Lessons Learned Using API 1163 for Metal-Loss ILIs

Tom Bubenik and Matt Ellinger
DNV USA

PPIM 2025
PIPELINE PIGGING & INTEGRITY MANAGEMENT
JANUARY 27-31 • HOUSTON

Organized by
CLARION
TECHNICAL CONFERENCES

https://doi.org/10.52202/078572-0088

https://doi.org/10.52202/078572-0088

Abstract

API developed Standard 1163, In-Line Inspection [ILI] Systems Qualifications, to help pipeline operators ensure successful ILI projects. The second edition of the standard is incorporated by reference into the U.S. Code of Federal Regulations Title 49, Parts 192 and 195. The third edition of API 1163 (September 2021) is not yet incorporated by reference. The third edition made many of the optional analyses in the second edition mandatory, thereby making the standard more prescriptive. In addition, PRCI has developed a guidance document and spreadsheet for conducting verifications and validations per the third edition of API 1163. The second and third editions are widely used in industry to verify and validate metal-loss ILIs.

API 1163 (third edition) Level 2 and Level 3 validations require an operator to compare field non-destructive evaluation (NDE) measurements to ILI-reported dimensions, such as depth, length, and width. Per the third edition of API 1163 Level 2 validations, the results of the comparisons are assigned one of three “Outcomes” that determine whether the ILI is rejected outright (Outcome 1), further evaluated (Outcome 2), or accepted (Outcome 3).

This paper presents 100 metal-loss ILI assessment case studies using the Level 2 requirements of API 1163, third edition and the PRCI guidance document/spreadsheet. It also discusses Level 3 results. The case studies involve actual ILI assessments that were performed and accepted by pipeline operators to ensure the integrity of their pipeline systems. The results show that API 1163 third edition is more stringent than the second edition.

This paper shows that:

- About 15% of the prior ILI assessments did not include enough field NDE measurements to conduct a Level 2 validation to API 1163, third edition. The operators that accepted these ILI results would need to collect more field NDE measurements to successfully validate their inspections or justify that the inspections met Level 1 validation criteria.
- About one-third of the prior ILI assessments failed the API 1163 Level 2 validation requirements (Outcome 1). Based on the case studies, additional field NDE measurements are required to validate the inspections, or the inspections could have been rejected.
- About half the prior ILI assessments required additional analyses (Outcome 2). Many of these did not pass the practical guidelines issued by PRCI for Level 2 validations.
- Less than 5% passed the Level 2 validation (Outcome 3), with a performance that generally exceeded the ILI service provider’s performance specification.

The PRCI guidance document and spreadsheet also provide an “equivalent” tolerance for Outcome 2 cases, and they give a method to estimate the actual ILI performance using a Level 3 approach for all outcomes. Lessons learned from these evaluations are discussed and guidelines for use are given.

The modifications in API 1163, third edition, are significant and are likely to broadly impact pipeline operators using ILI as the basis for integrity decision-making. The third edition is more stringent than the second edition, and it will likely require more digs and increased tolerances when assessing which anomalies threaten pipeline integrity. Guidance is provided in this paper regarding key terms an operator must define when using the third edition. Clearly defining the important terms will make a pipeline operator’s ILI response program more consistent and defendable.

 https://doi.org/10.52202/078572-0088

Introduction

The first edition of API Standard 1163, In-Line Inspection [ILI] Systems Qualifications, was published in 2005 as a tool for pipeline operators to ensure accurate and repeatable ILIs. API developed the standard with a group of pipeline operators, ILI companies, and engineering consultants. The goal was to standardize ILI language, clearly define roles and responsibilities, set minimum expectations for ILI system providers, and provide guidance for pipeline operators to select and use ILI systems. The standard is an umbrella document and is used with NACE Standard Practice SP0102 and ANSI/ASNT ILI-PQ, which were originally developed around the same time.

API 1163 covers ILI preparation and tool selection, performance standards, verifying and validating reported results, quality systems, and other aspects of ILI projects. It has evolved through two later editions, the second edition issued in April of 2013 and a third edition in September of 2021. The second edition was incorporated by reference into the U.S. Code of Federal Regulations Title 49, Part 192 in 2019 and Part 195 in 2017. The second edition provides optional validation analyses, updated guidelines, and performance criteria.

The third edition of API 1163 is not yet incorporated into the U.S. Code of Federal Regulations, but the authors understand PHMSA is evaluating the standard and considering its incorporation. The third edition made many of the optional analyses in the second edition mandatory, thereby making the standard more prescriptive. Perhaps most importantly, the third edition provides explicit requirements for Level 1, Level 2, and Level 3 ILI system validations. However, the third edition does not define some important terms referenced in the validation requirements, leaving them subject to interpretation by inspection stakeholders.

Subsequently, PRCI developed a guidance document and spreadsheet[*] for conducting validations under the third edition of API 1163. The document and spreadsheet are gaining widespread use in the pipeline and inspection industries. Revisions of the PRCI documents based on user feedback are underway and expected to finish in 2025.

References to API 1163 throughout the remainder of this paper are specific to the third edition.

API 1163 validation levels

The validation levels in the third edition of API 1163 provide a tiered approach to evaluating the accuracy of an inspection, starting with a relatively simple process for Level 1 validations, followed by progressively more advanced analyses in Level 2 and Level 3 validations. Figure 1 summarizes the various levels.

[*] PRCI Project IM-1-06, "API-1163 Section 8 - Level 1, 2 & 3 Research," which is available through the PRCI website (www.prci.org).

https://doi.org/10.52202/078572-0088

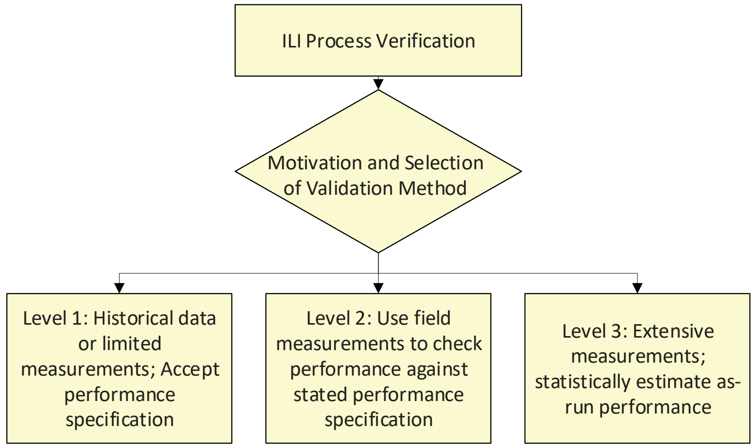

Figure 1. API 1163 Validation levels

API and PRCI ILI validation guidance

Level 1 guidance

Level 1 validations are allowed when:

- The pipeline being inspected has no significant anomalies (as discussed below),
- There are not enough ILI-reported anomalies to conduct a statistical validation,
- The pipeline being inspected is considered a low-risk pipeline, or
- The operator has significant experience using the same or similar ILI system on the same[†] or similar pipeline.

For Level 1 validations, API 1163 and the PRCI guidance document identify several key definitions the pipeline operator should make; however, API 1163 does not provide definitions for the key factors listed below. It is recommended that these parameters be defined within each operator's program for a consistent and defendable approach.

- Significant anomalies
- Similar pipelines
- Similar ILI systems
- Low risk

The following subsections offer some proposed definitions for consideration in support of consistent and defendable methodologies.

Significant anomalies

API 1163 states that "the pipeline operator shall define significant anomalies based on the operating conditions and risk posed by a pipeline segment." No further guidance is provided in API 1163 or in the PRCI guidance document.

[†] API 1163 does not explicitly state that a Level 1 validation can be based on inspections with similar ILI systems. Instead, it says the validation can be based on inspections with the same system, as discussed later in this paper.

 https://doi.org/10.52202/078572-0088

In the authors' experience, pipeline operators typically define a significant anomaly as one with a depth of 40% or 50% of the wall thickness (WT). Alternative definitions could be based on calculated failure pressures.

Similar pipelines

API 1163 says that "A similar line is one where the same level of performance can be reasonably expected based on past experience." API 1163 also states that several "examples of parameters that can be used to define similarity could be same or similar diameter, same or similar wall thickness range (within the same essential variable range used to qualify the performance specification), [and] seam type (when running a crack tool for long seam assessment). Consideration should be given to relevant parameters that affect ILI performance such as threat being assessed and should consider the specific tool technology as well as tool run dynamics considerations. For specific ILI systems, additional parameters such as pressure, tool velocity, product type, or threat characteristics may also be applicable."

The PRCI guidance document lists attributes that may affect measurement performance, including "wall thickness, seam type, the threat causing the anomaly, the anomaly morphology, anomaly depth range, pipeline pressure, ILI tool velocity, and transport product." To evaluate whether a pipeline is similar to another pipeline, the PRCI guidance document offers qualitative and quantitative methods of assessment.

The qualitative assessment is a checklist that gives the following considerations:

- "The morphology of the anomalies of interest based on a review of photographic record from excavation reports on the prior pipeline is the same as the new pipeline" ...
- "The complexity of anomalies is similar.
- The degradation mechanism causing the anomalies is the same (e.g., internal corrosion).
- The prior pipeline ILI run has been fully verified, performance has been validated using a Level 2 or Level 3 validation as described in these guidelines, and the measurement error is well-characterized.
- The prior pipeline ILI run tool speed is within the vendor's specification.
- The pipeline diameter and wall thickness are covered by the range of ILI essential variables defined by the vendor.
- The pipeline cleanliness is similar."

The quantitative assessment is much more complicated and calls for the use of an Anderson-Darling test[‡]. The Anderson-Darling test evaluates whether a sample is likely to have come from a population with a specific distribution, or alternatively, if two samples come from the same distribution. The test requires a minimum of eight points in each dataset. The authors are not aware of pipeline companies employing the Anderson-Darling test to determine similarities.

Similar ILI systems

API 1163 does not define similar inspection systems. Instead, it says that Level 1 validations can be performed "where an operator has extensive ILI experience using the same inspection system (i.e. the same tool, configuration, and analysis practice) on the pipeline being inspected or on similar lines." The PRCI guidance document allows Level 1 validations to be based, in part, on similar pipelines

[‡] Anderson, T.W., Darling, D.A., 1954. A Test of Goodness of Fit. J. Am. Stat. Assoc. 49, 765–769. https://doi.org/10.1080/01621459.1954.10501232

inspected with similar ILI systems. In addition, it says "the ILI tool vendor company, model, and software version [should be] substantially the same for both pipelines. When this criterion is used to determine *'similar pipelines'* for a Level 1 assessment (Section 4.3.3), the vendor need not be the same if the measurement performance is substantially the same."

Low (Levels of) risk

API 1163 states that a Level 1 validation when the anomaly population represents low levels of risk as determined or defined by the pipeline operator. No further guidance is provided.

The PRCI guidance document offers the following considerations regarding determining whether a pipeline is low risk:

- The pipeline has an assessed probability of failure and when combined with consequence, the risk falls below the operator's risk threshold and will remain below the threshold until the next planned inspection.
- The anomaly growth rate used to assess risk and rupture pressure ratio is validated and considers growth rate uncertainty, including the possibility of higher-than-expected growth.

Low risk might also be inferred where the number of anomalies identified by a verified run is so small that obtaining a statistically significant sample of validation data is not realistic.

The definition of low levels of risk or its implications is left to the pipeline operator in this paper.

Level 2 guidance

API 1163 Level 2 validations are based on a statistical approach called *null hypothesis testing*. Typically, *null hypothesis testing* uses statistical *inference* to assess whether a null hypothesis can be <u>rejected</u>. In API 1163, the null hypothesis is that the ILI met its performance specification, i.e., that the actual depths of ILI-reported anomalies are within the performance specification's tolerance (e.g., ±10% of the wall thickness) and certainty (80% of the time). So, the null hypothesis testing assesses whether to reject the hypothesis that the tool met its specification at a given confidence level. *Inference* allows the user to make a judgment (whether the performance specification can be rejected) about a population (all ILI anomalies on a pipeline) based on a sample (a subset of the population). Level 2 validations require field measurements, the number of which depends on whether the measured anomaly dimensions (e.g., depth) are within specification.

In API 1163, the concept of null hypothesis significance testing is expanded to help the user judge whether the performance specification could or should be <u>accepted</u> based on a sample of correlated ILI-reported and field-measured data. As such, Level 2 validation does not <u>prove</u> that an ILI met its performance specification, but it provides confidence that the specification was reasonable for the ILI being evaluated.

Level 2 validations in API 1163 estimate the range of certainties that could represent the ILI at its stated tolerance. Recall the ILI performance specifications typically state a tolerance and a certainty. The tolerance describes the error in ILI-reported depths and the certainty describes the likelihood or how frequently actual depths will be within the tolerance of the depth reported by the ILI. The field-measured depths are often considered actual depths, though consideration should be given to the field-measure error. The estimated range of the certainty has an upper and lower limit or endpoint, which is a function of the confidence level used in the analysis. If the certainty range includes or

 https://doi.org/10.52202/078572-0088

exceeds the stated certainty, the inspection has the potential to meet its performance specification (i.e., it is plausible that the tool met or exceeded the performance specification).

Some definitions are introduced in Level 2 assessments. These include:

- Confidence levels
- Single-sided and double-sided confidence intervals
- Upper and lower confidence bounds or endpoints
- Outcomes 1, 2, and 3
- Recommended lower confidence bound

The following subsections elaborate and offer additional information on the terminology introduced for Level 2 validations.

Confidence levels

The statistical calculations in API 1163, the PRCI guidance document, and the PRCI spreadsheet are based on confidence levels typically of 90% or 95%. Users sometimes confuse certainty levels with confidence levels. Certainty levels describe how often an ILI-reported depth is within the stated tolerance of the actual depth. For metal-loss anomalies, this is how often the ILI-reported depths are within ±10% of the field-measured depth. The stated certainty is typically 80%, meaning roughly eight of ten measurements will be within the stated tolerance. This generally applies to a randomly selected sample of data.

Confidence levels deal with repeated tests and describe how frequently different samples of data taken from the population of ILI-reported anomalies will meet some criteria. A sample is the ILI-reported anomalies selected by the pipeline operator for field measurements. Typically, this sample is less, maybe much less, than the total ILI population. The confidence level refers to how often a random sample of field measurements will meet the stated certainty level.

Single-sided (one-tail) and double-sided (two-tail) confidence intervals

Single and double-sided confidence intervals are statistical concepts associated with the certainty range of ILI survey results. As the confidence level increases, the certainty range gets larger. The certainty range can be single-sided, representing the likelihood the ILI-reported depth is greater than or equal to the actual depth minus the stated tolerance (i.e., it is generally conservative):

$$\text{ILI-reported depth} \geq \text{Actual depth} - 10\%\text{WT}$$

or it can be double- sided, stating whether the ILI-reported depth is within the tolerance of the actual depth (i.e., the absolute value of the ILI-reported depth minus the actual depth is within the tolerance provided in the ILI performance specification):

$$\text{Actual depth} + 10\%\text{WT} \geq \text{ILI-reported depth} \geq \text{Actual depth} - 10\%\text{WT}$$

Figure 2 illustrates single- and double-sided confidence intervals.

https://doi.org/10.52202/078572-0088

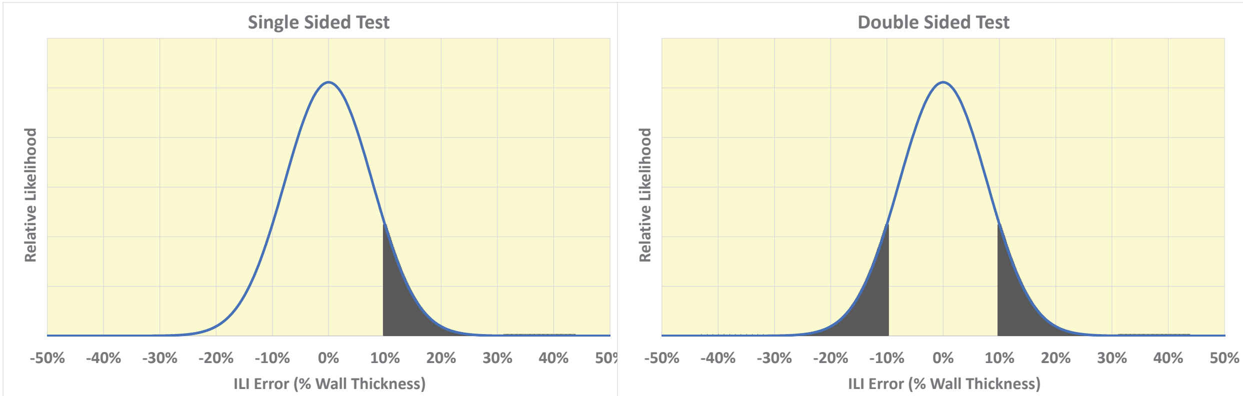

Figure 2. Single- and double-sided confidence interval

Upper and lower confidence bounds or endpoints

A Level 2 validation estimates a double-sided range of certainties associated with an ILI.[§] At a given confidence level, the upper confidence limit or endpoint represents the "best" the ILI may have performed, and the lower confidence limit or endpoint represents the "worst" the ILI may have performed. Thus, for example, if the range of uncertainties is [27%,73%], the actual certainty of all ILI-reported anomalies is between 27% and 73% at the confidence level chosen.

Outcomes 1, 2, and 3

Outcomes 1, 2, and 3 are related to where the tolerance stated in the ILI performance specification falls relative to the upper and lower confidence bounds or endpoints. Outcome 1 refers to the situation where the stated tolerance is above the upper and lower confidence bounds. Outcome 2 is when the stated tolerance is within the upper and lower confidence bounds. Outcome 3 is when the stated tolerance is less than the upper and lower confidence bounds. The three outcomes are shown in Figure 3.

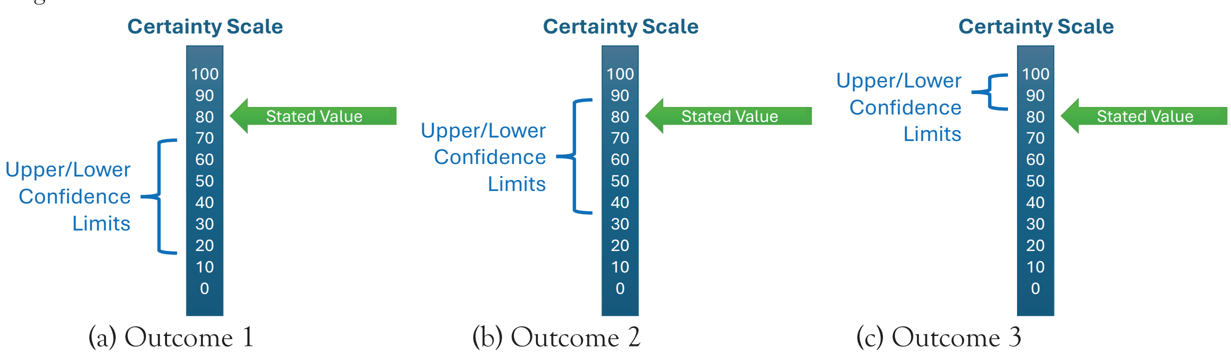

Figure 3. Three outcomes in Level 2 validations

API 1163 states "Outcome 1 generally means the ILI did not meet its performance specification, Outcome 2 means the ILI may have met its performance specification, and Outcome 3 means the ILI met or exceeded its performance specification."

[§] The third edition of API 1163 uses a double-sided confidence interval, while the second edition uses a single-sided confidence interval.

https://doi.org/10.52202/078572-0088

Recommended lower confidence bound

API 1163 states that a pipeline operator should establish an acceptable lower confidence bound in accordance with the operator's risk tolerance, but it does not offer further guidance. The PRCI guidance document recommends a lower confidence interval endpoint of 60%, which is implemented by default in the PRCI spreadsheet, with a two-sided confidence level of 90%. A lower endpoint of 60% implies the certainty associated with an ILI-stated tolerance (e.g., ±10% WT) may have been as low as 60%, versus a typical ILI performance specification value of 80%. The PRCI guidance document says that a 60% lower confidence interval endpoint corresponds to an "equivalent" tolerance of ±15.2%WT. Equivalent tolerances are discussed later in this paper.

Level 3 guidance

In API 1163, Level 3 validations are more sophisticated statistical methods meant to provide more accurate estimates of the PODs, POIs, and actual sizing accuracies. Two example methodologies are provided in a nonmandatory appendix (annex) within the third edition of API 1163. API 1163 states that Level 3 analyses can be used to estimate the actual ILI performance based on field-measured data.

Level 3 validations serve a different purpose than Level 2 validations. Level 3 validations seek to estimate the actual ILI performance based on field measurements. Level 3 validations use tolerance interval calculations that estimate the tolerance associated with the certainty level given in the ILI performance specification.

The safety factors implied in Level 2 and Level 3 validations differ because each validation level serves a different purpose. Level 2 validations allow the use of the ILI tool tolerance specifications even when the inspection may not have met the specification, while Level 3 calculates an "actual" tolerance. As a result, the Level 3 tolerances are generally larger than the ILI performance specification when the inspection just passes the Level 2 Outcome 2 validations considering the PRCI guidance document. The safety factor for use with Level 3 tolerances is not discussed in either API 1163 or the PRCI guidance document.

The two example methodologies included in the API 1163 appendix are statistical tolerance intervals and Bayesian inference. The statistical tolerance interval method is easier to implement in Excel and is incorporated into the PRCI guidance document and spreadsheet. The method estimates the actual ILI performance from the available field measurements. The estimated performance is expressed as a bias and an "actual" tolerance.

The tolerance interval in the PRCI spreadsheet is calculated at the certainty level stated in the ILI performance specification (80% by default). The interval is typically calculated as a single-sided confidence level of 95%. This means that at least 95% of random samples from the ILI-reported population are expected to be within the calculated tolerance interval 80% of the time (i.e., the absolute values of the random sampling of ILI-reported minus field-measured depths will be within the calculated tolerance interval 80% of the time).

Case studies

This paper details the analysis of 100 metal-loss ILI assessments using the Level 2 requirements of API 1163, third edition, and the PRCI guidance document/spreadsheet. It also discusses Level 3

https://doi.org/10.52202/078572-0088

tolerances calculated using the PRCI spreadsheet for the same 100 cases. The case studies involve actual ILI surveys previously performed by pipeline operators to ensure the integrity of their pipeline systems. All validations are for ILI-reported versus field-measured depths (i.e., lengths, widths, and failure pressures** were not considered).

The ILI surveys were performed between about 2010 and present, with an emphasis of the case studies on more recent inspections. The surveys are thought to represent typical pipeline company practices in terms of inspection tool selection, data analysis procedures, dig selection procedures, and field-measurement methods. All are metal-loss inspections conducted using either magnetic flux leakage (most cases) or ultrasonic wall thickness inspections (several cases). The number of field measurements per inspection ranged from a handful into the thousands and were analyzed in the condition as provided by the operator.

The authors used the ILI-reported and field-measured metal-loss depths in terms of percent wall thickness, changing the actual wall thickness to 0.250 inches to maintain anonymity. There are potential errors in some of the dig records associated with a handful of ILI response programs. In one case, the reported field depths are suspiciously close to the reported ILI depths; in the authors' experience, this is rarely true. In several cases, multiple field depths may have been correlated to a single ILI reported anomaly and vice versa. Nonetheless, the data were used as assembled.

The authors used the PRCI spreadsheet with minimal modifications. The ILI tolerance was assumed to be ±10% WT with a stated certainty of 80%. That is, it did not account for tolerances that might be larger for anomaly types, such as pinholes, pitting, and grooves or slots. As such, the validation results discussed in this paper are thought to be conservative. The calculations range was expanded to incorporate the field-measured depths, which totalled around 15,000.

Not enough data for analysis

Based upon the data provided and considering the methodology offered in API 1163, 13 of the 100 cases have fewer than six field measurements, which is less than the minimum number needed to just pass a Level 2 validation (i.e., Outcome 2 with a lower confidence endpoint of 60%). These cases would be candidates for a Level 1 validation if the worst reported anomaly were not significant in the pipeline operator's judgment or if the operator has extensive experience using the ILI system on the same or similar pipelines:

- Seven of the 13 cases had maximum measured field depths less than 40% WT and two had depths between 40% and 50% WT. These nine pipelines may not have other reported anomalies greater than 40% or 50% WT – i.e., they may not have other remaining "significant" defects.
- The other four cases had field-measured depths over 50% and would fail the Level 1 assessment without extensive experience.

In summary and based on these results, it is likely some ILI response programs will need to increase their field data collection programs to provide enough data for API 1163 validations. A minimum of

** Validating an inspection based on failure pressures rather than depths is more difficult because failure pressure is a function of both depth and length. The PRCI spreadsheet is not currently able to address failure pressure validations.

 https://doi.org/10.52202/078572-0088

six, and preferably more, field-measured depths should be obtained before attempting to perform a Level 2 validation under API 1163.

Level 2 validation – Outcome 1

Level 2 (and 3) validation calculations were conducted for the remaining 87 cases. Forty (40) produced Outcome 1, which indicates there is statistical evidence the ILI did not meet its performance specification. All these cases failed the upper endpoint criterion, and all but six failed the lower endpoint criterion. These are cases where the validation concluded there was less than a 5% chance the ILI met the performance specification. Nearly all cases had over 15 field measurements, and some had several thousand. Additional digs are not expected to change the outcome for these cases.

An approach to performing a Level 2 validation, not discussed in API 1163 or the PRCI documents, is to artificially increase (adjust) the ILI tolerance until the validation passes the endpoint criteria in Level 2. The tolerances needed to meet a Level 2 validation range from 11% to over 25%, as shown in Figure 4. In many cases, a tolerance of 15% or less resulted in a successful Level 2 validation. Because the adjusted tolerances exceed the stated tolerance of 10%, they are more conservative than the stated tolerance (which did not meet API 1163).

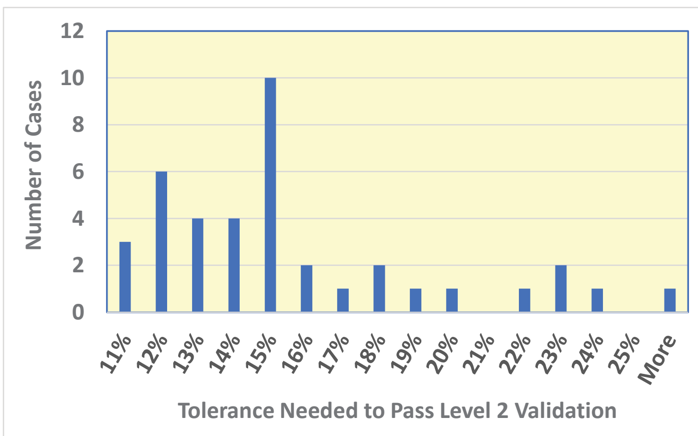

Figure 4. Tolerance needed to meet Level 2 validation requirements

In summary, the chances of reaching Outcome 1 are high, suggesting some or many ILI response programs will need to either use higher tolerances in integrity decision-making or, for cases with few field measurements, increase their field data collection efforts.

Level 2 validation – Outcome 2

Forty-four (44) cases corresponded to Outcome 2, where the performance specification may or may not have been met. Of these, 19 did not meet the lower endpoint criterion and do not meet the ILI performance specification per the PRCI guidance document and spreadsheet. Many of these cases failed because the confidence interval width was large due to a small number of data points. Sixteen (16) had less than 15 data points, suggesting additional field measurements could decrease the interval width and increase the chance of passing the endpoint criteria in a Level 2 validation.

https://doi.org/10.52202/078572-0088

The PRCI spreadsheet also provides an equivalent tolerance for Outcome 2 cases. The equivalent tolerances ranged from 10% to 56%, suggesting some of the ILIs were far from meeting their performance specification. These equivalent tolerances could be used in integrity decision-making, instead of using the ILI performance specification. It is noted that using higher tolerances may lead to more required remediation activities.

The equivalent tolerances for the 19 cases that did not meet the lower endpoint criterion but did meet the upper endpoint criterion ranged from 16% to 56% WT, again suggesting some of the ILI surveys were far from meeting their performance specification. The total tolerances needed to just meet the Level 2 endpoint criteria for the same cases ranged from 11% to 32% WT, a modest improvement for most compared to the equivalent tolerances. Figure 5 compares the equivalent tolerance and the tolerance needed to pass the Level 2 endpoint criteria. Using higher tolerances may lead to more required remediation activities.

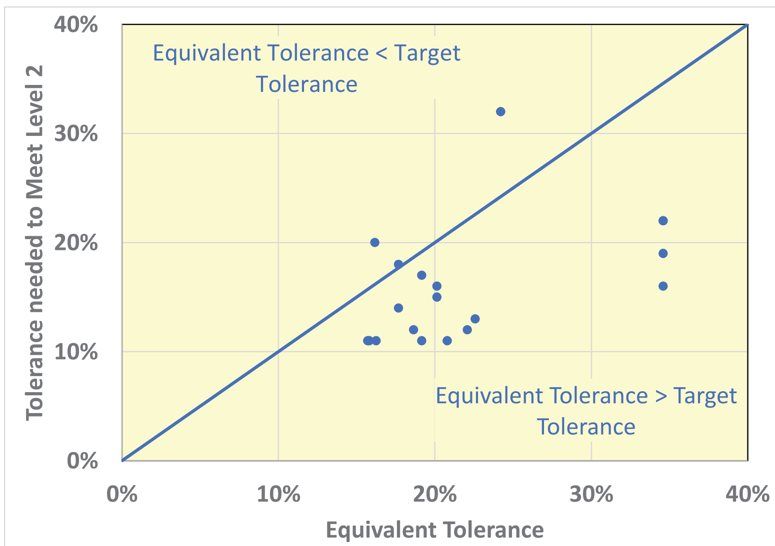

Figure 5. Tolerance Needed to Meet Level 2 versus equivalent tolerance for cases failing Level 2 validation criteria

The remaining 25 cases met the upper and lower endpoint criteria provided in the PRCI guidance document and spreadsheet and therefore passed the Level 2 validation. The number of field measurements in these cases ranged from 6 (the minimum needed for a Level 2 validation) to over 500. Per API 1163, a pipeline operator can use the tolerances given in the ILI performance specification for these cases.

An equivalent tolerance was also computed for the 25 cases that passed the Outcome 2 endpoint criteria for informational purposes. Here, the equivalent tolerances ranged from 10% to 15% WT, a modest increase from the assumed tolerance of ±10% WT. Figure 6 shows the relationship between the equivalent tolerance and the tolerance needed to just pass the Level 2 criteria. Here, nearly all the tolerances needed to just pass are equal to or less than the tolerance in the ILI performance specification (10%) and the equivalent tolerance.

 https://doi.org/10.52202/078572-0088

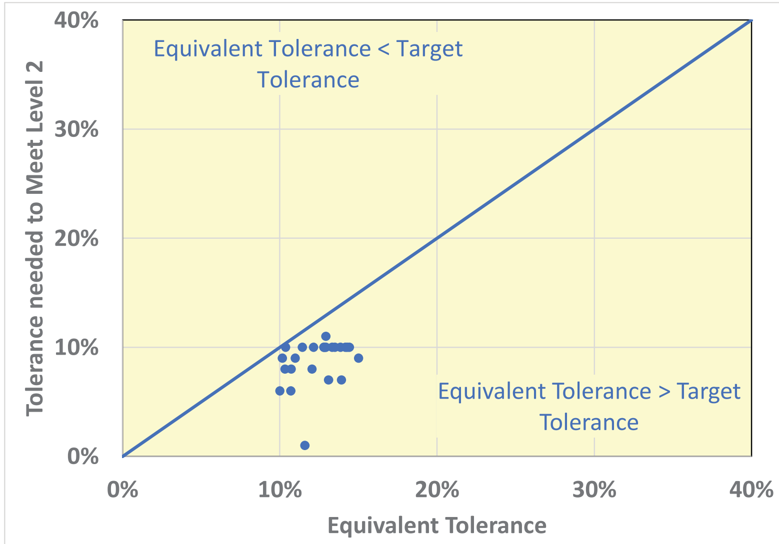

Figure 6. Tolerance needed to meet Level 2 versus equivalent tolerance for cases meeting Level 2 validation criteria

In summary, reaching Outcome 2 occurred for less than 50% of the case studies in which the results were previously determined to be acceptable. Of these, a sizable fraction did not pass the lower endpoint criterion provided in the PRCI guidance document. For these cases, using the tolerance needed to pass the endpoint criterion is expected to increase the number of anomalies requiring remediation. In cases with few field measurements, increasing field data collection efforts to provide more data for API 1163 validations could be helpful. For cases that met the lower endpoint criterion, the tolerance needed to just pass Level 2 is sometimes less than the ILI performance specification value. This could decrease the number of reported anomalies requiring remediation.

Level 2 validation – Outcome 3

Only three cases corresponded to Outcome 3, where the performance specification may have been exceeded per API 1163 and the PRCI guidance document and spreadsheet. This finding is unlikely. In these cases, the equivalent tolerances were higher than that given in the ILI performance specification (10% to 12% WT), while the tolerances needed to pass the Level 2 validation were less (6% to 8% WT).

Level 3 validation

The previous sections discussed Level 1 and 2 validations. The PRCI spreadsheet calculates Level 3 biases and tolerances regardless of the number of data points. The PRCI guidance document recommends at least ten correlations; more correlations provide narrower tolerances. More correlations also provide more accurate Level 3 results.

Level 3 biases for the 100 cases[††] represent the average or mean errors associated with ILI-reported depths, which are shown in Figure 7. Here, the biases reflect the amount added to the ILI-reported depth to match the actual depth (a positive bias means the ILI underreported the actual depth and vice versa). There is a slight tendency for the ILI surveys to undercall actual anomaly depths. The

[††] Including those with fewer than six data points needed for a Level 2 analysis.

https://doi.org/10.52202/078572-0088

average across all ILI surveys was calculated as approximately2% WT, and 11 ILI surveys showed absolute biases greater than or equal to 0.1 (10% WT). The largest bias (-30% WT) corresponded to an inspection with only four data points, but even inspections with over a thousand data points showed absolute biases of 10% WT or more.

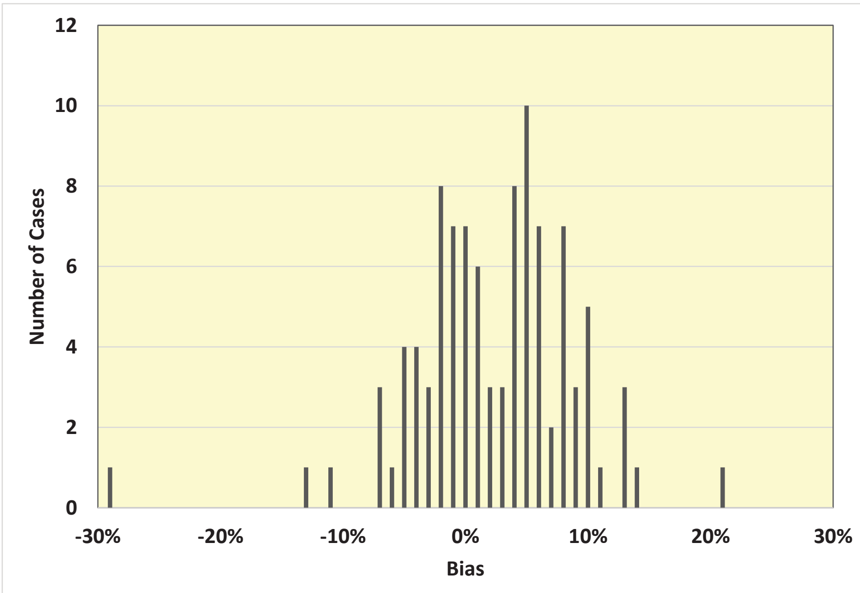

Figure 7. Level 3 biases for 100 example pipelines

Nonetheless, 59 out of 100 biases were less in absolute value than 0.05 (5% WT). This is consistent with the authors' experience that most ILIs are reasonably accurate. While the authors have seen ILI results with biases greater than 10% WT, they are unusual.

Figure 8 shows 95 of the 100 calculated Level 3 tolerances (five with tolerances above 50% WT are not displayed; the largest calculated Level 3 tolerance was 113% WT). All five correspond to cases with less than 10 correlations, the recommended minimum number of correlations for a Level 3 validation. Twelve cases above 30% WT had less than 10 correlations, suggesting that few correlations correspond to large Level 3 tolerances (as expected).

 https://doi.org/10.52202/078572-0088

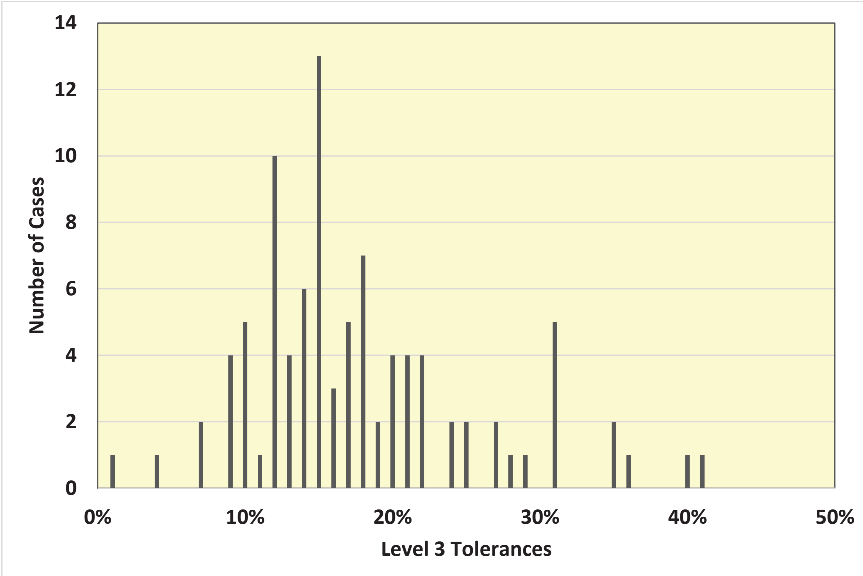

Figure 8. Level 3 tolerances

All but 12 of the Level 3 tolerances are greater than the assumed ILI tolerance of ±10% WT:

- The Level 2 Outcome 1 tolerances ranged from 7% WT to 82% WT
- The Level 2 Outcome 2 Fail tolerances ranged from 3% WT to 113% WT
- The Level 2 Outcome 2 Pass tolerances ranged from 1% WT to 16% WT and
- The Level 2 Outcome 3 tolerances were all 9% WT.

As previously discussed, factors of safety associated with the tolerances from Level 3 validations are not the same as those determined when passing the Level 2 endpoint criteria. The Level 3 tolerances tend to be larger than the tolerances needed to pass the Level 2 endpoint criteria, as shown in Figure 9.

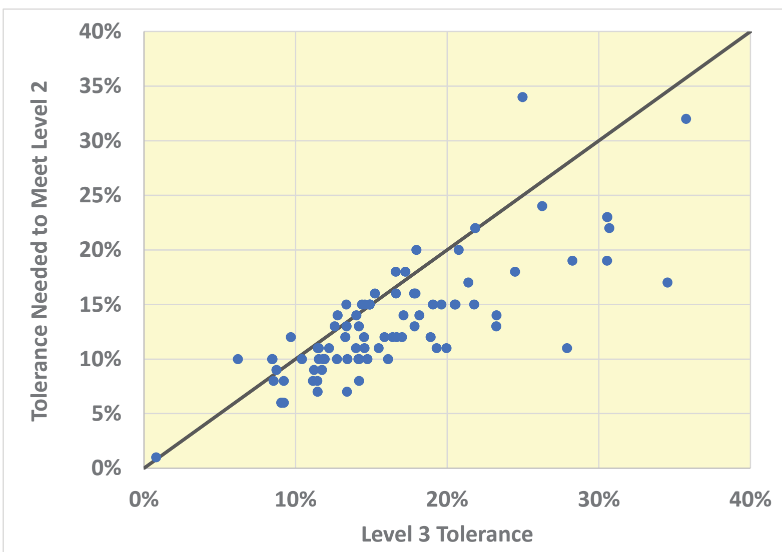

Figure 9. Tolerance needed to meet Level 2 validation versus tolerance from Level 3 validation

Conclusions

One hundred sets of data representing previously conducted ILI and field depth measurements were assembled and used to evaluate the Level 1, 2, and 3 validation approaches given in API 1163 (third edition) relative to historical approaches and ILI response programs. Some interesting results were observed:

- Expansion of field-measurement programs should be considered to collect more data for API 1163 validations. Out of the 100 cases evaluated for this paper, 13 had less than six field measurements, which is the minimum needed to successfully conduct a Level 2 assessment. These could be candidates for Level 1 validations, provided the worst reported anomalies were less severe than "significant" anomalies, as defined by the operator, or if the operator has extensive experience with a similar ILI system on similar pipelines.
- There is a significant chance a Level 2 validation will result in Outcome 1, where there is statistically significant evidence the ILI did not meet the performance specification, regardless of the number of field measurements made. When there is evidence the ILI did not meet its performance specification, higher tolerances will likely need to be considered in integrity decision-making for these cases. Either the equivalent tolerances given in the PRCI spreadsheet or the tolerances needed to meet a Level 2 validation could be used. The former is more conservative.
- Many ILI validations will result in Outcome 2, where the statistical conclusion is that the ILI may or may not have met its performance specification. A little over half of the Outcome 2 cases passed the Level 2 validation criteria, in which case the tolerances in the ILI performance specification can be used in integrity decision-making per API 1163. Alternatively, equivalent tolerances given in the PRCI spreadsheet or the tolerances needed to meet a Level 2 validation can be used. In many cases, the tolerances needed to meet a Level 2 validation will be less than the tolerance specified in the ILI performance specification.
- For Outcome 2 cases that failed the Level 2 validation criteria, higher tolerances should be considered in integrity decision-making. As before, equivalent tolerances given in the PRCI spreadsheet or the tolerances needed to meet a Level 2 validation can also be used. These tolerances usually equal or exceed the tolerance specified in the ILI performance specification.
- Very few cases passed the Level 2 validation with Outcome 3, which indicates the ILI may have exceeded its performance specification. In these cases, the tolerance needed to meet a Level 2 validation is generally less than that given in the performance specification.
- Level 3 validation tolerances offer an alternative to the Level 2 tolerances, but they are often larger than the tolerances needed to meet a Level 2 validation.

The modifications in API 1163, third edition, are significant and are likely to broadly impact pipeline operators using ILI as the basis for integrity decision-making. The third edition is more stringent than the second edition, and it will likely require more digs and increased tolerances when assessing which anomalies threaten pipeline integrity and calculating reassessment intervals. Using a tolerance that just passes the Level 2 validation generally provides a more realistic and less conservative tolerance than using a Level 3 tolerance.

https://doi.org/10.52202/078572-0088

https://doi.org/10.52202/078572-0088