thread audits 100% of the elements and acceptance criteria as outlined in the FDA guidelines.

SENSITIVITY

Rationale: These tests are critical for accurately and consistently detecting the lowest measured concentration in a biological matrix. This is important because bioanalytical data is used to evaluate efficacy and safety of the drug in preclinical and clinical studies. These tests play an important role in dosing decisions and treatment schedules for potent drugs.

Results: Red Thread audits 75% of the elements and acceptance criteria as outlined in the FDA guidelines.

The FDA's acceptance criteria for "The analyte response at the LLOQ should be ≥ five times the analyte response of the zero calibrator" requires access to raw data generated by the instrument. Since Red Thread does not have access to raw instrument data, it cannot audit the report for this acceptance criterion. If raw instrument data were to be included in the report, Ariadne can code to test for this criterion.

STABILITY

Rationale: These tests are critical to cover the time period and activities that occur during sample analysis.

Results: Red Thread audits 100% of the elements and acceptance criteria as outlined in the FDA guidelines.

"Our bioanalysts can now focus on resolving issues instead of trying to find errors in the data, all while working more closely with our CRO partners"

–Txheng Yang, Associate Director, Bioanalytical Development

CONCLUSION

Red Thread assists bioanalysts with their review and auditing of preclinical and clinical bioanalytical data to enable time savings, work as a force multiplier, and increase transparency between the sponsor and the CRO, while ensuring the highest levels of data compliance, quality, and integrity. Red Thread is designed to highlight whether data is compliant, non-compliant, or requires additional review by bioanalysts for ultimate data-driven decision-making. The application is not designed to remove the human from the process.

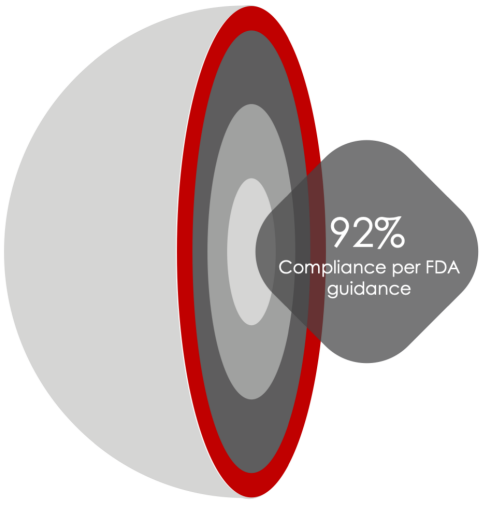
Expanding into areas with dire needs, Ariadne looks forward to extending the capabilities of Red Thread to immunogenicity, biomarker, early drug discovery, cell therapy, and gene therapy data for assay validation as well as sample analysis in the near future. Ariadne also plans on using further AI capabilities to better parse and read reports in different formats.

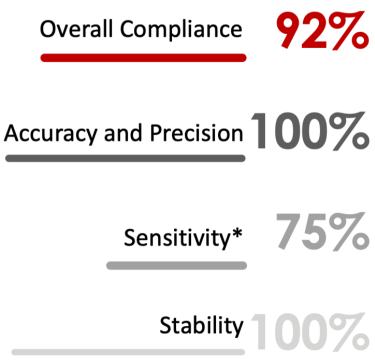
Arrowhead is leading the translation of scientific discovery into the development of new therapies to treat numerous and diverse diseases in need of adequate treatment options. Arrowhead's proprietary technology platform called TRiM™ leverages the Nobel Prize winning RNA interference pathway to potentially address diseases in the liver, lung, solid tumors, muscle, and beyond, with a goal of reaching a new tissue type every 18-24 months.

Red Thread's Compliance with FDA's 2018 Guidance for Small Molecule Validation

FDA Bioanalytical Parameter	FDA Guidelines from 2018 Small Molecule Validation Guidance	Red Thread Meets FDA Criteria?	Example Red/Yellow/Green Flags from Red Thread Output Report
Accuracy and Precision	Include 3 or more batches	Yes	Red Flag*: Table has fewer than 3 batches that met all the validation requirements for accuracy and precision
	Each batch should have 4 concentration levels (LLOQ, Low QC, Mid QC and High QC)	Yes	Red Flag*: [Batch 4] Intra-run (Within Run) Accuracy and Precision was not demonstrated at 3 µg/mL level
	Each concentration level for each batch should have at least 5 replicates	Yes	Red Flag: [Batch 5] Column: 1 µg/mL has fewer than 5 replicates
	The run should meet the calibration curve acceptance criteria and include the LLOQ calibrator.	Yes	Any failed run due to calibration standards not meeting criteria are reported in the calibration std table section of the report.
	Inter- and intra-batch accuracy should be ≤ 15% for nominal concentrations; except ≤ 20% at LLOQ	Yes	Red Flag: [Batch 3] %RE above 15% on Column: 300 ng/mL for concentration 214.510
	Inter- and intra-batch precision should be ≤ 15% CV, except ≤ 20% CV at LLOQ	Yes	Red Flag: [Batch 3] Column: 300 ng/mL: Overall %CV above 15%
Sensitivity	Check if the lowest nonzero standard on the calibration curve defines the sensitivity (LLOQ)	Yes	In the standard curve table Red Thread applies the LLOQ acceptance criteria to the lowest standard found in the Accuracy and Precision table
	The analyte response at the LLOQ should be ≥ five times the analyte response of the zero calibrator	No	At this time, Red Thread can't confirm these results without access to raw data. Also, Ariadne does not have access to samples of sensitivity raw data to code and evaluate data for this criterion.
	The accuracy should be ± 20% of nominal concentration (from ≥ five replicates in at least three runs)	Yes	Red Flag: [Batch 4] Column: 1 µg/mL has fewer than 5 replicates
	The precision should be ≤ 20% CV (from ≥ five replicates in at least three runs)	Yes	Green Flags: [Batch 1] Intra-run (Within Batch) Accuracy and Precision was demonstrated
Stability	For auto-sampler, bench-top, extract, freeze-thaw, stock solution and long-term stability, check if at least 3 replicates at LQC and HQC concentrations are present	Yes	Red Flag: [Batch 8 A: Fresh QC] Column: A 3 µg/mL: minimum requirement of N = 3 samples tested per concentration level was not met
	The accuracy (% nominal) at each level should be ± 15%	Yes	Yellow Flags: [Batch 7] Column: 3 ng/mL: Overall %RE above 10%

*These data yield themselves to subjective interpretation of the guidance. While Red Thread takes a conservative approach to avoid any findings being missed, these data may be acceptable per an organization's SOP. For more details, please refer to discussion point 1 in Additional Findings and Discussion section in Appendix.





*One of the four Sensitivity checks requires access to raw data, which is usually unavailable in the report

APPENDIX

ADDITIONAL FINDINGS AND DISCUSSION

As an additional step to ensure highest levels of data compliance, quality and integrity are provided by Red Thread, Arrowhead and Ariadne also performed an additional manual review of two of the reports (high range rat and high range monkey reports). Some key findings from this manual review and a few other discussion points from the data review of Red Thread output reports are discussed below:

1. Red Thread conservatively interprets regulatory guidance and develops its expert system rules to meet the highest levels of compliance to ensure no errors, findings, or trends are missed.

   In the Intra-Batch Precision and Accuracy of Quality Control Samples table, per the 2018 FDA guidance, at least three batches should demonstrate Accuracy and Precision (A&P) with four concentration levels (Lower Limit of Quantitation or LLOQ, Low Quality Control or LQC, Mid-Quality Control or MQC, and High Quality Control or HQC); five replicates at each of the four levels of concentration are required; and precision (%CV) should be ≤ 15% and accuracy (%RE) should be within ±15%. However, the guidance does not specify if the accuracy should be demonstrated on both, an individual replicate level and at the mean level of these five replicate measurements, or only at mean level of all five replicates (flagged yellow as data with no specific guidance lends itself to subjective interpretation). Red Thread takes the most conservative approach and requires that a minimum of five replicates meet acceptance criteria outlined for accuracy at both levels, individually and when reported as a mean. However, per a client's SOP, fewer than five replicates passing may be acceptable if the mean value demonstrates A&P.

   Thus, per Red Thread's conservative standards, only runs 1 and 3 demonstrated A&P for all 4 QC levels. For Batch 2, the implication would be that Accuracy was not demonstrated at the Low QC level because 2 of the 6 replicates had RE% values above ± 15%. Also, per 2018 guidance, all 4 levels are required for A&P experiment. Run 4 and 5 QC were batch performance (non-A&P) QCs and are not required to be included in the table. Consequently, accuracy and precision were not demonstrated across three independent batches. Per Red Thread's conservative standards, a resolution of this finding would entail repeating an A&P run with all four levels to ensure the lab had at least 3 examples where the LQC demonstrates sufficient A&P.

   However, per the CRO's Standard Operating Procedures (SOPs), a run is acceptable if 50% or more of all individual replicates demonstrate accuracy and the mean demonstrates appropriate accuracy and precision within the acceptance criteria described in the FDA guidance. The accuracy and precision for this data set is therefore demonstrated.

Batch No.	QC Samples, µg/mL			
	1	3	175	375
1 Fresh QC	0.921	3.110	170.532	373.479
	NV	2.961	170.786	366.744
	0.912	3.048	170.805	358.572
	0.940	3.014	170.928	361.209
	0.981	2.980	170.927	369.848
	1.093	3.007	171.080	356.993
Mean	0.969	3.020	170.843	364.474
SD	0.074	0.053	0.185	6.582
%CV	7.64	1.76	0.11	1.81
%RE	-3.06	0.67	-2.38	-2.81
N	5	6	6	6
Batch No.	QC Samples, µg/mL			
	1	3	175	375
2	0.946	2.894	178.493	363.023
	0.943	2.962	182.832	367.547
	0.945	2.996	180.843	371.363
	0.985	3.019	181.750	367.947
	0.985	2.475*	177.238	365.391
	0.974	2.179*	180.669	366.329
Mean	0.963	2.754	180.304	366.933
SD	0.021	0.346	2.078	2.795
%CV	2.13	12.58	1.15	0.76
%RE	-3.70	-8.19	3.03	-2.15
N	6	6	6	6
Batch No.	QC Samples, µg/mL			
	1	3	175	375
3	0.948	2.951	158.857	351.734
	0.947	3.075	177.627	373.517
	0.887	3.020	159.745	347.713
	1.046	2.908	184.806	373.210
	0.923	2.976	164.004	332.263
	0.963	2.956	184.854	380.183
Mean	0.952	2.981	171.649	359.770
SD	0.053	0.059	12.223	18.724
%CV	5.57	1.97	7.12	5.20
%RE	-4.77	-0.63	-1.91	-4.06
N	6	6	6	6
Batch No.	QC Samples, µg/mL			
	1	3	175	375
4 Fresh QC		2.967	174.237	357.020
		2.873	173.027	360.489
		2.965	175.591	360.623
		2.905	173.592	359.657
		2.872	173.413	358.196
		2.990	175.779	354.710
Mean	N/A	2.929	174.273	358.449
SD	N/A	0.052	1.163	2.299
%CV	N/A	1.77	0.67	0.64
%RE	N/A	-2.38	-0.42	-4.41
N	0	6	6	6
Batch No.	QC Samples, µg/mL			
	1	3	175	375
5		3.129	173.824	387.053
		2.930	170.327	387.879
		3.005	170.772	373.949
		3.060	170.659	383.300
		2.968	171.271	370.157
		2.917	174.898	387.708
Mean	N/A	3.002	171.959	381.674
SD	N/A	0.081	1.916	7.731
%CV	N/A	2.71	1.11	2.03
%RE	N/A	0.05	-1.74	1.78
N	0	6	6	6

Red Thread uses conservative interpretation where specific guidance is lacking

The data has been redacted and modified to highlight the data being discussed.

2. In the All Calibration Standards Table, there are many values approaching the acceptance criteria of ± 15% accuracy (%RE). When multiple values approach the outer limits of the acceptance criteria within a concentration level of a batch, this can indicate trends that should be explored, such as weakness in a method, issues with analyst performance, and/or contamination of samples or reagents (Yellow flag for data that is near failure for the same test multiple times during a study or validation indicates a possible weakness in the assay or study conduct issues by the scientific staff). Failing Calibration Standards during sample analysis increases the likelihood that batches will need to be repeated.

Yellow flags highlight near failure data indicating a possible weakness in the assay or study conduct issues by the scientific staff

Batch No.	Calibration Standards, µg/mL							
	1	2	10	30	100	200	400	500
1 (Fresh STD)	0.973	1.885	10.015	29.874	95.055	183.200	324.996*	394.052*
	1.005	2.144	11.122	33.871	106.910	201.455	362.368	460.210
2	0.973	2.027	10.925	33.475	106.029	196.004	353.302	453.320
	1.016	1.913	10.828	33.358	105.757	196.690	350.011	445.359
3	0.983	2.074	11.331	33.928	116.931*	193.267	355.338	453.432
	0.971	2.005	10.760	32.156	106.054	193.823	359.050	453.312
5	0.945	2.211	10.096	28.040	100.052	199.950	391.153	508.776
	0.955	2.192	10.189	28.301	100.076	199.861	391.761	508.220
8	0.939	2.013	9.714	28.050	98.830	194.668	404.120	521.016
	0.993	2.261	10.430	29.565	106.150	200.204	362.036	507.004
10	1.058	1.755	9.986	30.203	105.044	196.783	404.290	475.476
	1.040	1.836	10.132	31.563	104.648	199.419	410.138	485.191
Mean	**0.985**	**2.044**	**10.297**	**30.634**	**101.610**	**197.905**	**381.794**	**489.484**
SD	**0.040**	**0.139**	**0.483**	**2.266**	**3.961**	**5.041**	**20.972**	**27.485**
%CV	**4.05**	**6.79**	**4.70**	**7.40**	**3.90**	**2.55**	**5.49**	**5.62**
%RE	**-1.46**	**2.18**	**2.97**	**2.11**	**1.61**	**-1.05**	**-4.55**	**-2.10**
N	**18**	**18**	**18**	**18**	**17**	**18**	**17**	**17**

The table has been redacted and modified to highlight data being discussed.

3. During the manual review, a critical finding from the Bench-top Stability of Unextracted QC Samples, highlighted that upon failing of batch 4 for bench-top stability, a repeat experiment run (batch 8) was ran using the same conditions. Although, batch 8 passed, the CRO widened the acceptance criteria for batch 4 after performing bench-top stability for batches 4 and 8. The decision and note to file (NTF) for the expanded criteria was not defined prior to the start of the validation and applied to all validation experiments (as required by the regulatory guidance) but came after a failed Benchtop stability experiment. This makes an auditor suspicious that the criteria was changed just to avoid a failed experiment. This is a manual finding as Red Thread may not have enough context within the table to assess if there was a duplicate rerun or a rerun with a new set of test conditions. However, through the many red flags and yellow flags, the user is prompted to further investigate the data and decide on whether the data is acceptable or needs further work or rework to be addressed (data near failure implying high risk of future failures).

Yellow flags highlight near failure data indicating future risk

Red
[Batch 4 A: Fresh QC] %RE above 15% on Column: A 3 µg/mL for concentration 3.692
[Batch 4 A: Fresh QC] %RE above 15% on Column: A 3 µg/mL for concentration 3.569
[Batch 4 B: BTS QC] %RE above 15% on Column: B 3 µg/mL for concentration 3.492
[Batch 4 B: BTS QC] %RE above 15% on Column: B 3 µg/mL for concentration 3.636
[Batch 4 B: BTS QC] %RE above 15% on Column: B 3 µg/mL for concentration 3.838
[Batch 8 A: Fresh QC] Column: A 3 µg/mL : minimum requirement of N = 3 samples tested per concentration level was not met
[Batch 8 A: Fresh QC] Column: A 375 µg/mL : minimum requirement of N = 3 samples tested per concentration level was not met
[Batch 4 B: BTS QC] Column: B 3 µg/mL : Overall %RE above 15%
Bench-Top stability experiment for [Batch V04 B: BTS QC] failed

Yellow
[Batch 4 A: Fresh QC] %RE above 10% on Column: A 3 µg/mL for concentration 3.305
[Batch 4 B: BTS QC] %RE above 10% on Column: B 3 µg/mL for concentration 3.427
[Batch 4 A: Fresh QC] Column: A 3 µg/mL : Overall %RE above 10%

4. In Validation Assay Schedule, a Red Thread flag points out the possibility that the time period from Extraction date to Assay Date for Batch 8 (used for long term stability and bench top stability) exceeded the validated time period determined for the processed (extracted)/or reinjection stability. If Batch 8 exceeded that validated time period, it will need to be retested. This is a manual finding as some of the information for cross-checking is available in text form in Section 3, but no definitive information has been provided in the report (Data requiring additional context for decision making).

   While QC and QA teams are more interested in green and red flags to quickly understand if the data passes or fails and whether it meets regulatory guidelines, bioanalytical scientists reviewing the output are more interested in yellow flags because they highlight trends or findings that pertain to how robust a method is.
5. In Calibration Curve Parameters table, the correlation coefficient, r, was used instead of the coefficient of determination, r^2. Arrowhead noticed this deviation from previous reports received through their contracted CRO. Although both, r and r^2, are acceptable per regulatory guidance, the use of r^2 is considered best practice. The base model for Red Thread is developed using most conservative interpretations of the regulatory guidance, and since both, r and r^2, are acceptable, Red Thread can only highlight a shift from common practice when customized to do so per client request.