A statistical approach to material verification of expected grade through opportunistic field measurements

by Dr. Steven Palkovic, Parth Patel, Soheil Safari, and Dr. Simon Bellemare
Massachusetts Materials Technologies LLC, Waltham, MA, USA

Pipeline Pigging and Integrity Management Conference

Marriott Marquis Hotel, Houston, USA
February 17-21, 2020

Organized by
Clarion Technical Conferences *and* **Great Southern Press**

Material verification is the process of measuring material property data on in-service pipelines when existing records are not traceable, verifiable, and complete (TVC). The new regulations for gas transmission pipelines include §192.607 which provides requirements for a material verification process. This work describes a statistical approach to meet these provisions by achieving a 95% confidence level on material sampling and conservatively accounting for measurement uncertainty when performing a nondestructive evaluation (NDE) of material strength properties. This analysis allows for the determination of whether a pipeline segment exceeds the expected grade, is more conservative than the expected grade, or requires additional testing and the use of additional techniques to make a final determination. Strength measurements collected during integrity excavations on more than 200 pipe joints with several pipeline operators are used to present the implications of applying this process to a diverse network of pipeline segments.

1. Introduction

The Pacific Gas and Electric (PG&E) pipeline failure in San Bruno, CA highlighted the need to ensure traceable, verifiable, and complete (TVC) records and to reconfirm the maximum allowable operating pressure (MAOP) on critical pipeline assets [1]. The Pipeline and Hazardous Materials Safety Administration (PHMSA) later issued two advisory bulletins directing pipeline operators to use TVC records to integrate accurate data and information from their entire pipeline system for establishing MAOP [2,3]. These concepts were incorporated within the Notice of Proposed Rulemaking (NPRM) [4] that was subsequently discussed and revised through several Gas Pipeline Advisory Committee (GPAC) meetings [5] to clarify the language and implications of the new regulations.

On October 1, 2019, PHMSA published the final rule which included §192.607 that defines a process to verify material properties through either nondestructive or destructive measurements [6]. Material verification is required as part of opportunistic testing within ongoing integrity excavations and at above ground locations when material records are not TVC and when referenced by other sections of Part 192, including §192.619 MAOP, §192.624 MAOP reconfirmation, §192.632 engineering critical assessment for MAOP reconfirmation, and §192.712 analysis of predicted failure pressure. When material verification is required, the new regulations provide prescriptive sampling guidelines of one excavation per mile rounded up to the nearest whole number, with a maximum of 150 excavations for a line segment that is more than 150 miles long. In lieu of the prescribed requirements, operators may submit an alternative sampling plan that uses a statistical basis to verify the material properties of a pipeline segment with at least a 95% confidence level.

This paper describes a framework for an alternative sampling plan to compare measured data with an initial expectation of the pipeline material properties. For this application, the assertion being tested is that the material properties meet the specified minimum yield strength (SMYS) and ultimate tensile strength (UTS) requirements of an expected steel grade. This approach further considers uncertainty of nondestructive measurements of the strength properties that are allowed by §192.607 provided that the chosen process meets additional requirements for nondestructive methods. The statistical process is used for a case study on vintage pipe joints from the same line segment to illustrate how material properties can be verified at a specified confidence level through an increasing number of random samples. The same process is then utilized for hundreds of pipe joints from different line segments to estimate the minimum number of samples required to verify material properties for different steel grades, and to illustrate the benefits of using a more accurate method to measure strength properties. Additional considerations that have a statistical basis and impact the practicality and conservatism of the analysis are also discussed.

2. Summary of §192.607 Method Requirements

A significant addition to the regulations through §192.607 is the explicit allowance of nondestructive methods for the determination of steel grade through the measurement of the yield strength and ultimate tensile strength. This acknowledges recent advancements in nondestructive testing and industry efforts to validate these new techniques and processes. A nondestructive approach provides a cost-effective alternative to traditional destructive testing because it allows for in-situ measurements on an exposed portion of the pipe joint while it remains in-service. The nondestructive solution considered in this study is the Hardness, Strength, and Ductility (HSD) Tester which is the only commercial implementation of the contact mechanics technique known as frictional sliding. The HSD was determined to be, "the best performing technique with the lowest MAPE, highest correlation coefficients, and highest quantity of data within the specified error bands for both yield strength and ultimate tensile strength," based on a Pipeline Research Council International (PRCI) validation study of available nondestructive methods on 50 blind samples of vintage pipe joints covering a range of steel grades, geometry, and manufacturing processes [7].

The use of any nondestructive methods for material verification must have a procedure which meets several requirements that are defined in §192.607. These criteria are summarized in Table 1, along with the prescribed guidelines for a destructive approach that necessitates the removal of a full-round pipe cylinder for subsequent laboratory testing of tensile cut-outs. Table 1 also includes commentary on how the current implementation of the HSD meets these requirements.

Table 1: Summary of §192.607 requirements for material verification methods

Topic	§192.607 Requirement [6]	HSD Method
Destructive Procedure	§192.607(c)(2): For destructive tests, at each test location, a set of material properties tests for minimum yield strength and ultimate tensile strength must be conducted on each test pipe cylinder removed from each location, in accordance with API Specification 5L.	HSD tests can be performed at any exposed location without the need for pressure reduction, service interruption, or sample coupon extraction.
NDE Procedure	§192.607(c)(1): For nondestructive tests, at each test location, material properties for minimum yield strength and ultimate tensile strength must be determined at a minimum of 5 places in at least 2 circumferential quadrants of the pipe for a minimum total of 10 test readings at each pipe cylinder location.	The HSD tests 2 circumferential quadrants and gathers hundreds of measurements during each frictional sliding test. This data is analyzed to ensure at least 5 averaged readings at each quadrant, and a minimum of 10 for the pipe joint.
NDE Special Requirements	§192.607(d)(1): Use methods, tools, procedures, and techniques that have been validated by a subject matter expert based on comparison with destructive test results on material of comparable grade and vintage.	The HSD method was validated through PRCI NDE-4-8 which showed it was the best performing technique for measurement of yield and UTS [7].
	§192.607(d)(2): At each excavation, determination of material property values must conservatively account for measurement inaccuracy and uncertainty using reliable engineering tests and analysis.	Section 3.3 provides the current measurement uncertainty for the HSD, which has the lowest inaccuracy of all nondestructive methods per the results of PRCI NDE-4-8 [7].
	§192.607(d)(3): If nondestructive tests are performed to determine strength, the operator must use test equipment that has been properly calibrated for comparable test materials prior to usage.	The HSD is calibrated daily on a reference steel sample prior to testing on unknown materials.

3. Statistical Components of an Alternative Sampling Plan

The objective of a statistical analysis for material verification is to account for uncertainties from randomly sampling locations along a pipeline and to consider uncertainties from the method used to measure material strength properties to determine an estimate of the upper or lower bound strength properties of the entire population. In the following sections, the statistical basis for confidence intervals, hypothesis testing, and measurement uncertainty are introduced.

This analysis assumes that the pipeline material strength property distribution can be reasonably represented by a normal distribution that is characterized by a population mean μ and population standard deviation σ. A pipeline population is composed of hundreds of joints that were constructed and installed within a narrow timeframe and with similar characteristics. The new regulations include §192.607(e)(1) which provides guidelines for the determination of a population, and specifies the consideration of nominal wall thickness, grade, manufacturing process, and vintage as defined by pipes for which the manufacturing and construction dates are within 2 years [6]. This information can be obtained through a combination of construction records, material test records, in-line inspections (ILI), and direct assessments.

During a material testing program, a set of n random samples of the material strength $x_i = \{x_1, x_2, \ldots, x_n\}$ are obtained from the population. These measurements can be used to calculate the sample mean,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

and sample standard deviation,

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}. \tag{2}$$

It is assumed that the sample size is small compared to the size of the population and the population standard deviation is unknown, so that the Student's t distribution is used with degrees of freedom $v = n - 1$ to estimate the properties of the population. A comparison of the Student's t distribution and standard normal distribution is shown in Fig. 1 for varying degrees of freedom. These plots show that the t distribution is wider resulting in increased uncertainty at low sample sizes, and converges to the standard normal distribution as $n \to \infty$.

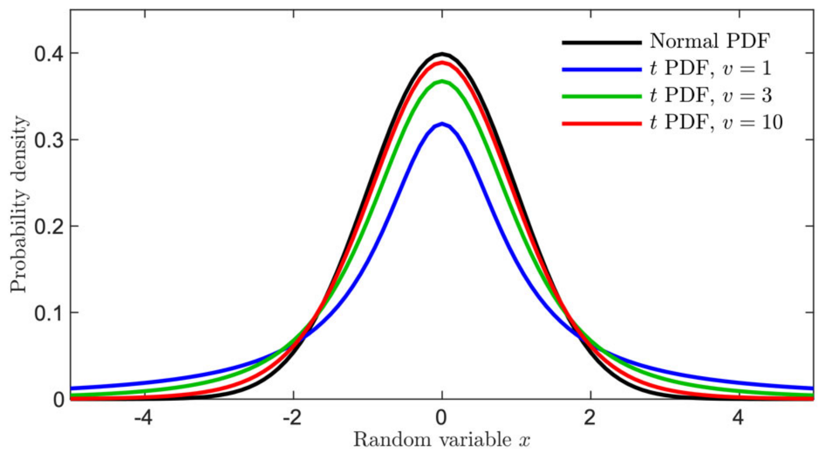

Fig. 1: Comparison of standard normal distribution and Student's t distribution for $v = 1, 3$ and 10.

3.1. Confidence Intervals

A confidence interval relates statistics obtained from random sampling to bounds of the population properties at a specified confidence level of $1 - \alpha$, where α denotes the significance level of the confidence interval (e.g. $\alpha = 0.05$ gives a 95% confidence level). The confidence level defines the probability that a computed interval will include a fixed population parameter. For this analysis we consider the population mean, meaning that for a 95% confidence level we would expect 95% of confidence interval estimates to include the population mean. This concept is illustrated in Fig. 2 for a significance level $\alpha = 0.05$ and population mean $\mu = 0$.

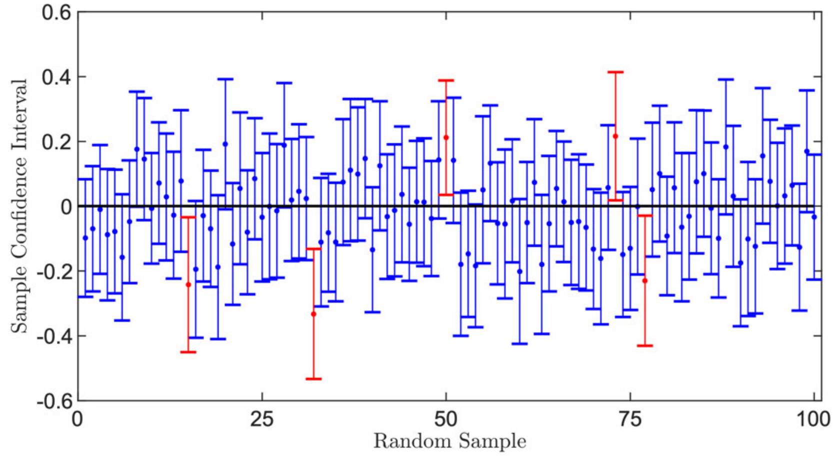

Fig. 2: Illustration of the confidence level of a confidence interval, where 95 out of 100 confidence intervals (95%) contain the population mean $\mu = 0$ for $\alpha = 0.05$.

The probability for the selected confidence level is obtained from the area under the Student's t-distribution up to a critical value $t_{(\alpha,v)}$, as shown in Fig. 3. The critical value $t_{(\alpha,v)}$ can be determined from a look-up table or inverse cumulative distribution function based on the significance level α and degrees of freedom $v = n - 1$. The upper and lower bound confidence interval estimates of the population mean can be determined as,

$$\mu_{lb} = \bar{x} - t_{(\alpha,v)} \frac{s}{\sqrt{n}} \quad \text{(lower bound)} \tag{3a}$$

$$\mu_{ub} = \bar{x} + t_{(\alpha,v)} \frac{s}{\sqrt{n}} \quad \text{(upper bound)} \tag{3b}$$

where $s/\sqrt{n}$ is the standard error of the sample mean. The uncertainty from random sampling is defined by the $t_{(\alpha,v)}\, s/\sqrt{n}$ term in Eqs. 3a and 3b, which is plotted in Fig. 4 for varying α, s and n. These results show that the sampling uncertainty is greatly reduced for the first 10 to 20 samples, after which the effect of increasing sample size is diminished. The plots also show that a larger sample standard deviation s decreases the chance that the sample mean $\bar{x}$ is a good estimate of the population mean, resulting in larger uncertainty.

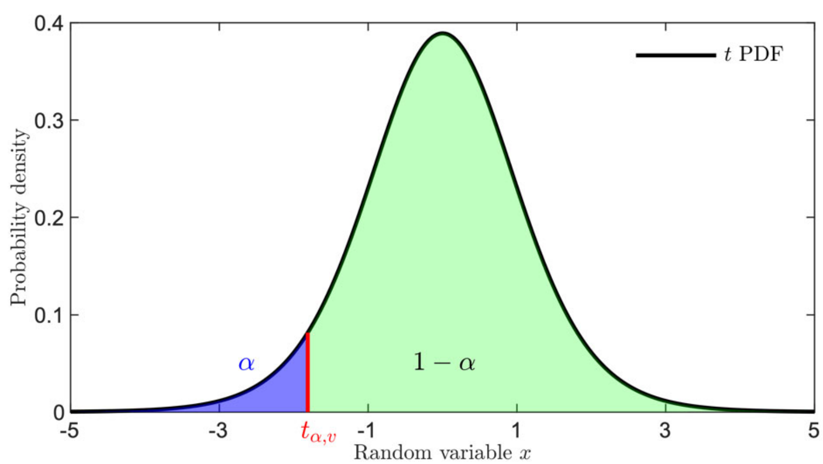

Fig. 3: Significance level α and confidence level $1 - \alpha$ probabilities obtained from the area under the Student's t-distribution up to the critical value of $t_{(\alpha,v)}$.

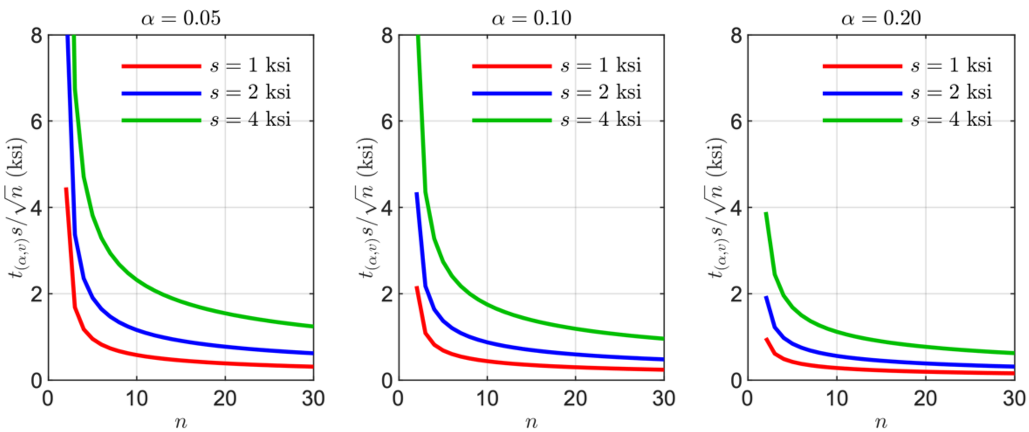

Fig. 4: Uncertainty of the mean for random sampling with different confidence levels $1 - \alpha$ and sample standard deviation s. The new §192.607 requires $\alpha = 0.05$.

3.2. Hypothesis Testing

Hypothesis testing compares randomly sampled measurements with a prior expectation to determine if the differences are statistically significant or if they could be due to chance. This requires a null hypothesis (H_0) which provides a belief about the properties of the population which can be written as,

$$H_0: \mu = \mu_0 \tag{4}$$

where μ_0 is the expected value of the population mean. The null hypothesis is compared to the measured data to test an alternative hypothesis (H_1) that can be used to assert,

$$H_1: \begin{cases} \mu > \mu_0, & \text{mean is greater than } \mu_0 \text{ (upper bound test)} \\ \mu < \mu_0, & \text{mean is less than } \mu_0 \text{ (lower bound test).} \end{cases} \tag{5}$$

This determination is made based on a calculated test statistic that is dependent on the assumed distribution of the measured data. For the Student's t distribution, the t-test is used which is given by,

$$t_s = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{6}$$

The t-statistic t_s is used to calculate a p-value that defines the conditional probability of obtaining a test statistic as extreme or more extreme than the computed value if the null hypothesis is true. As shown in Fig. 5, the p-value is given by the area under the t distribution at the critical value defined by t_s, and is dependent on the degrees of freedom $n - 1$ and whether the alternative hypothesis is testing an upper or lower bound. The p-value can be written as $P_{H_0}(x \leq t_s)$ for a lower bound test, and $P_{H_0}(x \geq t_s)$ for an upper bound test. In practice, p-values can be determined from a look-up table or the cumulative distribution function. A smaller p-value indicates there is less evidence to support the null hypothesis. If the p-value is less than a chosen significance level α, than the null hypothesis is rejected in favor of the alternative hypothesis. The significance level α is equivalent to the parameter used to determine a confidence level for a one-sided confidence interval. This means that the p-value can be used to monitor changes in the confidence level between the sample mean and expected population parameter as more samples are added to the analysis.

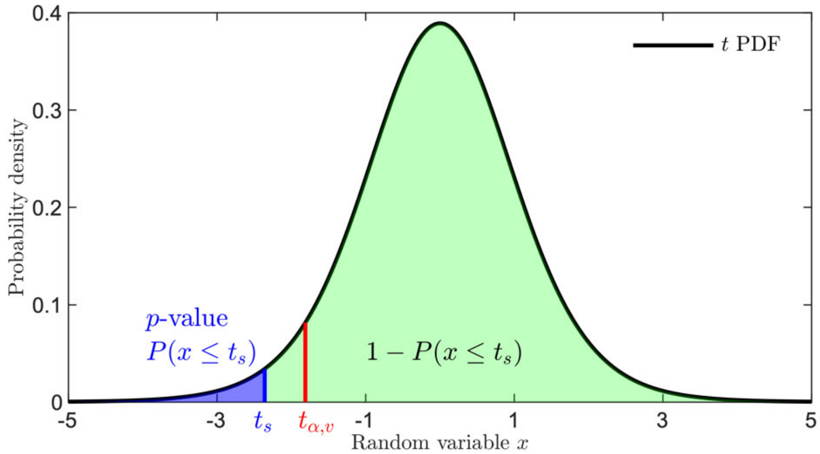

Fig. 5: Lower bound p-value probability determined from area under the Student's t-distribution up to t_s. For this example, $t_s < t_{(\alpha,v)}$, meaning that p-value $< \alpha$, and the null hypothesis would be rejected.

3.3. Measurement Uncertainty

An additional source of uncertainty that is independent of sampling is from the methodology used to measure material strength properties. For nondestructive techniques like the HSD, the measurement uncertainty can be determined from the prediction interval which provides the confidence for a predicted response based on prior performance of the same methodology on a database of samples with similar characteristics. For material verification, a one-sided prediction interval is used to determine an upper or lower bound estimate of the prediction. Similar to confidence intervals, the size of the prediction interval is set by the significance level α_m, which provides a corresponding confidence level of $1 - \alpha_m$. The measurement uncertainty given by the prediction interval is referred to as U_m in this study.

The current performance of the HSD is summarized by the unity plots in Fig. 6, which compare the nondestructive strength measurement of yield strength and UTS from the HSD with conventional laboratory tensile test measurement on the same samples. The current database consists of 167 unique pipe joints including seamless, flash-welded, electric-resistance-welded (ERW), and submerged-arc-welded (SAW) manufacturing processes that cover a broad range of pipe vintages and steel grades. The mechanical properties for this database include 0.5% total elongation under load (EUL) yield strength spanning from 29 to 80 ksi (200 to 550 MPa), UTS measurements ranging from 50 to 104 ksi (340 to 720 MPa), and yield/UTS ratios ranging from 0.59 to 0.96. The unity plots in Fig. 6 also show the one-sided prediction interval for varying α_m, with these tabulated prediction intervals also provided in Table 2. This performance requires the use of a calibrated HSD unit with trained and certified HSD technicians that adhere to all test procedures. Included within these prediction intervals is variation within the laboratory tensile test benchmark which is known to exhibit differences between independent test labs and for measurements on different circumferential quadrants of a pipe joint. The authors recommend the use of a confidence level of 80% ($\alpha_m = 0.20$) based on prior usage of the industry for statistical analysis of ILI results through API 1163 [8] and the reasonable size of the prediction intervals that balances conservatism and practicality. The influence of this parameter when applying a statistical process to measured data is shown in Section 4.2.

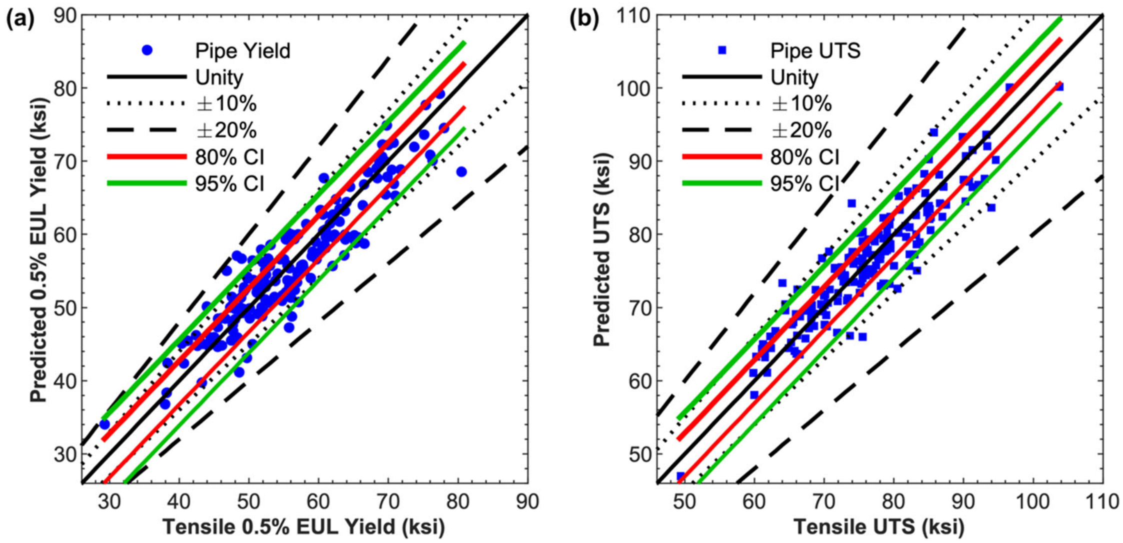

Fig. 6: Unity plots and measurement uncertainty of the HSD for (a) yield strength and (b) UTS.

Table 2: Measurement uncertainty at different confidence levels for the HSD for a database of 167 pipe joints.

Confidence Level, $1-\alpha_m$ (%)	0.5% EUL Yield Strength (ksi)	UTS (ksi)
60	0.9	0.9
65	1.4	1.3
70	1.9	1.8
75	2.4	2.3
80	3.0	2.9
85	3.7	3.5
90	4.6	4.4
95	5.9	5.6

A last consideration for strength measurement uncertainty is whether there are any systematic differences between the nondestructive method and conventional laboratory measurement. A histogram of the percent error between HSD strength predictions and the laboratory tensile test are shown in Fig. 7. These results show that errors are normally distributed and centered around 0%, indicating that there is no systematic over or under estimation of the material strength that should also be included in the analysis of strength measurement uncertainty.

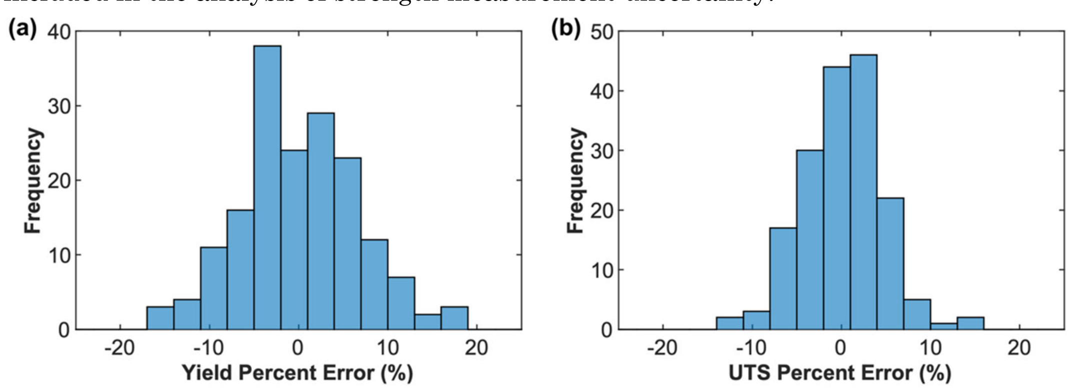

Fig. 7: Histogram of the percent error between the HSD and laboratory tensile tests for measurement of the material (a) yield strength and (b) UTS.

4. Alternative Sampling Plan

In this section, the general concepts for quantifying statistical uncertainties are applied to material verification. It is assumed that there is an expected grade that supports existing calculations of the pipeline MAOP. The statistical analysis will account for sampling and measurement uncertainties to make a comparison of the bounds of the pipeline mean properties with the specified minimum strength requirements (μ_0) of the expected grade based on API 5L [9]. Note that if the prescribed sampling of §192.607(e)(2) is followed, an alternative sampling plan is not required and only the measurement uncertainty has to be considered when comparing the measured properties to the expected grade.

Two equivalent statistical approaches for an alternative sampling plan are to (i) estimate the bounds on the population properties at a specified confidence level and compare to the grade requirements, and (ii) calculate the p-value based on the comparison of the sample measurements with the null hypothesis $H_0 = \mu_0$, and reject the null hypothesis in favor of the alternative hypothesis if the p-value is less than α. The parameters that influence this analysis are the significance level from random sampling α which §192.607(e)(5) requires to be 0.05, the significance level of the strength measurement α_m which controls the size of the prediction interval U_m, and statistics measured from samples of the population that are summarized through the sample mean $\bar{x}$, sample standard deviation s, and number of samples n. Table 3 summarizes the three potential outcomes of this statistical analysis, and the calculated criterion to reach each result. These outcomes are further summarized as follows:

- **Population exceeds minimum requirements ($\mu > \mu_0$)**: The measured strength property distribution is greater than the minimum requirements of the expected grade at the specified confidence level. Further sampling of the pipeline is not required because the material strength properties have been verified.
- **Population more conservative than minimum requirements ($\mu < \mu_0$)**: The measured strength property distribution is more conservative than the minimum requirements of the expected grade at the specified confidence level. An expanded sampling plan is required per §192.607(e)(4), and the material properties will need to be updated with a more conservative estimate.
- **Analysis is inconclusive**: Given the sample mean and uncertainty, the population cannot be statistically confirmed as being above or below the minimum grade requirements at the specified confidence level. This outcome does not indicate that the expected grade is incorrect. If n is low, additional testing of pipe joints within the population will reduce the uncertainty from sampling and may change the outcome to statistically confirming the material grade requirements.

Table 3: Summary of outcomes for material verification. The criteria for confidence intervals and hypothesis testing are equivalent if the same significance level α and measurement uncertainty U_m are considered.

Outcome	Criterion for Confidence Intervals		Criterion for Hypothesis Testing	
$\mu > \mu_0$	$\bar{x} - U_m - t_{(\alpha,v)}\, s/\sqrt{n} > \mu_0$	(7a)	$P_{H_0}\left(x \geq t_s - U_m\sqrt{n}/s\right) < \alpha$	(8a)
$\mu < \mu_0$	$\bar{x} + U_m + t_{(\alpha,v)}\, s/\sqrt{n} < \mu_0$	(7b)	$P_{H_0}\left(x \leq t_s + U_m\sqrt{n}/s\right) < \alpha$	(8b)
Analysis is Inconclusive	$\bar{x} - U_m - t_{(\alpha,v)}\, s/\sqrt{n} < \mu_0$ and $\bar{x} + U_m + t_{(\alpha,v)}\, s/\sqrt{n} > \mu_0$	(7c)	$P_{H_0}\left(x \geq t_s - U_m\sqrt{n}/s\right) > \alpha$ and $P_{H_0}\left(x \leq t_s + U_m\sqrt{n}/s\right) > \alpha$	(8c)

A proposed material verification process that is based on these potential outcomes is illustrated by the flow chart in Fig. 8. The sampling process stops when an outcome has been statistically confirmed as being above or below grade. If the population is below grade, an expanded sampling plan is required which is outside the scope of this analysis but would presumably include many of the same parameters and could incorporate existing concepts outlined in ASME CRTD Vol. 91 to determine a conservative lower bound strength estimate [10]. When the outcome cannot be statistically confirmed as above or below grade, §192.607 would require continued testing until the prescribed sampling requirements from §192.607(e) of 1 excavation per mile rounded up to the nearest whole number with a maximum of 150 excavations are met. This acknowledges there are situations where the actual material properties are close to the minimum strength requirements and no amount of additional testing will change the outcome because of the measurement uncertainty and diminishing influence of n on the sampling uncertainty (see Fig. 4). More practical bounds could be determined based on a criterion applied to the diminishing changes in sampling uncertainty with increasing sample size, or a more sophisticated analysis based on a database of pipeline materials with similar characteristics. Another potential solution is to consider additional techniques to substantiate partial records, such as the comparison of chemistry data from field measurements with documented material test records to approach TVC requirements.

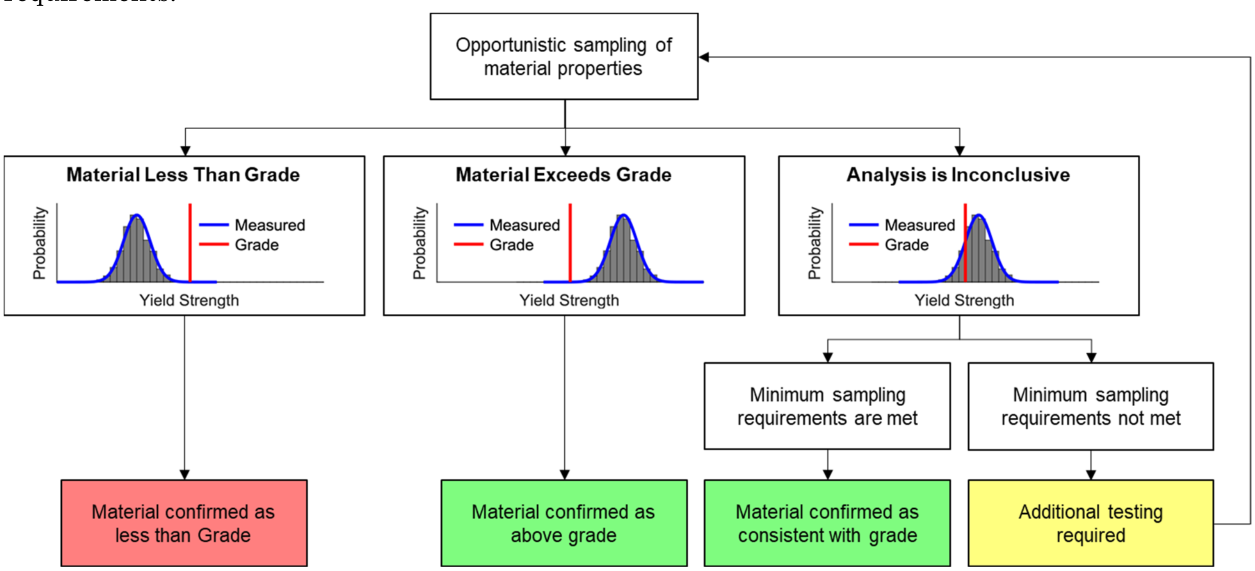

Fig. 8: Process flow chart for material verification of a pipeline population.

4.1. Application to a Vintage Pipeline

As a demonstration of the analysis process, consider 16 samples that were obtained from a 12 inch diameter ERW line segment that was installed in the early 1950s. For simplicity, the analysis considers only the yield strength of the material, which has a SMYS requirement of $\mu_0 = 42$ ksi for the expected grade of X42. Table 4 shows the individual yield strength measurements x that were opportunistically collected along the pipeline, and the resulting analysis considering $\alpha = 0.05$ for sampling uncertainties, and $\alpha_m = 0.20$ for measurement uncertainties. These calculations are cumulative, meaning that each row includes all prior measurements to simulate the effect of changing uncertainty during the sampling program. Table 4 utilizes measurement uncertainties for the HSD provided in Table 2, lower and upper bound uncertainties from Eqns. (7a) and (7b), and probabilities evaluated from Eqns. (8a) and (8b).

Table 4: Example calculations for a line segment with an expected grade of X42. The shaded cells represent samples collected prior to confirmation of the material grade.

Sample No.	x (ksi)	$\bar{x}$ (ksi)	s (ksi)	U_m (ksi)	$t_{(\alpha,v)}\, s/\sqrt{n}$ (ksi)	Eq. 7a (ksi)	Eq. 7b (ksi)	Eq. 8a (%)	Eq. 8b (%)
1	50.9	---	---	---	---	---	---	---	---
2	46.3	48.6	3.3	3.0	14.5	31.0	66.1	18.2	81.8
3	52.8	50.0	3.3	3.0	5.6	41.4	58.6	6.1	93.9
4	53.8	50.9	3.3	3.0	3.9	44.0	57.8	1.9	98.1
5	53.4	51.4	3.1	3.0	2.9	45.5	57.4	0.5	99.5
6	51.6	51.4	2.8	3.0	2.3	46.2	56.7	0.1	99.9
7	52.2	51.6	2.5	3.0	1.9	46.7	56.4	0.0	100.0
8	53.7	51.8	2.5	3.0	1.7	47.2	56.5	0.0	100.0
9	54.5	52.1	2.5	3.0	1.5	47.6	56.7	0.0	100.0
10	50.3	51.9	2.4	3.0	1.4	47.6	56.3	0.0	100.0
11	54.9	52.2	2.4	3.0	1.3	47.9	56.5	0.0	100.0
12	51.3	52.1	2.3	3.0	1.2	47.9	56.3	0.0	100.0
13	49.2	51.9	2.4	3.0	1.2	47.7	56.1	0.0	100.0
14	49.3	51.7	2.4	3.0	1.1	47.6	55.9	0.0	100.0
15	52.7	51.8	2.3	3.0	1.1	47.7	55.8	0.0	100.0
16	50.0	51.7	2.3	3.0	1.0	47.7	55.7	0.0	100.0

The analysis results as a function of the number of samples measured are plotted in Fig. 9. The uncertainties from sampling and the strength measurement are shown in Fig. 9(a), which are used to determine the bounds of the mean strength of the population from confidence intervals that are plotted in Fig. 9(b). These results show that the lower bound mean strength exceeds the expected SMYS and positively verifies the material after 4 samples. When there was less than 4 samples, the analysis would be inconclusive due to high sampling uncertainty. The same outcome, but using the hypothesis testing approach, is demonstrated in Fig. 9(c) where the p-value defining the probability from a lower bound and upper bound test is shown as a function of the number of samples. The probability is below the specified significance level of $\alpha = 0.05$ after 4 samples, which means that the population yield strength exceeds the expected SMYS at the specified confidence level.

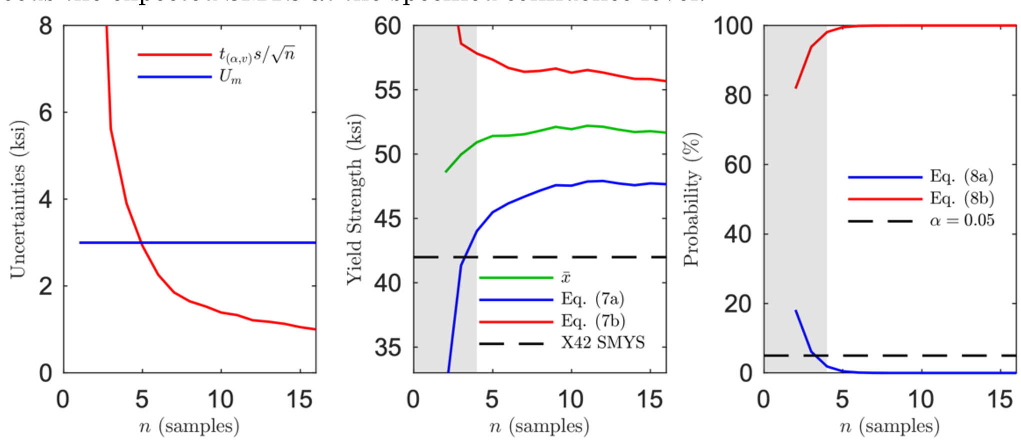

Fig. 9: Statistical analysis to verify an expected SMYS of 42 ksi considering $\alpha_m = 0.20$ and $\alpha = 0.05$. (a) Uncertainty as a function of number of samples tested. (b) Sample mean $\bar{x}$ and confidence intervals of the population mean as a function of samples tested. (c) Null hypothesis p-value for hypothesis testing. The shaded area represents samples prior to statistical verification of the expected SMYS.

4.2. Application to Many Pipelines

In order to demonstrate the implications of this analysis, it is useful to consider the application of the methodology on pipelines of varying grade, vintage, and manufacturing process. Multiple histograms for different API 5L grades are provided in Fig. 10 for HSD yield strength measurements on over 200 pipe joints where an expected grade was shared with MMT. These results show that the measured yield strength will generally exceed the SMYS requirements of the expected grade because pipeline steel is manufactured to ensure it exceeds the minimum value. Of the 3 samples that were measured below API 5L requirements, the 2 grade X60 measurements were confirmed through tensile tests and the grade B pipe had been expanded due to insufficient strength properties. For the subsequent statistical analysis, the measurements were grouped into line segments when multiple samples were obtained from the same population as defined by §192.607. The sample size for a unique population varied from 1 to 30 samples, with a median size of around 3 samples.

.

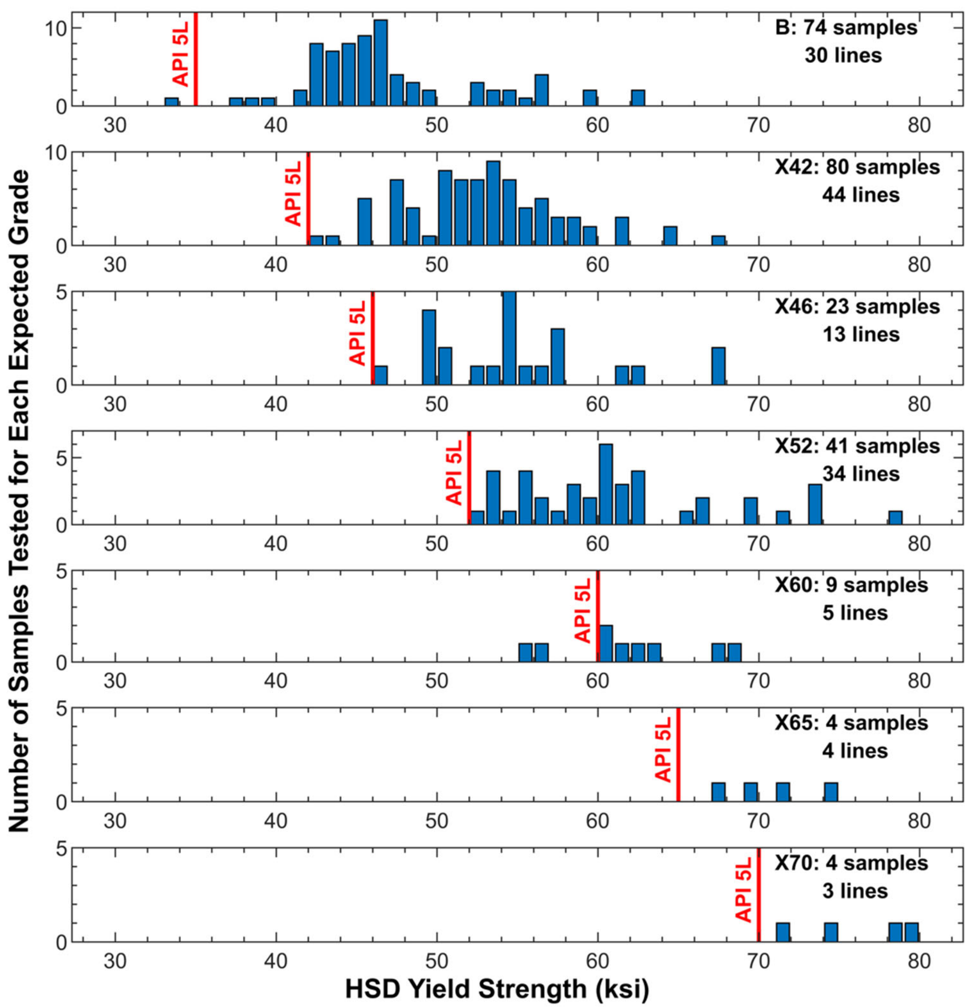

Fig. 10: Comparison between expected grade and actual NDE yield values from HSD testing on samples obtained from various pipeline segments. The number of unique line segments based on populations of pipe joints sharing the same nominal characteristics are specified for each grade.

An estimate of the proportion of potential analysis outcomes is determined by accounting for the measurement uncertainty U_m that is set by α_m. This will identify a subset of populations where the sample mean is close to the grade requirements and $|\bar{x} - \mu_0| \leq U_m$, meaning that the analysis outcome will always be inconclusive regardless of how many samples are collected. The remaining data set can theoretically be statistically confirmed as above or below grade if enough samples are collected to reduce the sampling uncertainty to meet the criterion given in Table 3. Applying this process to the data set results in the distribution of outcomes shown in Fig. 11 for $\alpha_m = 0.20$ and $\alpha_m = 0.05$. These plots indicate that 90% of pipe samples can be statistically confirmed as either above or below grade with $\alpha_m = 0.20$, but those proportions diminish to 73% with a lower significance level of $\alpha_m = 0.05$. Figure 11 also shows that the percentage of pipes that can be statistically confirmed is reduced with increasing grade. For the current data set, higher grade materials have less samples for analysis and therefore the percentage of different outcomes is less representative than the lower grades with a larger number of unique line segments. The use of a nondestructive methodology that is less accurate than the HSD will have the same effect as decreasing α_m, meaning that less pipeline populations can reach statistical confirmation at the specified confidence level.

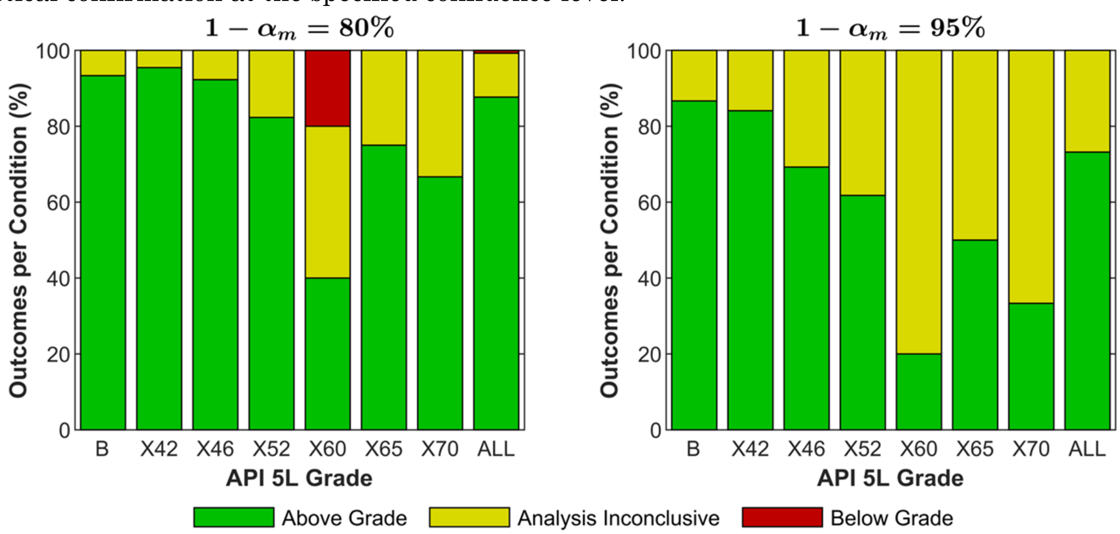

Fig. 11: Proportion of material verification analysis outcomes for different API 5L grades and α_m. The "All" condition represents the entire data set without accounting for grade.

The data set can also be used to estimate the number of samples that will be required to reach a statistical confirmation of the measured properties for the different line segments. For this analysis, only the data where $|\bar{x} - \mu_0| > U_m$ is considered, meaning that a statistical determination can theoretically be determined if enough sampling is performed. For line segments containing only 1 sample, a standard deviation of $s = 3$ ksi is assumed when determining the sampling uncertainty given by $t_{(\alpha,\nu)}\, s/\sqrt{n}$. Prior testing has shown that typical values of s range from 2 to 4 ksi. The analysis results are provided in Fig. 12, which shows the overall distribution of estimated samples in Fig. 12(a), and a summary of the 25%, 50%, and 75% quartile for each grade and overall data set in Fig. 12(b). The analysis is performed for a measurement uncertainty $\alpha_m = 0.20$ and $\alpha_m = 0.05$ to show the influence of measurement confidence level and accuracy. There are some data points where $\bar{x} \pm U_m$ is close to the grade requirements which would require hundreds to thousands of samples to reach statistical confirmation of grade. For practical considerations, this results in an inconclusive analysis outcome, and for this study a maximum of 150 samples was assumed based on the maximum prescribed sampling guidelines of §192.607. However, Fig. 12 shows that at least 75% of pipelines can reach statistical confirmation with 2-4 samples for $\alpha_m = 0.20$, and 2-5 samples for $\alpha_m = 0.05$. The number of samples required is somewhat dependent on the material grade, especially for the highest quartile of samples, but these sample estimates are less representative for larger grades because of the limited data that is currently available. The number of samples in Fig. 12 also provides a lower bound estimate of the minimum length of a pipeline to achieve a reduction in the required number of digs to confirm material properties compared to the prescribed sampling requirement of 1 dig per mile. The results suggest that segments greater than 2 miles should be considered.

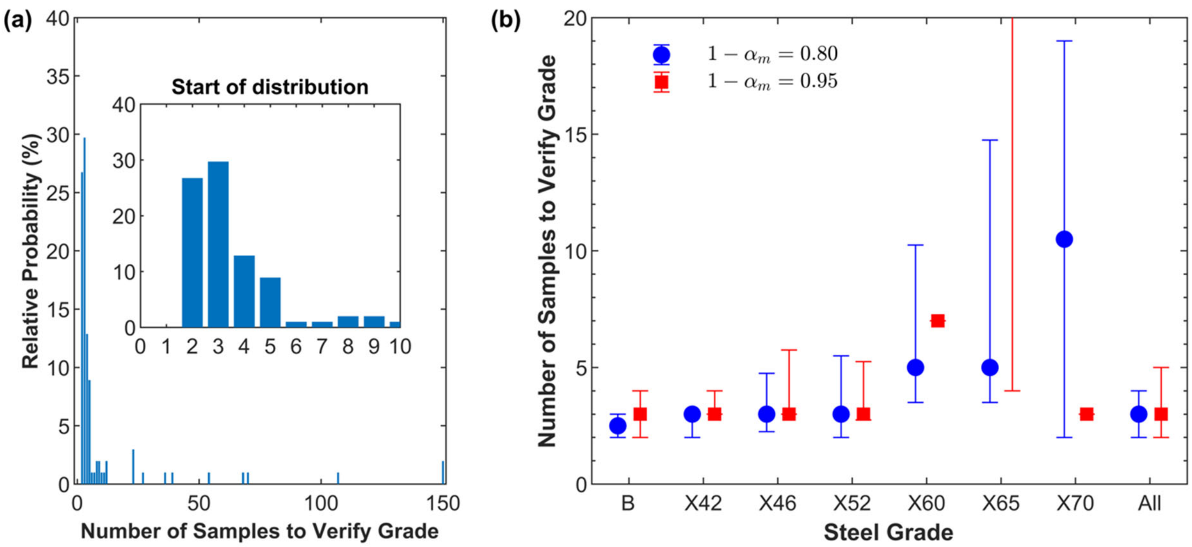

Fig. 12: Expected number of samples to statistically confirm a pipeline as above or below grade. (a) Distribution for all samples showing a large peak at low sample numbers and a long tail. (b) Distribution summary for different material grades and the overall data set. The lower error bar represents the bottom quartile, the data point is the median, and the upper error bar is the top quartile. If no error bar is shown, then the quartile is the same value as the median. Data points that were omitted had a required number of samples greater than 40. The analysis considers $s = 3$ ksi, $\alpha = 0.05$, and α_m of 0.20 and 0.05.

To summarize the analysis findings, the distribution of potential analysis outcomes is largely dependent on the uncertainty of the strength measurement, which requires a reasonable choice of α_m and an accurate nondestructive methodology. For outcomes where $|\bar{x} - \mu_0| > U_m$, the majority of pipelines can be statistically confirmed with less than 5 samples collected because of the large difference between the as manufactured and minimum required strength properties of pipeline steel. However, for the remaining data sets where the measured properties are close to the grade requirements and the analysis is inconclusive, the number of samples collected would be based on the sampling frequency of §192.607(e)(2) unless a more practical bound could be justified. This illustrates the cost-effectiveness of a more accurate nondestructive method, because the reduced uncertainty results in a higher proportion of outcomes that can be statistically confirmed with a smaller number of digs than what would otherwise be required by prescribed sampling.

5. Discussion

The statistical approaches detailed above include the basic components required for material verification of an expected grade. The framework can be adjusted to provide additional conservatism or practicality through the inclusion of additional statistical concepts. This section discusses some of these potential considerations and provides additional analysis that shows how these changes would affect the number of samples required to reach statistical confirmation with the same set of measurements that was considered in Section 4.2.

Expanded Uncertainty and Coverage Factors

Individual measurements on a sample pipe joint exhibit uncertainty due to material variation and other factors that are not considered in the proposed analysis. This additional uncertainty could be considered through a coverage factor applied to the sample standard error by replacing the $s/\sqrt{n}$ term with $ks/\sqrt{n}$, where k is a coverage factor that is dependent on the significance level and sample size but will typically vary between 2 and 3. The consideration of coverage factors would increase the sampling uncertainty, leading to a need for additional testing to reach statistical confirmation of the measured properties as above or below grade. This is reflected in Fig. 13, which shows the median number of samples that need to be tested based on the prior dataset of HSD measurements, but now considering varying coverage factors. A coverage factor $k = 1$ is equivalent to the prior analysis shown in Fig. 12.

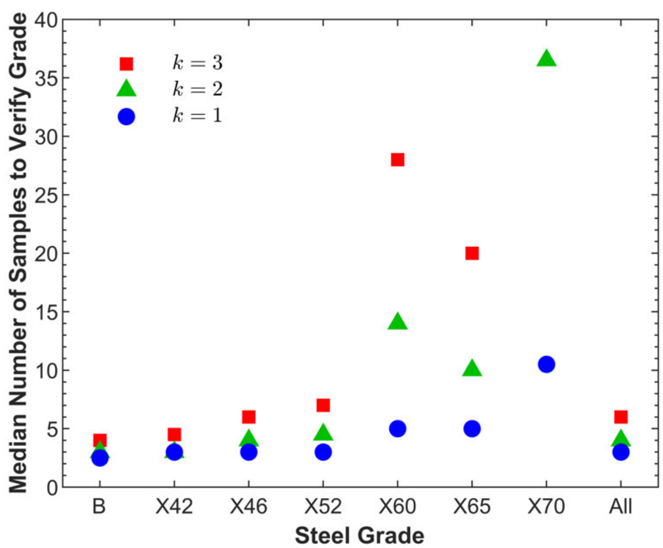

Fig. 13: Expected median number of samples to statistically confirm the material as above or below grade. This analysis considers $s = 3$ ksi, $\alpha = 0.05$, $\alpha_m = 0.20$, and varying coverage factors k. The data point missing for grade X70 had a median value of 76 samples that was excluded from the plot.

Tolerance Intervals and the Population Standard Deviation

The proposed analysis compares the upper and lower bounds of the population mean with the minimum strength requirements of the expected grade. Every population described by a normal distribution has a variability that is characterized by its standard deviation. This could be considered by using a tolerance interval instead of a confidence interval, which uses an additional analysis input parameter α^* that defines the probability of exceeding a given strength for a pipe joint sampled from the population. In practice, α^* sets the critical value z_{α^*} of a normal distribution, and the additional uncertainty is given by $z_{\alpha^*}\sigma$, where σ is the standard deviation of the population. With this approach, σ could be determined through an upper bound confidence interval from the asymmetric Chi-square distribution. A similar approach has been applied through ASME CRTD Vol. 91 to determine a lower bound hardness estimate for field assessments of pipelines [10].

A tolerance interval approach is illustrated in Fig. 14 which uses the measured data from the case study in Section 4.1. Considering $\alpha = 0.05$, $\alpha_m = 0.20$, and $\alpha^* = 0.10$, after 16 samples the lower bound population mean given by Eq. 7a is $\mu_{lb} = 47.7$ ksi and the upper bound population standard deviation is $\sigma_{ub} = 3.3$ ksi. The critical value of $z_{\alpha^*} = 1.28$ is used to further reduce the lower bound population mean to the minimum strength value given by $x_{min} = \mu_{lb} - z_{\alpha^*}\sigma_{ub} = 43.5$ ksi. Applying this approach to the cumulative data in Section 4.1 would result in 9 samples being required to verify that the measured properties exceed the expected SMYS of 42 ksi for the given confidence level and probability of exceedance. This can be compared to the 4 samples that were required when only the mean of the population was considered.

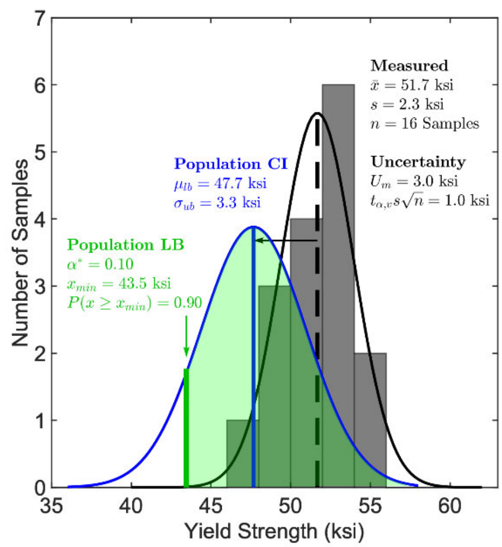

Fig. 14: Analysis of prior case study data using a tolerance interval. This analysis considers $\boldsymbol{\alpha = 0.05}$, $\boldsymbol{\alpha_m = 0.20}$, and $\boldsymbol{\alpha^* = 0.10}$.

As shown in the prior example, including the population standard deviation through a tolerance interval would greatly increase the sampling uncertainty, leading to an increase in the number of samples required to statistically confirm the measured strength as above or below the expected grade. A comparison of the median number of samples required to verify the material grade for different values of α^* with the upper bound standard deviation σ determined using a significance level of $\alpha = 0.05$ is provided in Fig. 15. Note that $\alpha^* = 0.50$ considers only the population mean because $z_{\alpha^*} = 0$, so the median values are the same as the prior analysis summarized in Fig. 12.

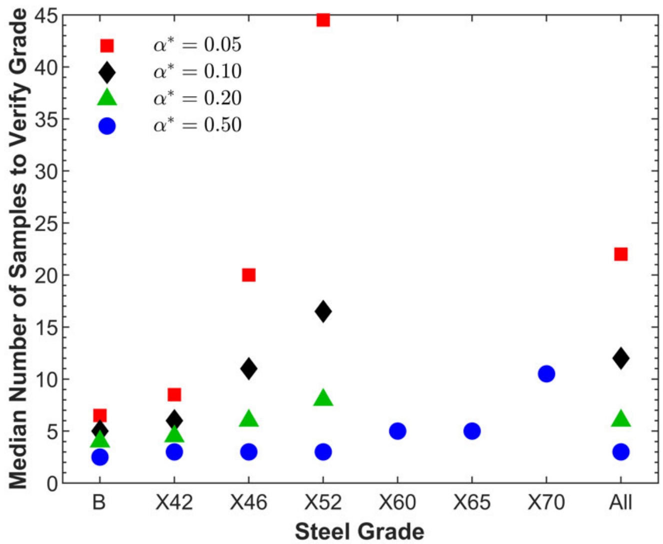

Fig. 15: Expected median number of samples to statistically confirm the material as above or below grade. This analysis considers $s = 3$ ksi, $\alpha = 0.05$, $\alpha_m = 0.20$, and varying α^*. The data points missing had median values between 75 and 150 samples and were excluded from the plot.

Type II Errors in Hypothesis Testing

The significance level α in hypothesis testing defines the probability of rejecting the null hypothesis when the null hypothesis is actually true, or a Type I error. Hypothesis testing can also be used to determine the probability of a Type II error β, that defines the chance of accepting the null hypothesis when it is actually false. The differences between these errors are illustrated in Fig. 16, where the distribution of the t-statistic given by Eq. (6) gives α under the null hypothesis, and β under the alternative hypothesis.

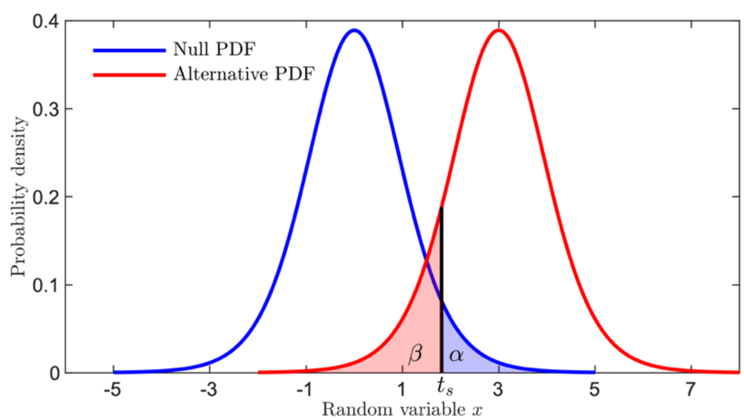

Fig. 16: Illustration of the probability of Type I (α) and Type II (β) errors based on the position of t_s under the null and alternative hypothesis, respectively.

If the analysis was modified to consider a maximum value of β to limit the risk of Type II errors (e.g. $\beta = 0.05$), it would require additional sampling or a larger difference between the sampled distribution and the expected material properties to meet these additional requirements. This was considered in the analysis shown in Fig. 17, which provides the median number of samples that need to be tested based on the prior dataset of HSD measurements on different grades, but now considering varying β criterion. Considering no β criterion provides the same median values as the analysis in Fig. 12.

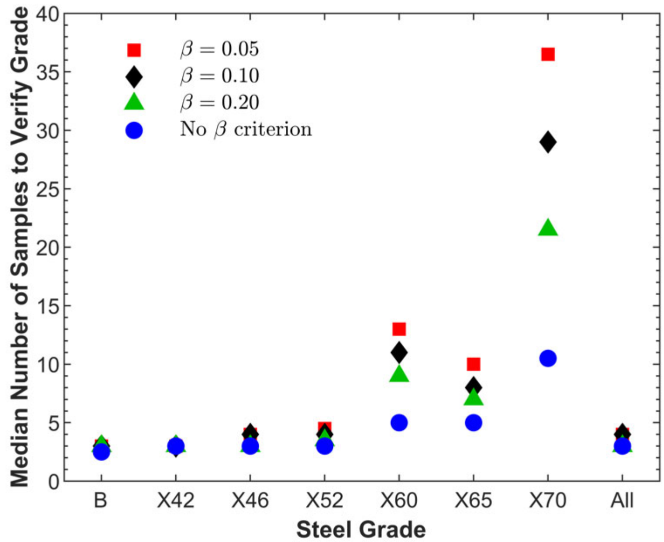

Fig. 17: Expected median number of samples to statistically confirm the material as above or below grade. This analysis considers $s = 3$ ksi, $\alpha = 0.05$, $\alpha_m = 0.20$, and varying β.

Segmenting a Validation Database by Grade and Vintage

The measurement uncertainty of the HSD provided in Section 3.3 is applicable to a database that considers pipes ranging from Grade A to X70, manufacturing dates ranging from 1910 to 2015, and all pipe manufacturing processes. Multiple uncertainties could be determined by segmenting the database into different grades, vintages, and manufacturing processes. This is appropriate if further analysis of the database indicates statistically significant differences when controlling for these potential variables, and if there is evidence that the sample of pipes within that segment is representative of unknown pipe samples that are not included in the database. This analysis is ongoing for the HSD and the current database of 167 unique pipe joints. Changes to the measurement uncertainty will impact the proportions of pipeline segments where no statistical confirmation of grade can be achieved.

Outliers

This analysis has assumed that all of the samples belong to the same distribution. There may exist samples that are outliers or have characteristics that make them unique to the other samples in the population. Outliers can be identified within a dataset through several analysis methods, including normal distribution z-scores, deviations from the median, probabilistic modeling, and regression analysis. A pipe sample that is determined to be an outlier and is more conservative than the rest of the population may require additional assessment or remedial actions. The ability to differentiate different populations within a pipeline network and identify potential outliers is best accomplished through ILI methodologies.

Engineering Use of NDE Data

Testing on over 200 pipe joints has shown that the vast majority of measured properties exceeds the SMYS (see Fig. 10). As shown in Fig. 11, using an 80% confidence level for measurement uncertainty with an accurate methodology allows for the statistical confirmation of assets that have an actual yield strength below the minimum requirement for the expected grade as validated by tensile testing of cut-outs from the same samples. Therefore, consistently applying an 80% confidence level for measurement uncertainty better allows for achieving the goal of reducing the risk of failure for a network of transmission assets, as opposed to a larger confidence level that diminishes the ability to make statistical conclusions. The 80% confidence level is also utilized for analysis of In-Line Inspection (ILI) data of the same pipeline network per API 1163 [8]. Ultimately, considering flaw sizing from ILI with accurate material properties for condition assessment, fitness for service, and MAOP re-confirmation will allow for the prioritization of repairs and the reduction of risk for systems that are impacted by the combination of these geometrical and material properties.

6. Conclusions

This work describes an analysis approach that provides a statistical basis to perform material verification based on the requirements of §192.607. An alternative sampling plan must account for uncertainties from randomly sampling locations along a pipeline segment and uncertainty in the measurements of the material strength properties from the method used. If the prescribed sampling requirements of §192.607(e)(2) are followed, only the measurement uncertainty per §192.607(d)(2) must be considered when comparing the measured properties to the expected grade. For an alternative sampling plan, the number of digs to reach statistical significance cannot be pre-determined. Thus, it is advantageous to implement a more accurate nondestructive method to lower measurement uncertainty per §192.607(e)(2) and significantly reduce the number of digs required to reach a conclusive result at the specified confidence levels. Additional testing up to a maximum number of required samples or the consideration of additional material information is suggested when the analysis cannot reach a statistical confirmation. This acknowledges that there are instances where the actual material properties are close to the grade requirements and statistical confirmation at the desired confidence level cannot be achieved.

The essential analysis inputs are the significance levels α and α_m which determine the confidence level for random sampling and strength measurements, respectively. The significance for random sampling $\alpha = 0.05$ has been set by the requirements of an alternative sampling plan in §192.607, but it is recommended to use a significance level of $\alpha_m = 0.20$ for measurement of strength uncertainty. This is justified based on its use in existing industry analysis for ILI and the reasonable size of the one-sided prediction intervals that still allows for statistical determinations. Decreasing α_m or using a less accurate methodology will greatly increase the proportion of outcomes where no statistical confirmation can be achieved no matter how much sampling is performed.

Every analysis needs to balance conservatism and practicality to provide a methodology that is safe but not overly burdensome. The discussion section described many additional considerations that can have an influence on the outcome of the statistical analysis, and the case studies demonstrate the value of applying a proposed process on a large database of results to understand the implications of any proposed methodology.

References

[1] NTSB, "Pacific Gas and Electric Company Natural Gas Transmission Pipeline Rupture and Fire, San Bruno, California, September 9, 2010," *Pipeline Accident Report NTSB/PAR-11/01*, Aug 2011.

[2] DOT, "Pipeline Safety: verification of records," Docket No. PHMSA-2012-0068, May 2012.

[3] DOT, "Pipeline Safety: Establishing maximum allowable operating pressure or maximum operating pressure using record evidence, and integrity management risk identification, assessment, prevention, and mitigation," Docket No. PHMSA-2010-0381, Jan 2011.

[4] DOT, "Pipeline Safety: Safety of gas transmission and gathering lines, Notice of proposed rulemaking," Docket No PHMSA-2011-0023, Apr 2016.

[5] DOT, "Safety of gas transmission and gathering pipelines – Gas Pipeline Advisory Committee Meeting," Docket: PHMSA-2011-0023, Dec 2017.

[6] DOT, "Pipeline Safety: Safety of Gas Transmission Pipelines: MAOP Reconfirmation, Expansion of Assessment Requirements, and Other Related Amendments," *Federal Register Vol. 84, No. 190*, Oct 2019.

[7] Amend, Riccardella and Dinovitzer, "Material Verification – Validation of In Situ Methods for Material Property Determination," Catalog No. PR-335-173816, May 2018.

[8] American Petroleum Institute, "API STD 1163 – In-line Inspection Systems Qualification Standard," *Second Edition*, 2013.

[9] American Petroleum Institute, "API Spec 5L – Specification for Line Pipe," *Forty-sixth Edition*, 2018.

[10] Clark and Amend, "Applications guide for determining the yield strength of in-service pipe by hardness evaluation," ASME CRTD Vol. 91, 2009.