PRACTICAL APPROACH TO CALCULATING PROBABILITY OF FALSE ACCEPT FOR DECISION RULES IN CONFORMITY ASSESSMENT

Manuscript Info

……………………..

Manuscript History

Received: xxxxxxxxxxxxxxxx

Final Accepted: xxxxxxxxxxxx

Published: xxxxxxxxxxxxxxxx

Abstract

………………………………………………………………

The decision rule, as defined in the ISO/IEC 17025 standard, is a rule that describes how measurement uncertainty is taken into account when stating conformity with a specified requirement (1). Essentially, it is a rule which is based on the idea of how much risk should be accepted when starting to use an item. In other words, the decision rule is fundamentally concerned with the question of what is the acceptable level of risk. The concept of risk here represents the probability of accepting an item that actually doesn't conform with specification (false accept) or the probability of rejecting an item that actually conforms with specification (false reject). Risk, by its nature, can never be reduced to zero, it can only be close to zero. Due to its consequences, false accept of an item is a more undesirable situation than false reject. Therefore, when a facility decides that an item is suitable for use, the acceptable level of probability of false accept which is also known as specific risk should be determined (2). This acceptable level is the facility's choice and may vary depending on the business objective.

The aim of this article is to analyze a conformity assessment of an item based on sample calibration results by using binary statement (pass-fail) decision rule, to calculate acceptable specific risk levels when starting to use an item and to serve as a supplementary document to conformity assessment documents.

……………………………………………………………………………………………….....

Introduction

Conformity assessment is an activity used in testing, inspection and calibration processes and is used to verify whether products, materials, services and systems meet the expected conditions within the framework of standards, regulations and legal requirements. This process, which aims to have mission of ensuring consumer confidence, quality of life and safety, plays a decisive role in the global economy as it involves the acceptance or rejection of items. Risk analysis has a direct impact on business decisions and financial and reputational outcomes (3).

First, the terminology used in this document must be defined. Although all conformity assessment guidance documents express the same thing, different terminology is used for the same meanings. For example, the user determines the lower and/or upper limit for conformity assessment. The limit here is expressed with different expressions in the documents (2)(3)(4)(5)(6)(7)(8). The expressions used in the guidance documents are given in Table 1.

ILAC-G8:09/2019	Tolerance Limit (Specification Limit)	Tolerance Interval (Specification Interval)	Acceptance Limit	Acceptance Interval
BIPM JCGM 106:2012	Tolerance Limit	Tolerance Interval	Acceptance Limit	Acceptance Interval
OIML G 19, 2017	Maximum Permissible Error (MPE) Limit	Conformance Zone	MPE Guard Band	Shifted Conformity Boundaries
EUROLAB. *Technical Report No.01/2017*	Tolerance Limit	Tolerance Interval	Acceptance Zone Limit	Acceptance Zone
UKAS LAB 48, 2021	Tolerance Limit	Tolerance Interval Specification	Acceptance Limit	Acceptance Interval
ASME B89.7.3.1-2001	Specification Limit	Specification Zone	Stringent Acceptance Zone Limit	Acceptance Zone
Eurachem/CITAC Guide, 2021	Specification Limit	Specification Zone	Acceptance Limit	Acceptance Zone

Table 1: Terminology of Conformity Assessment Guidelines

This article will use the terms "tolerance limit (specification limit), tolerance interval (specification interval), acceptance limit and acceptance interval."

If we want to evaluate the conformity of an item with a specification according to the measurement result and measurement uncertainty, we should basically consider 4 types of possible situations (2)(4)(9).

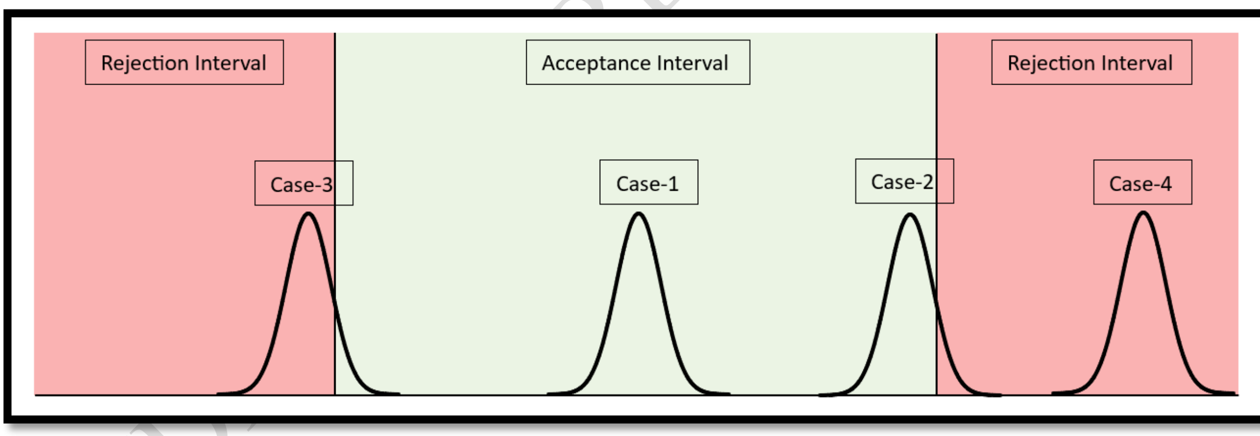

Figure 1: Conformity Assessment Criteria

Based on the case shown in Figure 1, where the measurement result must be given with an expanded measurement uncertainty value U (k = 2, interval with ≈95 % level of confidence);

- In case 1, we can say that the item is accepted as conforming by taking measurement uncertainty into account.
- In case 4, we can say that the item is accepted as non-conforming by taking measurement uncertainty into account.
- In cases 2 and 3, we must use mathematical models to decide whether the item is accepted as conforming or non-conforming.

In case 2, even if a large section of the result appears as conforming, the non-conforming section must be evaluated along with the risk factor. The percentage of the non-conforming section needs to be calculated. This calculated value will give us the probability of false accept which is commonly abbreviated as PFA. The facility will decide whether the item is accepted as conforming or non-conforming for use based on an acceptable level of specific risk. For example, if the facility's maximum target PFA value is 2 % and the calculated PFA value based on the measurement result is less than 2 %, the item will be considered as conforming. Otherwise, it will be considered as non-conforming.

The importance of PFA is best explained in NASA's 1342 Metrology document (10):

"Certain negative consequences may arise because of false accepts. Test process false accepts can lead to reduced end-item capacity or capability, mission loss or compromise, loss of life, damaged corporate reputation, warranty expenses, shipping and associated costs for returned items, loss of future sales, punitive damages, legal fees, etc."

The calculation of the PFA value is related to the z-score which is a subject of the science of statistics (2). If we call the measured value x, the arithmetic mean of the data μ and the standard deviation of the data σ;

$$Z = \frac{x - \mu}{\sigma}$$

If we call our variable value x and assume it is between -∞ and +∞, our probability function is as follows (11):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

To find the area under the probability curve, we use integration and this area is equal to 1 (100 %).

$$\int_{-\infty}^{+\infty} f(x)\mathrm{dx} = 1$$

That probability density function is known as normal distribution. If we standardize that function as the arithmetic mean of the data μ=0 and the standard deviation of the data σ=1, then it would be called standard normal distribution (also expressed as z-distribution). Standard normal distribution assumes that population standard deviation is known. Based on statistics, standard deviation would be obtained if the measurement were repeated an infinite number of times (11). Since it is not possible, t-distribution (also expressed as student's distribution) which is based on the sample standard deviation is used.

t-distribution is a distribution which is similar to z-distribution. Basically, *t* approaches and gets similar to *z* when sample size (*n*) rises and they are equal when *n* is infinite (see Figure 2)(12). When the t-distribution is used, we use something called degrees of freedom (df=n-1). When the degrees of freedom (df) are equal to or greater than 30, *t* gets close to *z* and *z* can be used in place of *t* for that sample size. While performing calibration, since we do not usually take ≈30 repeatability measurements for one calibration point, the t-score representing the t-distribution should be used rather than the z-score representing the z-distribution. However, according to metrology practices and some EURAMET, DKD guidance documents (e.g. EURAMET cg-18 appendix B), 10 measurement repeatability observations are accepted as sufficient reliability.

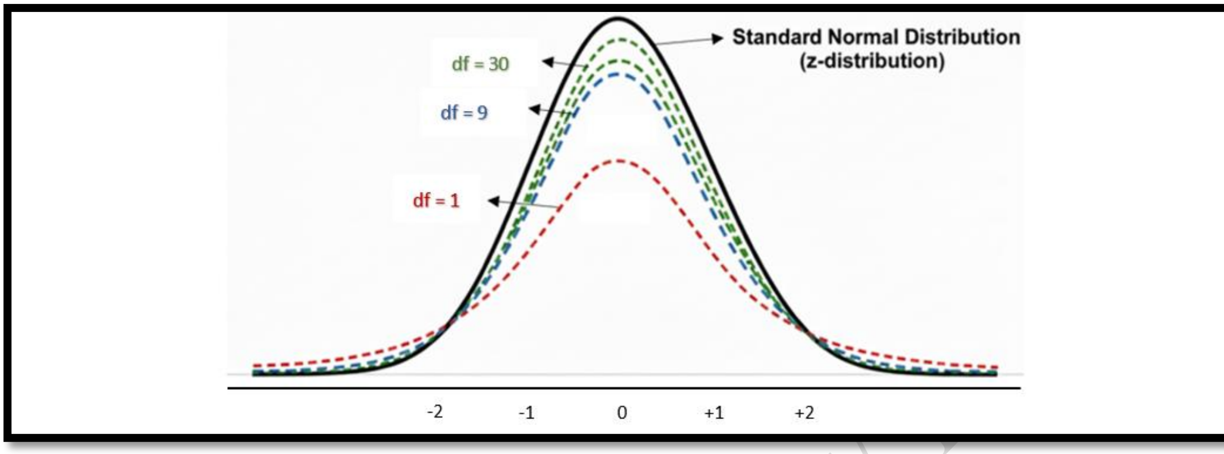

Figure 2: z-distribution vs. t-distribution

Sufficient reliability depends on the degrees of freedom and when the number of repeatability measurement isn't less than 10, sufficient reliability can be assumed (13)(14). Therefore, we can use z-score and, in metrology world, we name z-score as the coverage factor which is abbreviated as k_p. The most commonly used values are as follows (11)(15):

- k_p = 1 provides an interval with a level of confidence of 68,3 %
- k_p = 1,96 provides an interval with a level of confidence of 95 %
- k_p = 2 provides an interval with a level of confidence of 95,45 %
- k_p = 3 provides an interval with a level of confidence of 99,73 %

As it is described in BIPM GUM guidance; in most cases, it is difficult to distinguish between an interval with a level of confidence of 95 % and 96 % (11). The reason is that it is only an approximation because coverage probability of 95 % means that one chance in 20 that the value of the measurand *Y* lies outside the interval and probability of 96 % means that one chance in 25. For metrology practices, it is often adequate that taking k_p= 2, U=2u which defines an interval with a level of confidence of approximately 95 %.

Interval with levels of confidence (coverage probability) associated with the coverage factors are determined (see Figure 3). It is assumed that the measurement result is within the region which is calculated based on the z-score.

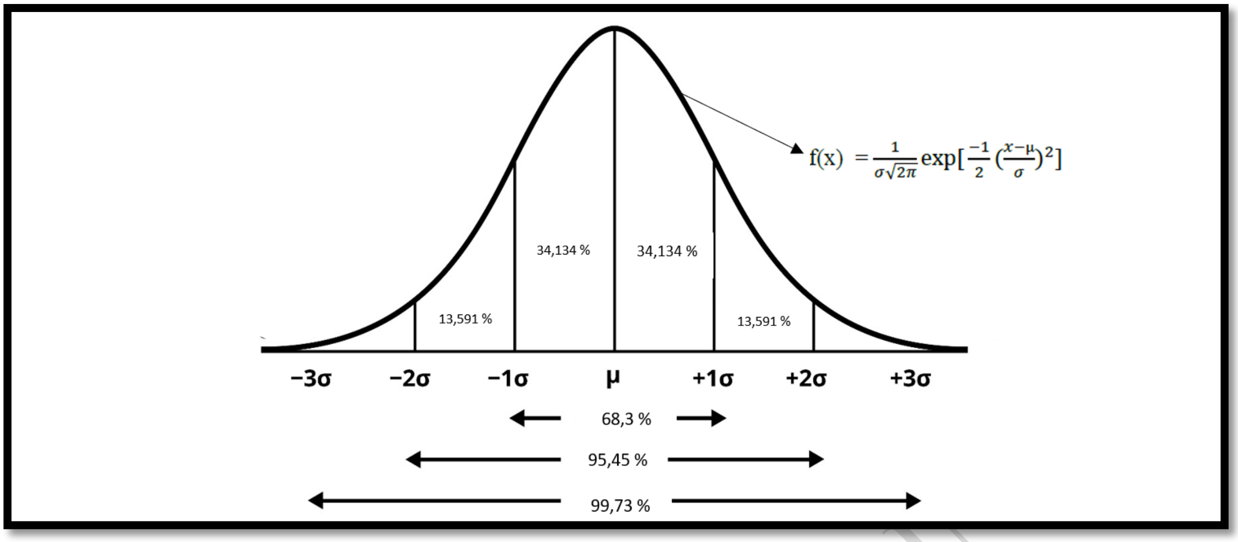

Figure 3: Coverage Probabilities

When the decision rule is applied, in other words, when we take the measurement uncertainty into account in the measurement result, the probability of results that are close to the tolerance limit (specification limit) are more likely to be considered as non-conforming. To prevent this, we need to subtract the value of w = rU from the tolerance limit value. The symbol 'w' here is considered a type of safety factor and is called the guard band. Thus, the tolerance limit is replaced by the acceptance limit and the tolerance interval is replaced by the acceptance interval (2)(4). In this formula, the value of r indicates how we incorporate uncertainty into the result (the ratio according to uncertainty). For example, if we subtract the uncertainty from the tolerance limit by r = 1 (w = U), maximum PFA is 2,275 %. The maximum risk values according to the selected r value are given in Table 2.

In the UKAS LAB 48 document, the guard band is shown as $w = k_w u$. In the ILAC-G8 document, it is shown as w = rU (4)(6). Although both essentially express the same concept, it should be noted that the coefficients differ due to the approach. When the expanded uncertainty U is evaluated as two times standard uncertainty (U = 2u), the coefficients in Table 2 can be used for the maximum PFA value.

	UKAS LAB48	ILAC-G8
PFA_max (%)	k_w	$r = k_w / 2$
0,000001%	5,6120	2,8060
0,1000%	3,0902	1,5451
1,0000%	2,3263	1,1632
2,0000%	2,0537	1,0269
2,2750%	2,0000	1,0000
2,5000%	1,9600	0,9800
4,5500%	1,6901	0,8451
5,0000%	1,6449	0,8224
10,0000%	1,2816	0,6408
20,0000%	0,8416	0,4208
30,0000%	0,5244	0,2622
40,0000%	0,2533	0,1267
50,0000%	0,0000	0,0000

Table 2: Guard Band Coefficients

For example, if our expanded uncertainty value is U = 2u and we aim for a maximum PFA value of 5 % in our measurement results according to a calibration report;

- We subtract 0,8224 x U value (or in other way to express as 1,6449 x u value) from our tolerance limits and determine whether our measurement results are inside within this interval (acceptance interval). To clarify something in this example; the value of 0,8224 x U is equal to the guard band w. In other words, we are not actually reducing the uncertainty value from 1U to 0,8224U. The uncertainty value is still 1U but we accept the guard band as 0,8224U to ensure a maximum PFA of 5 %.It should also be pointed out that if the measurement result approaches the reference point,probability of false accept becomes lower than PFAmax.

Current approaches in conformity assessment include usually simple acceptance (w = 0), binary statement (pass-fail), and non-binary statement (pass-conditional pass-fail-conditional fail). The binary statement rule is evaluated in this article. To apply the binary statement rule, the information required from the user is the tolerance value and the maximum PFA value. According to this approach, statement of conformity is reported as pass if the measurement result is less than acceptance limit, otherwise a fail decision is given.

Note: In this article, probability of false accept (PFA) refers to specific risk. There is also another risk type which is known as global risk. To put it simply, specific risk deals with individual measurement results, while global risk deals with average measurement results. Global risk is not the subject of this article.

Material And Method

We assume that the thermometer is calibrated according to Euramet cg-8 (16) and general approach on its uncertainty is applied according to the BIPM GUM and EA-4/02 uncertainty documents (11)(17). An example calibration certificate is given in Table 3 below. We also assume that the user has set PFA< 2,275 % and tolerance ±2 °C. In this case, if the PFAmax value is as 2,275 %, we assume that the guard band w is equal to 1U according to the coefficients in Table 2.

Reference	Measured	Deviation	Expanded Uncertainty (U, k=2)
100	101,5	1,5	0,25
200	201,5	1,5	0,5
300	301,5	1,5	1
400	401,5	1,5	1,5

Table 3: Sample Thermometer Calibration Results (in degree Celsius)

Based on the calibration results, we determine the safe zone, or acceptable interval, by subtracting the guard band value from the tolerance limit value. For example, if we calculate for a reference value of 100 °C;

- Upper tolerance limit is 102 °C
- Lower tolerance limit is 98 °C
- For PFAmax of 2,275 %, the guard band w value is 1U (Table 2).
- Upper acceptance limit is 102 °C – 1 x 0,25 °C = 101,75 °C
- Lower acceptance limit is 98 °C + 1 x 0,25 °C = 98,25 °C

According to the calculation above, the values for the other points are given in Table 4.

Reference	Measured	Deviation	Expanded Uncertainty (U, k=2)	Lower Specification	Lower Acceptance Limit	Upper Acceptance Limit	Upper Specification
100	101,5	1,5	0,25	98,000	98,250	101,750	102,000
200	201,5	1,5	0,5	198,000	198,500	201,500	202,000
300	301,5	1,5	1	298,000	299,000	301,000	302,000
400	401,5	1,5	1,5	398,000	399,500	400,500	402,000

Table 4: Tolerance Limit and Acceptance Limit (in degree Celsius)

If we make a conformity assessment according to PFAmax of 2,275 % and tolerance ±2 °C;

- The measured value for point-1 is within the acceptance interval [98,25 °C…101,75 °C]. The statement of conformity is reported as pass.
- The measured value for point-2 is in acceptance limit [198,5 °C…201,5 °C]. The statement of conformity is reported as fail because there is no equality in the statement of PFAmax of 2,275 %.
- The measured value for point-3 is outside the acceptance interval [299 °C…301 °C]. The statement of conformity is reported as fail.
- The measured value for point-4 is outside the acceptance interval [399,5 °C…400,5 °C]. The statement of conformity is reported as fail.

For a binary statement (pass-fail), it is sufficient to say that the measurement results are acceptable if they are within the defined acceptance interval, otherwise unacceptable. This is because we are setting an upper limit for risk according to the selected PFAmax value. In our example, the meaning of the guard band w value of 1U is that if the measurement results are within the acceptance interval, which replaces the tolerance interval, the maximum probability of false accept is obtained as 2,275 %. In addition, if we set the guard band w value as 0,8224U instead of 1U, the maximum probability of false accept would be obtained as 5% (see Table 2). Furthermore, if we want to see the point-based PFA value in the results, we need to perform a conformance probability calculation. The conformance probability, P_C value, is calculated according to the following formula(2):

$$P_C = \Phi ((T_U - y) / u) - \Phi ((T_L - y) / u)$$

P_C = Conformance Probability, T_L = Lower Tolerance Limit, T_U = Upper Tolerance Limit, y = Measured Value, u = Standard Uncertainty (k=1),

The P_C formula is calculated using Microsoft Excel as follows:

- To find the value of $\Phi ((T_U - y) / u)$;

=NORM.DIST(upper tolerance limit (T_U); average measured value (y); combined standard uncertainty (u, $k=1$); TRUE)

- To find the value of $\Phi ((T_L - y) / u)$;

=NORM.DIST(lower tolerance limit (T_L); average measured value (y); combined standard uncertainty (u, $k=1$); TRUE)

The results calculated according to the P_C formula are given in Table 5. The calibration results are visualized in Figures 4, 5, 6, and 7.

Reference	Measured	Deviation	Expanded Uncertainty (U, k=2)	Lower Specification	Lower Acceptance Limit	Upper Acceptance Limit	Upper Specification	Conformance Probability	Non-Conformance Probability
100	101,5	1,5	0,25	98,000	98,250	101,750	102,000	99,997%	0,003%
200	201,5	1,5	0,5	198,000	198,500	201,500	202,000	97,725%	2,275%
300	301,5	1,5	1	298,000	299,000	301,000	302,000	84,134%	15,866%
400	401,5	1,5	1,5	398,000	399,500	400,500	402,000	74,751%	25,249%

Table 5: Conformance Probability Results

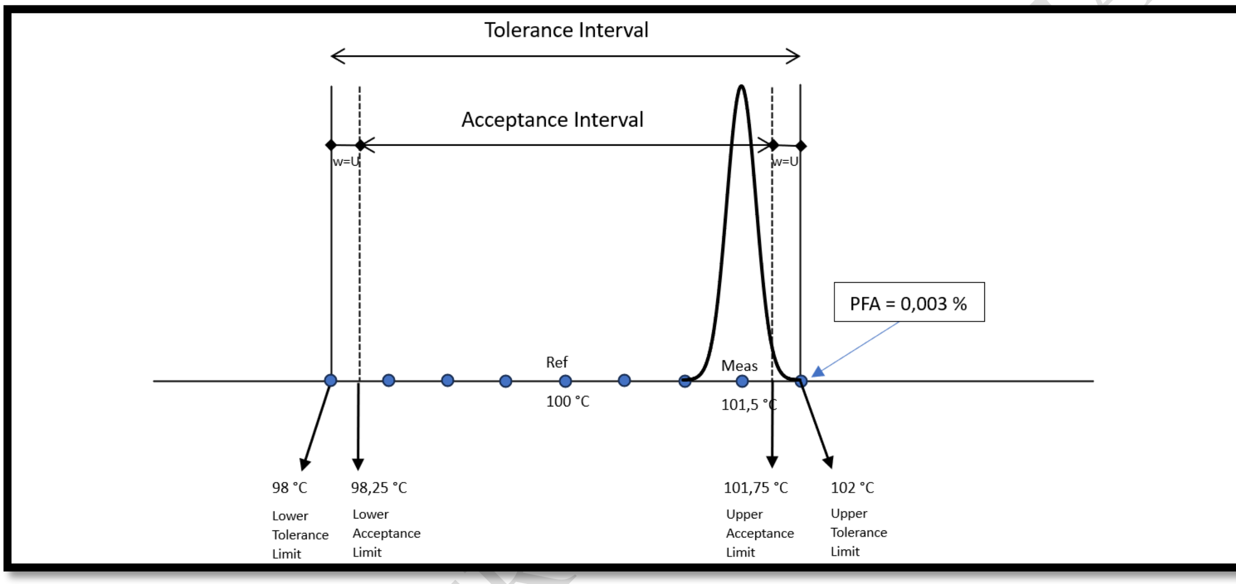

Figure 4: Calibration Point-1 Results

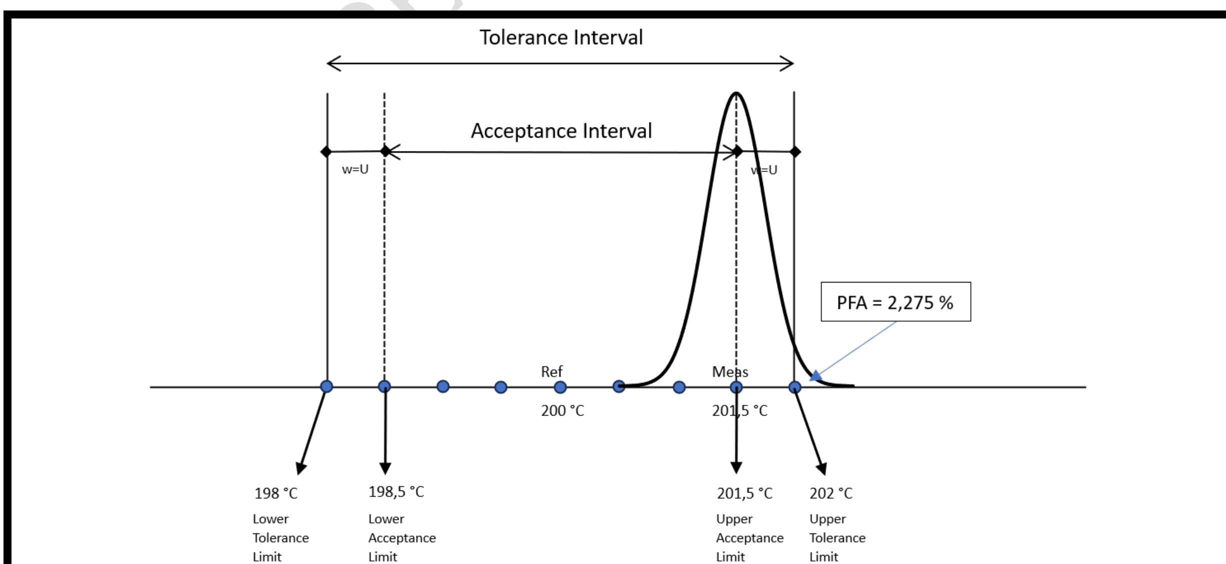

Figure 5: Calibration Point-2 Results

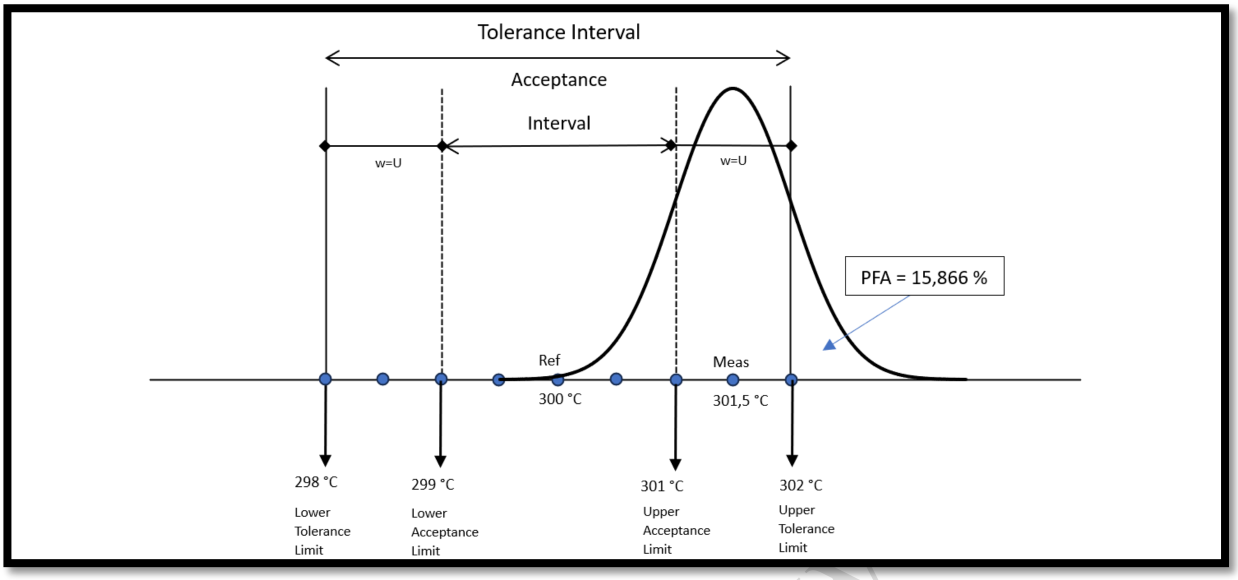

Figure 6: Calibration Point-3 Results

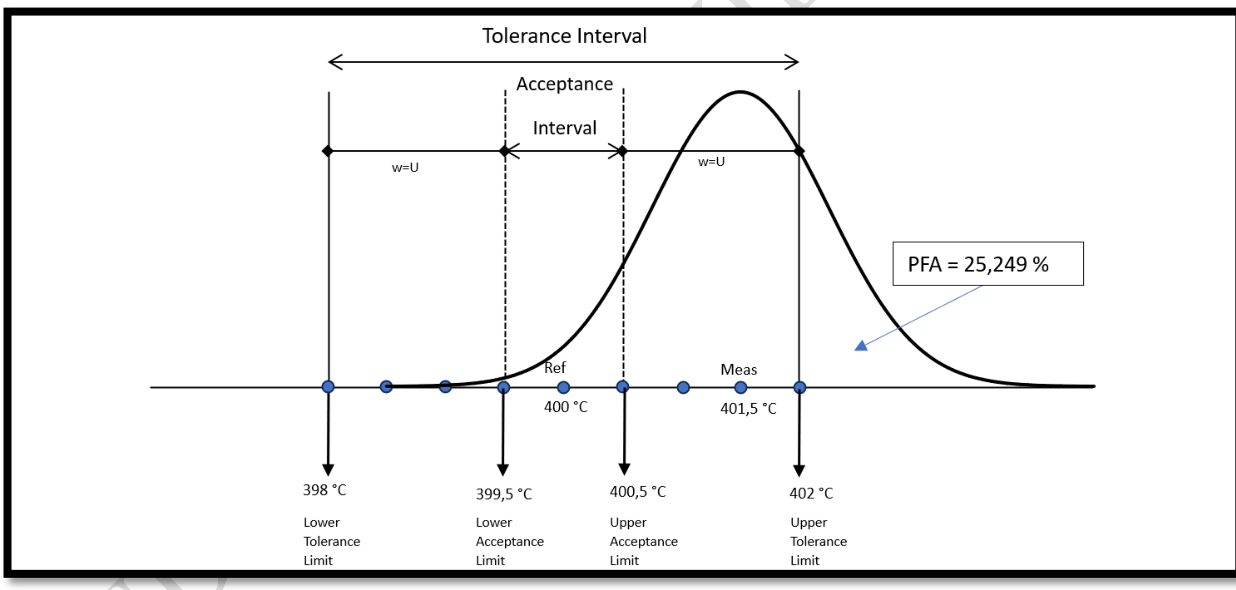

Figure 7: Calibration Point-4 Results

Discussion And Conclusion

It is a common misunderstanding that stating only tolerance would be enough to make conformity assessment. Stating desired PFA is not often considered. For our scenario in this article, the user aims PFA < 2,275 %. As a result, the first point conforms with specification, while the other points do not. If the user aims PFA < 10 %, second point would conform with specification too. Consequently, it is obvious that the conformity assessment not only depends on tolerance but also desired PFA value.

ILAC-G8 tells us that guard band w is equal to r x U. In this context, r is related to probability of false accept. If we choose r is 1, which means w = 1 x U, that would be equal to PFAmax of 2,275 %. In other words, if measurement values are inside of acceptance interval (specification limit minus w), the PFAmax value would be 2,275 %. The key point is that 2,275 % value would occur in exact acceptance limit values. If the results are less than acceptance limits, risk would be lower. In addition, if we choose r is 0,5 which means w = 0,5 x U, that would be equal to PFAmax of 15,866 % and if r is zero, PFAmax value is 50 % which means measurement could be out of tolerance as 50 % of probability. These situations are shown in figure 8, 9, 10.

To summarize PFAmax value, if the measurement is exactly at the acceptance limit, the area outside the distribution tail area will correspond exactly to the desired maximum risk (18).Again, it should be pointed out that if the measurement result approaches the reference point,probability of false accept becomes lower than PFAmax.

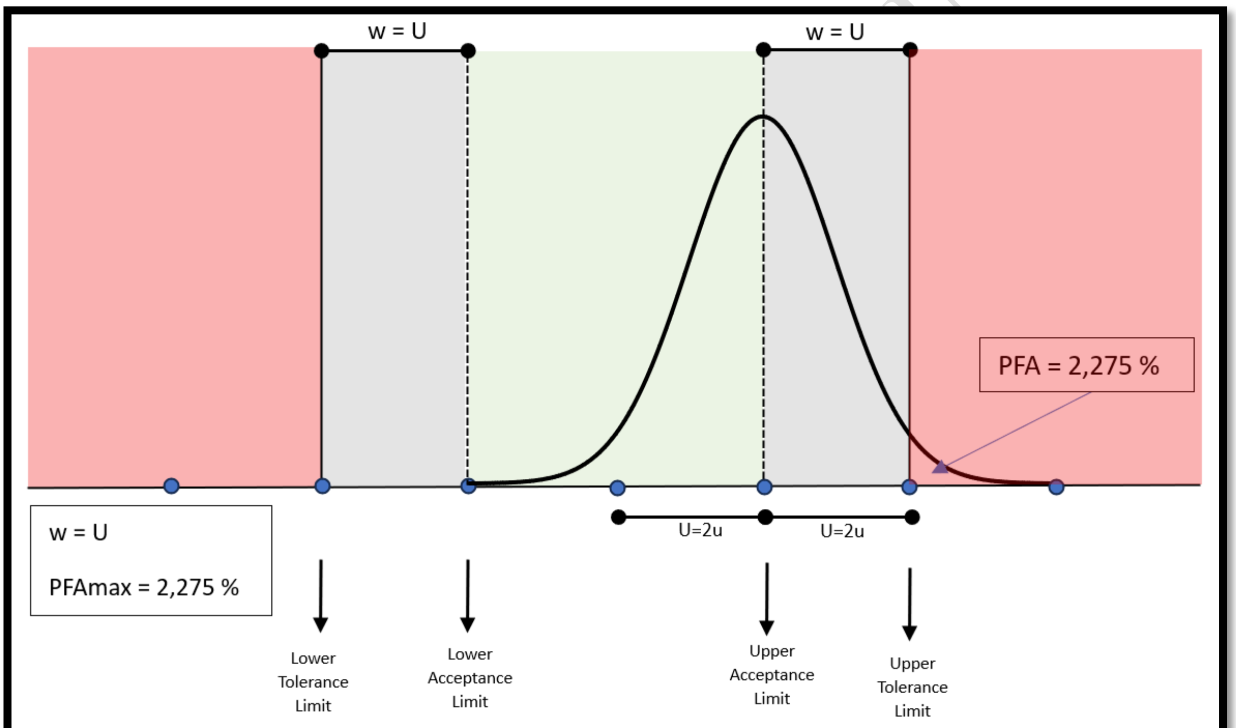

Figure 8: w = 1U

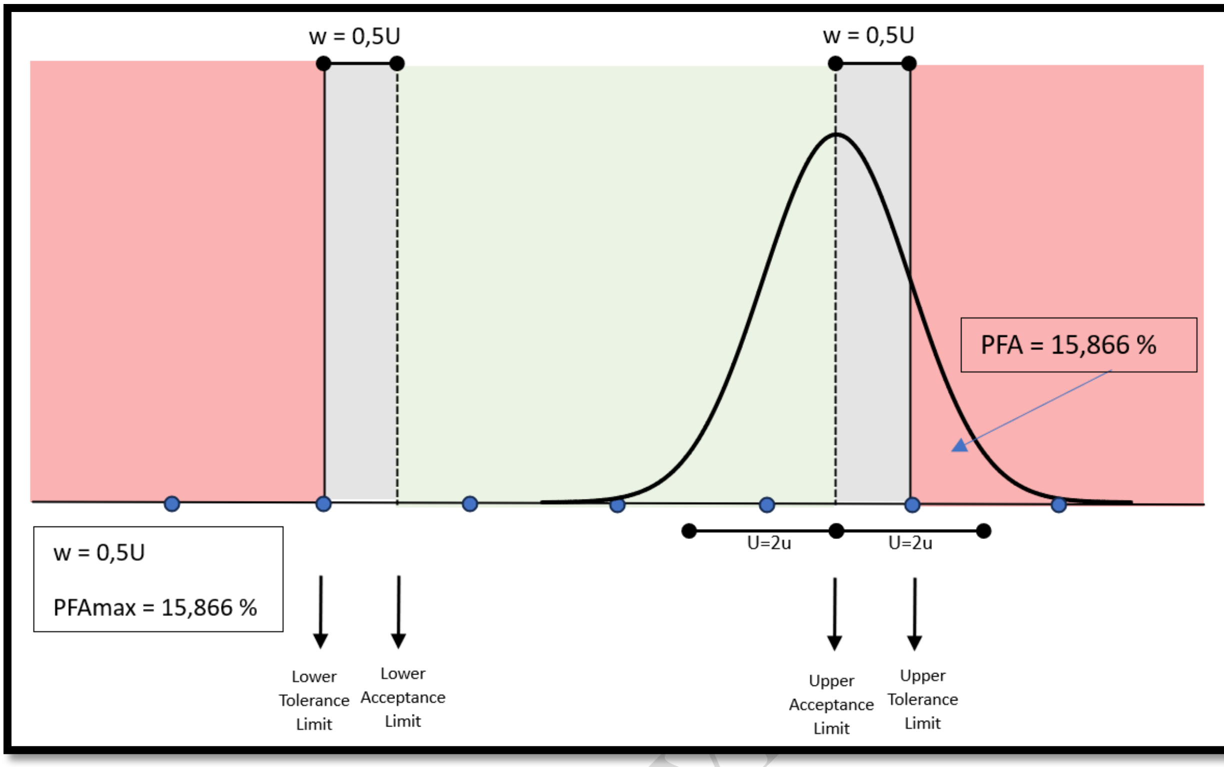

Figure 9: w = 0,5U

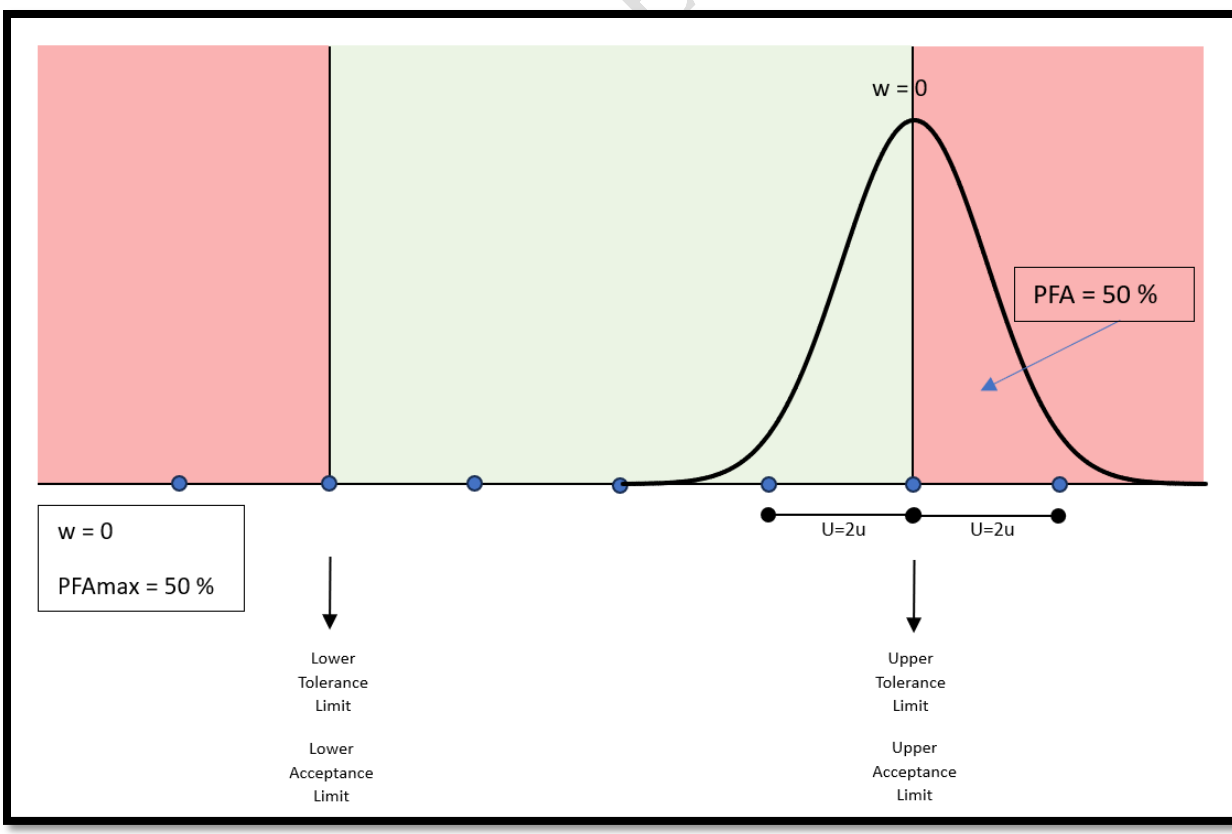

Figure 10: w = 0

BIPM GUM document points out that the evaluation of uncertainty is neither a routine task nor a purely mathematical one(11), rather, it is fundamentally a matter of technical expertise, critical judgment, and intellectual integrity.

Based on that idea, in our calibration scenario, if we evaluate the results according to only tolerance, all points would be considered as conforming. The key point is that when we add the specific risk value to our equation, in other words when we take measurement uncertainty into account for conformity, only the first point is considered as conforming, while the others are not. Therefore, users should ask simply these questions. What is my tolerance? What is my risk associated with the probability of false accept? Should my statement be binary statement (pass-fail) or non-binary statement (pass-conditional pass-fail-conditional fail)? Based on these questions, results would be accepted as conforming or non-conforming.

Source

1. **ISO/IEC 17025:2017.***General requirements for the competence of testing and calibration laboratories.* 2017.

2. **BIPM, Joint Committee for Guides in Metrology.***JCGM 106:2012, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment.* 10/2012.

3. **EUROLAB.***Technical Report No.01/2017, Decision Rules Applied to Conformity Assesment.* November 2017.

4. **International Laboratory Accreditation Cooperation.***ILAC-G8:09/2019, Guidelines on Decision Rules and Statements of Conformity.* s.l. : 09/2019, 09/2019.

5. **OIML, INTERNATIONAL ORGANIZATION OF LEGAL METROLOGY.***OIML G 19 Edition 2017 (E), The role of measurement uncertainty in conformity assessment decisions in legal metrology.* 2017.

6. **UKAS.***LAB 48, Decision rules and statements of conformity.* Edition 5 July 2024.

7. **The American Society of Mechanical Engineers.***ASME B89.7.3.1-2001, GUIDELINES FOR DECISION RULES: CONSIDERING MEASUREMENT UNCERTAINTY IN DETERMINING CONFORMANCE TO SPECIFICATIONS.* March 18 2002.

8. **Eurachem / CITAC Guide.***Use of Uncertainty Information in Compliance Assessment.* Second edition (2021).

9. **Arpacık, Hakan.***CONFORMITY ASSESSMENT OF TEMPERATURE MEASURING DEVICES USED IN FORENSIC.* Istanbul : İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA, ADLİ TIP VE ADLİ BİLİMLER ENSTİTÜSÜ, June 2019.

10. **National Aeronautics and Space Administration (NASA).***Reference Publication 1342, Metrology — Calibration and Measurement Processes Guidelines.* June 1994.

11. **BIPM, Joint Committee for Guides in Metrology.***JCGM 100:2008 (GUM 1995 with minor corrections) Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement, First Edition,.* September 2008.

12. **Gatzeva, Mariana.***Simple Stats Tools.*

13. **EURAMET.***Calibration Guide – 18, Calibration of Non-Automatic Weighing Instruments.* 11/2015.

14. **Deutscher Kalibrierdienst (DKD).***DKD-R 5-1 Calibration of resistance thermometers.* 11/2023.

15. **The University of Arizona.***STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.*

16. **EURAMET.***Calibration Guide – 8, Calibration of Thermocouples.* 02/2020.

17. **European Accreditation.***EA-4/02, Evaluation of the Uncertainty of Measurement in Calibration.* 04/2022.

18. **Henry Zumbrun-Morehouse Instrument Company, Greg Cenker-Indysoft, Dilip Shah-E=mc³ Solutions.***Decision Rule Guidance.* 04-2024.