Reducing Appraisal Costs in Computer System Validation (CSV): A Risk-Based Auditing Approach Using DMAIC Methodologies

Jenylis Vega Quintero
Master of Engineering in Manufacturing Engineering
Advisor: María García Sandoval, Ph.D.
Polytechnic University of Puerto Rico
Graduate Project EXPO, May 2026

Abstract *— Computer System Validation (CSV) in the pharmaceutical industry often relies on test-everything approaches, resulting in excessive Appraisal Costs. This research applies to the Lean Six Sigma DMAIC framework to optimize the validation lifecycle by transitioning to a risk-based approach aligned with FDA Computer Software Assurance (CSA) and ISPE GAMP 5 R2 guidelines. A retrospective analysis of 11 protocols established a historical baseline of 336.0 hours across 42 execution days, requiring 8.20 hours per identified issue. A 2025-specific dataset of 144.0 baseline hours was used for back-testing. Results demonstrated that the risk-based framework reduces validation effort to 103.7 hours (a 28% reduction), saving $2,015 per project cycle while maintaining a 100% detection rate for all 28 critical defects.*

Keywords: *Appraisal Costs, Computer System Validation, Lean Six Sigma, Risk-Based Auditing.*

INTRODUCTION

The pharmaceutical industry operates in one of the most strictly regulated environments globally, overseen by agencies such as the FDA. To maintain patient safety and data integrity, Computer System Validation (CSV) is mandatory. However, traditional validation often relies on a one-size-fits-all approach, testing every system's function regardless of risk.

This research identifies a Cost of Quality crisis where Appraisal Costs, the resources spent on testing and auditing, have become an operational bottleneck. As companies move toward Pharma 4.0, the administrative burden of over-processing documentation creates significant waste. By applying the Lean Six Sigma DMAIC framework, this study proposes a shift from exhaustive testing to a risk-based auditing approach that optimizes resources without compromising regulatory compliance.

LITERATURE REVIEW

Organizations frequently underestimate hidden costs in software validation. While scrap or recalls are easily tracked, the administrative burden of over-processing, such as redundant testing or witnessing, is often buried within operational budgets. Schiffauerova and Thomson [1] suggest that disproportionate capital is spent on Appraisal (inspection) rather than Prevention (designing quality into the system).

Historically, the pharmaceutical industry adopted a risk-averse, one-size-fits-all approach, resulting in massive volumes of documentation and extended project timelines. However, software risk is not uniform; different components have varying failure probabilities. Integrating Risk Analysis early in the design phase is significantly more efficient than exhaustive testing at the end of the lifecycle [2].

J.M. Juran [3] defines appraisal costs as those associated with measuring and inspecting products to ensure conformance to quality standards. This study adopts the GAMP 5 guide's advocacy for a risk-based approach [4], shifting effort from exhaustive testing to focused analysis of high-risk features.

J. M. Juran divided quality costs into four categories: Prevention, Appraisal, Internal Failure, and External Failure. This research focuses on Appraisal Costs, which represent the resources spent on testing, inspecting, and auditing systems to ensure they meet requirements. A disproportionate amount of money is often spent on Appraisal activities that do not necessarily correlate with higher defect detection.

A common misconception is that Lean Six Sigma applies only to production floors. Recent literature supports its application to Quality Assurance and Auditing [5], [6]. By applying the Measure and Analyze phases to the audit workflow, teams can identify bottlenecks, such as redundant reviews or unclear criteria, that inflate audit cycle times.

While often applied to production, the DMAIC framework is an effective mechanism for re-engineering the validation lifecycle itself. By treating validation as a process, efficiency gains seen in manufacturing lines can be realized in document execution.

METHODOLOGY

The primary methodology utilized is the DMAIC (Define, Measure, Analyze, Improve, Control) cycle, which provides a structured, data-driven path for process improvement.

- Define Phase: The scope was defined as the validation lifecycle within a pharmaceutical manufacturing site. The primary goal was to reduce appraisal effort while maintaining a 100% detection rate for critical defects aligned with FDA Computer Software Assurance (CSA) [7] and ISPE GAMP 5 R2 [4] guidelines.
- Measure Phase: A baseline was established using historical data from 11 protocols executed between 2024 and 2025. Key metrics included total hours spent on script execution, defect rates, and the ratio of critical to non-critical exceptions.
- Analyze Phase: Statistical tools, including Pareto Charts and Ishikawa (Fishbone) Diagrams, were applied to identify the root causes of validation delays. This phase differentiates between value-added testing and non-value-added over-processing.
- Improve Phase: A Risk-Based Validation Framework was developed. This framework was validated through a retrospective simulation (back-testing) of 2025 project data to determine whether reduced testing for low-risk items would affect safety.
- Control Phase: A control plan involving updated Standard Operating Procedures (SOPs) aligned with ICH Q9 (R1) [8] was proposed to sustain the improvements.

Framework Replication Protocol

To allow validation engineers and quality managers to replicate this risk-based auditing optimization model within alternative regulated environments, a standardized five-step sequence must be followed:

- Historical Baseline Mapping: Define the target system boundaries and aggregate historical verification logs to quantify total baseline validation execution hours and calculate process efficiency (Hours / Issue).
- Exception Categorization: Perform a Pareto analysis on historical system exceptions to distinguish non-critical clerical or protocol generation errors from true technical anomalies.
- Risk Level Alignment: Map system test functions through an objective matrix evaluating System Criticality and Function Complexity to categorize testing boundaries into High or Low risk.
- Regulated Effort Scaling: Apply a targeted 50% execution and review cycle compression exclusively to verification tests meeting Low-Risk criteria, while maintaining 100% testing rigor for High-Risk data integrity functions.
- Back-Testing Simulation: Execute a retrospective back-test against past data to mathematically verify that the compressed testing strategy successfully captures 100% of historical critical defects, confirming safety before standard operating procedure deployment.

Objective Risk Classification Boundaries

To ensure audit repeatability and eliminate subjectivity during protocol scaling, risk levels are defined using a structured multi-factor matrix

balancing system complexity against product safety impact:

- Low-risk (50% Effort Reduction): Includes protocols for indirect-impact systems, cosmetic UI verifications, or standard reporting functions where failure does not compromise data integrity or patient safety.
- High-risk (100% Rigor Retained): Includes all functions related to data calculation, SCADA-to-PLC communication, security access levels, and audit trail functionality.

RESULTS AND DISCUSSION

The process requires approximately 8.20 hours per identified issue (336.0 total hours across 41 issues). This measured total of 336.0 execution hours highlights a significant Opportunity Cost, where specialists are dedicated to repetitive, manual verification rather than high-value defect prevention or process optimization.

Regulatory Support

The implementation of a risk-based model is supported by the FDA CSA guidance, which encourages the least burdensome testing for non-critical software features.

Hypothesis Validation

Statistical analysis allows for the rejection of the null hypothesis (H_0), providing empirical evidence that a risk-prioritized approach is significantly more efficient ($p<0.05$) than the current baseline. To validate the significance of the appraisal cost reduction, a paired t-test was performed using Minitab. The analysis compared historical execution hours with simulated risk-based model hours across the 11 protocols.

The baseline descriptive metrics used to establish this comparison are compiled in Figure 1 and Figure 2, which outline the shift in execution totals and mean hours between the two models.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Traditional_Hours (Actual)	11	13.091	1.729	0.521
Risk_Based_Hours (Simulated)	11	9.427	1.131	0.341

Figures 1
Summary of Descriptive Statistics Comparing Traditional and Risk-Based Models

Statistics

Variable	Mean	StDev	Sum
Traditional_Hours (Actual)	13.0909	1.72942	144
Risk_Based_Hours (Simulated)	9.42727	1.13057	103.7

Figure 2
Summary of Descriptive Statistics Comparing Traditional and Risk-Based Models

The inferential analysis of this dataset is displayed in the statistical hypothesis test output in Figure 3. The test yielded a T-value of 14.69 and a P-value of 0.000. Since the p-value is less than the significance level ($\alpha=0.05$), the null hypothesis (H_0: $\mu_{difference}=0$) is rejected in favor of the alternative hypothesis (H_1: $\mu_{difference}>0$).

Test

Null hypothesis	H_0: μ_difference = 0
Alternative hypothesis	H_1: μ_difference > 0

T-Value	P-Value
14.69	0.000

Figure 3
Statistical Hypothesis Test Results for the Reduction of Appraisal Hours

The precise mathematical estimate of the paired difference and its corresponding confidence thresholds are illustrated in Figure 4, establishing a 95% Lower Bound for the population mean difference of 3.211 hours.

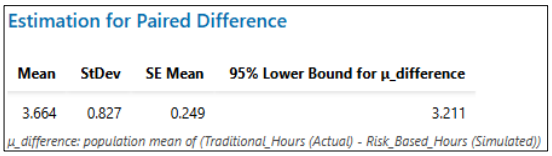

Figure 4
Mean Difference and 95% Confidence Interval for Appraisal Hour Savings

This statistical significance is further reinforced by the Boxplot of Differences shown in Figure 5, which visually maps out the tight distribution of appraisal hour reductions and demonstrates that the savings are highly consistent across the sample protocols.

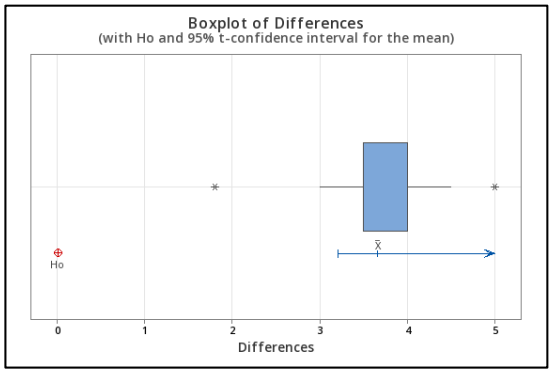

Figure 5
Distribution of Appraisal Hour Reductions with 95% Confidence Interval

Pareto Analysis and Root Cause Identification

Of the 41 issues identified, 28 were classified as Critical (68%) and 13 as non-critical (32%). The Pareto Analysis confirmed that a small number of categories, specifically Protocol Generation Errors (n=16) and System Issues (n=7), are responsible for most validation delays. The complete frequency distribution and impact levels of these validation exceptions are detailed in Table 1.

Table 1
Frequency and Distribution of Validation Exceptions

Exception Category	Frequency (n)	Impact Level
Protocol Generation Errors	16	Non-Critical
System Issues	7	Critical
Human Execution Errors	10	Mixed
Communication Timeouts	8	Critical

The cumulative impact and distribution of these exception categories are shown in Figure 6 below.

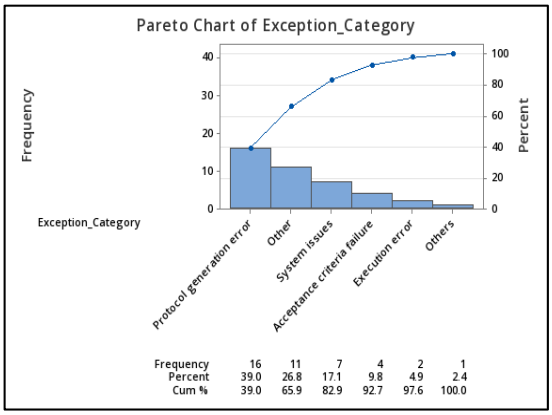

Figure 6
Frequency and Cumulative Percentage of Exceptions in Pharmaceutical CSV Projects

The distribution of validation exceptions provides the why behind the proposed framework, using the Pareto principle to separate the vital few from the trivial many.

- The Vital Few: The data indicates that Protocol Generation Errors (n=16) are the most frequent cause of exceptions. While these errors consume significant appraisal time during the review and correction phase, they are classified as non-critical because they do not impact system functionality or patient safety.
- Technical vs. Process Exceptions: System Issues (n=7) and Communication Timeouts (n=8) represent the technical risks inherent to industrial automation and SCADA systems. These are the areas where the 100% appraisal rigor must be maintained to ensure FDA and GAMP 5 compliance.
- Risk Discrepancy Analysis: A major finding in this analysis is Risk Discrepancy, the observation that non-critical documentation slips often receive the same level of auditing rigor as critical software bugs. This one-size-fits-all approach leads to inflated project budgets and extended timelines without a corresponding increase in quality.
- Shift Toward Prevention: By identifying that a high volume of exceptions stems from Detailed Design Specification (DDS) errors, the research

justifies a shift in strategy. Instead of spending hours in the Appraisal phase (testing), organizations should redirect these resources to the Prevention phase to eliminate documentation defects before validation protocols are even generated.

Root Cause Categorization

The root causes of the 28 Critical Defects were categorized into three primary areas.

- Machine/Software Risks: Communication timeouts between PLC and SCADA and configuration mismatches in the software were identified as technical risks inherent to automation.
- Method/Design Risks: Errors in the Detailed Design Specification (DDS) and tight thresholds for non-critical alarms represent preventable waste.
- Personnel Risks: Human errors during manual evidence handling and execution steps within the validation system contribute significantly to non-critical delays.
- The Prevention Shift: By identifying DDS errors as a root cause for critical defects, this research justifies moving resources from Appraisal (testing at the end) to Prevention (improving design earlier) to reduce the total Cost of Quality.

Retrospective Simulation Impact

The simulation applied a 50% reduction in effort to protocols categorized as Low Risk. The results demonstrated:

- Efficiency Gain: Total effort for the 2025 sample was reduced from 144.0 to 103.7 hours, a 28% improvement.
- Financial Benefit: Using a conservative industry rate of $50/hour, this represents a savings of $2,015 per project cycle.
- Safety Retention: All 28 critical defects were successfully identified despite the reduction in test volume, providing empirical proof that the risk-based approach is safe.

A direct comparative breakdown of these traditional metrics against the simulated risk-based model performance is presented in Table 2.

Table 2
Comparative Analysis of Traditional vs. Risk-Based Validation Performance

Metric	Traditional Model (Actual)	Risk-Based Model (Simulated)	% Improvement
Appraisal Hours (2025)	144.0	103.7	28% Reduction
Critical Defects Found	28	28	0% Loss (Safe)
Calculated Cost ($50/hr)	$7,200	$5,185	$2,015 Savings

The comparative data presented in Table 2 provide the primary quantitative evidence that a risk-based approach is both economically viable and operationally safe.

- Appraisal Effort Reduction: The traditional test-everything model required a baseline of 144.0 execution hours for the 2025 dataset. By applying the risk-based framework, which reduces execution cycles by 50% for low-risk protocols, the effort dropped to 103.7 hours.
- Efficiency Metric: This represents a 28% improvement in resource utilization. In a Lean Six Sigma context, this reduction directly addresses the waste of over-processing, which occurs when testing rigor exceeds the actual risk level of the system component.
- Economic Impact: For this simulation, a conservative industry rate of $50/hour was used, representing the combined labor costs of a CSV Specialist and a Quality Reviewer in the Puerto Rico pharmaceutical sector. The resulting savings of $2,015 per project cycle demonstrate that efficiency can be achieved without capital investment, simply by re-engineering the validation process.
- Regulatory Safety Margin: Crucially, the Critical Defects Found metric remained at 100% (28 out of 28 defects). This confirms that the 50% reduction in effort was successfully

targeted at non-value-added activities, ensuring that critical data integrity risks were still rigorously verified.

CONCLUSION

The application of a Lean Six Sigma DMAIC framework to the Computer System Validation (CSV) lifecycle has provided a clear, data-driven roadmap for operational excellence in pharmaceutical manufacturing. This research successfully bridged the gap between theoretical quality management and the practical, highly regulated demands of software validation.

Summary of Findings

The initial baseline assessment revealed a highly inefficient traditional process, characterized by a test-everything approach that required a total of 336.0 hours to identify 41 issues. This equals approximately 8.20 hours of execution effort per defect, which empirically indicates the process is over-processing under Lean Six Sigma methodologies. The Pareto analysis of 41 validation exceptions confirmed that although 68% of issues were critical, they were heavily concentrated in high-risk technical areas such as system communication and software configuration. Conversely, non-critical errors were primarily due to documentation slips and protocol-generation errors.

The retrospective simulation of the 2025 dataset showed that, by using a risk-prioritized framework, the site could reduce total validation effort by 28%. This improvement resulted in projected savings of $2,015 per project cycle. Most importantly, the simulation demonstrated a 100% detection rate across all 28 critical defects, providing empirical evidence to satisfy both financial and regulatory stakeholders.

Although the statistical evaluations provide robust evidence of optimization, certain limitations of this study must be addressed. The quantitative baseline and retrospective simulation profiles relied on a sample of 11 validation protocols from a 2025 dataset at a single pharmaceutical manufacturing site. While these protocols represent critical automation layers, data collection boundaries were constrained by corporate confidentiality restrictions and site-specific computerized architectures. Consequently, while the systematic DMAIC re-engineering logic is replicable across the industry, the exact resource compression metric (28%) and financial return may fluctuate depending on company-specific software configurations, legacy document baselines, and regional professional labor rates.

Research Contributions

The primary contribution of this study is the development of an operational Risk-Based Validation Framework that translates the theoretical guidelines of GAMP 5 and Pharma 4.0 into measurable results. This study provides Quality Project Managers with a validated strategy for justifying a reduction in test volume by focusing resources on high-risk system functions. Furthermore, the research offers empirical evidence that over-testing does not yield higher quality; rather, optimized testing provides a safer, more transparent compliance profile for FDA audits.

Recommendations and Future Work

To ensure the sustainability of these efficiency gains, the following actions are recommended for implementation at the manufacturing site:

SOP Integration: The current Standard Operating Procedures (SOPs) must be updated to mandate a formal Risk Assessment before protocol generation, ensuring that testing rigor is proportional to the criticality of the system function.

Prevention Over Appraisal: Future quality initiatives should target the Prevention phase of Juran's Cost of Quality model. Specifically, reducing errors in Detailed Design Specifications (DDS) can eliminate defects before the expensive validation phase begins.

KPI Monitoring: It is recommended to implement Appraisal Hours per Protocol as a permanent Key Performance Indicator (KPI) to track process efficiency and identify future bottlenecks.

Study Limitations

The quantitative baseline and retrospective simulation profiles relied on a sample of 11 validation protocols from a 2025 dataset at a single pharmaceutical manufacturing site. While these protocols represent critical automation layers, data collection boundaries were constrained by corporate confidentiality restrictions and site-specific computerized architectures. Consequently, while the systematic DMAIC re-engineering logic is replicable across the industry, the exact resource compression metric (28%) and financial return may fluctuate depending on company-specific software configurations, legacy document baselines, and regional professional labor rates.

REFERENCES

[1] A. Schiffauerova and V. Thomson, "Cost of quality: a survey of models and best practices," International Journal of Quality & Reliability Management, vol. 23, no. 6, pp. 647-659, 2006.

[2] D. Verdon and G. McGraw, "Risk analysis in software design," IEEE Security & Privacy, vol. 2, no. 4, pp. 79-84, July-Aug. 2004.

[3] J. M. Juran and A. B. Godfrey, "Quality Costs", in Juran's Quality Handbook, 5th ed., New York, NY, USA: McGraw-Hill, 1999, pp. 8.1-8.26.

[4] International Society for Pharmaceutical Engineering (ISPE), GAMP 5: A Risk-Based Approach to Compliant GxP Computerized Systems, 2nd ed., Tampa, FL, USA: ISPE, 2022.

[5] F.A. Hasan, M. S. Muda, and M. R. Mustapha, "An empirical analysis of the role of internal auditing in Lean Six Sigma," The Social Sciences, vol. 12, no. 8, pp. 1448-1455, 2017.

[6] G. S. Ghaleb, "The application of DMAIC Six Sigma methodology to control the quality of internal audit performance: A case study," Res Militaris, vol. 12, no. 2, pp. 885-896, Summer-Autumn 2022.

[7] Food and Drug Administration, Guidance for Industry: Computer Software Assurance for Production and Quality Management System Software, Silver Spring, MD, USA: FDA, 2026.

[8] International Council for Harmonization, ICH Q9(R1): Quality Risk Management, Geneva, Switzerland: ICH, 2023.