EPRI | ELECTRIC POWER RESEARCH INSTITUTE

Nondestructive Evaluation: Probabilistic Analysis of Performance Demonstration Ultrasonic Flaw Detection and Through-Wall Sizing Results for Reactor Pressure Vessel Inspections (Revision 2)

2018 TECHNICAL REPORT

Nondestructive Evaluation: Probabilistic Analysis of Performance Demonstration Ultrasonic Flaw Detection and Through-Wall Sizing Results for Reactor Pressure Vessel Inspections (Revision 2)

3002013319

Final Report, July 2018

EPRI Project Manager
T. Seuaciuc-Osorio

All or a portion of the requirements of the EPRI Nuclear Quality Assurance Program apply to this product.

YES (NO)

ELECTRIC POWER RESEARCH INSTITUTE
3420 Hillview Avenue, Palo Alto, California 94304-1338 ▪ PO Box 10412, Palo Alto, California 94303-0813 ▪ USA
800.313.3774 ▪ 650.855.2121 ▪ askepri@epri.com ▪ www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

NOTE

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

ACKNOWLEDGMENTS

The Electric Power Research Institute (EPRI) prepared this report.

Principal Investigators
T. Seuaciuc-Osorio
M. Dennis
C. Latiolais
F. Yu

This report describes research sponsored by EPRI.

EPRI acknowledges the support of Larry Becker for his assistance in reviewing the statistical data. EPRI also thanks Patrick Heasler of Pacific Northwest National Laboratory for the use of his R script in determining the model parameters necessary for the data analysis.

This publication is a corporate document that should be cited in the literature in the following manner:

Nondestructive Evaluation: Probabilistic Analysis of Performance Demonstration Ultrasonic Flaw Detection and Through-Wall Sizing Results for Reactor Pressure Vessel Inspections (Revision 2). EPRI, Palo Alto, CA: 2018. 3002013319.

ABSTRACT

This report presents a statistical analysis of detection and through-wall sizing performance of inside surface examinations of reactor pressure vessels under the American Society of Mechanical Engineers Boiler & Pressure Vessel Code, Section XI, Appendix VIII, Supplements 4 and 6, based on data from the Performance Demonstration Initiative program at the Electric Power Research Institute. Models to describe the probability of detection and the sizing performance are described, as well as other performance parameters. Detection analysis shows a very good performance in both cases and illustrates the screening ensured by the qualification program. Supplement 4, however, shows better detection capabilities for inside surface examinations; this is expected, given that it addresses the inner volume of the component, while Supplement 6 applies to the outer examination volume. Analysis of the sizing performance shows the same previously observed trend to oversize small flaws and undersize large flaws; this bias is much more pronounced in Supplement 4, which leads to low probability of acceptable sizing (PAS)—the probability that a flaw will be detected and sized to be within a given tolerance of its true size—for flaw sizes away from the transition size at which oversizing (small flaws) gives way to undersizing (large flaws). This low PAS for Supplement 4 measurements and the trend to undersize large flaws, however, do not cause safety concerns because the probability of rejection (PR)—the probability that a flaw will be detected and measured to be equal to or larger than a given acceptance criterion—shows a better than 99% chance of correctly rejecting flaws in the region where undersizing takes place. On the other hand, the trend to oversize small flaws might be of financial concern, because it leads to a high PR for flaws below the acceptance criterion.

Keywords

Nondestructive evaluation reliability
Probabilistic analysis
Probability of detection
Reactor pressure vessel
Ultrasonic examination performance

Deliverable Number: 3002013319

Product Type: Technical Report

Product Title: Nondestructive Evaluation: Probabilistic Analysis of Performance Demonstration Ultrasonic Flaw Detection and Through-Wall Sizing Results for Reactor Pressure Vessel Inspections (Revision 2)

PRIMARY AUDIENCE: Nuclear nondestructive evaluation industry personnel

SECONDARY AUDIENCE: Personnel involved in assessing the reliability of performance demonstration activities

KEY RESEARCH QUESTION

Analyze the detection and through-wall sizing performance of inside surface examinations of reactor pressure vessels under the American Society of Mechanical Engineers (ASME) Boiler & Pressure Vessel Code (BPVC), Section XI, Appendix VIII, Supplements 4 and 6, in order to provide insight into the reliability of these inspections.

RESEARCH OVERVIEW

Relevant data for inside surface examinations under ASME BPVC, Section XI, Appendix VIII, Supplements 4 and 6, were retrieved from the PDI program at EPRI. A traditional logistic statistical model was adopted to describe probability of detection, and classical regression analysis was applied to model the sizing performance. Based on these two models, new probabilistic parameters that combined both analyses were developed to quantify the inspection performance.

KEY FINDINGS

- Detection capabilities under both supplements is better than 80%.
- The performance screening achieved by the qualification program is seen by comparing the detection performance when only data from candidates who have passed the exam are included to the performance when the analysis also includes candidates who have failed the exam.
- There is a trend to oversize small flaws and undersize large flaws. Although present in both supplements, this bias is more pronounced for Supplement 4 examinations.
- Using 0.15in (3.8mm) as an acceptance criteria for Supplement 4 examinations, the probability that a flaw of this size or larger would be detected and sized at or above 0.15in (3.8mm) is better than 80% despite the trend to undersize small flaws.

WHY THIS MATTERS

The results of this research provide insight into the reliability of inside surface examinations of reactor pressure vessels under ASME BPVC, Section XI, Appendix VIII, Supplements 4 and 6. Its value is in demonstrating the screening effect of the qualification test and the reliability of the examinations to minimize safety concerns.

HOW TO APPLY RESULTS

The results presented in this report can be used to support issues related to the reliability of reactor pressure vessels inspections. Additionally, the methodology introduced in this report can be applied to similar analyses for different inspections in order to provide insight into their reliability.

LEARNING AND ENGAGEMENT OPPORTUNITIES

- Readers interested in applying the methodology described in this report for other applications can contact EPRI for further support.

EPRI CONTACTS: Thiago Seuaciuc-Osorio, Technical Leader, tsosorio@epri.com

PROGRAMS: Nuclear Power, P41 and Nondestructive Evaluation, P41.04.01

IMPLEMENTATION CATEGORY: Category 2

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com

RECORD OF REVISION

Revision	Description of Change
0	Original Issue (EPRI PID 1025787)
1	This revision addresses comments raised by a review of the original issue by the U.S. Nuclear Regulatory Commission (NRC). The main changes are summarized as follows: • All graphs dependent on the current POD or sizing model results were truncated at 0.1in flaw depth to avoid misinterpretation of the data. • Further explanation and clarification on topics identified by the review have been provided. • General edits of the body of the text to align the discussion with the modifications.
2	Corrected Equation 3-1 on page 3-1.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1
INTRODUCTION

This report provides an analysis of detection and through-wall sizing performance based on data from the Performance Demonstration Initiative (PDI) program at the Electric Power Research Institute (EPRI) for inside surface examinations of the reactor pressure vessel under the American Society of Mechanical Engineers Boiler & Pressure Vessel Code, Section XI, Appendix VIII, Supplements 4 and 6.

To some extent, this report follows the same framework of the 2004 EPRI report *Reactor Pressure Vessel Inspection Reliability Based on Performance Demonstrations* (1007984) [1]. However, not all of the analysis presented in the 2004 report is performed in this report; the most notable aspects included in the 2004 report that are not present in this report are the inclusion of outside surface examinations and the analysis of post-inspection distributions. Outside surface examinations are not included because pressurized water reactor vessels, which are the main focus of this report, are primarily examined from the inside surface.

Additionally, there are some differences in the methodology applied in each report and in the parameters analyzed. The sizing model, for instance, is different (although both lead to the same conclusions), and this report introduces the concepts of probability of acceptable sizing (PAS) and probability of rejection (PR) instead of the parameters probability of correct sizing and probability of correct rejection introduced in EPRI report 1007984.

To the extent that both reports have similar content, this report serves as a supplement to *Reactor Pressure Vessel Inspection Reliability Based on Performance Demonstrations*.

Explanation of Analyzed Data and Data Summary

As mentioned, this report analyzes data from the PDI program relating to inside surface examinations of reactor pressure vessels under Supplements 4 and 6. The data for each supplement will not be combined; they will be treated separately and independently. Because the inspection is from the inside surface, all procedures employ automated ultrasonic techniques. The data presented here are current as of July 13, 2011.

For detection performance analysis, the data for each supplement are grouped in two categories. *Passed* includes only candidates who have passed the qualification exam, and *passed plus failed* includes candidates who have passed the qualification exam and those who have failed the exam while missing one single flaw.

This analysis method is similar to the one adopted in EPRI report 1007984; however, the analysis performed therein includes data from candidates who have failed while missing two flaws. The total number of relevant detection measurements in the database and in each category is listed in Table 1-1 for each supplement. Based on the number of passed candidates, the amount

of data has increased approximately 65% for Supplement 4 and 58% for Supplement 6 from January 2004 (the time of the analysis in EPRI report 1007984) to July 2011 (according to EPRI report 1007984, there were 717 data points for passed candidates for inside surface inspections under Supplement 4 and 658 for Supplement 6).

In the sizing performance analysis, only data from candidates who have passed the through-wall extent sizing qualification test are included. The acceptance criterion for depth sizing is based on the candidate's root mean square error (RMSE) value: for Supplement 4, the acceptance criterion is 0.15 in. (3.8 mm) RMSE; for Supplement 6, it is 0.25 in. (6.4 mm) RMSE. The total number of relevant sizing measurements for each Supplement is listed in Table 1-2. Although there has been a considerable increase in the number of through-wall sizing measurements as well, the 2004 report does not provide a count for measurements from the inside surface only; the combined total (inside and outside surface examinations) in EPRI report 1007984 for through-wall sizing measurements is 961 for Supplement 4 and 1008 for Supplement 6. The current data show higher counts even when only inside surface examinations are considered.

Table 1-1
Number of detection measurements

Supplement	P	P+F	All
4	1186	1405	1568
6	1043	1448	1767

P = Passed; P+F = Passed plus Failed

Table 1-2
Number of through-wall sizing measurements

Supplement	Passed	All
4	1079	1542
6	786	1749

Data Security

Given the nature of the PDI program, all data from its database are secured and confidential. Any compromise of data security would jeopardize the program and cause great financial loss to the industry.

To preserve the confidentiality of the data, data points are not shown and all graphs have been extrapolated beyond the actual range of available data points. The ranges in flaw size displayed in the current report have been chosen so that the relevant curve characteristics are observable while maintaining data security, and they do not represent the actual range present in the database. As a consequence, interpretation of data near the ends of the range is not recommended; in particular, interpretations of data below a flaw size of 0.1 in. (2.54 mm) are not recommended.

2
PROBABILITY OF DETECTION

Probability of detection (POD) is typically defined as a measure of the probability that a flaw of a given size will be detected during an in-service inspection.

POD Model

The most common model for POD is a two-parameter, single-variable logistic model on the flaw size in the form

$$POD(s) = \frac{e^{(\beta_1 + \beta_2 s)}}{1 + e^{(\beta_1 + \beta_2 s)}} \qquad \text{Eq. 2-1}$$

where s represents the flaw size (independent variable), and β_1 and β_2 are the unknown model parameters to be determined. The parameters β_1 and β_2 in Equation 2-1 are estimated by the maximum likelihood method using a generalized linear model. For that end, the software package R is used with a script developed by Patrick Heasler of Pacific Northwest National Laboratory. The complete R script, along with a detailed explanation of the method and examples of the output it produces, can be found in EPRI's 2017 report 3002010988, *Materials Reliability Program: Development of Probability of Detection Curves for Ultrasonic Examination of Dissimilar Metal Welds (MRP-262, Revision 3): Typical PWR Leak-Before-Break Line Locations* [2]. In addition to providing the estimation of the POD model parameters, the R code also provides upper and lower confidence bounds calculated by the log-likelihood method; as typical, the choice was to use 95% confidence level bounds.

As detailed in the 2017 report, a solution to Equation 2-1 will not exist if any of the following conditions are present:

- The data set consists solely of detects.
- The data set consists solely of misses (non-detects).
- Detections and misses in the data set can be perfectly separated by a particular flaw size.

However, the data did not show any of these singular cases, and the proposed methodology successfully modeled the four analyzed scenarios.

Supplement 4

As previously mentioned, the detection data for each supplement were analyzed in two ways: considering only passed candidates (designated by P) and considering also those failed candidates that had a single miss (P+F).

The resulting POD curves for Supplement 4 in each of these cases are seen in Figure 2-1 and Figure 2-2, with their corresponding 95% confidence level bounds. In addition, Figure 2-3 compares the resulting POD curves including only passed candidates and passed plus failed candidates. Table 2-1 lists the model parameters and their respective standard deviations (σ) for all cases analyzed, and Table 2-2 provides the sizes at which each modeled POD reaches 99% and is close to the asymptotic limit of 100%.

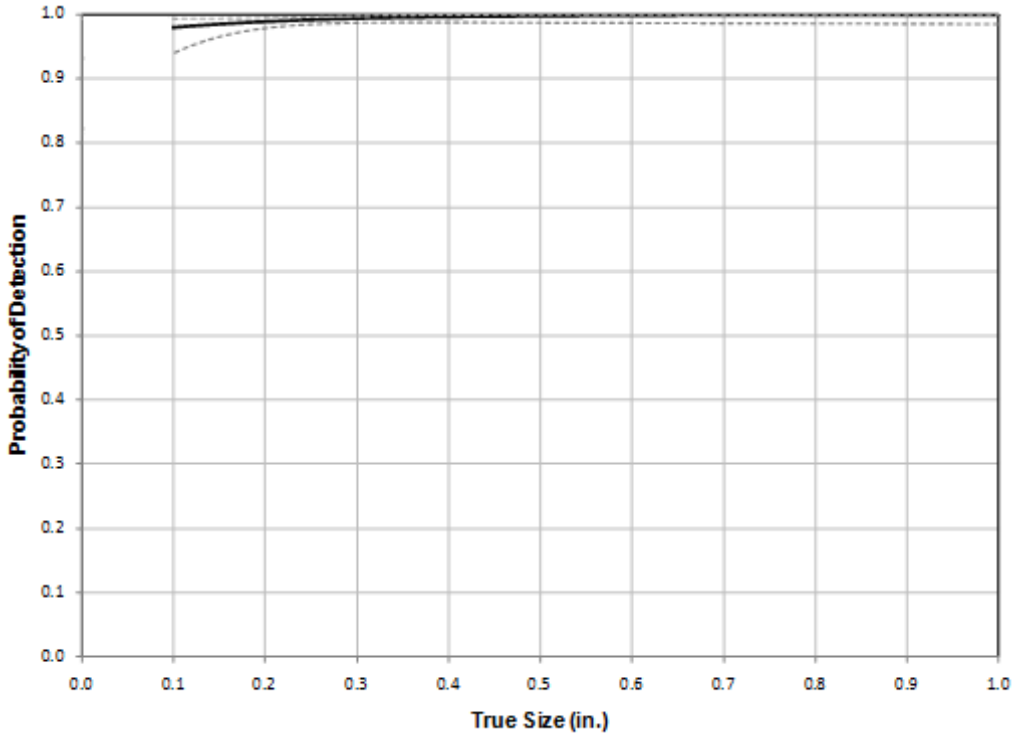

Figure 2-1
POD curve and 95% confidence bounds for Supplement 4, passed candidates only

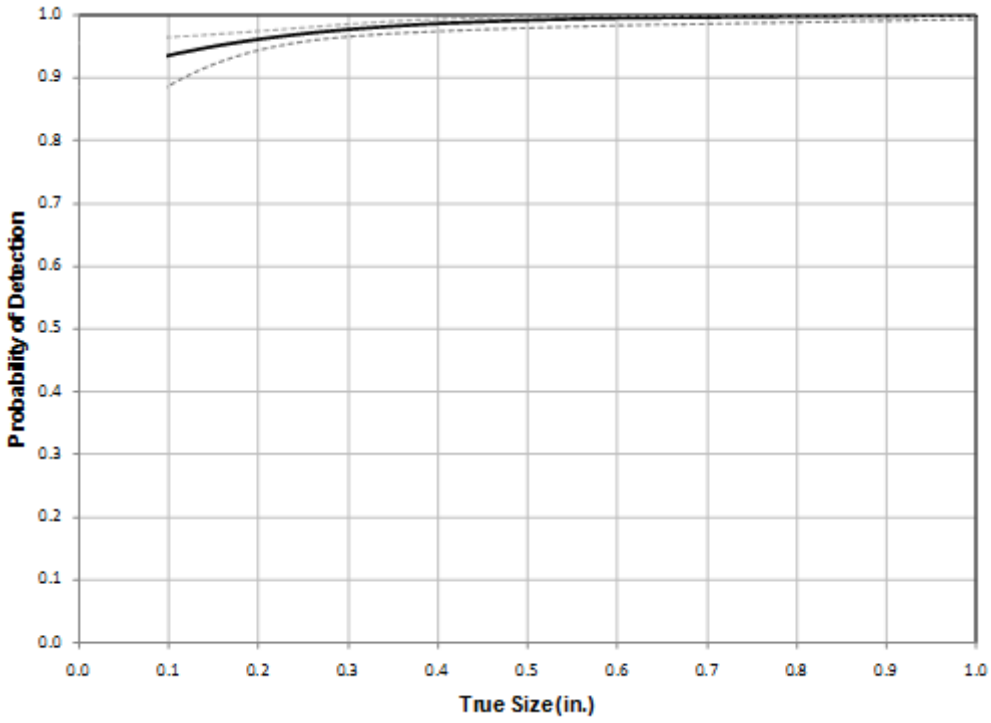

Figure 2-2
POD curve and 95% confidence bounds for Supplement 4, passed plus failed candidates

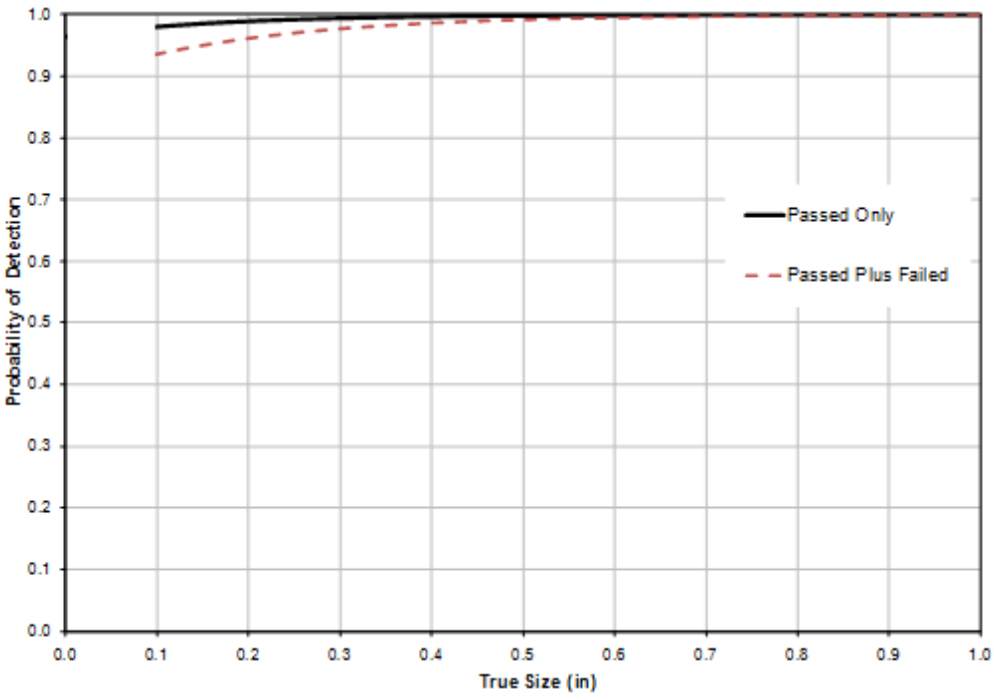

Figure 2-3
POD curves for Supplement 4

Table 2-1
POD model parameters and their respective standard deviations

Case	β_1	$\sigma_{\beta 1}$	β_2	$\sigma_{\beta 2}$
Supplement 4, P	3.28	0.92	6.23	3.69
Supplement 4, P+F	2.14	0.47	5.43	1.80
Supplement 6, P	2.68	0.62	4.00	1.62
Supplement 6, P+F	1.72	0.34	3.23	0.82

P = Passed; P+F = Passed plus Failed

Table 2-2
Flaw sizes for 99% POD for each case

Case	Size at Which POD is 99%	
	in.	mm
Supplement 4, P	0.21	5.3
Supplement 4, P+F	0.45	11.4
Supplement 6, P	0.48	12.2
Supplement 6, P+F	0.89	22.6

P = Passed; P+F = Passed plus Failed

It is seen that, considering passed candidates only, the POD is 99% for flaws as small as 0.21 in. (5.3 mm) (even the lower 95% confidence bound shows a very high POD at this size); when failed candidates are considered as well, the same happens for flaws 0.45 in. (11.4 mm.) in size, demonstrating the qualification process screening. Figure 2-3 compares both POD curves and illustrates well the results of the screening.

Also, because the PDI program was not originally designed to support statistical analysis, Code restrictions put a lower limit on the flaw size. This leads to a lack of data on the small flaw region (the region of the sharp increase in the POD curve) that contributes to the wider confidence bounds for smaller flaws seen in Figure 2-1 and Figure 2-2. As previously mentioned, data in this area are extrapolated and interpretation should be limited.

The curves in Figure 2-1 through Figure 2-3 in this report should be compared with the curves seen in Figures 3-1 through 3-3 in EPRI report 1007984; for convenience, the POD curves obtained in EPRI report 1007984 are duplicated and can be seen in Figure 2-4. It is noted that in EPRI report 1007984 the POD curves for passed candidates have been constrained to be zero for zero flaw size; this constraint has not been applied here, and the results shown in Figure 2-1 and Figure 2-3 are directly obtained from the model parameters as listed in Table 2-1, according to Equation 2-1 (the confidence bounds are outputs from the R code used [2]). In both cases, interpretation of the results for flaws smaller in size than 0.1 in. (2.54 mm) is not recommended.

Comparing the POD curves for the passed candidates in Figure 2-3 and Figure 2-4, it is noted that in both cases POD approaches 100% at a size of approximately only 0.2 in. (5.1 mm), although the results from EPRI report 1007984 show a slightly higher POD at this flaw size, which can be a result of the constraint applied. The results considering failed candidates from EPRI report 1007984 show a lower POD for smaller flaws; this is expected, because in EPRI report 1007984 failed candidates considered in the analysis were allowed two misses, whereas in the present study they are allowed a single miss. The model fit in EPRI report 1007984 rises to the asymptotic limit (100% POD) at a slightly faster pace.

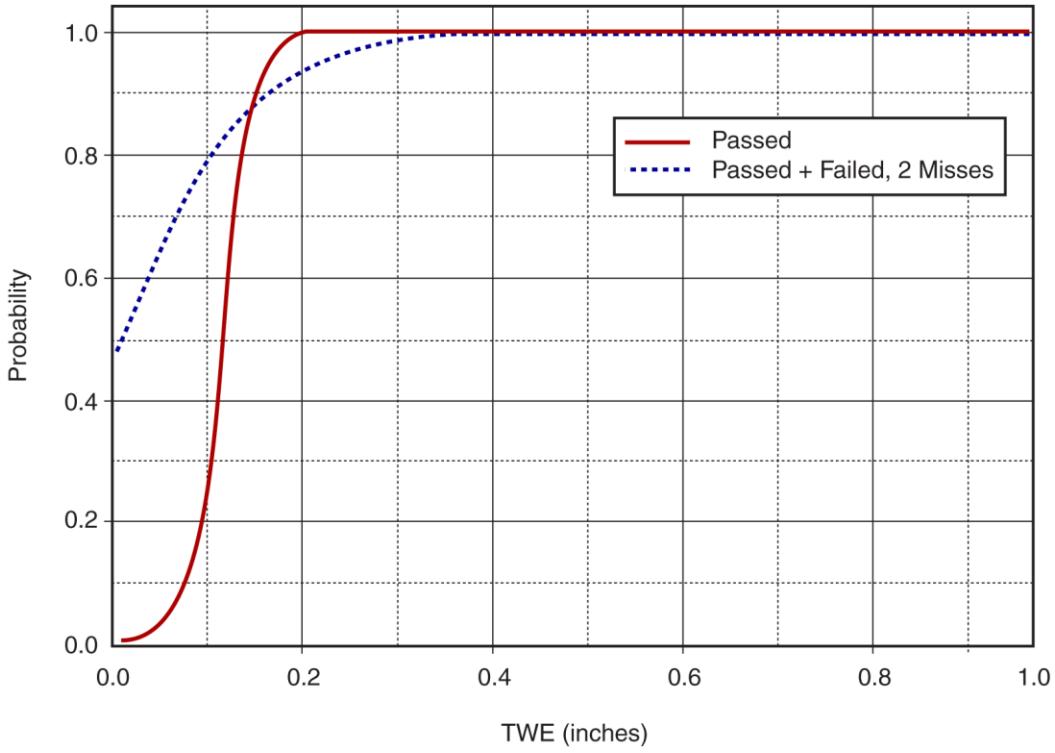

Figure 2-4
POD curves for Supplement 4 previously obtained in EPRI report 1007984. Note TWE = through-wall extent.

Supplement 6

Similar analysis was performed for Supplement 6, and the resulting POD curves can be seen in Figure 2-5 through Figure 2-7. The model parameters are found in Table 2-1.

The same wider confidence bounds are seen for smaller flaws, and POD reaches 99% at a flaw size of 0.48 in. (12.2 mm) when only passed candidates are considered and a flaw size of 0.89 in. (22.6 mm) when failed candidates are included (see Table 2-2).

In EPRI report 1007984, the corresponding figures are numbered 3-9, 3-10, and 3-11. Both POD curves from EPRI report 1007984 are also duplicated for convenience in Figure 2-8. Once again, the POD for passed candidates only from EPRI report 1007984 is slightly higher, reaching almost 100% at a flaw size of approximately 0.3 in. (7.6 mm). The behavior of the curves including failed candidates for Supplement 6 is also similar to the one observed for Supplement 4: the POD curve obtained in EPRI report 1007984 is lower for smaller flaws because more failed attempts were permitted, but it rises to the asymptotic limit faster.

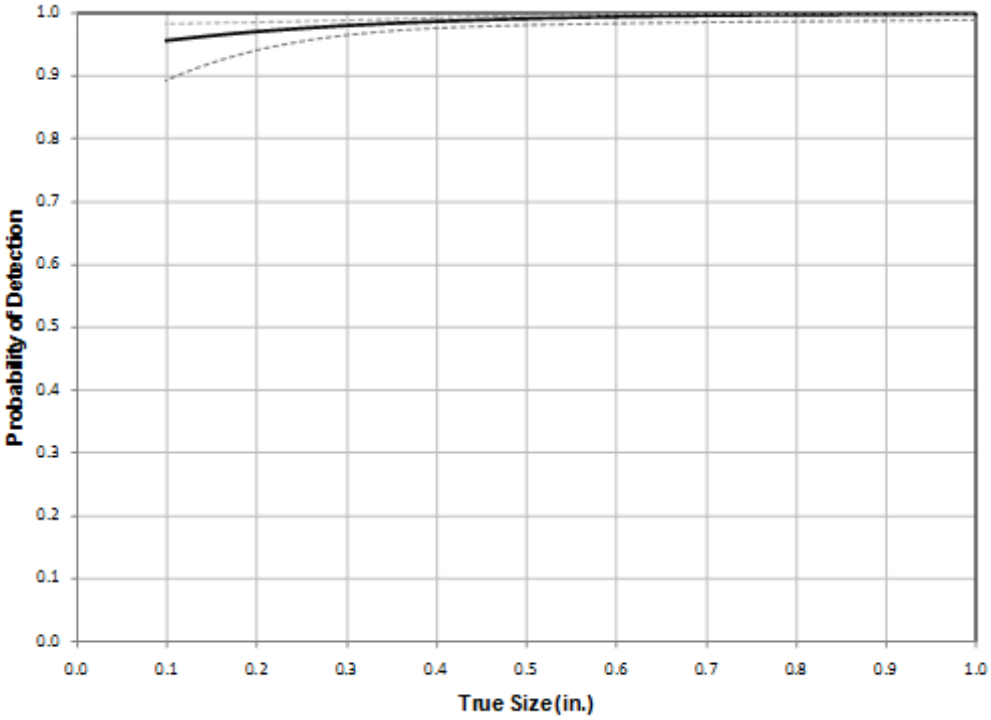

Figure 2-5
POD curve and 95% confidence bounds for Supplement 6, passed candidates only

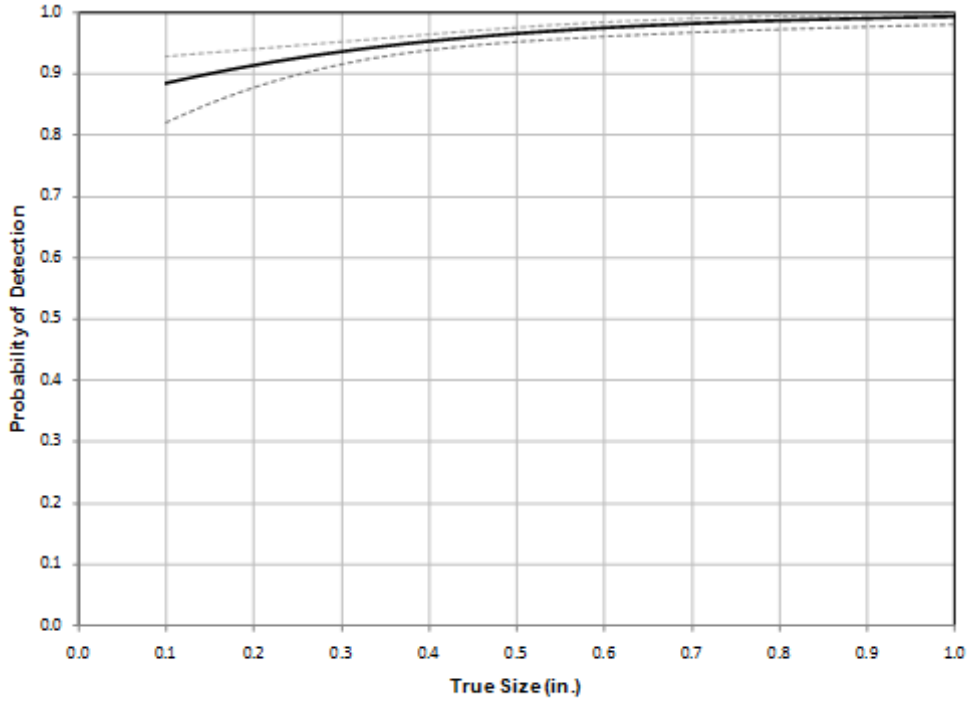

Figure 2-6
POD curve and 95% confidence bounds for Supplement 6, passed plus failed candidates

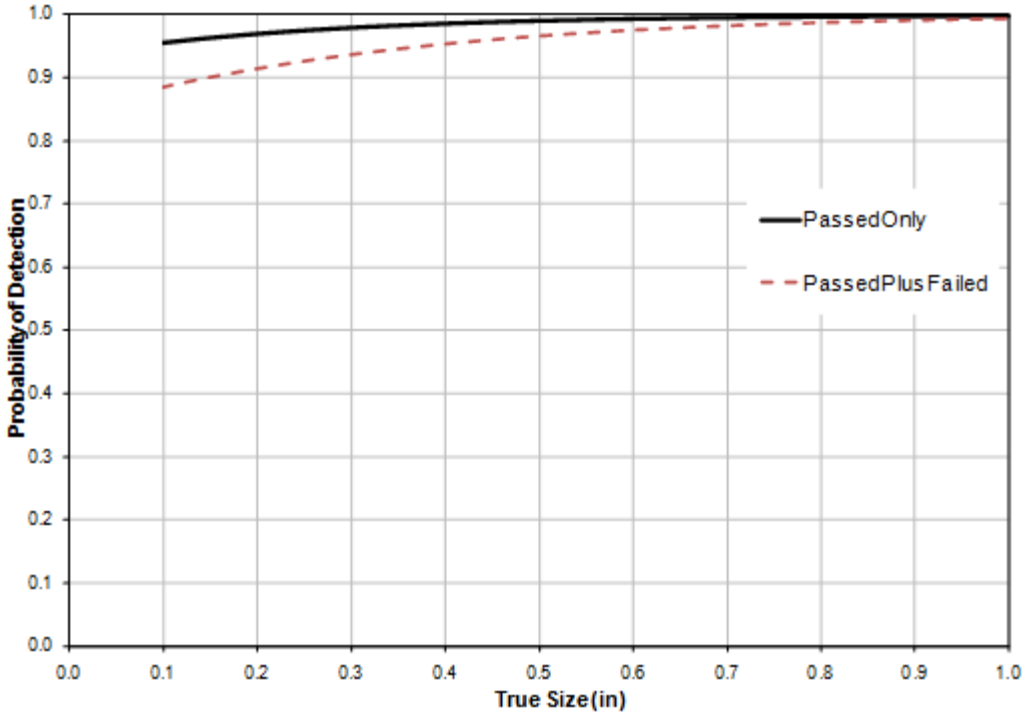

Figure 2-7
POD curves for Supplement 6

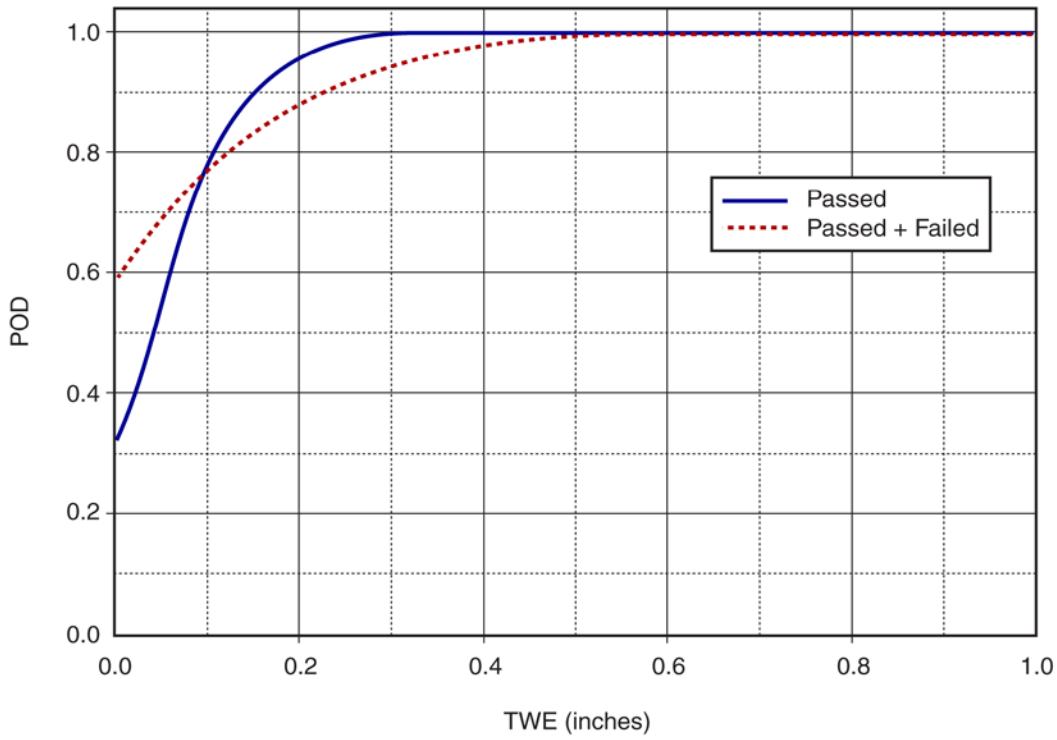

Figure 2-8
POD curves for Supplement 6 previously obtained in EPRI report 1007984. Note TWE = through-wall extent.

Comparing the results obtained for Supplements 4 and 6, the current results show the same characteristics as the ones from EPRI report 1007984: Supplement 4 inspections show a better performance regarding POD. Analysis of Table 2-2 reveals that the curves for Supplement 4 reach 99% POD at considerably smaller sizes than those for Supplement 6; in fact, even the Supplement 4 POD curve including the failed candidates reaches 99% at a smaller flaw size than the Supplement 6 POD curve with only passed candidates (0.45 in. [11.4 mm] versus 0.48 in. [12.2 mm]). This behavior is expected for inspection from the inside surface, because Supplement 4 is concerned with the inner region and Supplement 6 is concerned with the outer region of the component, meaning that the volume of interest is farther from the inspection point. Analogously, as observed in EPRI report 1007984, the opposite is true for inspections from the outside surface: Supplement 6 inspections show a better performance than Supplement 4 in that case. The current work, though, does not include analysis of outside surface examination qualification results.

3
SIZING PERFORMANCE

Overall Sizing Performance

The acceptance criterion for depth sizing is based on the candidate's root mean square error (RMSE) value, which is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - s_i)^2} \qquad \text{Eq. 3-1}$$

where M_i represents the measured size of the i-th flaw, s_i is its true size, and n is the number of measurements included. The acceptance criterion for Supplement 4 is 0.15 in. (3.8 mm) RMSE; for Supplement 6, it is 0.25 in. (6.4 mm) RMSE.

Figure 3-1 illustrates the RMSE performance of all candidates (including those who failed) in through-wall extent sizing examinations, and Table 3-1 lists the corresponding basic statistics; Figure 3-2 shows the performance of passed candidates only.

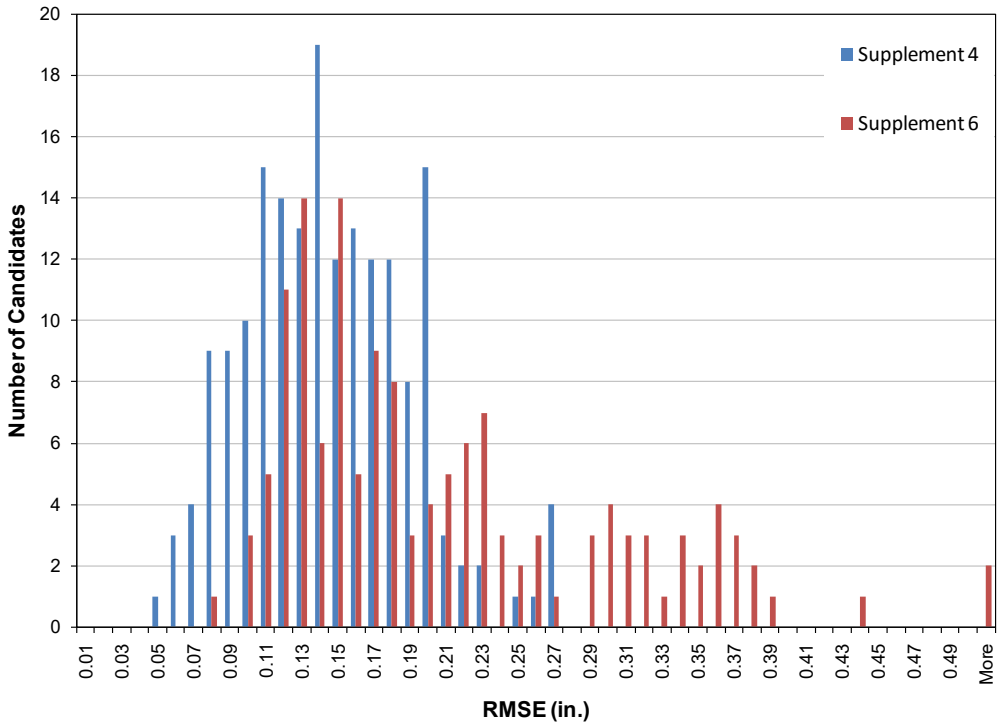

Figure 3-1
RMSE sizing performance of all candidates (passed and failed)

Table 3-1
Basic statistics for candidate RMSE distribution (including failed candidates)

Case	Mean	Standard Deviation	Number of Data Points	Error in Mean
Supplement 4	0.140	0.046	182	0.003
Supplement 6	0.205	0.098	142	0.008

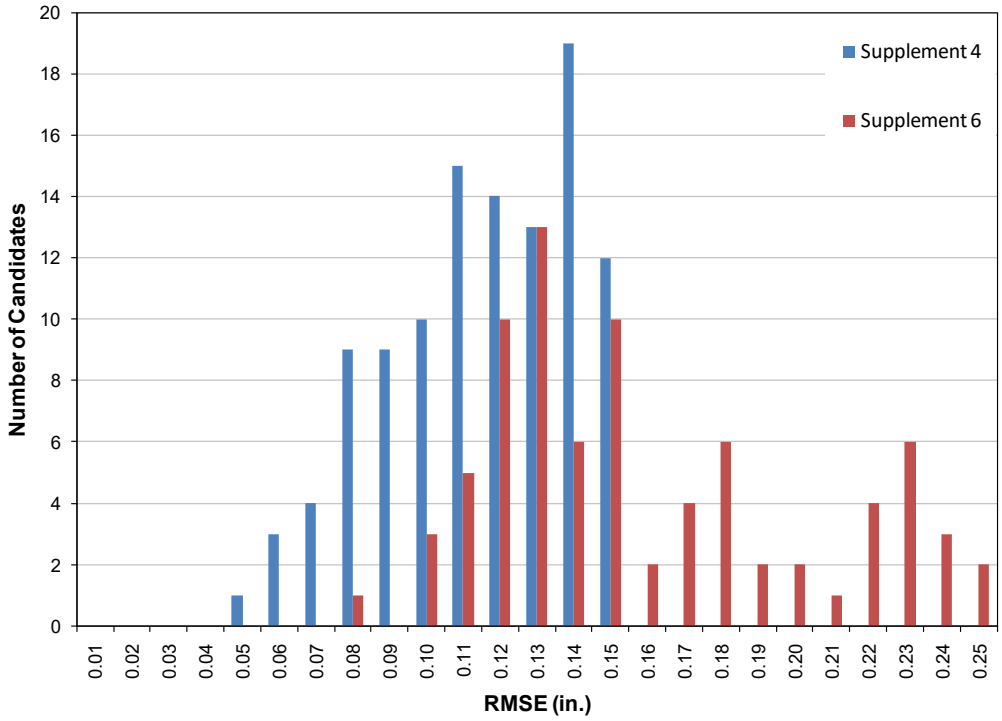

Figure 3-2
RMSE sizing performance of passed candidates only

In addition, Figure 3-3 shows the RMSE normal distribution for each supplement based on the mean and standard deviation values listed in Table 3-1. It is seen that Supplement 4 sizing shows a much smaller standard deviation, leading to a narrower distribution. Also, in the case of Supplement 4, the mean is much closer to the acceptance criterion (93%) than in the case of Supplement 6 (82%).

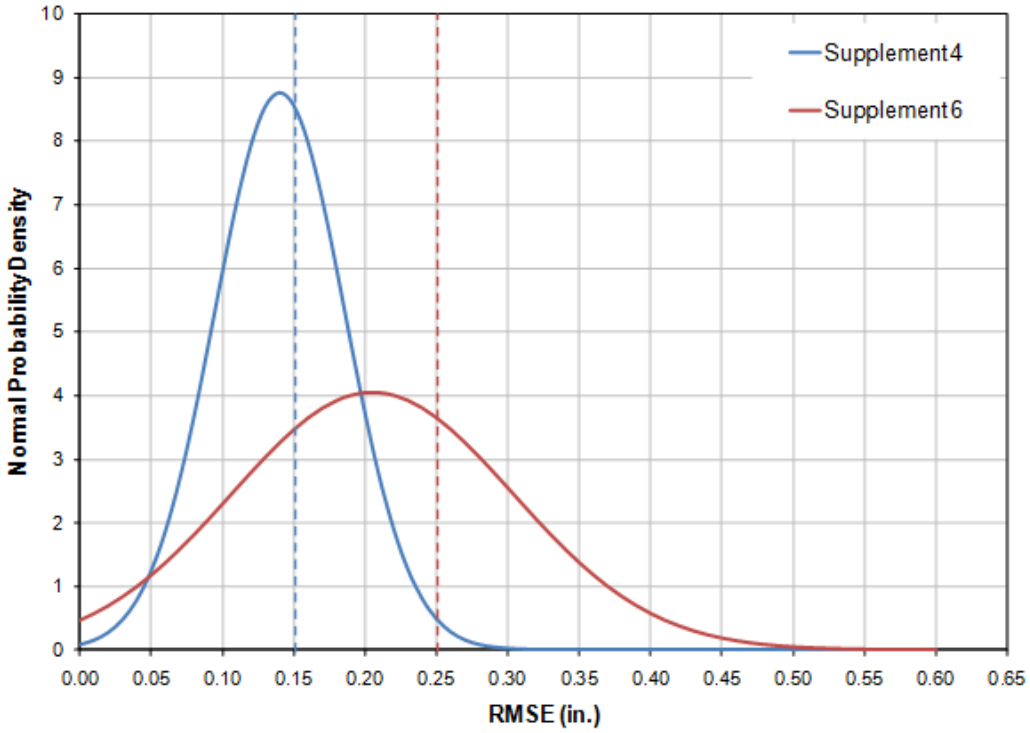

Figure 3-3
Normal distribution for the RMSE for each supplement and respective passing criteria

Sizing Model

The sizing model assumes a linear relationship between true and measured sizes of the form

$$M(s) = a + bs + \varepsilon \qquad \textbf{Eq. 3-2}$$

where M is the measured size, s is the true size, a and b represent the model parameters to be determined by a linear regression, and ε is an error term assumed to be normally distributed and has zero mean. Classical regression analysis then provides estimates for the model parameters (intercept, a, and slope, b). Explicitly, the equations used in the model to determine the parameters are

$$b = \frac{n\sum_{i=1}^{n} M_i s_i - \sum_{i=1}^{n} M_i \sum_{i=1}^{n} s_i}{n\sum_{i=1}^{n} s_i^2 - \left(\sum_{i=1}^{n} s_i\right)^2} \qquad \textbf{Eq. 3-3}$$

$$a = \frac{\sum_{i=1}^{n} M_i - b\sum_{i=1}^{n} s_i}{n} \qquad \textbf{Eq. 3-4}$$

where n is the number of data points in the set and M_i and s_i form the i-th pair of measured and true sizes, respectively.

Supplement 4

For the sizing analysis, only data from candidates who have passed the sizing qualification exam are considered. The model results for Supplement 4 can be seen in Figure 3-4, and Figure 3-5 shows the predicted error as a function of the true size.

As seen previously in EPRI report 1007984[1], there is a clear trend to overestimate smaller flaws and underestimate larger ones. In this case, this trend is so pronounced that for flaws larger than about 0.85 in. (21.6 mm), the true line (that is, the line on which measured and true sizes are the same) is outside the 95% confidence bounds of the model.

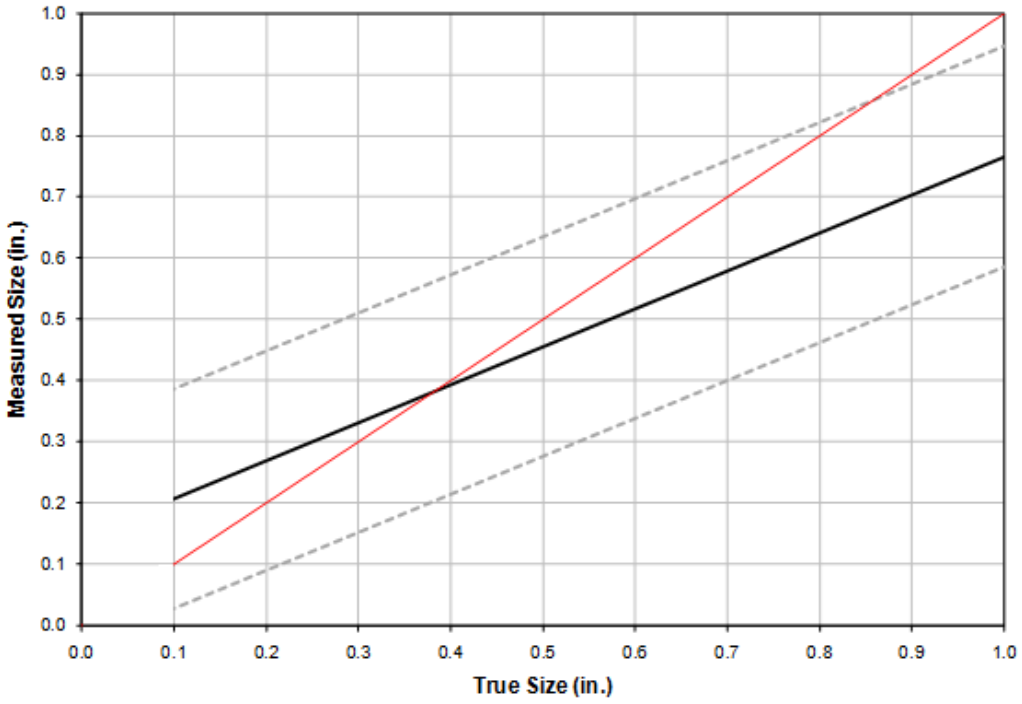

Figure 3-4
Sizing model results for Supplement 4 (black) with 95% confidence bounds (gray) and truth line (red)

[1] It should be noted that the sizing analysis performed in EPRI report 1007984 not only follows a different methodology but also includes inside and outside surface examinations indiscriminately.

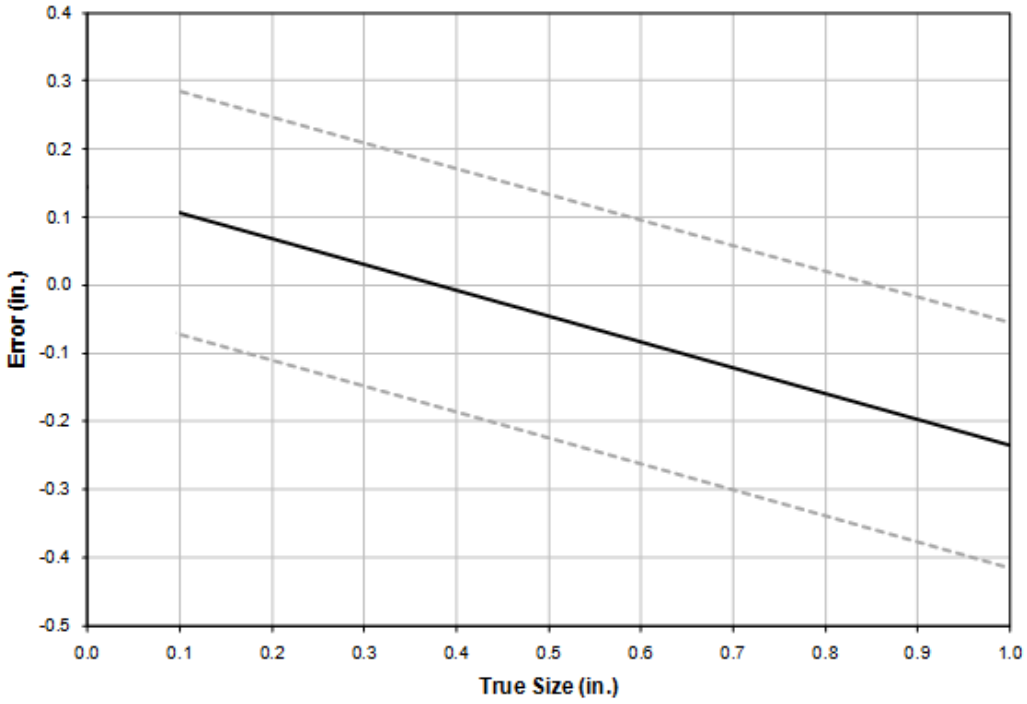

Figure 3-5
Predicted sizing error for Supplement 4 with 95% confidence bounds

Supplement 6

Similar analysis for Supplement 6 yields the regression seen in Figure 3-6 and Figure 3-7.

The same trend to oversize small flaws and undersize large flaws seen for Supplement 4 is observed here, but the bias for Supplement 6 is much smaller than for Supplement 4: notice how much closer the sizing model (black line) is to the truth line (red) in Figure 3-6 when compared with Figure 3-4, or note the much slighter slope in the predicted error for Supplement 6 when compared with Supplement 4 (see Figure 3-5 and Figure 3-7). Furthermore, for Supplement 6 the truth line is always within the 95% confidence bounds in the considered interval. This characteristic was observed in EPRI report 1007984.

On the other hand, the much wider spread for Supplement 6 seen in Figure 3-3 (or the higher standard deviation in Table 3-1) has a visible effect in the error as well: notice how much wider are the 95% confidence bounds for the predicted error for Supplement 6 (see Figure 3-7) when compared with those of Supplement 4 (see Figure 3-5). This fact also was observed in EPRI report 1007984.

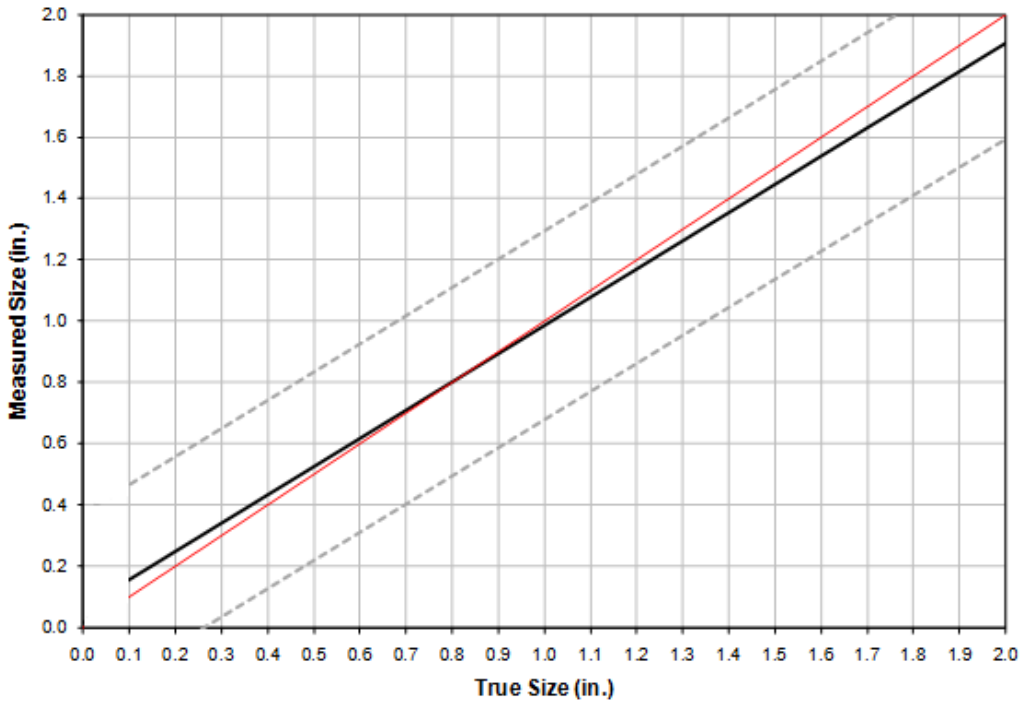

Figure 3-6
Sizing model results for Supplement 6 (black) with 95% confidence bounds (gray) and truth line (red)

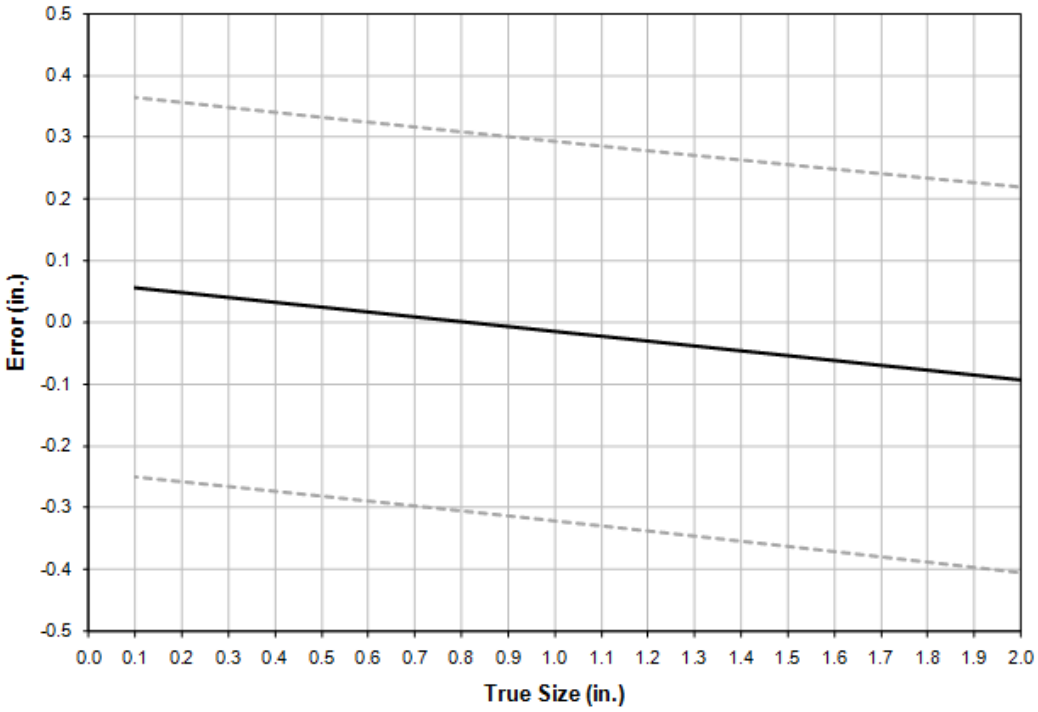

Figure 3-7
Predicted sizing error for Supplement 6 with 95% confidence bounds

In summary, the notable difference between the sizing performance in Supplement 4 and Supplement 6 is the fact that Supplement 6 shows a much smaller bias (trend to oversize small flaws and undersize large flaws) but has a much broader spread, leading to wider confidence bounds. In addition, the overall RMSE confirms that Supplement 4 candidates pass with a much narrower margin (RMSE of 0.113 in. [2.9 mm] out of 0.15 in. [3.8.mm], or 75.33%) than Supplement 6 candidates (RMSE of 0.159 in. [4.0 mm] out of 0.25 in. [6.4 mm], or 63.60%).

4
PROBABILITY OF ACCEPTABLE SIZING

Definition

Probability of Acceptable Sizing (PAS) is defined as the probability that a flaw of a given size is detected and sized within a determined tolerance from its true size. Following this definition, the PAS is calculated by

$$PAS_{\alpha}(s) = POD(s) \times \int_{s-\alpha}^{s+\alpha} N\left(x, \mu(s), \sigma(s)\right) dx \qquad \text{Eq. 4-1}$$

where α is the sizing tolerance, $N(x,\mu,\sigma)$ represents the normal distribution of mean μ, and standard deviation σ calculated at x. Taking the mean to be the predicted measured size from the sizing model ($a+bs$) and the standard deviation to be the square root of the variance in the prediction of the measured size, the integral in Equation 4-1 is the probability that the flaw size will be measured between $s-\alpha$ and $s+\alpha$. Because detection is a necessary condition for sizing (a flaw cannot be sized unless it is detected), this probability is then multiplied by the POD, yielding the PAS as defined.

It is noted that the PAS can be calculated at different tolerance levels (α parameter); in this report, α is chosen to be the passing criterion for the supplement in question—that is, α is 0.15 in. (3.8 mm) for Supplement 4 applications and 0.25 in. (6.4 mm) for Supplement 6.

Supplement 4

Applying the definition of PAS in Equation 4-1 to Supplement 4 with a tolerance of 0.15 in. (3.8 mm), the curve shown in Figure 4-1 is obtained.

It is seen that the PAS peaks at approximately 90% around 0.4 in. (10.2 mm); this is the size for which the bias is null—that is, it is the transition point from oversizing small flaws to undersizing large flaws. This transition point is seen in Figure 3-4 as the crossover point of the black curve (sizing model) and the red curve (truth); alternatively, it is seen in Figure 3-5 as the point where the error is zero. As the flaw size moves away from this point to either side, the PAS falls considerably, because flaws are either oversized (small flaws, to the left of the peak in the PAS) or undersized (large flaws, to the right of the peak in the PAS).

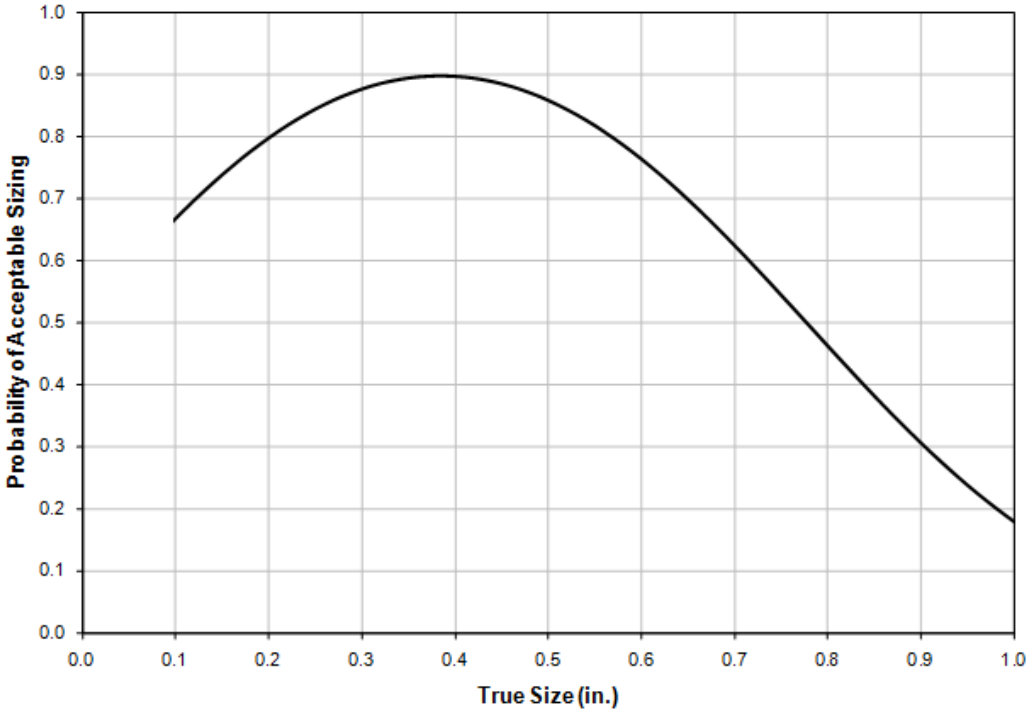

Figure 4-1
Probability of Acceptable Sizing for Supplement 4 with 0.15-in. (3.8-mm) tolerance

Supplement 6

Similar analysis for Supplement 6 with a tolerance of 0.25 in. (6.4 mm) yields the PAS curve seen in Figure 4-2. The PAS peaks nearly at 90% around 0.8 in. (20.3 mm), which again is the transition point for Supplement 6 from oversizing to undersizing (crossover point between model and truth curves in Figure 3-6 or point of zero error in Figure 3-7). However, the PAS decline to either side of the peak is not nearly as pronounced in this case, and the PAS is always above 80% in the range considered.

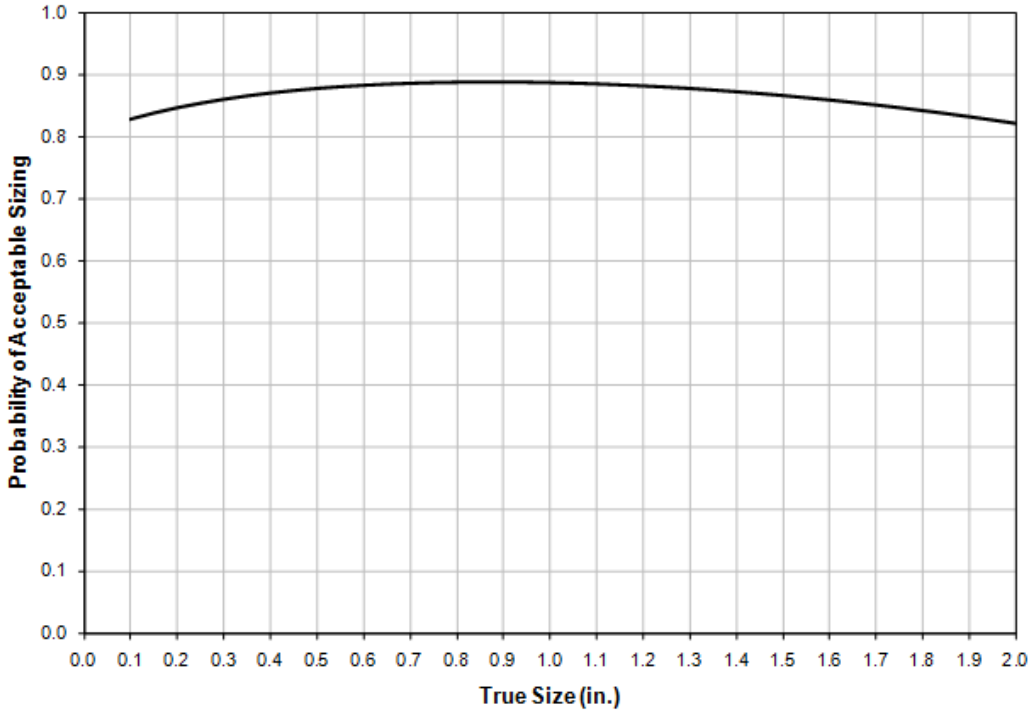

Figure 4-2
Probability of Acceptable Sizing for Supplement 6 with 0.25-in. (6.4-mm) tolerance

The great difference in the PAS curves for Supplement 4 (see Figure 4-1) and Supplement 6 (see Figure 4-2) is caused by the difference in the bias and spread (deviation) observed for each supplement (the bias can be graphically seen either as the difference between the model and truth curves in Figure 3-4 and Figure 3-6 or as the slope of the error curve in Figure 3-5 and Figure 3-7; the deviation, or spread, can be seen graphically as the width of the confidence bounds in Figure 3-5 and Figure 3-7).

To understand how the difference in the bias and spread observed for each supplement results in the great differences seen in the PAS curves for Supplement 4 (Figure 4-1) and 6 (Figure 4-2), consider the two following affirmations:

- A larger bias leads to a mean that is farther from the true size at either end of the size range; and
- A smaller standard deviation leads to a narrower distribution of measurements around the mean.

Supplement 4 measurements exhibit an unfortunate combination when compared to Supplement 6: a larger bias with a smaller standard deviation, which means that the measurements are more tightly distributed around a mean that is farther from the true value; this leads to a tendency in which the true value of the flaw is virtually not included in the measurement distribution for Supplement 4, thus leading to smaller PAS values. Conversely, for Supplement 6 the bias is much smaller and the standard deviation is larger, leading to a broader distribution around a mean that is already close to the true value; therefore, in this case the true size is well within the distribution of measurements. Figure 4-3 illustrates this fact using the Supplement 4 predicted measured size distribution for a true size of 1 in. (25.4 mm) and the Supplement 6 predicted

measured size distribution[2] for a true size of 2 in. (50.8 mm) (these true sizes are at the end of the range for each case). As observed, the true size of 1 in. (25.4 mm) for the Supplement 4 curve is practically outside the narrow, highly biased distribution, leading to the low PAS at this size; on the other hand, a true size of 2 in. (50.8 mm) is well within the broader Supplement 6 distribution with its lower bias, leading to a considerably high PAS even at the end of the range.

Graphically, one can see that the low PAS in Figure 4-1 corresponds to the size intervals where the model line (black curve) is farther away from the truth line (red curve) in Figure 3-4 or the interval where the absolute error is larger in Figure 3-5; additionally, the PAS is lower for larger flaws because the error is greater in this area. Similarly, inspection of Figure 3-6 reveals that the model curve (black) is close to the truth curve (red) over the entire range for Supplement 6; indeed, Figure 3-7 shows that the mean error in this case is at most 0.1 in (2.5 mm), which is only a fraction (40%) of the 0.25 in (6.4 mm) tolerance used for the PAS calculation for Supplement 6. For Supplement 4, however, Figure 3-5 shows that the magnitude of the mean error exceeds the PAS tolerance of 0.15 in (3.8 mm) at a flaw size around 0.75 in (19.1 mm), leading to the low PAS values observed for larger flaw sizes.

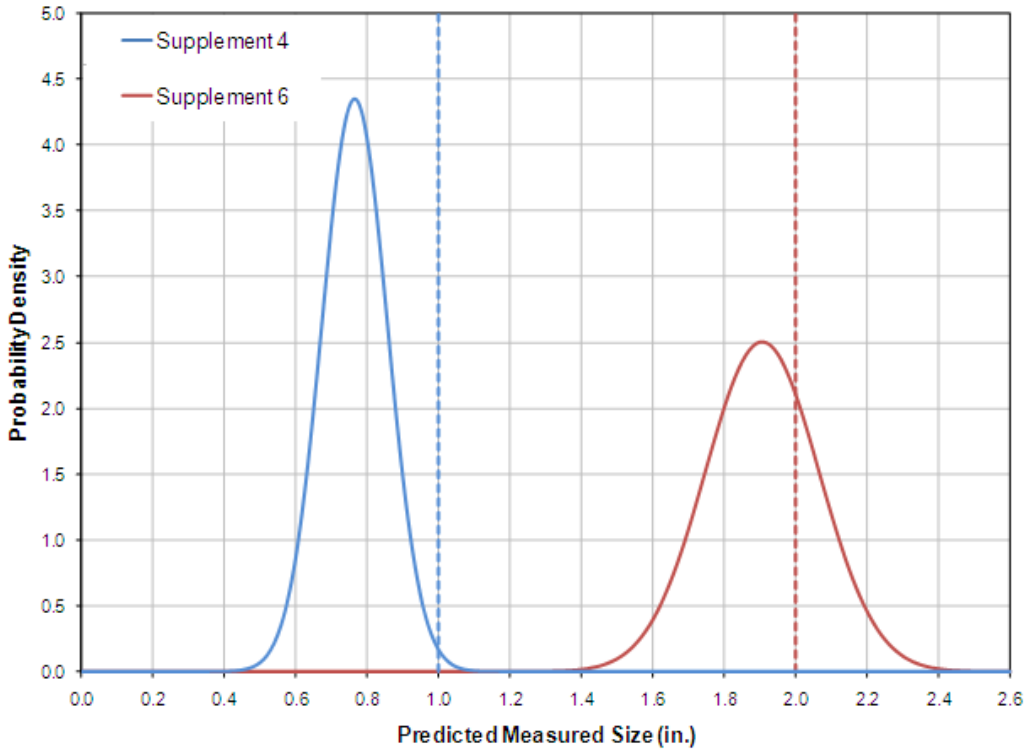

Figure 4-3
Predicted measured size distributions for Supplement 4 for 1 in. (25.4 mm) true size and Supplement 6 for 2 in. (50.8 mm) true size

[2] The distribution of predicted measure size for each Supplement in Figure 4-3 is a normal distribution with mean and standard deviation obtained from the sizing model for each Supplement at the given sizes, that is, 1.0in (25.4mm) for Supplement 4 and 2.0in (50.8mm) for Supplement 6. Graphically, the mean can be seen as the value of the black curve at a true size of 1.0in in Figure 3-4 for Supplement 4, and at a true size of 2.0in in Figure 3-6 for Supplement 6. The standard deviation for each Supplement is proportional to the width of the confidence bounds in each of these figures at the given sizes.

5
PROBABILITY OF REJECTION

Definition

Probability of Rejection (PR) is the probability that a flaw of a given size will be detected and sized to be larger than an acceptance criterion. Following this definition, the PR is defined in a manner similar to the PAS by

$$PR_{\delta}(s) = POD(s) \times \int_{\delta}^{\infty} N(x, \mu(s), \sigma(s))\, dx \qquad \text{Eq. 5-1}$$

where δ is the acceptance criterion adopted and, again, $N(x,\mu,\sigma)$ is a normal distribution of mean μ and standard deviation σ. In the model, the mean is taken as the predicted measured size ($a+bs$) and the standard deviation is the square root of the variance in the prediction of the measured size.

Analogously to the analysis performed in EPRI report 1007984 [1], this report provides the PR for Supplement 4 with an acceptance criterion of 0.15 in. (3.8 mm).

Supplement 4

Applying Equation 5-1 to Supplement 4, the PR curve seen in Figure 5-1 is obtained. The vertical dashed line represents the adopted acceptance criterion (0.15 in. [3.8 mm]); a flaw of this size has an approximately 82% chance of being rejected. This also means that a flaw of any size equal to or larger than 0.15 in. (3.8 mm) (acceptance criterion) has at least an 82% chance of being correctly reported. As expected, the PR increases with size.

The region to the left of the vertical dashed line represents flaws that are detected but incorrectly reported to be larger than the acceptance criterion (analysis and interpretation for flaws smaller than 0.10 in. [2.54 mm] is not recommended). To the extent that analysis is permitted, the trend to oversize flaws has an economically detrimental impact: the curve in Figure 5-1 shows that flaws that are 0.10 in. (2.54 mm) in size (and therefore are below the acceptance threshold) have an approximately 72% chance of being incorrectly reported. On the other hand, safety concerns are minimized by this same trend.

Additionally, it is observed that the trend to undersize large flaws does not cause a safety concern with the adopted acceptance criterion of 0.15 in. (3.8 mm): as covered previously and seen in Figures 3-4 and 3-5, under Supplement 4, flaws start to be undersized at approximately 0.4 in. (10.2 mm) in size; at this size and above, Figure 5-1 shows a PR of better than 99% in spite of the undersizing trend.

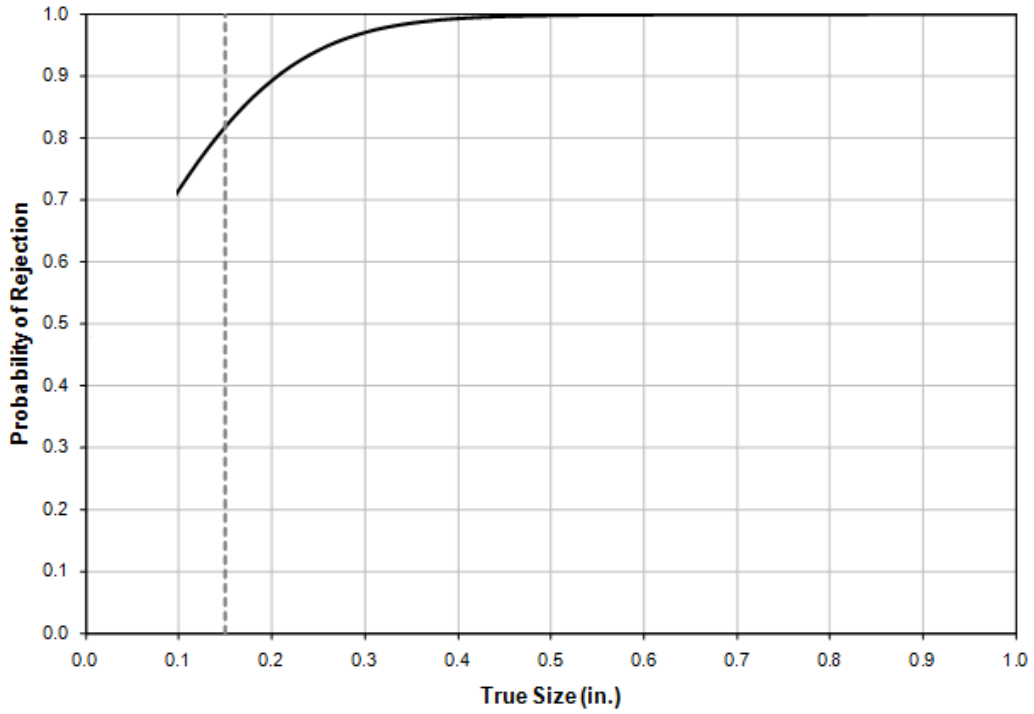

Figure 5-1
Probability of Rejection for Supplement 4 with an acceptance criterion of 0.15 in. (3.8 mm)

6
CONCLUSIONS

This report provides an analysis of available performance demonstration detection and through-wall sizing data for inside surface reactor pressure vessel inspections under the American Society of Mechanical Engineers Boiler & Pressure Vessel Code, Section XI, Appendix VIII, Supplements 4 and 6. Curves and model parameters for POD and sizing performance are provided. Further performance indicators that combine detection and sizing capabilities—that is, the PAS and PR—are defined and calculated as well.

Analysis of detection data has shown that candidates who have passed the qualification exam show a very good performance, achieving 99% POD for flaws as small as 0.21 in. (5.3 mm) for Supplement 4 and 0.48 in. (12.2 mm) for Supplement 6. In addition, comparing the performance of passed candidates with that of failed candidates who have missed only one flaw shows the improved performance of passed candidates, illustrating the screening ensured by the qualification process. It is also observed that Supplement 4 shows a better detection performance than Supplement 6; this is expected, because inspection is done from the inside surface and while Supplement 4 includes the inner volume of the component, Supplement 6 considers the outer (and therefore more remote) volume of the component. Lastly, because of the lack of data for small flaws, interpretation of data below 0.1 in. (2.54 mm) is not recommended; this limitation in the range of validity of the POD curve is extended to the other measures and curves in this report.

Study of the sizing performance shows that candidates pass Supplement 4 criterion only narrowly, while the Supplement 6 pass criterion is met by passed candidates with greater ease. Also, the error distribution for Supplement 4 is narrow, while for Supplement 6, it is considerably broader. Notably, the same trend first observed in EPRI report 1007984 [1]—that candidates tend to oversize small flaws and undersize large flaws—is also observed here. This bias is much more pronounced for Supplement 4 measurements.

The Probability of Acceptable Sizing (PAS)—that is, the probability that a flaw of a given size is detected and measured to be within a certain tolerance of the true value—is also provided. In this work, the tolerances have been adopted to be the same as the pass criterion for each case—that is, 0.15 in. (3.8 mm) for Supplement 4 and 0.25 in. (6.4 mm) for Supplement 6. Results show that although the PAS curve for both supplements peaks at approximately 90% at the transition size at which oversizing is replaced by undersizing (approximately 0.4 in. [10.2 mm] for Supplement 4 and 0.8 in [20.3 mm] for Supplement 6), the PAS curve for Supplement 4 falls much more drastically to either side of this point (being as low as 20% for flaws approximately 1 in. [25.4 mm] in size) than the PAS curve for Supplement 6, which is above 80% for the entire range considered. This is caused by the much larger bias and tighter distribution observed in Supplement 4 measurements.

The Probability of Rejection (PR), which measures the probability that a flaw of a given size is detected and measured to be equal to or larger than a given acceptance criterion, is also provided for Supplement 4 for a criterion of 0.15 in. (3.8 mm). In accordance with this criterion, results show that there is at least an 82% chance that a flaw 0.15 in. (3.8 mm) or greater would be detected and rejected (and the PR increases with flaw size). Although the trend to oversize small flaws leads to conservative errors and minimizes safety concerns, it is economically detrimental: the PR shows approximately a 72% chance of incorrectly reporting a flaw 0.10 in. (2.54 mm) in size (analysis below this size is not recommended, and this in itself is a limiting case). On the opposite end, the trend to undersize large flaws is observed not to be a safety concern: Supplement 4 (which shows the worst bias) measurements start to undersize flaws approximately 0.4 in. (10.2 mm) in size; with an acceptance criterion of 0.15 in. (3.8 mm) at this size the PR is already higher than 99% so that flaws are rejected even though they are undersized.

7
REFERENCES

1. *Reactor Pressure Vessel Inspection Reliability Based on Performance Demonstrations*. EPRI, Palo Alto, CA: 2004. 1007984.
2. *Materials Reliability Program: Development of Probability of Detection Curves for Ultrasonic Examination of Dissimilar Metal Welds (MRP-262, Revision 3): Typical PWR Leak-Before-Break Line Locations*. EPRI, Palo Alto, CA: 2017. 3002010988

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI members represent 90% of the electric utility revenue in the United States with international participation in 35 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; and Lenox, Mass.

Together...Shaping the Future of Electricity

Programs:

Nuclear Power

Nondestructive Evaluation

3002013319

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com