

This article evaluates the time (labor effort) required to validate computer or software systems as a function of the applied validation strategy.

Cost and Benefit Analysis of Validation Strategies

by Kent Lohrey

Introduction

Each additional day in the marketplace of patent protected sales for a pharmaceutical product or medical device can have significant revenue impact for a company, sometimes on the order of millions or tens of millions of dollars per day. While the duration of patent protection is clearly defined, the portion of that time period where a product is available for sale or generating revenue is variable. One factor influencing how soon a product can reach the market place (or how long the patent protected sales period lasts) is the time spent in development, deployment,

and validation of any computer or software system required to produce the pharmaceutical or medical product or the clinical trial supplies required to get the product to the point where it can be sold. Some production lines also require new technologies and computer systems once product sales have started. Increasing product demand can require additional production capacity, which can drive changes to the manufacturing systems. As a result, there is significant pressure on the delivery of new computer or software systems that support or provide the capability to deliver these revenue-creating products. Regardless of business pressures,

Table A. Protocol details by customer, project, protocol, and strategy, including system type and total test items per protocol.

Customer	Project	Protocol	Type of System	Always (ALW)	Only When Different (OWD)
A	1	1	extrusion	1760	
		2	extrusion	916	
		3	compounding	932	
B	1	1	building management/room monitoring	1761	
C	1	1	process analytical technology	554	
		2	process analytical technology	2694	
		3	process analytical technology	1270	
		4	process analytical technology	527	
D	1	1	building management/room monitoring	1444	
E	1	1	building management/room monitoring	2634	
F	1	1	solution preparation/tablet coating		6304
		2	solution preparation/tablet coating		5171
		3	solution preparation/tablet coating		7076
		4	cream production		3521
		5	solution preparation/tablet coating		7486
		6	building management/room monitoring		2533
		7	purified water production and distribution		1176
G	1	1	Chromatography		1008
7 total	13 total	18 total	Average:	1449	4284

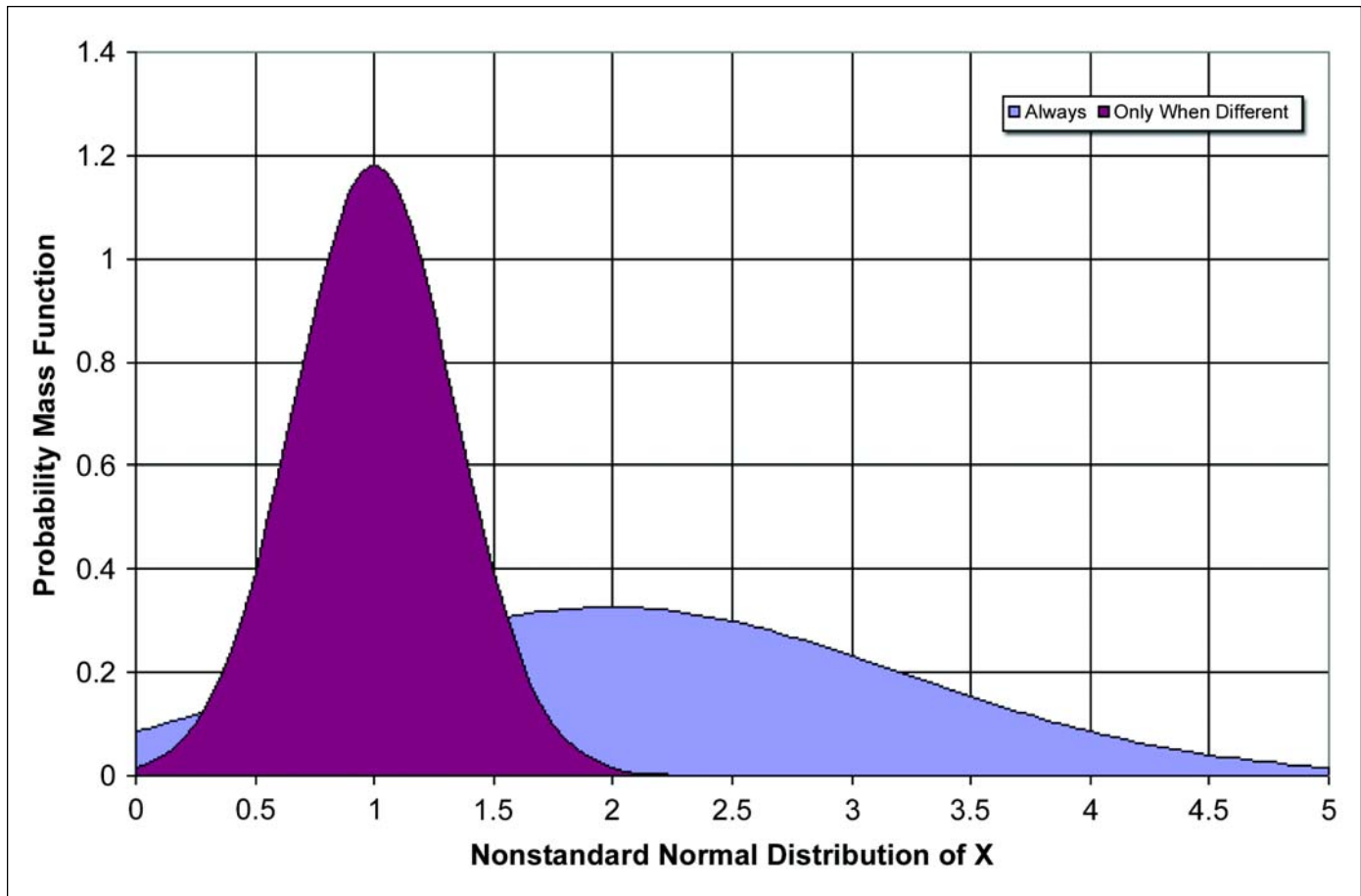


Figure 1. Normal distribution of scaled data (scaled OWD mean = 1.00, scaled ALW mean = 1.99).

these systems also must be deployed in a manner satisfying all applicable regulatory requirements, including software validation.

Companies have chosen to apply different validation strategies that are conducive to rapidly delivered and regulatory compliant systems. This article evaluates the time (labor effort) required to validate these computer or software systems as a function of the applied validation strategy. This is accomplished by comparing testing metrics collected from the execution of many protocols on several system validation projects where different strategies were applied by different companies. Analysis of this data quantifies the impact of the differing strategies. Advantages and disadvantages of each strategy are discussed in the context of regulatory requirements, and some conclusions are suggested to consider when setting validation strategy for future projects.

All of these validated systems controlled FDA regulated activities for pharmaceutical or medical device manufacturers. Most of the systems were manufacturing control systems (e.g., extrusion, Process Analytical Technology (PAT), solution preparation, and tablet coating). The remaining systems included building management systems, room monitoring systems, and purified water production and distribution. All of these systems were delivered for items already in production except for the PAT system which was used to produce clinical trial supplies.

Data

Data from validation test execution has been collected for a variety of purposes, including evaluating project or task efficiency and estimating future work. This data also can be used to compare validation strategies applied by different companies, as each company has its own method to satisfy the regulatory requirements for system validation, while attempting to meet business needs.

The data includes the cost of testing time in units of hours. This eliminates influences on the data, and the corresponding conclusions, due to different rates or hourly charges related to resources on different projects. Time is an acceptable unit for comparing different validation strategies as testing cost in dollars is directly proportional to testing time, meaning that an increase in time will create an increase in costs.

Test execution is quantified by calculating the average number of hours used to execute each test item or testing time per test item. This calculation requires dividing the total hours used to execute a test protocol by the total number of test items within the protocol. The test execution time includes all of the following tasks: creating test conditions, observing results, assessing the results (pass or fail), documenting actual results, writing deviations (test discrepancies), and implementing the resolutions defined within deviations. Deviations include specification changes, protocol

changes, system changes, and retesting or additional testing required.

The data used in this analysis are from 18 separate test protocols executed as part of 13 control and information system projects where hardware and software elements of a system were validated. These projects were performed with a total of seven different companies. Table A includes important project and protocol attributes for the data. Common elements of all of these validation studies included:

- protocols developed to be consistent with **Good Automated Manufacturing Practice (GAMP4)** principles
- cGMP documentation rules in effect
- structured protocols derived from the validation company (contractor) test template
- clearly defined and consistent test instructions and acceptance criteria (for example, in some cases, the exact same user interface test instruction language was used for different protocols with different customers)
- pass or fail assessment made for each test item
- deviations required for specification, system (hardware or software), or protocol changes that are needed to address failed test items

Significant differences were present in the validation strategies employed. These differences surrounded two fundamental choices in the strategies used, when to document actual results and when to use a risk-based approach to testing.

Two different actual results strategies were used on these protocols:

- actual results recorded at all times (five companies, 10 protocols), referred to from this point on as ALW for the Always recorded strategy
- actual results recorded only when the actual results were different than expected (two companies, eight protocols), referred to from this point on as OWD for Only When Different strategy

Two different risk strategies were used on these protocols:

- testing 100 percent of design specification content (five companies, 12 protocols) or
- applying a risk-based (less than 100 percent) test approach (two companies, six projects, and six protocols)

The same set of companies, projects, and protocols was used to analyze both major strategies. Individual companies consistently used the same actual results strategy for all of their projects. Some of the companies applied only one of the risk-based validation strategies, while others used both risk-based strategies depending on the specific system and project. Each strategy is addressed separately below.

All test execution time included in this analysis was expended by either contractor employees or employees from the customer companies. The total test execution time for each protocol was obtained from a combination of time sheets

	Always (ALW)	Only When Different (OWD)
Count of Protocols	10	8
Mean	1.99	1.00
Standard Deviation	1.22	0.34
Minimum	0.91	0.61
Maximum	4.60	1.59

Table B. Scaled data comparing testing time impact of actual results strategies (values divided by OWD mean).

and test activity reports. All contractor time was documented on time sheets reporting testing hours on a daily basis. The hours reported on these contractor time sheets also were submitted to and approved by the customer companies through approval of a daily activity report. This approval step and a customer's financial incentive to only pay for work performed ensured accuracy in this time sheet data. The daily activity reports, generated by the contractor, also documented when customer employees assisted with test execution activities as defined above. The accuracy of this total test time component is robust, but not as robust as the time sheet data because these reports were based on contractor observations, not direct input from the customer employees. As a result, some inaccuracy is possible in the customer time contribution to the total time. The extent of this possible inaccuracy is unknown, but mitigated by the following:

- Only eight of the 18 protocols included customer time. On these protocols, the customer time averaged less than one third of the total time spent on a protocol.
- The contractor employees were responsible for coordinating all test activities, regardless of who performed them.
- Most of the customer contribution was performed in combination with the contractor employees or performed independently, but in the same room as the contractor and when the contractor was present.

Validation Strategy: Actual Results Documentation

Data

The actual results data analysis produced a mean test execution time per test item metric and a standard deviation for the protocols within each strategy. The units for these values are hours/item. Dividing the mean and standard deviation from each population by the mean for the OWD population scaled the data, changing the values from hours/item to a percentage of the mean for the OWD population. For example, the OWD mean value scaled results in a value of 1.00 (the mean divided by itself). The mean for the ALW population is 199% of the OWD mean, reported as a value of 1.99 in Table B (ALW mean divided by OWD mean = 1.99). Comparison of the scaled test time per test item for the ALW and OWD populations shows the ALW method requires double the execution time per test item, compared to the OWD method.

The scaled mean and standard deviation results are represented in Table B. Table B also includes the scaled mini-

imum and maximum test execution time per test item values found within each population.

The scaled test time per test item results are presented in Figure 1. This figure uses a normal distribution of the data shown in Table B to illustrate the difference between the two populations. Not only was the ALW mean much greater than the OWD mean, but the standard deviation also was much greater (scaled value of 1.22 compared to 0.34 or approximately 3.6 times greater). This shows the execution time for the ALW tests was far more variable than the OWD tests. Stated another way, the mean test time per test item in the OWD population was much more consistent. It is likely that the ALW population variability was driven primarily by differences in the types of actual results required within the ALW population. For example, some ALW tests only required printing and referencing a screen capture, which requires much less time to document than re-writing the entire set of expected results in the protocol as was required for some of the other ALW tests.

Strategy Comparison

The different actual results strategies have important similarities and differences. The differences create advantages and disadvantages for each strategy as summarized in Table C.

Applicable Regulatory Requirements or Guidance

Current Good Manufacturing Practice (cGMP) requirements are defined in 21CFR Part 820. These requirements include

at least two references to the results of validation activities. The regulation for production and process controls defines a requirement for results documentation during validation of automated processes as follows: “Automated processes. When computers or automated data processing systems are used as part of production or the quality system, the manufacturer shall validate computer software for its intended use according to an established protocol. All software changes shall be validated before approval and issuance. These validation activities and results shall be documented.”¹

Similarly, the cGMP regulation for process validation defines the requirement for results documentation within process validation as follows: “The validation activities and results, including the date and signature of the individual(s) approving the validation and where appropriate the major equipment validated, shall be documented.”²

These requirements call for documentation of validation results. The requirements do not specify instructions for application of these requirements or if these results requirements apply specifically to actual test results. Therefore, individual companies must interpret the requirements and decide how the requirements can, or should, be implemented. The number of companies and protocols within the data analyzed in this article is only a small portion of the pharmaceutical and medical device industries. However, these companies do offer some insight into how regulated companies have attempted to implement validation strategies to address this result requirement.

The companies choosing the ALW method often explained this choice as being based on the presence of actual

	Always (ALW)	Only When Different (OWD)
Similarities	<ul style="list-style-type: none"> • Testing is the same: test instructions, initial conditions, and expected results • Does not take varying priorities of tests into account • Relies on the integrity of the personnel involved 	
Differences	<ul style="list-style-type: none"> • Every test requires the tester to document actual results 	<ul style="list-style-type: none"> • The tester is only required to document actual results when they differ from expected results
Advantages	<ul style="list-style-type: none"> • Evidence of actual results is provided for every test • Reviewers (including auditors) have more insight into test results • Availability of evidence may eliminate need for witness to testing • Can include some objective evidence (such as screen prints or reports) for every test of a specific type 	<ul style="list-style-type: none"> • Evidence is collected and documented when dictated by actual results (as part of a deviation) • Less time is spent recording actual results, requiring less time and cost (labor, either internal or vendor) • Less documentation requires less time by reviewers
Disadvantages	<ul style="list-style-type: none"> • Increases opportunities for human error (such as writing actual results incorrectly) • Some actual results can be generated by more than one set of instructions and conditions and recording only actual results provides no insight into this part of the test • The method of collecting evidence can create issues (for example some companies save screen prints as .jpg files - a format that can be easily edited - the use of electronic files for evidence may constitute use of electronic records to satisfy regulatory requirements, possibly invoking 21CFR Part 11 for these actual results) • More time is spent recording actual results, requiring more time and cost (labor, either internal or vendor) • More documentation requires more time by reviewers • Provides inconsistent levels of evidence when different types of evidence are used (such as written vs. screen prints) 	<ul style="list-style-type: none"> • No objective evidence is collected for successful or passed tests (such as screen prints or reports) so no evidence is available for reviewers or auditors on these tests • May require additional resources for witnessed testing

Table C. Comparison of different actual results strategies.

	100 percent Test	Risk-based
Similarities	<ul style="list-style-type: none"> System design included GMP-critical and non-GMP features and functions 	
Differences	<ul style="list-style-type: none"> All specified design elements are fully tested 	<ul style="list-style-type: none"> Some specified design elements are fully tested, others are partially tested
Advantages	<ul style="list-style-type: none"> System is comprehensively tested creating less risk of even minor issues going unnoticed No justification needed for testing reductions (less documentation) Less perceived risk by individual reviewers and approvers 	<ul style="list-style-type: none"> Reduces execution time and paperwork generated by testing, reducing time, and cost Risk assessment exercise focuses project and personnel on highest priorities
Disadvantages	<ul style="list-style-type: none"> Increases execution time and paperwork generated by testing which increases time and cost Treats all design features and tests as equal in importance or priority Can prevent inclusion of some useful features due to associated testing costs 	<ul style="list-style-type: none"> System is not comprehensively tested, increasing the risk of some issues going unnoticed Justification needed for testing reductions (additional documentation) Perceived risk by some reviewers or approvers as they approved the reduction in testing Requires additional design considerations to support risk strategy May need additional design work to limit access to non-GMP functions through security (user level) requirements May need additional design work to create features that are universal and can be reused on different systems (designing to satisfy multiple systems)

Table D. Comparison of different risk strategies.

results, for all tests, generating a higher confidence during internal reviews. These documented results also were cited as a critical component when defending the documentation in any audit activities.

Those applying the OWD method decided documenting that a test passes is the functional equivalent of writing down actual results when the results are the same as expected results. The test performer is documenting the actual results without rewriting them in the protocol by indicating that a test passed. In these protocols a “pass” assessment was defined as the actual results matched the expected results.

Conclusions

Within the limited population used in this analysis, companies more frequently chose the ALW method, by a ratio of 5:2 in this sampling. Validation strategy discussions with the companies using the ALW method revealed a very strong belief that anything less than always documenting actual results for each test was inviting regulatory failure. In many cases, validation personnel called on their experience with past regulatory audits to explain the necessity for their chosen strategy. These experiences hinged on a greater comfort level that auditors had expressed with the actual result documentation provided for every test. These companies did not cite specific regulatory requirements as part of the rationale for choosing the ALW strategy.

Even within the ALW strategy companies, there was varying confidence placed on different types of actual results. Objective forms of results evidence like screen captures or prints generated a much higher level of confidence than the written, subjective observations of a test performer. As noted in Table C, even verifiable objective evidence like screen prints have limitations in the insight or value they can provide to a reviewer. The fact that a screen print cannot provide specific definition of the actions taken to generate the

actual result prevents even this objective method of capturing results from providing a faultless illustration of all critical aspects of the test. The chain of evidence used to support a pass or fail assessment on an individual test can only be as strong as the weakest link. If only the test results are documented with evidence, written or otherwise, the unsupported or weak link in the chain of evidence is still relying on the integrity of the tester to have used the instructions and initial conditions provided to generate this documented result. If a company must rely on the tester’s integrity for the instruction part of the test, then is it possible or reasonable to rely on the same integrity for the result?

This was a central part of the OWD strategy justification for individuals within the two companies not using the ALW method. Both of these companies believed it was completely reasonable to rely on the tester for the accuracy of both the instruction and result portion of a pass or fail assessment. Both also concluded the regulatory requirements did not necessitate documenting actual results when the results were a match with the expected results defined in the protocol. Additional results documentation and the associated effort did not provide a significant compliance advantage, in their opinion. However, recording actual results was an obligation on these protocols and provided additional results evidence when the expected results defined in a protocol were not observed. In this scenario, the actual results were documented as a deviation to the protocol. For example, if a valve graphic turned the wrong status color when a valve alarm occurred, the deviation would document the behavior found, any required corrective action (e.g., specification or software change), and any retesting required.

Like the ALW method, the OWD strategy also has disadvantages. Screen prints, and other evidence like reports, can offer objective evidence that can support a test assessment of “pass” in a visual way that can be very powerful. Not using

evidence like this for passing tests denies future reviewers of evidence that could support at least the result portion of the tester's assessment. Reduced evidence for reviewers and auditors is a disadvantage in the OWD method that could offset some of the time and cost savings offered by this method.

Both strategies present disadvantages or challenges that must be carefully considered before choosing a validation strategy. The impact to schedule and cost is significant with the ALW method taking double the test time per test item for test execution. A lack of actual results evidence for passed tests in the OWD method, even for the most critical tests, could invite or influence a future regulatory review.

One possible solution to these challenges is to apply a risk-based approach to the need for actual results. A hybrid strategy could be adopted to require the ALW method for only the highest risk items. This limits cost and schedule impacts of the ALW method, while still providing evidence of actual results for every high-risk test. The OWD method could be applied for non-critical tests. For example, in a system where temperature is a critical process variable, temperature alarms are likely to impact the quality, safety, or efficacy of the product. As a result, these critical alarms deserve a high degree of scrutiny. At the same time, system usability features such as colors displaying device status do not have a direct impact on the product, requiring less emphasis on the verification of these system functions. Using this mixed strategy requires the instructions and acceptance criteria to clearly define where the different actual results methods apply and how they must be implemented. This hybrid and risk-based approach to actual results recognizes that all test items are not equal in terms of the functions or requirements they address while allowing companies to limit the time and cost impacts of validation testing.

Validation Strategy: Risk-Based Testing

Data

The risk-based data analysis focused entirely on those projects that applied some level of reduced testing based on risks evaluated within the system. These projects fit into two distinctly different types of risk-based approaches. In both types, these systems were designed to support the application of a risk-based strategy. Reviewing the specific system design specifications and counting the individual design elements and conditions not tested quantified the amount of reduced testing.

Five of the risk-based protocols applied a strategy of reduced testing for the control system software by minimizing testing of maintenance only functions. The design included some windows containing content only required for maintenance purposes. Access to these windows was restricted to prevent system operators from accessing these functions. These windows also contained no GMP critical information. All GMP critical information and process control was included in other portions of the applications. These restrictions supported an approach to test a representative sample of the functions included in windows like a variable

frequency drive status window. Testing 100 percent of these window features would have added on average approximately 13 percent more test items to validation tests that already averaged nearly 6000 test items per protocol.

One of the projects in the data analysis applied a different risk-based strategy. In this project, a control system application was developed for use on a number of different systems that had many common components. This application was designed generically in the windows that were used on each of the systems (such as the security and alarm summary windows). The application was used to control a suite of air extrusion systems. Some were single extruders and others were co-extruders (two extruders). The components of the single extruder were the same as the first or primary extruder in the co-extrusion systems. This allowed the primary extruders and single extruders to be controlled through the exact same set of user interface objects that were designed and programmed the same, except for linking to different field equipment. The design also disabled any window features that did not apply on a specific system. Secondary extruder objects were disabled when the application was installed on a single extruder system.

The first of these systems tested included almost all of the features common to all systems (a co-extruder) and was executed fully. The second installation of this application applied a risk-based approach to testing by not repeating tests of unchanged functions. For example, the access limits defined for a setpoint entry object in the user interface were entirely a function of the user interface objects, not the system attached to the software. These features were not retested. Full testing was limited to features not previously tested and those software components interacting with the specific system devices. This included testing of analog inputs like temperature, line speed, and pressure. Outputs like those related to starting and stopping devices also were tested. The protocol for this second installation included approximately 900 tests. Testing 100 percent of the software features, including those common features, would have required more than 1600 additional tests, an increase of more than 150 percent.

Strategy Comparison

The different risk strategies have important similarities and differences. The differences create advantages and disadvantages for each strategy as summarized in Table D.

Applicable Regulatory Requirements or Guidance

Current Good Manufacturing Practice (cGMP) requirements defined in 21CFR Part 820.70 also apply to this risk-based strategy scenario. Part (i) for automated processes requires companies to "validate computer software for its intended use."³ This regulation could be interpreted to demand 100 percent testing of all specified software elements as additional language in this section specifically directs validation of all software changes.

The door to the use of a risk-based strategy was opened in

2002 when the FDA announced a new initiative to enhance the pharmaceutical GMP rules and regulations. “The first goal will be to enhance the focus of the Agency’s cGMP requirements more squarely on potential risks to public health, by providing additional regulatory attention and agency resources on those aspects of manufacturing that pose the greatest potential risk.”⁴

The risk-based strategies used by the projects analyzed actually applied a minimal amount of risk when identifying testing that could be reduced. Risk-based decisions were limited to areas that did not affect electronic records or electronic signatures and software features that did not influence or alter product safety, quality, or efficacy. This avoided the aspects of manufacturing that pose a great degree of potential risk, minimizing the need for FDA scrutiny of these risk-based decisions.

Conclusions

More than half the projects included in this data analysis did not apply a risk-based approach. These companies and projects shared a common element: individuals were more comfortable with the 100 percent testing method. Validation strategy and test protocol approvers perceived risk-based validation as a potential compliance risk. This, in turn, was seen as a personal risk if they were associated with approving a potential compliance risk and a future regulatory activity questioned that choice.

Both of the risk-based approaches applied by the organizations in this sample delivered noticeable time and cost results for the companies. In the case of the maintenance function approach, the applications were able to include helpful monitoring and troubleshooting capabilities with minimal additional testing. These additional features will aid the company during operation and maintenance activities for years to come. Data from these projects demonstrate how risk-based testing allows an increase in application features for the same or less testing time than a 100 percent tested application. These projects could have pursued an even greater savings, beyond the average 13 percent, by applying the risk-based approach to testing of other non-critical features within the applications that were not isolated from critical functions as well as the maintenance functions.

The second risk-based approach, multiple installations of the same application, cut the number of test items by more than 50 percent on the second installation. The software was considered a custom configured application (GAMP category 5) for this company, which typically requires validation of the complete system.⁵ Through careful design choices and use of common user interface objects, the company reduced validation testing time and assumed very little compliance risk in the process.

The risk-based method focused these project teams and testing resources on the highest priority aspects of the specific systems which impacted quality, safety, or efficacy of the product. These critical features were fully tested. In these organizations, the project decision-makers were encouraged

and supported in the risk-based work.

Those companies applying a risk-based approach were able to validate their systems, while avoiding significant testing time which would have been required by using the 100 percent testing method. This time savings was achieved also while maintaining a strong position for any future compliance reviews through full testing of all critical functions. These projects proved a risk-based approach could provide regulatory compliance and reduce testing time (costs) simultaneously.

Discussion

The different validation strategies discussed above each have advantages and disadvantages. While some strategies may appear to be more commonly accepted, the more commonly used testing strategies – to always document actual results and to apply no risk-based reductions – drive validation testing time and costs up as indicated by this data analysis. Full or limited use of the other emerging strategies can generate schedule and cost benefits that merit consideration by companies needing to design, validate, and deploy systems within their own budgetary environment and regulatory history.

Considering these strategy options and their tangible impact on time and cost is likely to either generate more confidence in the current methods applied by a company or provide ideas for changes to future validation strategies. Regardless of the validation strategy chosen, clearly defining and documenting the strategy applied will provide the basis for validation decisions and support the defense of the applied strategies in any future regulatory compliance evaluations or audits.

References

1. 21 CFR Part 820.70 (i).
2. 21 CFR Part 820.75 (a).
3. 21 CFR Part 820.70 (i).
4. FDA Press Release, “FDA Unveils New Initiative To Enhance Pharmaceutical Good Manufacturing Practices,” August 21, 2002
5. *GAMP® 4, Good Automated Manufacturing Practice (GAMP®) Guide for Validation of Automated Systems*, International Society for Pharmaceutical Engineering (ISPE), Fourth Edition, December 2001, Appendix M4 - Guideline for Categories of Software and Hardware, p. 4. www.ispe.org.

About the Author



Kent Lohrey attended Princeton University, earning a Bachelor of Science and Engineering (BSE) in mechanical and aerospace engineering (with honors). While at Princeton, he was a cadet in the Air Force Reserve Officer Training Corps and received a commission in the United States Air Force. Following graduation, Lohrey reported to the Space and Missile Center, Los Angeles Air Force Base.

During his time in the Air Force, Lohrey held a variety of positions, including the liquid propulsion engineer for the TitanII and TitanIV launch vehicles where he supported the launch of the Cassini spacecraft mission to Saturn. After the Air Force, he joined Accenture as a process consultant. While at Accenture, he led the design and validation of controlled document management systems for pharmaceutical companies to comply with 21 CFR Part 11 requirements for electronic records and electronic signatures. Lohrey joined the validation department of Total Systems Design (TSD), Inc., a control system integrator, in 2002 and he is currently the Validation Program Manager. In this role, Lohrey sets validation strategy for the company, manages all validation projects, writes and executes validation protocols, and searches for opportunities to improve validation methods or tools. His most significant accomplishment at TSD is successfully developing and deploying a process and tools to automate the creation of test plan content from design specifications. Kent is a member of ISPE and was elected to membership in Sigma Xi, the Scientific Research Society, for his research work as a Princeton undergraduate. He can be contacted by telephone at: +1-610-857-1666 or by e-mail at: kent.lohrey@totalsystemsdesign.com. 