Standard Practice for Probability of Detection Analysis for Hit/Miss Data[1]

This standard is issued under the fixed designation E2862; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice defines the procedure for performing a statistical analysis on nondestructive testing hit/miss data to determine the demonstrated probability of detection (POD) for a specific set of examination parameters. Topics covered include the standard hit/miss POD curve formulation, validation techniques, and correct interpretation of results.

1.2 The values stated in inch-pound units are to be regarded as standard. The values given in parentheses are mathematical conversions to SI units that are provided for information only and are not considered standard.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E1316 Terminology for Nondestructive Examinations

2.2 *Department of Defense Handbook:*

MIL-HDBK-1823A Nondestructive Evaluation System Reliability Assessment[3]

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*

3.1.1 *analyst, n*—the person responsible for performing a POD analysis on hit/miss data resulting from a POD examination.

3.1.2 *demonstrated probability of detection, n*—the calculated POD value resulting from the statistical analysis on the hit miss data.

3.1.3 *false call, n*—the perceived detection of a discontinuity that is identified as a find during a POD examination when no discontinuity actually exists at the inspection site.

3.1.4 *hit, n*—an existing discontinuity that is identified as a find during a POD demonstration examination.

3.1.5 *miss, n*—an existing discontinuity that is missed during a POD examination.

3.1.6 *probability of detection, n*—the fraction of nominal discontinuity sizes expected to be found given their existence.

3.2 *Symbols:*

3.2.1 a—discontinuity size.

3.2.2 a_p—the discontinuity size that can be detected with probability p.

3.2.2.1 *Discussion*—Each discontinuity size has an independent probability of being detected and corresponding probability of being missed. For example, being able to detect a specific discontinuity size with probability p does not guarantee that a larger size discontinuity will be found.

3.2.3 $a_{p/c}$—the discontinuity size that can be detected with probability p with a statistical confidence level of c.

3.2.3.1 *Discussion*—$a_{p/c}$ is calculated by applying a statistical uncertainty bound to a_p. The uncertainty bound is a function the amount of data, the scatter in the data, and the specified level of statistical confidence. The resulting value represents how large the discontinuity with POD equal to p could be when uncertainty associated with estimating a_p is accounted for. Hence $a_{p/c} > a_p$. Note that POD is equal to p for both $a_{p/c}$ and a_p. a_p is based solely on the hit/miss data resulting from the examination and represents a snapshot in time, whereas $a_{p/c}$ accounts for the uncertainty associated with limited sample data.

4. Summary of Practice

4.1 This practice describes step-by-step the process for analyzing nondestructive testing hit/miss data resulting from a POD examination, including minimum requirements for validating the resulting POD curve.

4.2 This practice also includes definitions and discussions for results of interest (for example, $a_{90/95}$) to provide for correct interpretation of results.

---

[1] This practice is under the jurisdiction of ASTM Committee E07 on Nondestructive Testing and is the direct responsibility of Subcommittee E07.10 on Specialized NDT Methods.

Current edition approved Jan. 15, 2012. Published February 2012. DOI:10.1520/E2862-12.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from Standardization Documents Order Desk, DODSSP, Bldg. 4, Section D, 700 Robbins Ave., Philadelphia, PA 19111-5098, http://dodssp.daps.dla.mil.

5. Significance and Use

5.1 The POD analysis method described herein is based on a well-known and well established statistical method. It shall be used to quantify the demonstrated POD for a specific set of examination parameters and known range of discontinuity sizes when the initial response from a nondestructive evaluation inspection system is ultimately binary in nature (that is, hit or miss). This method requires that a relationship between discontinuity size and POD exists and is best described by a generalized linear model with the appropriate link function for binary outcomes.

5.2 Prior to performing the analysis it is assumed that the discontinuity of interest is clearly defined; the number and distribution of induced discontinuity sizes in the POD specimen set is known and well-documented; discontinuities in the POD specimen set are unobstructed; the POD examination administration procedure (including data collection method) is well-defined, under control, and unbiased; and the initial response is ultimately binary in nature (that is, hit or miss). The analysis results are only valid if convergence is achieved and the model adequately represents the data.

5.3 The POD analysis method described herein is consistent with the analysis method for binary data described in MIL-HDBK-1823A, which is included in several widely utilized POD software packages to perform a POD analysis on hit/miss data. It is also found in statistical software packages that have generalized linear modeling capability. This practice requires that the analyst has access to either POD software or other software with generalized linear modeling capability.

6. Procedure

6.1 The POD analysis objective shall be clearly defined by the responsible engineer or by the customer.

6.1.1 The analyst shall obtain the hit/miss data resulting from the POD examination, which shall include at a minimum the documented known induced discontinuity sizes, whether or not the discontinuity was found, and any false calls.

6.2 The analyst shall also obtain specific information about the POD examination, which shall include at a minimum the specimen standard geometry (for example, flat panels), specimen standard material (for example, Nickel), examination date, number of inspectors, type of inspection method (for example, line-of-site Level 3 Fluorescent Penetrant Inspection), and pertinent comments from the inspector(s) and test administrator.

6.3 Prior to performing the analysis, the analyst shall conduct a preliminary review of the POD examination procedure and resulting hit/miss data to identify any examination administration or data issues. The analyst shall resolve any issues prior to conducting the POD analysis. Examples of examination administration or data issues and possible resolutions are:

6.3.1 If problems or interruptions occurred during the POD examination that may bias the results, the POD examination should be re-administered. If this occurs, it shall be documented in the report.

6.3.2 If a discontinuity was missed because it was obstructed (such as a clogged discontinuity), the discontinuity shall be removed from the POD analysis since there was not an opportunity for the discontinuity to be found. If a discontinuity is removed from the analysis, the specific discontinuity and rationale for removal shall be documented in the final report.

6.3.3 POD cannot be modeled as a continuous function of discontinuity size if there is a complete separation of misses and hits as crack size increases. If a complete separation of misses and hits is present in the data, the POD examination may be re-administered. If this occurs, it shall be documented in the report. If a complete separation of misses and hits occurs on a regular basis, the specimen set should be examined for suitability as a POD examination specimen set.

6.3.4 POD cannot be modeled as a continuous function of discontinuity size if all the discontinuities are found or if all the discontinuities are missed. If this occurs, the specimen set is inadequate for the POD examination.

6.4 The analyst shall use a generalized linear model with the appropriate link function to establish the relationship between POD and discontinuity size. For application to POD, the generalized linear model with discontinuity size as the single predictor variable is typically expressed as $g(y) = b_0 + b_1 \bullet a$ or $g(y) = b_0 + b_1 \bullet ln(a)$, where a or $ln(a)$ is the continuous predictor variable, b_0 is the intercept, b_1 is the slope, y is the binary response variable, and $g(\bullet)$ is the function that "links" the binary response with the predictor variable. If predictor variables other than discontinuity size are quantifiable factors, a generalized linear model with more than one predictor may be used.

6.5 The analyst shall choose the appropriate link function based on how well the model fits the observed data. MIL-HDBK-1823A discusses four different link functions (Logit, Probit, Log-Log, Complementary-LogLog) and describes methods for selecting the appropriate one. In general, the logit and probit link functions have worked well in practice for modeling hit/miss data.

6.6 Only hit/miss data for induced discontinuities shall be used in the development of the generalized linear model. False call data shall not be included in the development of the generalized linear model.

6.7 The analyst shall conduct the analysis using software that has generalized linear modeling capabilities.

6.8 After running the analysis, the analyst shall verify that convergence has been achieved. The resulting POD curve shall not be used if convergence has not been achieved.

6.9 After verifying convergence, the analyst shall use at a minimum the informal model diagnostic methods listed below to assess the reliability of the model and verify that the model adequately fits the data.

6.9.1 If included in the analysis output, the analyst shall check the number of iterations it took to meet the convergence criterion. If more than twenty iterations were needed to reach convergence, the model may not be reliable. A statement indicating that convergence was achieved and the number of iterations needed to achieve convergence shall be included in the report.

6.9.2 The analyst shall visually assess the shape of the POD curve. (POD curves tend to be s-shaped.)

6.9.3 The analyst shall visually assess how well the POD curve fits the data by comparing how well the range over which the POD curve is rising matches the range over which misses begin to overlap with and transition to hits as discontinuity size increases.

6.9.4 The analyst should also compare an empirical POD curve to the POD curve based on the generalized linear model. The empirical POD curve shall be used for validation purposes only. It shall not be used as a substitute for a POD curve resulting from a hit/miss analysis.

6.9.4.1 To create an empirical POD curve, divide the discontinuity sizes into bins. For example, (0.010 in., 0.020 in.), (0.020 in., 0.030 in.), …, (0.100 in., 0.110 in.), etc. ((0.0254 cm, 0.0508 cm), (0.0508 cm, 0.0762 cm), …, (0.2540 cm, 0.2794 cm), etc.). For each bin, calculate the total number of discontinuities contained in the bin and how many were detected. Calculate the empirical POD in each bin by dividing the number detected in the bin over the total number of discontinuities in the bin. Plot the empirical POD versus the midpoint of the bin to obtain the empirical POD curve. Overlay the POD curve based on the generalized linear model on the empirical POD curve to assess how well the generalized linear model fits the data by how well it matches the empirical POD curve.

6.9.5 If applicable, the analyst shall visually assess the shape of the confidence bound on the POD curve. The confidence bound should roughly follow the same shape as the POD curve. If the confidence bound flares out significantly on either or both ends or intersects the x-axis, the confidence bound should be viewed as suspect and may not be reliable.

6.9.6 The analyst should assess the impact of data that appears to be outlying (for example, an early hit in the small size range or a late miss in the large size range) by removing the outlying value from the data and re-running the analysis to assess its influence on the shape of the POD curve and confidence bound (if applicable). Both analysis results (with and without the outlying data) shall be included in the report along with a discussion of the impact to the POD curve and confidence bound (if applicable).

6.9.7 The analyst shall analyze any false call data and shall report the false call rate at the 50 %, 90 %, and 95 % level of statistical confidence. Acceptable false call rates shall be determined by the responsible engineer or by the customer.

6.9.7.1 The false call rate shall be defined as the number of false calls divided by the number of opportunities in the specimen set that do not contain discontinuity.

6.9.7.2 What constitutes a false call shall be clearly defined by the responsible engineer or by the customer.

6.9.7.3 What constitutes an opportunity in the specimen set that does not contain a discontinuity shall be clearly defined by the responsible engineer or by the customer.

6.9.7.4 The Clopper-Pearson binomial method for constructing confidence intervals for proportions should be used to calculate the false call rate at the 50 %, 90 % and 95 % level of statistical confidence. The Clopper-Pearson upper 100•(1-α)% confidence bound for p is:

$$P_U = \left\{ 1 + \frac{n - x}{(x+1) \cdot F_{(1-\alpha,\ 2x+2,\ 2n-2x)}} \right\}^{-1}$$

where $F_{(1-\alpha,\ 2x+2,\ 2n-2x)}$ is the F-statistics with degrees of freedom (2x+2, 2n–2x) and $P[F < F_{(1-\alpha,\ 2x+2,\ 2n-2x)}]=1-\alpha$. This method is consistent with that used in MIL-HDBK-1823A.

7. Report

7.1 At a minimum the following information about the POD analysis shall be included in the report.

7.1.1 The specimen standard geometry (for example, flat panels).

7.1.2 The specimen standard material (for example, Nickel).

7.1.3 Examination date.

7.1.4 Number of inspectors.

7.1.5 Type of inspection method (for example, line-of-site Level 3 Fluorescent Penetrant Inspection).

7.1.6 Any comments from the inspector(s) or test administrator.

7.1.7 The documented known induced discontinuity sizes.

7.1.8 Which discontinuities were found and which were missed.

7.1.9 Any false calls.

7.1.10 The selected link function.

7.1.11 The generalized linear model coefficients.

7.1.12 The variance-covariance matrix (if included in the software output).

7.1.13 A statement indicating that convergence was achieved.

7.1.14 The number of iterations needed to achieve convergence if included in the output.

7.1.15 A plot of the resulting POD curve and confidence bound (if applicable).

7.1.16 Specific results of interest as required by the analysis objective (for example, $a_{90/95}$).

7.1.17 A statement about the model diagnostic methods used and conclusions.

7.1.18 Any deviations from the POD examination procedure or standard POD analysis.

7.1.18.1 If the POD examination was re-administered, the original results and rationale for re-administration shall be documented in the report.

7.1.18.2 If a discontinuity is removed from the analysis, the specific discontinuity and rationale for removal shall be documented in the final report.

7.1.18.3 If the impact of outlying data was assessed, the results shall be included in the report along with an explanation.

7.1.19 Summary of false call analysis, including the following.

7.1.19.1 Definition of what constitutes a false call.

7.1.19.2 Definition of what constitutes an opportunity in the specimen set that does not contain a discontinuity.

7.1.19.3 False call rate at the 50 %, 90 %, and 95 % level of confidence.

7.1.20 Name of analyst and company responsible for the POD calculation.

8. Keywords

8.1 hit/miss analysis; Probability of Detection; POD; POD analysis; penetrant POD

ANNEX

(Mandatory Information)

A1. TERMINOLOGY

A1.1 Definitions:

A1.1.1 *binary response, n*—a response variable with only two possible outcomes.

A1.1.1.1 *Discussion*—The response from a POD examination on a manual fluorescent penetrant inspection system, for example, is binary. The discontinuity is either found or it is missed.

A1.1.2 *generalized linear model, n*—a class of statistical regression models, which uses a nonlinear link function to model a response variable whose distribution is a member of the exponential family (for example, exponential, normal, binomial) as a linear function of one or more predictor variables.

A1.1.2.1 *Discussion*—Generalized linear models are the basis for the hit/miss POD analysis method described in MIL-HDBK-1823A. The discontinuity size is the continuous predictor variable. The binary outcome is whether or not the discontinuity was found during the POD examination.

A1.1.3 *statistical confidence, n*—the long run frequency associated with the ability of the statistical method to capture the true value of the parameter of interest.

A1.1.3.1 *Discussion*—Statistical confidence is a probability statement about the statistical method used to estimate a parameter of interest—for example, the probability that the statistical method has captured the true capability of the inspection system. The opposite of statistical confidence can be equated to risk. For example, a statistical confidence level of 95 % implies a willingness to accept a 5 % risk of the statistical method yielding incorrect results—for example, there is a 5 % risk that the wrong conclusion has been drawn about the capability of the inspection system.

A1.1.4 a_{90}—the discontinuity size that can be detected with 90 % probability.

A1.1.4.1 *Discussion*—The value for a_{90} resulting from a POD analysis is a single point estimate of the true value based on the outcome of the POD examination. It does not account for variability due to sampling or inherent variability in the inspection system, which is always present.

A1.1.5 $a_{90/95}$—the discontinuity size that can be detected with 90 % probability with a statistical confidence level of 95 %.

A1.1.5.1 *Discussion*—The value for a_{90} resulting from a POD analysis is an estimate of the true a_{90} based on the outcome of the POD examination. If the examination were repeated, the outcome is not expected to be exactly the same. Hence the estimate of a_{90} will not be the same. To account for variability due to sampling, a statistical confidence bound with a 95 % level of confidence is applied to the estimated value for a_{90} resulting in an $a_{90/95}$ value. POD is still 90 %. The 95 % refers to the ability of the statistical method to capture (or bound) the true a_{90}. That is, if the examination were repeated over and over under the same conditions, the value for $a_{90/95}$ will be larger that the true a_{90} 95 % of the time. In practice the POD examination will be conducted once. Using a 95 % confidence level implies a 95 % chance that the $a_{90/95}$ value bounds the true a_{90} and a 5 % risk that the true a_{90} is actually larger than the $a_{90/95}$ value.

A1.1.6 $a_{90/50}$—the discontinuity size that can be detected with 90 % probability with a statistical confidence level of 50 %.

A1.1.6.1 *Discussion*—Using a 50 % confidence level implies a 50 % chance that the $a_{90/50}$ value bounds the true a_{90} and a 50 % risk that the true a_{90} is actually larger than the $a_{90/50}$ value. Given this, $a_{90/50}$ is really the same as a_{90}.

A1.1.7 *statistical confidence bound*—a one-sided or two-sided bound around a single point estimate representing the variability due to sampling.

A1.1.7.1 *Discussion*—In accordance with the formula in MIL-HDBK-1823A, $a_{p/c}$ is a one-sided upper confidence bound on a_p. $a_{p/c}$ represents how large the true a_p could be given the statistical uncertainty associated with limited sample data. In general, statistical uncertainty decreases as sample size increases. That is, given an infinite amount of data (for example, an infinite number of flaw sizes adequately distributed across a POD specimen set), $a_{p/c}$ will approach a_p because the statistical uncertainty goes away. It is important to note that a statistical confidence bound on a_p only accounts for variability due to sampling. It does not account for inherent process variability. In order to capture inherent process variability, a tolerance bound should be used. As opposed to a confidence bound, a tolerance bound will always differ from the point estimate because process variability cannot be eliminated by increasing the sample size.

APPENDIX

(Nonmandatory Information)

X1. POD ANALYSIS PROCESS

X1.1 Fig. X1.1 shows a flowchart of POD Analysis for hit/miss data.

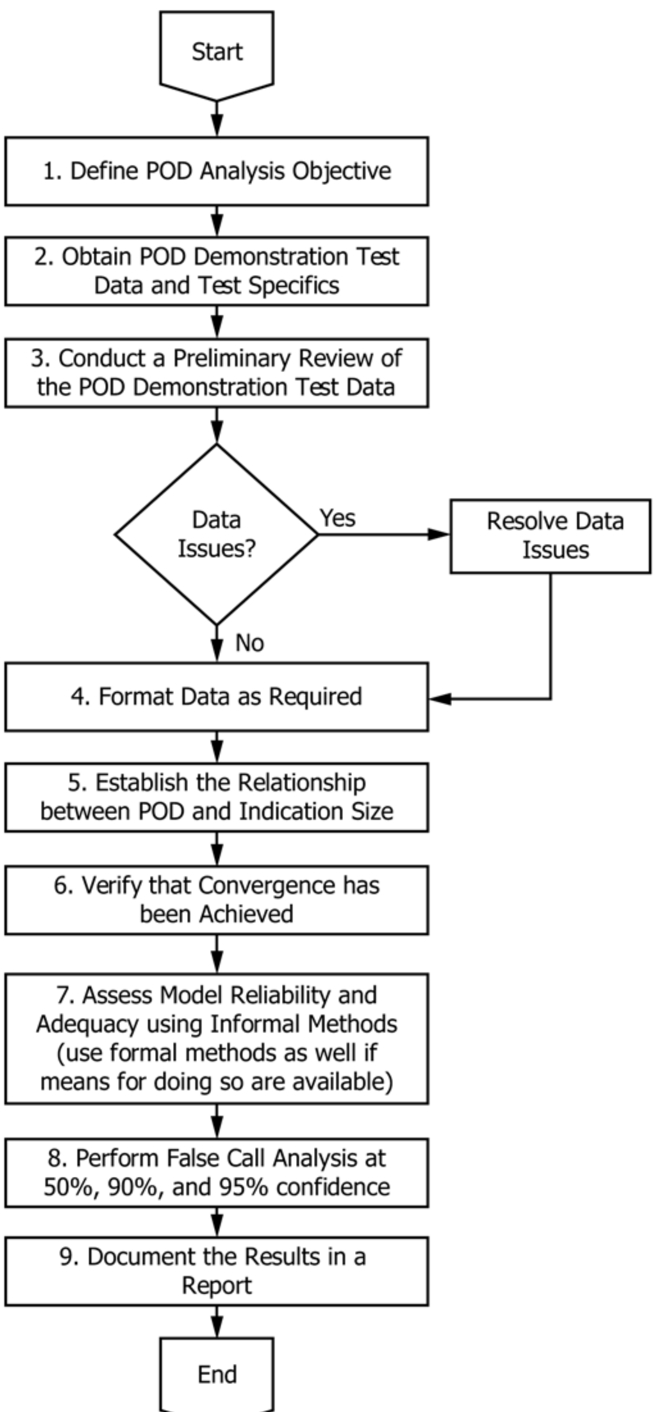

FIG. X1.1 Flowchart of POD Analysis for Hit/Miss Data

X1.2 Additional commentary on the POD analysis process as illustrated in Fig. X1.1 and its significance.

X1.2.1 In general, the objective of a POD analysis is to determine the relationship between discontinuity size and POD. Based on the established relationship, the objective may be to determine the discontinuity size that can be detected with a given probability *p* and specified statistical confidence level *c*, denoted $a_{p/c}$. It is important for the analyst to have a clear understanding of the specific analysis objective prior to performing the analysis.

X1.2.2 The model coefficients do not have a closed form solution. As such, an iterative numerical procedure is required to solve the system of equations from which the estimates of the model coefficients are derived. The procedure iterates until a convergence criterion is met, at which point estimates of the model coefficients are obtained from the last iteration. The analysis results are not valid unless the convergence criterion is met. Even if the analysis software outputs model information, the results shall not be used if the convergence criterion has not been met. Prior to performing the analysis, a preliminary review of the hit/miss data resulting from the POD examination can reveal whether or not failure to meet the convergence criteria may be an issue. If there is no overlap between misses and hits when the discontinuity sizes are sorted in ascending order, then the convergence criteria will not be met. If the responses are all misses or all hits, then the convergence criteria will not be met.

X1.2.3 Prior to performing the POD analysis, the analyst shall format the data as required by the software used to conduct the analysis.

X1.2.3.1 A hit is typically coded as a 1.

X1.2.3.2 A miss is typically coded as a 0.

X1.2.3.3 For some software this may require the analyst to perform a transformation of the predictor variable prior to running the analysis. For example, the natural log of discontinuity size is often used as the predictor variable since it forces the POD curve to pass through the origin, which is interpreted as zero POD for a discontinuity of size 0. If the natural log of discontinuity size is used as the predictor variable, then the analyst may need to create a new variable column for the natural log of discontinuity size prior to running the analysis.

X1.2.4 POD specific software or statistical software is commonly used to perform an analysis on hit/miss data in order to establish a functional relationship between POD and discontinuity size. Though the software performs the complex calculations, it does not check the validity of analysis inputs or outputs. The analyst is responsible for ensuring that the analysis inputs (for example, data, model formulation) are correctly specified and that the underlying model assumptions hold. Treating the software as a "black box" can lead to seriously misleading conclusions about the inspection capability of the system. Hence, it is critical that the practitioner have a basic understanding of the complete analysis process, including the underlying statistical methods and techniques for validating the results.

X1.2.5 Generalized linear models are the traditional statistical models used to describe the relationship between continuous variables (such as discontinuity size) and binary outcomes (such as hit or miss). For binary outcomes, the form of a generalized linear model with a single predictor variable is $g(y) = b_0 + b_1 \bullet x$, where x is the continuous predictor variable, b_0 is the intercept, b_1 is the slope, y is the binary response variable, and $g(y)$ is the function that "links" the binary response with the predictor variable. This model is the basis for the hit/miss analysis method as described in MIL-HDBK-1823A. Note that there can be one or more predictor variables in a generalized linear model. However, for POD applications there is often only a single predictor variable—discontinuity size or a function of discontinuity size (such as the natural log) since that is typically the only known physical characteristic of the discontinuity. In general, a generalized linear model is the appropriate statistical model for relating hit/miss data and flaw size since it restricts POD predictions to be between 0 and 1.

X1.2.6 The procedure states that if more than twenty iterations were needed to reach convergence, the model may not be reliable. This criterion was selected to be consistent with several well known software packages. The criterion of twenty is used in Minitab statistical software and PODv3. *mh1823* uses a criterion of twenty-five.

X1.2.7 Other methods exist for determining the demonstrated POD for hit/miss data. However, caution should be used with methods that yield only a point estimate and not an entire POD curve. With these methods it is not possible to assess the affect that size has on POD.

X1.2.8 If available, the analyst should also use the standard statistical model diagnostic methods recommended for generalized linear models for binary outcomes. Most statistics software packages with generalized linear modeling capabilities have the standard model diagnostic methods built-in. A description of the methods can be found in most statistics text books that cover categorical data analysis or general linear models. However, the diagnostic methods should be used with caution as they may not be reliable under certain conditions.

X1.2.9 The method for constructing a confidence interval for a binomial proportion proposed by Clopper and Pearson (1934) is an exact method for estimating confidence bounds for a proportion p based on x "failures" in a sample size of n where p is estimated as x/n. The upper 100•(1–α)% confidence bound for p is:

$$P_U = \left\{ 1 + \frac{n - x}{(x+1) \cdot F_{(1-\alpha,\ 2x+2,\ 2n-2x)}} \right\}$$

where $F_{(1-\alpha,\ 2x+2,\ 2n-2x)}$ is the F-statistics with degrees of freedom (2x+2, 2n–2x) and $P[F < F_{(1-\alpha,\ 2x+2,\ 2n-2x)}] = 1-\alpha$. MIL-HDBK-1823A refers to a false call analysis as "analyzing hit/miss noise." The MIL-HDBK-1823A false call analysis example supposes no false calls out of 150 opportunities, that is, $x = 0$, $n = 150$. For the 50 %, 90 %, and 95 % confidence levels, the false call rate at the 50 %, 90 %, and 95 % is calculated as follows.

X1.2.9.1 50 % confidence:

$$P_U = \left\{ 1 + \frac{150 - 0}{(0+1) \cdot F_{(0.50,\ 2\cdot(0)+2,\ 2\cdot(150)-2\cdot(0))}} \right\}^{-1} = \left\{ 1 + \frac{150}{0.695} \right\}^{-1} = 0.0046$$

X1.2.9.2 90 % confidence:

$$P_U = \left\{ 1 + \frac{150 - 0}{(0+1) \cdot F_{(0.90,\ 2\cdot(0)+2,\ 2\cdot(150)-2\cdot(0))}} \right\}^{-1} = \left\{ 1 + \frac{150}{2.320} \right\}^{-1} = 0.0152$$

X1.2.9.3 95 % confidence:

$$P_U = \left\{ 1 + \frac{150 - 0}{(0+1) \cdot F_{(0.95,\ 2\cdot(0)+2,\ 2\cdot(150)-2\cdot(0))}} \right\}^{-1} = \left\{ 1 + \frac{150}{3.026} \right\}^{-1} = 0.0198$$