UKAS

A world of confidence

LAB 48

Edition 4, April 2022

Decision Rules and Statements of Conformity

Contents

Changes since edition 3

Addition of section on Decision Rules - Basics

Removal of Appendix B (renaming of subsequent Appendices and relabelling of equations)

Minor corrections:

1. Correction to axis labels for figures in examples 7 and 8
2. Various typographical and editorial amendments

Introduction

The general requirements that testing and calibration laboratories have to meet if they wish to demonstrate that they operate to a quality system, are technically competent and are able to generate technically valid results are all contained within ISO/IEC 17025:2017. This international standard forms the basis for international laboratory accreditation and in cases of differences in interpretation remains the authoritative document at all times.

Additional guidance for the purposes of accreditation is provided by ILAC in the form of policy requirements and guidance. In particular, ILAC-G8:09/2019 'Guidelines on Decision Rules and Statements of Conformity' provides an overview of the requirements stated in ISO/IEC 17025:2017 that concern statements of conformity and describes how certain Decision Rules can be selected and how uncertainty can (and must) be taken into account by either 'direct' or 'indirect' means. It also provides a limited number of worked examples.

The stated purpose of ILAC-G8 is to provide:

"… an overview for assessors, laboratories, regulators and customers concerning decision rules and conformity with requirements. It does not enter into the details regarding underlying statistics and mathematics but refers readers to the relevant literature. This means that some laboratories, their personnel and their customers may be required to improve their knowledge related to decision rule risks and associated statistics."

This UKAS guidance document, LAB 48, provides some supporting material and some additional guidance examples to assist in that process.

Further guidance can be found in standards such as ISO/IEC Guide 98-4:2012 (JCGM 106), ISO 10576-1:2003, and ISO 14253-1:2017.

The material provided here is quite varied, but it is not intended to cover all possible decision scenarios, rather it is intended to demonstrate various principles. As acknowledged in ILAC G8 "decision rules can be not only very different but also very complicated." In keeping with the diverse nature of practical conformity decision scenarios the format and structure of the examples is also intentionally diverse.

The main body of this document begins with an overview of the basics of Decision Rules with brief examples. There then follows a variety of more fully developed examples demonstrating how decision rules might be defined under various practical scenarios. Later examples demonstrate how conformance probability and specific risk may be calculated in various situations and how these relate to decision rules.

Finally, several Appendices are provided, including: a glossary of some of the terminology used in the main body examples; an overview of how conformance probability and risk can be calculated using standard Excel Worksheet functions in situations where the measurement uncertainty can be described by a Gaussian (also referred to as 'normal') probability density function (PDF) or a PDF based upon a t-distribution (the same approach remains valid for other PDFs); calculation of guard band factors; and concluding with an explanation of why simple acceptance criteria on their own cannot define a valid decision rule.

Decision Rules - Basics

What is a Decision Rule?

A **Decision Rule** describes the agreed process for making a conformity decision.

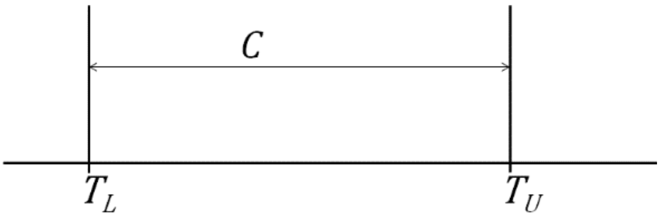
The rule explains how to use a measured value to decide whether a specification has been met and it explains the role of measurement uncertainty in reaching that decision.

Specifications

Specifications describe the desired characteristics of some quantity of interest (the **measurand**).

They can be thought of as describing the requirement for the ‘true’ value of the quantity.

Specifications can be ‘two-sided’, describing a **tolerance interval** *C* between upper and lower **tolerance limits**, T_L and T_U

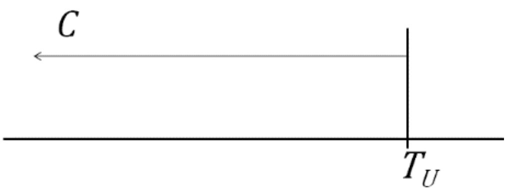


Examples:

The calibration error is to be 0 mm ± 0.02 mm

The laboratory temperature is to be between 18 °C and 22 °C

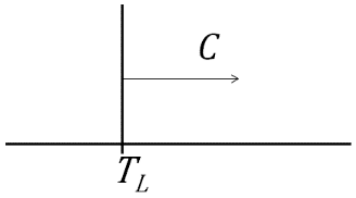
Specifications can also be ‘single sided’, describing just an upper limit



Examples:

The maximum storage temperature for a product is 4 °C

The maximum weight of a vehicle is 2200 kg or a lower limit



Examples:

The minimum height for an amusement ride is 1.2 m

The minimum weight of marmalade in a jar is 250 g

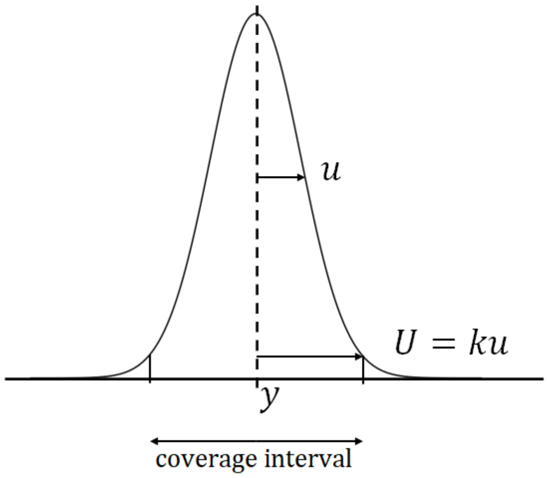
Measurements

Measurements are affected by a variety of influences that lead to uncertainty in the result (such as environmental conditions, limited information, and random effects).

A **measured value** therefore provides only an estimate of the 'true' value for the quantity of interest.

The **measurement uncertainty** for the measured value characterises the likely range of values for the 'true' value. This range of values is often described in terms of a **coverage interval** at a selected **coverage probability** i.e., the coverage interval describes the range in which there is a defined probability (usually 95 %) of locating the 'true' value, given the measured value *y* and its uncertainty.

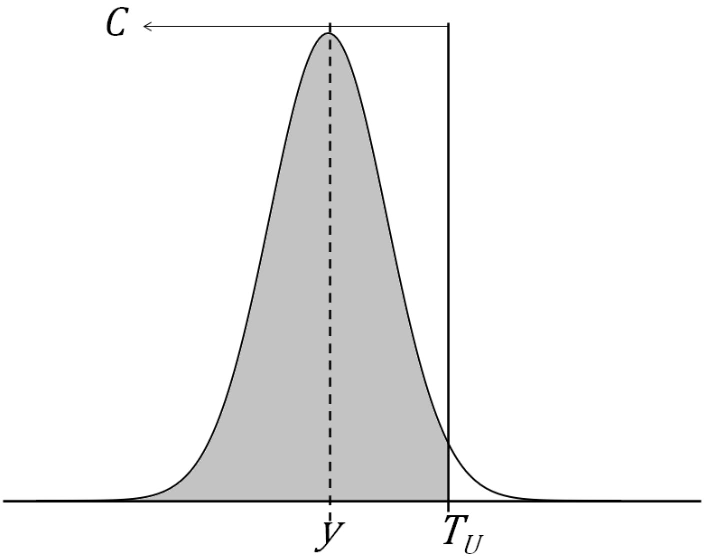
This measurement information is characterised by a **probability density function**, **PDF**, which describes the relative likelihood of different 'true' values. (Depicted here as a Gaussian distribution, however the same principles apply for all PDFs).



Using the known properties of the PDF, the **standard uncertainty** u can be expanded by a **coverage factor** k to calculate an expanded uncertainty $U = k.u$ which in turn defines a coverage interval, between $(y - U)$ and $(y + U)$.

How does measurement uncertainty affect conformity decisions?

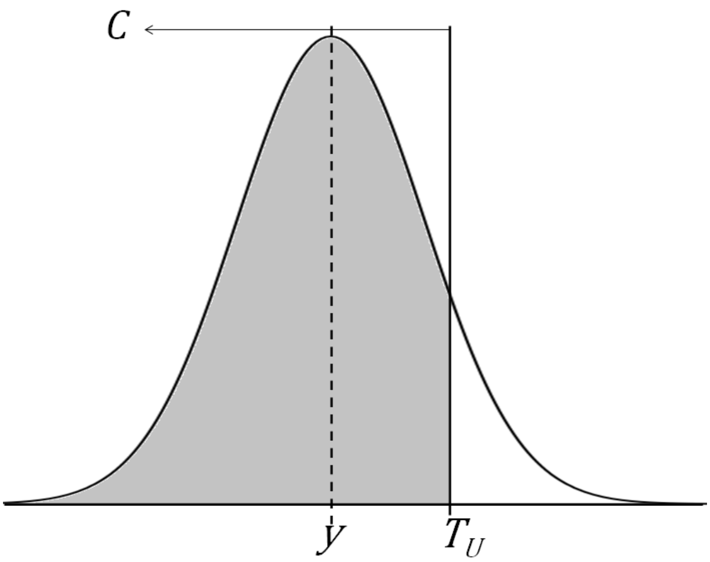
The concepts of a tolerance interval for conforming values of the measurand, and a PDF for the estimate (measured value) can be combined, as shown in the figure below…



In this example, the shaded region of the PDF is within the tolerance interval and represents possible conforming values of the measurand given the observed measurement result.

The unshaded region represents non-conforming values of the measurand that can similarly *also* be attributed to the measurand.

If the measurement uncertainty is larger, then a larger proportion of non-conforming 'true' values could have been responsible for the measured value.



In this example, as the measured value is within the range of values required of the 'true' value, it might be decided to accept that the result indicates conformance. In such a case the **risk** associated with the decision, i.e., the chance of making a false decision is noticeably higher for the case with the larger uncertainty. Clearly, if measurement uncertainty is not considered at *any* stage of a decision process the risk remains undefined (or uncontrolled) and the conformity decision is worthless.

Decision Rules

Risk in conformity decisions is controlled by agreeing a **Decision Rule** (DR).

There are two common types of Decision Rule: those based upon **simple acceptance** criteria, and those based upon **guard bands**. In both cases, the rule defines a range for measured values that are considered to indicate conformance, this is known as the **acceptance interval**

Rules involving Simple Acceptance

Rules based upon Simple Acceptance criteria (sometimes called "shared risk") all equate the **acceptance interval** with the **tolerance interval**.

Measurement uncertainty is taken into account by defining constraints that must be met *before* the simple acceptance decision can be made. These constraints are chosen to somehow limit the risk associated with a decision.

The role of measurement uncertainty in this type of decision process is to act as a pre-condition for the use of simple acceptance criteria.

For a two-sided specification, the constraint often takes the form of a specified minimum **measurement capability index,** $\boldsymbol{C_m}$ which is usually defined in terms of the tolerance interval, and the 95 % coverage interval as $\boldsymbol{C_{95}}$

$$C_{95} = (T_U - T_L)/(2.U_{95\%})$$

To avoid any ambiguity when used, C_{95} should be defined along with the accompanying decision rule in contract agreements and in reports and certificates.

In practice, various other parameters and differing terminology, such as Test Uncertainty Ratio (TUR), may also be used provided they contain the measurement uncertainty.

Example of a Rule based upon Simple Acceptance criteria with a limit on capability index:

PASS: when measured temperature is between 18 °C and 22 °C, AND $C_{95} \geq 5$

FAIL: otherwise

Alternatively, the constraint can be expressed as an upper limit on the measurement uncertainty.

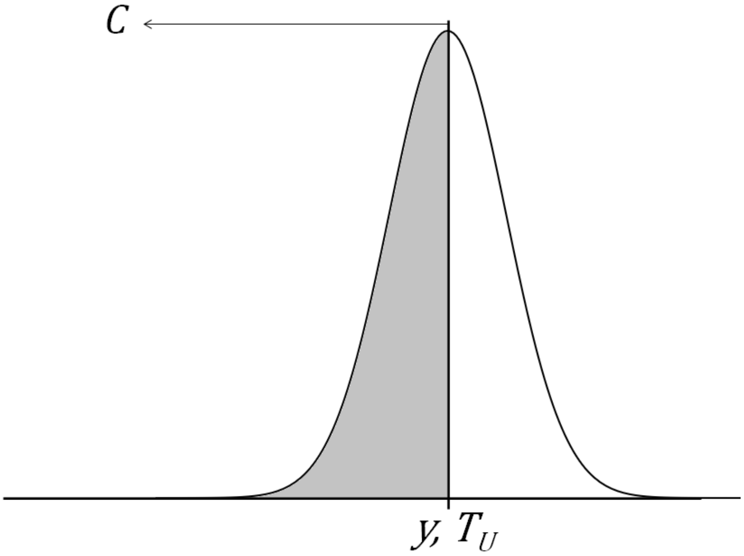
Example of a Rule based upon Simple Acceptance criteria with a limit on expanded uncertainty:

PASS: when measured value is below 2200 kg, AND expanded uncertainty $U_{95\%} \leq 100$ kg

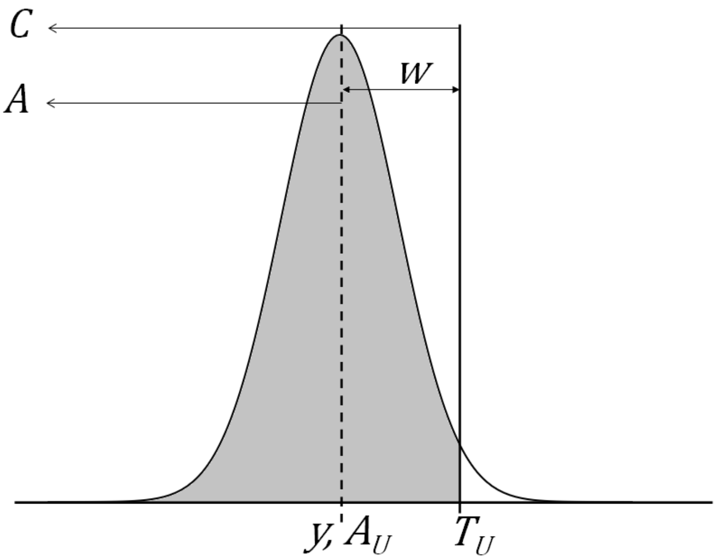
FAIL: otherwise

Rules involving Guard Bands

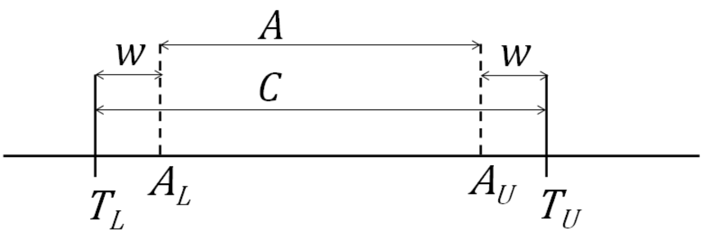
Under a decision rule based upon Simple Acceptance criteria, measured values up to and including the tolerance limits are taken to indicate conformance, with an associated risk of false decision of 50 % for values at the tolerance limit (and potentially higher risk for two-sided specifications).



In the case of a large capability index, this risk may only apply to a small proportion of all allowed values, nevertheless there will be situations where this risk is too high. In these circumstances the acceptance interval can be reduced by an amount known as a **guard band**, so that the maximum risk of false acceptance is reduced to a desired level as depicted below for the case of a single-sided specification.



Guard bands can be applied to both single-sided and two-sided specifications.



Typically, guard bands are defined to have a width w equal to the 95 % coverage interval $U_{95\%}$

$$w = U_{95\%}$$

for which there is only a 2.5 % risk of false acceptance for both single-sided specifications and for two-sided specifications where $C_{95} \gtrsim 1.5$

<u>Example of a Rule based upon Guard Bands for a two-sided specification:</u>

Specification: calibration error is to be 0 mm ± 0.02 mm

Acceptance interval: 0 mm ± 0.018 mm, for a guard band $w = U_{95\%}$=0.002 mm

PASS: when measured value is within the acceptance interval

FAIL: otherwise

<u>Example of a Rule based upon Guard Bands for a single sided specification:</u>

Specification: maximum storage temperature for product is 4.0 °C

PASS: when measured value is no larger than $(4.0\,°\mathrm{C} - U_{95\%})$

FAIL: otherwise

Binary and multi-state decision rules

The conformity decisions seen so far are known as **binary decisions** as only two possible outcomes are defined.

Both Simple Acceptance and Guard Banded decision rules can be written in terms of multiple different outcomes. They can also be represented in other formats

<u>Example of a rule with several possible outcomes.</u>

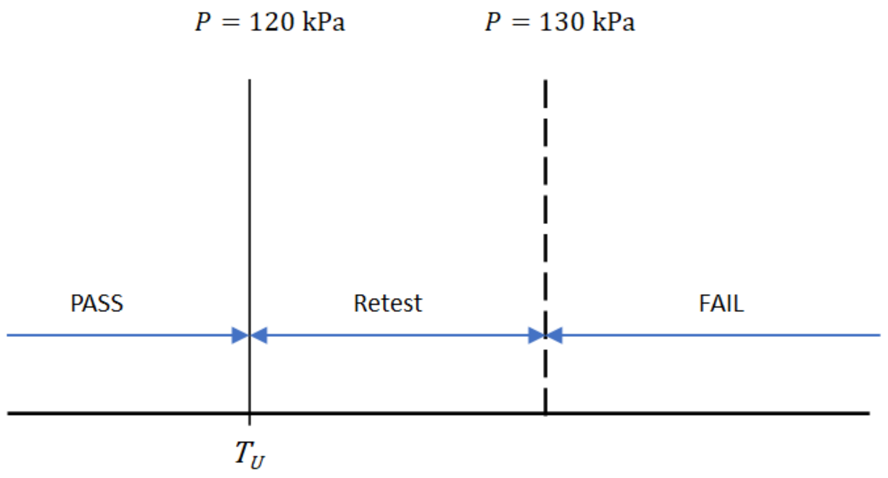
Specification: Pressure ≤ 120.0 kPa

PASS: when measured pressure P ≤ 120.0 kPa, AND $U_{95\%}$ ≤ 2.0 kPa,

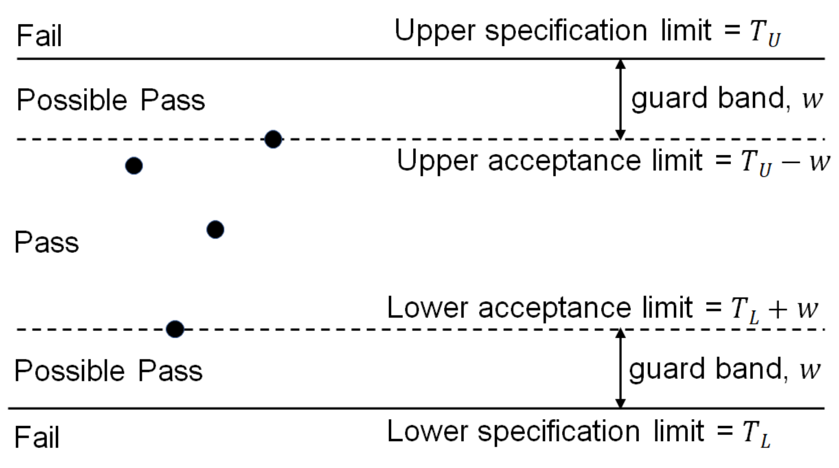
FAIL: when measured pressure P > 130 kPa, AND $U_{95\%}$ ≤ 2.0 kPa,

otherwise “Retest”

In graphical form this decision rule could be represented as



<u>Example of guard banded rule with several possible outcomes</u>



Decisions for 'qualitative' tests

A conformity decision is not always made purely on the basis of a measured quantity value.

Instead, many tests involve some form of **examination** (or inspection), sometimes referred to as a 'qualitative' test, to which the strict VIM concept of measurement uncertainty does not readily apply. For example, when the outcome is perhaps the assignment of a **nominal property** (e.g., colour, shape, sequence) or some other characteristic, such as position on an **ordinal** scale (e.g., Rockwell C, Richter, Beaufort).

That isn't to say that measurement uncertainty doesn't play a role in making conformity decisions based on the outcome of such tests… in most circumstances, examinations are performed under defined conditions that, in order to verify that these are being achieved, are *themselves* subject to measurement and require some form of rule that explains the role of measurement uncertainty for those quantities.

Decision rules may also include a requirement upon the **examination reliability**

Example of a Rule based upon Simple Acceptance for a 'qualitative' test:

DR: Classification of the perceived odour of the test item may be reported when:

a. The sample has been obtained and prepared according to method XYZ
AND
b. Prior to testing the indicated sample temperature has remained between 18 °C and 22 °C, measured with capability index $C_{95} \geq 5$; for a period of (20 ± 1) min measured with expanded uncertainty $U_{95\%}$ ≤ 10 s.
AND
c. The examiner demonstrates a reliability of ≥ 99.5 % in a daily proficiency test as defined in BSxxxx

'Difficult' decisions: inadequate standards

One of the greatest source of difficulty in making accreditable conformity decisions arises from inadequate test standards. In many published standards there is no decision rule. In fact, in many standards there is no mention of measurement uncertainty.

There are many possible reasons for this: the standard may predate the GUM (1995) and widespread use of the 'uncertainty framework'; for some reason the authors of the standard may have chosen not to state their requirements or assumptions about the uncertainty that would be achieved in conducting the tests with specified equipment; or the standard may simply be deficient.

Whatever the reason, ISO/IEC 17025:2017 and ILAC-G8:09/2019 still require measurement uncertainty to be taken into account, whether directly or indirectly.

In these circumstances the requirements of the inadequate standard must be supplemented by additional requirements so that a valid decision rule is established. This is generally a matter of contract review, for example agreement that the specification will be tested by a simple acceptance rule with a defined minimum capability index.

Example of a test standard supplemented with additional requirements to form a Decision Rule:

Suppose that a standard BS XXX describes a test requirement with "measurement error limits stated in Table XYZ" in which the role of measurement uncertainty is not defined.

This could be incorporated into a valid decision rule following agreement between the lab and the customer on an acceptable measurement capability C_m

For example,

Specification: measurement error limits $\pm T$ stated in BS XXX, Table XYZ (e.g., ± 0.6 °C)

PASS: Simple Acceptance criteria for measured error t when $(-T \leq t \leq T)$, AND $C_m \geq 6$

FAIL: Otherwise

Alternatively, it could be incorporated into a valid decision rule by agreeing upon an acceptable limit to the expanded uncertainty, e.g., $U_{\max} = 0.10$ °C

For example,

Specification: measurement error limits $\pm T$ stated in BS XXX, Table XYZ

PASS: Simple Acceptance criteria for measured error t when $(-T \leq t \leq T)$, AND $U_{95\%} \leq 0.10$ °C

FAIL: Otherwise

‘Difficult’ decisions: the problem with ‘accuracy’

The term ‘accuracy’ is probably intended by authors of standards and guidance to describe the expected quality of a measurement. Unfortunately, however, its use often creates immediate problems.

Most obviously, difficulties arise due to the various meanings attached in common practice to the word ‘accuracy’. Sometimes the intended use can be inferred from the context in which it is used, but often the meaning is ambiguous. For example, if a requirement states that a weighing instrument ‘should be capable of measuring loads up to 4 kg with an accuracy of 0.1 kg’ this ‘accuracy’ requirement might commonly be interpreted as referring to the size of the measurement error, to the resolution of the display, or to the measurement uncertainty.

More fundamentally, from a metrological (GUM) standpoint, uncertainty and accuracy are entirely different concepts. This distinction is clear from the definitions, notes and annotations of the International Vocabulary of Metrology (VIM) in which accuracy is described as a qualitative concept.

For example, in the HTML version of the VIM it is stated that:

> ‘Historically, the term ‘measurement accuracy’ has been used in related but slightly different ways. Sometimes a single measured value is considered to be accurate (as in the VIM definition), when the measurement error is assumed to be small (in magnitude). In other cases, a set of measured values is considered to be accurate when both the measurement trueness and the measurement precision are assumed to be good. Sometimes a measuring instrument or measuring system is considered to be accurate, in the sense that it provides accurate indications. Care must therefore be taken in explaining in which sense the term ‘measurement accuracy’ is being used. In no case is there an established methodology for assigning a numerical value to measurement accuracy.’

This isn’t simply a hypothetical problem or an issue of semantics - the GUM framework and standards such as ISO/IEC 17025, ISO 15189 are concerned with measurement uncertainty, not ‘accuracy’.

In practice, faced with a specification stated in terms of 'accuracy' it must be established through contract review, how the term 'accuracy' is to be interpreted. In addition, if not also stated, it must be established how measurement uncertainty is to be taken into account when specifications are tested using measured values.

Is a conformity decision needed?

Conformity decisions are not a default requirement of ISO/IEC 17025. Consequently, in many measurement situations, such as apply for most calibrations, there is no *a priori* requirement for a conformity decision to be made or reported.

In these circumstances, it is for the customer to decide whether or not a result meets their requirements. Therefore, if the customer does *not* require a conformity decision to be reported there is no need to identify a specification or to agree a decision rule.

In such cases, specifications may (if requested) be stated in a report (as a statement of fact), together with the measurement results, provided that there is no associated suggestion or implication of a conformity statement. In a general sense this approach has the advantage that the customer makes their decision on the acceptability of the result at their convenience, rather than committing to having their decision recorded on the test or calibration report at the time of measurement.

Example scenarios and calculations

The following examples present a variety of decision scenarios that might be encountered in practice. Later examples demonstrate how various calculations can be performed.

Example 1: Test standard is a “validated” method

Certain types of test are conducted using what is termed a ‘validated’ method or procedure. These range greatly in robustness from methods validated by the collaborative study approach described in ISO 5725, through to ‘industry accepted’ methods based on *ad hoc* accepted norms.

The degree to which uncertainty has already been incorporated into a method or standard may be clear and explicit. For example a method uncertainty may be provided, which simply needs to be combined with lab-specific factors into a final measurement uncertainty.

Alternatively, the method may be characterised in terms of precision and trueness, requiring an approach such as that described in ISO 21748 to evaluate the measurement uncertainty. For example, in the field of regulated environmental testing (such as MCERTS soil and water testing) the measurement uncertainty is usually taken into account by demonstrating consistency with published performance characteristics for precision (repeatability and/or reproducibility) and trueness (method and/or lab bias) that have been established during method validation exercises.

Example:

A customer specifies that the flowrate for a sample of powder must be above a particular limit, when measured using a certain test method. The powder flow test method for measuring the flowrate has been subjected to collaborative study in accordance with ISO 5725-2. The method has been published, together with values for its repeatability, reproducibility, and trueness.

A lab performs the test and confirms that it achieves (or betters) these values. It also confirms that there are no other influences on the outcome that are not adequately covered by the collaborative study.

In doing so the lab has implicitly confirmed that the measurement uncertainty[1] is within some (albeit unstated) limit.

The role of measurement uncertainty is considered to be defined for the performance of the test and can therefore form the basis of a Decision Rule with *indirect* account of measurement uncertainty.

Decision rule:

For a test performed according to published method ABC, with repeatability and bias consistent with the published method values, and no additional influences being identified…

PASS: when flowrate measured using the powder flow test method is greater than 0.45 kg/s

FAIL: otherwise.

[1] The measurement uncertainty in this case is evaluated by so-called ‘top-down’ approach. See for example the guidance provided by ISO 21748:2017 for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation.

Example 2: Measurement uncertainty and ‘qualitative’ tests

As was described earlier… many tests involve some form of examination (or inspection) where the outcome of the test is a nominal property (e.g., colour, species, sequence of markers…). Others might involve establishing a position on an ordinal scale (e.g., Rockwell C, Richter, Beaufort, octane…).

In the case of these ‘qualitative’ tests the strict VIM concept of measurement uncertainty does not readily apply. That is not to say that measurement uncertainty doesn't play a role in such tests… in fact, in most cases, such tests are performed under defined conditions that are themselves subject to measurement.

Example 2a

A test requires an inspector to examine the colour of a fluid sample after preparation according to a defined procedure and processing in an oven at (40 ± 1) °C for between 1 hr and 1 hr 5 min.

For this test, the examination of the colour involves a subjective judgement from a trained and competent examiner whose reliability can be established by proficiency testing. However, the oven temperature and elapsed time are both measurable quantities for which a value and a measurement uncertainty can be established.

In practice, demonstrating that such conditions have been met usually involves a form of conformity decision in which measurement uncertainty must play a part (a measurement uncertainty of say 0.05 °C for the oven temperature clearly presents less risk of nonconformance than a measurement uncertainty of 0.5 °C) and so a Decision Rule must be agreed.

For example:

Conformance is demonstrated if measured oven temperature has remained in the range (40 ± 1) °C, AND expanded measurement uncertainty U_{95} ≤ 0.05 °C

Example 2b

Suppose that the packaging for transportation of a fragile item is to be tested by packing a certain type of glass bottle and then, under specified conditions, dropping the package before unpacking and inspecting the bottle for damage.

Various rules could be established that take account of measurement uncertainty (that has already been evaluated by the laboratory[2]).

The specification and decision rule might for example be defined as follows:

Specification for test on integrity of packaging containing a glass bottle:

The packaged bottle should remain intact when dropped under the following conditions - height h in range 0.99 m to 1.05 m; temperature T in range 18 °C to 23 °C

Decision Rule

PASS: if bottle is unbroken, AND

measurement conditions conform to Simple Acceptance criteria for h and T (i.e., 0.99 m $\leq h \leq$ 1.05 m; 18 °C $\leq T \leq$ 23 °C), AND

provided also that $u(h) \leq$ 0.5 cm, $u(T) \leq$ 0.5 °C;

FAIL: otherwise.

The decision rule could alternatively have been expressed in terms of conformance probability for the test conditions, e.g.

PASS: if bottle is unbroken, AND

conformance probability p_c > 99 % for test conditions h and T;

FAIL: otherwise.

(See Appendix B for calculation of p_c)

Example 2c

(Reproduced from Basics section)

<u>Example of a Rule based upon Simple Acceptance for a ‘qualitative’ test:</u>

DR: Classification of the perceived odour of the test item may be reported when:

a. The sample has been obtained and prepared according to method XYZ,
AND
b. Prior to testing the indicated sample temperature has remained between 18 °C and 22 °C, measured with capability index $C_{95} \geq 5$; for a period of (20 ± 1) min measured with expanded uncertainty $U_{95\%}$ ≤ 10 s.
AND
c. The examiner demonstrates a reliability of ≥ 99.5 % in a daily proficiency test as defined in BSxxxx

[2] “Already evaluated”, since it is an accreditation requirement to evaluate the uncertainty of all key measurements.

Example 3: Test scenario in which a customer asks a laboratory to “ignore uncertainty”

At some point it is possible that a customer will approach a laboratory and ask them to make a conformity decision that “ignores uncertainty”.

Accredited reporting of the outcome for such a decision is not permitted by ISO/IEC 17025:2017 nor by ILAC-G8:09/2019 which requires that “measurement uncertainty is accounted for” (directly or indirectly) when conformity decisions are made. (See Appendix D for further explanation of why rules that take no account of uncertainty are not appropriate.)

The laboratory therefore needs to establish how their customer would like them to proceed.

Fortunately, in practice most customers usually *do* have some, albeit perhaps unrecognised or unstated expectation concerning the ‘reliability’ of the measurement they are asking for. (Would they really be happy with uncertainty of say 10, or a hundred, or a thousand times the specification?)

Example

Suppose that a laboratory is approached to test the breaking strain of a sample of thread. The customer declares that the thread is required to remain intact for loads up to 10 N and states that they would like the laboratory to ‘ignore uncertainty’ since there is no uncertainty requirement mentioned in the associated standard.

During contract review, the laboratory responds by explaining that, for the decision to be reported under their accreditation, uncertainty *cannot* be ignored. The laboratory also explains that, being accredited for the test, they have already established that the applied load can be measured with an expanded uncertainty of better than 0.1 N (k = 2 for approximately 95% coverage probability). Also, to reduce the risk of false acceptance, the laboratory proposes to apply a measured load of 10.1 N

The customer confirms that, in choosing an accredited provider, *they had in fact already assumed that the uncertainty would be appropriate for the test* and that they are therefore content to have the measurement performed under these conditions. The customer also confirms that they would like a binary, PASS/FAIL decision.

Therefore, in this case the outcome might be…

Agreed and reported specification: Conforming thread remains intact under load of 10.1 N

Agreed and reported Decision Rule:

PASS: if the thread remains intact under an applied load of 10.1 N,

AND

the expanded uncertainty (k = 2 for approximately 95% coverage probability) of the measured load is no larger than 0.1 N.

FAIL: otherwise

Reported decisions:

Thread remains intact for load L = 10.1 N: PASS

Thread is damaged by load L = 10.1 N: FAIL

The conformance probability for this result can be calculated using (B.2)

$p_c = 1 - \text{NORM.DIST}(10, 10.1, 0.1/2, \text{TRUE}) = 0.97725$

i.e. the probability of false acceptance (*PFA*) is (B.7)

$PFA \ = \ 1 - p_c = 2.3\%$

Example 4: Test standard does not mention measurement uncertainty

As already stated, it is commonplace for a testing standard to make no mention of measurement uncertainty. There are many possible reasons for this: the standard may predate the GUM (1995) and widespread use of the 'uncertainty framework'; for some reason the authors of the standard may have chosen not to state their requirements or assumptions about the uncertainty that would be achieved in conducting the tests with specified equipment; or the standard may simply be deficient.

Whatever the reason, ISO/IEC 17025:2017 and ILAC-G8:09/2019 require measurement uncertainty to be taken into account, whether directly or indirectly (not least so that the conformity decision is metrologically traceable).

At first sight this seems to present a problem, however the situation is similar to that described in the previous example in which a customer asks a laboratory to "ignore uncertainty".

Example

Suppose that a laboratory is asked to perform a test (or calibration) described in a standard 'ABC123' which defines a hierarchy of equipment and specifies equipment 'accuracy' requirements in terms or 'maximum permissible error' E but does not mention measurement uncertainty. The customer states that they would like the laboratory to 'ignore uncertainty' as there is no uncertainty requirement stated in the standard.

During contract review, the laboratory responds confirming that they are able to meet the 'accuracy' requirements and perform the relevant measurements, but for conformity statements to be reported under their accreditation the measurement uncertainty *cannot* be ignored.

The laboratory explains however, that in this case the expanded measurement uncertainty $U_{95\%}$ is usually less than 1/5 of the width of the specification ($(T_U - T_L) = 2E$) i.e., a measurement capability index $C_{95} = \frac{(T_U - T_L)}{2.U_{95\%}} > 2.5$ is routinely achieved.

The customer confirms that the proposed measurement capability is appropriate for their requirements. The customer also confirms that they would like a binary, PASS/FAIL decision.

In this case the agreed Decision Rule might be:

PASS: indicates that a measurement error conforms with the relevant ('accuracy') requirements of the testing standard, AND $C_{95} > 2.5$

FAIL: otherwise.

Example 5: Double sided tolerance limit, DR: p_c ≥ 95%

A customer's acceptance criteria (specification) for a 2 MPa pressure transducer is that "calibration errors should be no larger than 0.5% of nominal full scale" but they have not specified a decision rule.

The laboratory therefore proposes the following rule:

DR: At each measured calibration pressure, report as "Pass" when there is at least 95% probability that the error meets specification. Otherwise report as "Fail".

A set of calibration results can then be reported as follows:

Specification: Calibration errors should be no more than ±0.5% of nominal 2 MPa full scale value.

Decision Rule: At each measured calibration pressure, report as "Pass" when there is at least 95% probability that the error meets specification. Otherwise report as "Fail".

Results:

Indicated pressure p_{ind} /MPa	Transducer error $e_{\%\text{FS}}$ /%	Decision	Conformance probability
1.995	0.25	Pass	0.994
1.494	0.30	Pass	0.977
0.993	0.35	Fail	0.933
0.492	0.40	Fail	0.841
0.083	0.35	Fail	0.933
-0.006	0.30	Pass	0.977

Where, at each reference pressure, p_{ref}, the transducer error is calculated from

$$e_{\%\text{FS}} = \frac{100 \times (p_{\text{ref}} - p_{\text{ind}})}{2\ \text{MPa}}$$

In use, corrected pressure, $p = p_{\text{ind}} + e_{\%\text{FS}} \times \frac{2}{100}$

Expanded measurement uncertainty for e is $U(e)$ = 0.004 MPa (= 0.2% FS).

The reported expanded uncertainty $U(e)$ is based on a standard uncertainty multiplied by a coverage factor *k* = 2, providing a coverage probability of approximately 95%.

In this example the conformance probability has been calculated for each measurement of transducer error with standard uncertainty u = 0.1% FS.

For example, conformance probability for p_{ind} = 1.995 MPa is evaluated from (B3)

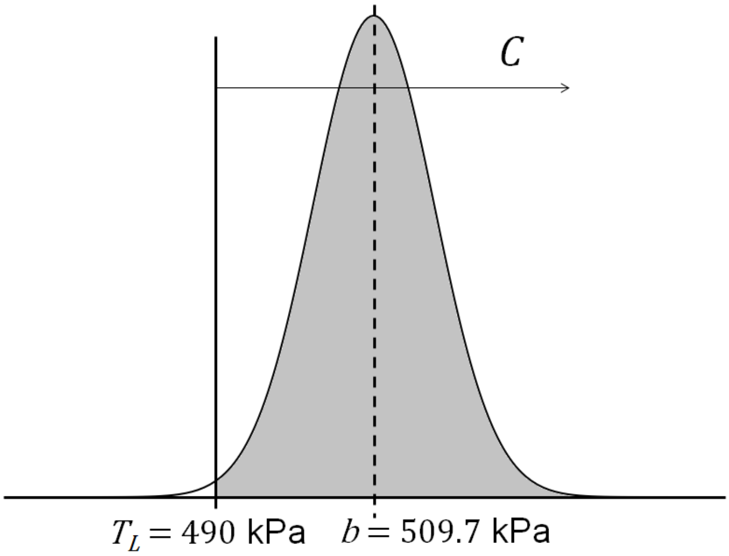
$p_c = \text{NORM.DIST}(T_U, e_{\%\text{FS}}, u, \text{TRUE}) - \text{NORM.DIST}(T_L, e_{\%\text{FS}}, u, \text{TRUE})$ i.e.

$p_c = \text{NORM.DIST}(0.5, 0.25, 0.1, \text{TRUE}) - \text{NORM.DIST}(-0.5, 0.25, 0.1, \text{TRUE}) = 0.994$

(See Appendix B for details of this calculation)

Example 6: Single sided lower tolerance limit (JCGM106 7.3.3 Ex 2)

A metal container is destructively tested using pressurized water in a measurement of its bursting strength B. The measurement yields a best estimate b = 509.7 kPa, with associated standard uncertainty u = 8.6 kPa. The container specification requires $B \geq$ 490 kPa, which is a lower limit on the bursting strength.



The conformance probability p_c is therefore (B.2)

$p_c = 1 - \mathrm{NORM.DIST}(490, 509.7, 8.6, \mathrm{TRUE}) = 0.99$

i.e. the conformance probability for this container is 99%

If a decision is taken to *accept* it as conforming the probability of false acceptance is (B.7)

$PFA = 1 - p_c = 1\%$

Possible Decision Rules for this conformity decision might therefore be defined in terms of p_c or PFA, for example:

ACCEPT: when $p_c \geq 95\%$

REJECT: otherwise

or equivalently

ACCEPT: when $PFA \leq 5\%$

REJECT: otherwise

This result might be reported as:

ACCEPT, with conformance probability of 99% which meets the acceptance criterion of $p_c \geq 95\%$

or equivalently

ACCEPT, with probability of false acceptance of 1% which meets the acceptance criterion of $PFA \leq 5\%$

Supposing instead that b = 495.2 kPa…

in this case

$p_c = 1 - \mathrm{NORM.DIST}(490, 495.2, 8.6, \mathrm{TRUE}) = 0.73$

This result might therefore be reported as:

> REJECT, with a conformance probability of only 73% which does not meet acceptance criterion of $p_c \geq 95\%$

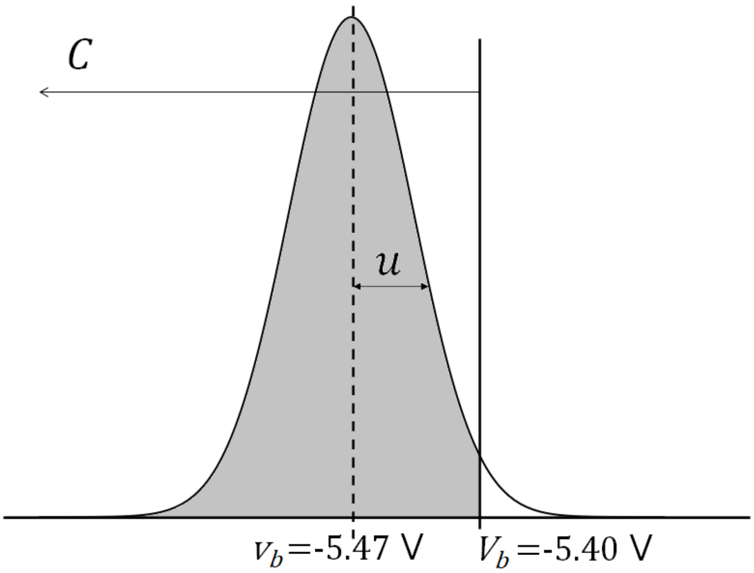
or

> REJECT, unable to meet PFA requirements

Example 7: Single sided upper tolerance limit (JCGM106 7.3.3 Ex 1)

The breakdown voltage V_b of a Zener diode is measured, yielding a best estimate $v_b = -5.47$ V with associated standard uncertainty $u = 0.05$ V.

Specification of the diode requires $V_b = -5.40$ V, which is an upper limit on the breakdown voltage.



The conformance probability p_c is represented by the portion of the PDF within conformity interval C where (B.1)

$p_c = \mathrm{NORM.DIST}(-5.4, -5.47, 0.05, \mathrm{TRUE}) = 0.92$

i.e., the conformance probability for this diode is 92%

If a decision is taken to accept it as conforming the probability of false acceptance is (B.7)

$PFA = 1 - p_c = 8\%$

Suppose that diodes will be accepted with a conformance probability of no worse than 95 % and rejected if 90 % or lower. Otherwise, their status is undetermined, and they are set aside. Possible Decision Rules for this conformity decision might therefore be defined in terms of p_c or PFA, for example:

ACCEPT: when $p_c \geq 0.95$; ($PFA \leq 5\%$)

REJECT: when $p_c \leq 0.90$; ($PFA \geq 10\%$)

UNDETERMINED otherwise

This result for the example above might be reported as:

UNDETERMINED, with a conformance probability of 0.92 which does not meet criteria for acceptance ($p_c \geq 0.95$) or rejection ($p_c \leq 0.90$)

Example 8:
Single sided upper tolerance limit, DR: Pass when *PFA* ≤ *PFA* max

Suppose that in the testing of Zener diode breakdown voltage as described previously, a probability of false acceptance of up to 0.5% is allowed.

Suppose also that the measurement uncertainty is the same, u = 0.05 V for *all* measurements of breakdown voltage made using this system.

In this case we can establish a *fixed value for an upper acceptance limit* A_U corresponding to PFA_{max} = 0.5% . Where (B.3)

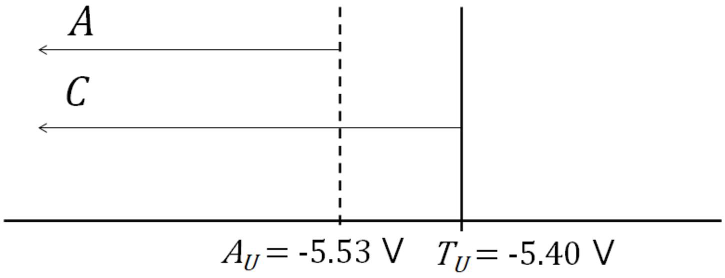
$A_U = T_U - k_w . u$

The required guard band factor is calculated (C.1) to be

$k_w = \mathrm{NORM.S.INV}(1 - PFA_{max}) = \mathrm{NORM.S.INV}(0.995) = 2.58$

therefore

A_U = -5.53 V



The region between A_U and T_U is known as a 'guard band'.

Now when any measurements are performed (with u = 0.05 V), all that is required is to test whether the measured value is within the acceptance interval ($v_b \leq$ -5.53 V) to accept a diode as conforming.

Possible Decision Rules for this conformity decision might therefore be defined in terms of PFA or p_c or A_U for example:

PASS: when $PFA \leq 0.5\%$, or equivalently

PASS: when $p_c \geq 0.995$, or equivalently

PASS: when the measured value does not match or exceed an upper acceptance limit A_U, which is defined in terms of the upper tolerance limit T_U and a guard band that is calculated to ensure a conformance probability of at least 99.5%,

FAIL: otherwise

Results might be reported as:

PASS, with $PFA \leq 0.5\%$, or equivalently

PASS, with $p_c \geq 0.995$, or equivalently

PASS, measured value does not exceed the upper acceptance limit, or

FAIL, unable to meet conformity requirements

Example 9:
Single sided upper tolerance limit, DR: Accept when *PFA* ≤ *PFA* max

A machine is designed to shred pruned tree branches up to a diameter of 50 mm. Larger diameter branches will go through the machine, but the owner of the machine does not wish this to happen more frequently than 10% of the time. He therefore uses a simple calliper to measure the diameter with a standard uncertainty of $u = 5$ mm.

What limit should he place on measured diameter? In other words, what size guard band should be applied?

The owner wishes to only falsely accept (i.e. attempt to shred an oversized branch) 10% of the time, i.e.

PFA_{max} = 0.1

Upper tolerance limit is T_U = 50 mm

So (B.3)

$A_U = T_U - k_w . u$

where k_w is found by using the Table in Appendix C, or by calculation (C.1)

$k_w = \text{NORM.S.INV}(1 - 0.1) = \text{NORM.S.INV}(0.9) = 1.28$

Hence

A_U = 50 – 1.28 x 5 = 43.6 mm

The owner should only accept branches measured to have a diameter of 43.5 mm or less.

Possible Decision Rules for this conformity decision might be defined in terms of PFA, for example:

ACCEPT: when measured diameter < 43.5 mm, for $PFA < 10\%$

REJECT: otherwise

Example 10:
Single sided lower tolerance limit, DR: Accept when *PFA* ≤ *PFA* max

In some situations, we may be more interested in *not* rejecting potentially conforming items i.e. we are prepared to Accept, even when the chance of falsely accepting is high. (This is a so-called *relaxed acceptance* scenario, in which values outside the tolerance interval are accepted).

For example, a gold miner performs initial grading by measuring the apparent density of each ore sample. Gold ore has a typical density of 19320 kg m^{-3}.

Because of the potential value of the ore he is happy to bear the cost associated with a high probability of false acceptance at this stage of his process, up to a maximum of 99.5%.

Possible Decision Rules for this conformity decision might therefore be defined in terms of PFA or p_c, for example:

ACCEPT: when $PFA \leq 99.5\%$

REJECT: otherwise

or equivalently

ACCEPT: when $p_c \geq 0.5\%$,

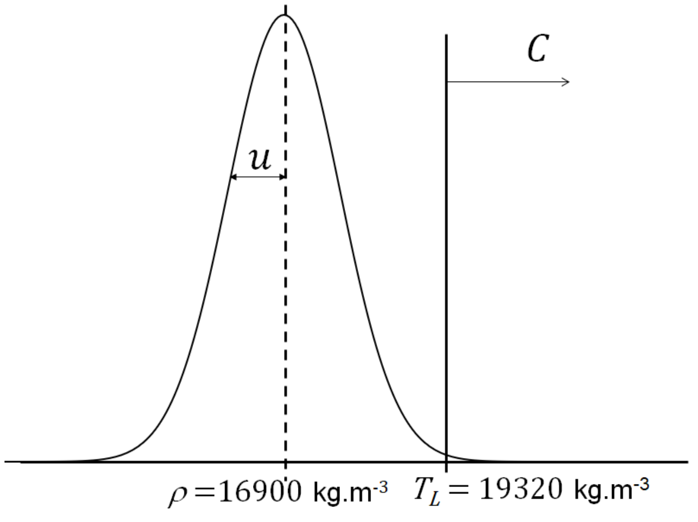
REJECT: otherwise

Suppose that a sample has an apparent density ρ = 16900 kg m^{-3} with an associated standard uncertainty u = 1000 kg m^{-3} the miner calculates (B.2) that

$$p_c = 1 - \text{NORM.DIST}(19320, 16900, 1000, \text{TRUE}) = 0.8\%$$

and (B.7)

$$PFA = 1 - p_c = 99.2\%$$



The result for this particular sample might then be reported as:

Accepted as conforming, having a probability of false acceptance of no more than 99.5%

or

Accepted as conforming, having a conformance probability of at least 0.5%

If a second sample has an apparent density ρ = 16500 kg m^{-3}, also with an associated standard uncertainty of u = 1000 kg m^{-3} the miner calculates that

$$p_c = 1 - \mathrm{NORM.DIST}(19320, 16500, 1000, \mathrm{TRUE}) = 0.2\%$$

$$PFA = 1 - p_c = 99.8\%$$

This result might then be reported as:

Rejected as non-conforming, unable to meet PFA requirements

or

Rejected as non-conforming, having a conformance probability of less than 0.5%

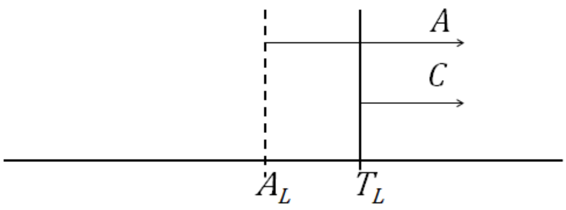
If the standard uncertainty of the process is always u = 1000 kg m^{-3} the miner can establish a guard band, i.e. calculate a fixed value for a lower acceptance limit A_L corresponding to PFA_{max} = 99.5% i.e. (using C.2 and C.1)

$$A_L = T_L + k_w . u$$

$$k_w = \mathrm{NORM.S.INV}(1 - PFA_{max}) = \mathrm{NORM.S.INV}(0.005) = -2.58$$

hence

A_L = 16744 kg m^{-3}



Now, when any measurements are performed (with u = 1000 kg m^{-3}), all that is required is to test whether the result is within the acceptance interval ($\rho \geq$ 16744 kg m^{-3}) to accept a sample for further grading.

A possible Decision Rule for this conformity decision might be defined as

ACCEPT: when the measured value exceeds a lower acceptance limit A_L, which is defined in terms of the lower tolerance limit T_L and a guard band that is calculated to ensure a conformance probability of at least 0.5%

REJECT: otherwise

A corresponding conformity statement could then be

ACCEPT, the measured value meets or exceeds the lower acceptance limit

Example 11: Single sided upper tolerance limit, DR: Accept when *PFA* ≤ *PFA* max (JCGM106 8.3.3.2 Ex 1)

In highway law enforcement, the speed of motorists is measured by police using devices such as radar and laser guns. A decision to issue a speeding ticket, which may potentially lead to an appearance in court, must be made with a high degree of confidence that the speed limit has actually been exceeded.

Using a Doppler radar, speed measurements in the field can be performed with a relative standard uncertainty $u(v)/v$ of 2% in the interval 50 km/h to 150 km/h. Knowledge of a measured speed v in this interval is assumed to be characterised by a normal PDF with expectation v and standard deviation 0.02 v.

Under these conditions one can ask, for a speed limit of $T_L = v_0$ = 100 km/h, what threshold speed v_{max} (acceptance limit, A_L) should be set so that for a measured speed $v \geq v_{max}$ the probability that v_0 is exceeded is at least 99.9%?

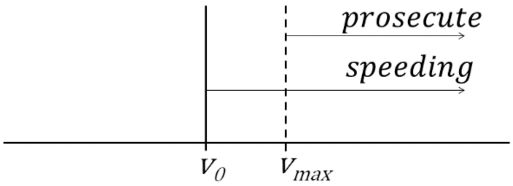
In this example, the tolerance interval corresponds to speeding motorists. To minimize the risk of false prosecution the test requires PFA_{max} = 0.001. Therefore, at the acceptance limit the probability must be p_c = 0.999

Possible Decision Rules for this conformity decision might therefore be defined in terms of PFA or p_c, for example:

Prosecute:	when $PFA \leq 0.1$%, or equivalently
Prosecute:	when $p_c \geq 0.999$, or
Prosecute:	when measured speed $v \geq v_{max}$ (the probability that v_0 is exceeded is at least 99.9%)
Reject:	otherwise

The guard band factor is (C.1)

$$k_w = \text{NORM.S.INV}(1 - PFA_{max}) = \text{NORM.S.INV}(0.999) = 3.09$$



Lower limit of speeding motorists is $T_L = v_0$ = 100 km/h and (C.2)

$$A_L = T_L + k_w.u$$

hence

$$v_{max} = v_0 + k_w \times (0.02 \times v_{max})$$

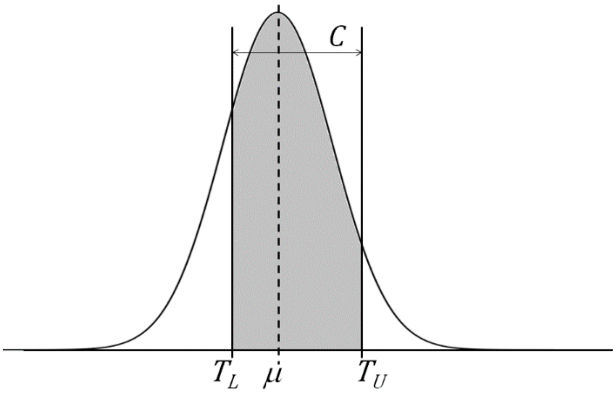
therefore

$$v_{max} = \frac{v_0}{1-0.02k_w} = \frac{100}{1-0.062} \approx 107 \text{ km/h}$$

To ensure that on average only 0.1% of drivers are falsely prosecuted the detected speed should be in excess of 107 km/h.

Example 12: Double sided tolerance limits (JCGM106 7.4)

A sample of SAE Grade 40 motor oil is required to have a kinematic viscosity at 100 °C of no less than 12.5 mm²/s and no greater than 16.3 mm²/s. The kinematic viscosity of the sample is measured at 100 °C, yielding a best estimate μ = 13.6 mm²/s and associated standard uncertainty u = 1.8 mm²/s.



The conformance probability p_c is represented by the portion of the PDF within the tolerance interval C i.e. (B.3)

$p_c = \mathrm{NORM.DIST}(16.3, 13.6, 1.8, \mathrm{TRUE}) - \mathrm{NORM.DIST}(12.5, 13.6, 1.8, \mathrm{TRUE}) = 0.66$

i.e. the conformance probability for this oil sample is 66 %

If a decision is taken to accept it as conforming the probability of false acceptance is (B.7)

$PFA = 1 - p_c = 34\%$

Possible Decision Rules for this conformity decision might therefore be defined in terms of p_c or PFA, for example:

ACCEPT: when $p_c \geq 0.6$, or

ACCEPT: when $PFA \leq 40\%$

REJECT: otherwise

This result might then be reported as:

Conforming, having a conformance probability of 66 %

or

Conforming, having a probability of false acceptance of 34 %

However, if the associated standard uncertainty was u = 2.2 mm^2/s this would instead result in a conformance probability of

$$p_c = \mathrm{NORM.DIST}(16.3, 13.6, 2.2, \mathrm{TRUE}) - \mathrm{NORM.DIST}(12.5, 13.6, 2.2, \mathrm{TRUE}) = 0.58$$

i.e. the conformance probability for this (same) oil sample is 58 %

If a decision is taken to accept it as conforming, the probability of false acceptance is

$$PFA = 1 - p_c = 42\%$$

Applying the same Decision Rule, this result might then be reported as:

Not conforming, having a conformance probability of only 58 %

or

Not conforming, unable to meet PFA requirements

Suppose that instead of suggesting a Gaussian distribution for likely values of the measurand, the uncertainty evaluation indicates that a t-distribution is more appropriate, having say $\nu = 3$ degrees of freedom. For the original example above, the conformance probability is now found (B.6) to be

$$p_c = \mathrm{T.DIST}\left(\left(\frac{16.3-13.6}{1.8}\right), 3, \mathrm{TRUE}\right) - \mathrm{T.DIST}\left(\left(\frac{12.5-13.6}{1.8}\right), 3, \mathrm{TRUE}\right) = 0.593$$

(which in this case, according to the possible Decision Rule proposed, would change the conformity decision from Accept to Reject).

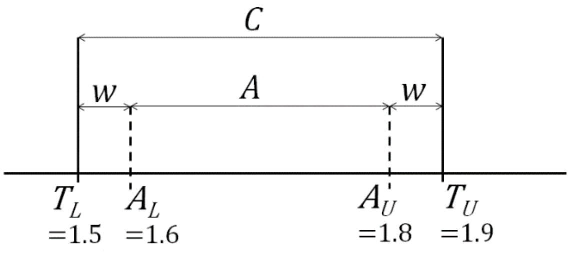
Example 13: Double sided tolerance limit, DR1: $w = 2u$; DR2: 'Constrained' Simple Acceptance with $u \leq$ a limit

Suppose that requirements for a material specify that the surface roughness r for a sample should be in the range $(T_L = 1.5) \leq r \leq (T_U = 1.9)$

Two possible decision rules that might be considered are…

DR1: A guard band of $w = 2.u$ at each side of the tolerance interval. i.e. accept as conforming all results r where $A_L \leq r \leq A_U$ (with limits $A_L = (T_L + w)$ and $A_U = (T_U - w)$)

For standard uncertainty u = 0.05 this corresponds to acceptance when 1.6 ≤ r ≤ 1.8



DR2: Simple Acceptance $(A_L = T_L) \leq r \leq (A_U = T_U)$, AND $u \leq 0.05$; or equivalently

Simple Acceptance $(A_L = T_L) \leq r \leq (A_U = T_U)$, AND Measurement Capability index $C_{95} \geq 2$, where $C_{95} = (T_U - T_L)/(2.U_{95})$

Outcomes for some possible measured values r with u = 0.05

r	Decision DR1	Decision DR2	$PFA = 1 - p_c$ (associated with PASS decisions)
1.7	PASS	PASS	0.01%
1.75	PASS	PASS	0.14%
1.8	PASS	PASS	2.3%
1.85	FAIL	PASS	16%
1.9	FAIL	PASS	50%
>1.9	FAIL	FAIL	

Note that:

DR1 - Risk can be stated as "PFA no worse than 2.3%" for all measured values 1.6 ≤ r ≤ 1.8

(For a PFA 'no worse than 5%' , narrower guard bands of 1.645 u could be applied)

DR1 – rejection rate is higher than DR2

DR2 – Risk can be stated as "PFA no worse than 50%" for all measured values 1.5 ≤ r ≤ 1.9

The choice of Decision Rule may depend for example upon how important it is to maintain low PFA, or how important it is to maintain a low rejection rate.

Example 14: Inspection of levels (conformity decisions for discrete measurements)

In its simplest form, the basic GUM law of propagation of uncertainties (LPU) approach to uncertainty evaluation is based upon two *assumptions*: the Central Limit Theorem applies i.e. the 'output' probability density function is taken to be Gaussian for the combination of 'input' quantities; and the variance in the output (the square of the standard uncertainty) is the sum of variances for the input quantities.

When these two assumptions apply, calculating the probability of conformity with a specification is usually a matter of establishing the proportion of the 'output' Gaussian distribution that overlaps the specification.

It is often incorrectly assumed that GUM LPU *always* applies or that it is 'close enough' that it can always be used. In fact, this is not the case and various methods are available to establish a 'better' understanding or representation of the uncertainty (e.g., Welch-Satterthwaite approach for dominant type A contributions with low degrees of freedom).

The GUM allows for other situations to apply and allows other statistically valid means of evaluation within the general GUM framework. Such an approach is necessary in the case highlighted below. The approach makes use of the probabilistic nature of uncertainty evaluation and allows for the discrete nature of the measurements.

Example measurement and conformity scenario

Suppose that a measurement can have only discrete values on a progressive scale of distinct levels. For example, visual evaluation of colour-fastness for a textile sample when compared against a reference scale, and the specification is stated in terms of acceptable levels.

<u>Example a:</u>

Suppose that the proficiency of the examiner has been demonstrated to be 100 % during a pre-test exercise (e.g., by successfully sorting a selection of reference items into the correct order). Suppose further that, when performing the examination, any remaining uncertainty is entirely determined by the ability of the examiner to resolve adjacent levels, so that when the result is level 'm' there is an equal probability of the 'true' level being $(m-1)$, m, or $(m+1)$

Specification: A conforming result will be at or between levels a and b

Decision Rule: A Simple Acceptance rule is to be applied. In addition, measurement uncertainty must be "entirely determined by the ability to resolve *adjacent* levels i.e. if measurement result is level 'm' then there is an *equal probability* of the 'true' level being $(m-1)$, m, or $(m+1)$"

Numerical example:

Suppose a colour fastness scale is defined as (0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0…)

Suppose also that the specification is that result must be 2.0 ± 0.5, i.e., conforming values are 1.5, 2.0, or 2.5

Now suppose that measurement result is 1.5

As this result is within specification the result 'conforms' (Simple Acceptance criteria).

If the Decision Rule is provided *<u>to</u>* the laboratory there is no (ISO 17025:2017) requirement upon them to evaluate the associated risk. The result can simply be reported as 'conforming' in terms of the associated specification and Decision Rule provided that the stated uncertainty criteria are met.

If, however the Rule is defined *by* the laboratory then the Risk is established as follows… for the observed result (1.5) there are three possible 'true' values that are *equally probable* according to our knowledge of the uncertainty. These are (1.0, 1.5, 2.0). Of these possible values two are conforming (1.5 and 2.0) and one is not (1.0). The probability of conformity is therefore 2/3 i.e., 66.7%, and the probability of false acceptance is 1/3 i.e., 33.3%.

Suppose instead that the result was 2.0. In this case, the three possible 'true' values that are equally probable according to our knowledge of the uncertainty are (1.5, 2.0, 2.5). Of these possible values all three are conforming so the probability of conformity is therefore 100%.

This probability of conformity of course depends upon the fact that the uncertainty is "entirely determined by the ability to resolve *adjacent* levels". If there is any possibility that the uncertainty could be larger the '100%' claim cannot be made (although it may in practice be 'approximately 100%').

For completeness, if the result was 2.5 the three possible 'true' values that are equally probable according to our knowledge of the uncertainty are (2.0, 2.5, 3.0). Of these possible values two are conforming. The probability of conformity is therefore 2/3 i.e., 66.7%, and the probability of false acceptance is 1/3 i.e., 33.3%.

On average, for all conforming results this example has probability of conformity, p_c of 78% i.e., a PFA of 22%.

Other possible scenarios might exist.

Example b:

As for Example (a) except that uncertainty is such that the observed level m is twice as likely as an adjacent level: $p(m-1) = 0.25$, $p(m) = 0.5$, $p(m+1) = 0.25$

The numerical example then gives for conforming results:

Result	p_c	PFA
1.5	75%	25%
2.0	100%	0%
2.5	75%	25%

On average, for all conforming results this example has probability of conformity, p_c of 83% i.e. a PFA of 17%

Example c:

As for Example (a) except that the specification now only permits two acceptable levels (1.5, 2.0)

The numerical example then gives for conforming results:

Result	p_c	PFA
1.5	67%	33%
2.0	67%	33%

On average, for all conforming results this example has probability of conformity, p_c of 67% i.e. a PFA of 33%

Example d:

As for Example (b) except that the specification now only permits two acceptable levels (1.5, 2.0)

Numerical example then gives for conforming results:

Result	p_c	PFA
1.5	75%	25%
2.0	75%	25%

On average, for all conforming results this example has probability of conformity, p_c of 75% i.e. a PFA of 25%

Appendix A: Glossary

Terminology used in this document is consistent with ISO/IEC Guide 98-4:2012 (JCGM 106).

T_U upper limit of conformity

T_L lower limit of conformity

C tolerance interval, corresponding to conforming values for a *measurand*, usually described in terms of a ‘tolerance’ or ‘specification’

C_m measurement capability index

C_{95} measurement capability index, defined in terms of expanded uncertainty U_{95} $C_{95} = (T_U - T_L)/(2.U_{95\%})$

A_U upper limit of acceptance

A_L lower limit of acceptance

A acceptance interval, corresponding to *measured values* that are accepted as demonstrating conformity for the measurand.

Y variable used to represent a measurand

η variable describing possible values of a measurand Y

y_m measured estimate of the value of the measurand

u, u_m standard uncertainty associated with the measured quantity

$U_{95\%}$ expanded uncertainty for 95 % coverage probability

PFA probability of false acceptance, sometimes known as ‘consumers risk’

PFR probability of false rejection, sometimes known as ‘producers’ risk’

p_c conformance probability

k_w guard band factor, used to define a guard band w as a multiple of standard uncertainty $w = k_w.u$

CAUTION: The guard band factor k_w should not be confused with the coverage factor (often written as k, k_p, k_{95}) that is used to establish an expanded uncertainty (although in practice it may be numerically equal).

Decision Rule (DR): documented rule that describes how measurement uncertainty will be accounted for with regard to accepting or rejecting an item, given a specified requirement and the result of a measurement

Simple Acceptance (SA): condition under which an Acceptance Interval is defined to be the same as a tolerance interval, $A = C$. Simple Acceptance *on its own* does not constitute a Decision Rule (as explained in Appendix D)

Guard band: interval between conformity limit and acceptance limit, usually defined by some multiple of the standard uncertainty with the purpose of limiting the risk of false acceptance.

Appendix B: Conformance probability and risk

The values of Y that fall within the tolerance interval C represent conforming values of the measurand that could have given rise to the measurement result. For a Gaussian distribution, the area under the PDF defined by these values is the conformance probability p_c

$$p_c = \int_C g(\eta; y_m, u_m) d\eta$$

For example, in the figure below the *shaded* region of the PDF is within the tolerance interval and represents *conforming* values of the measurand that can be associated with the measurement result. Whereas the unshaded region represents *non-conforming* values of the measurand that can similarly *also* be attributed to the measurement result.

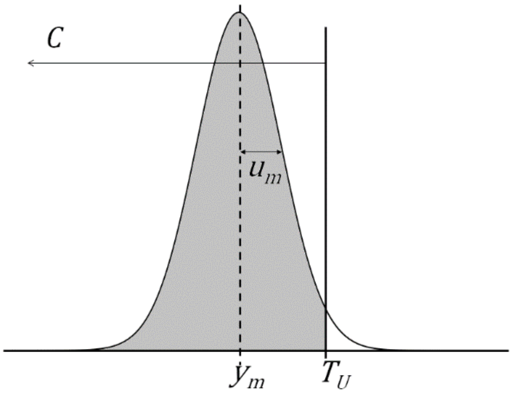


Figure 1: A measured value within a tolerance interval that is defined by a single upper limit

Definite integrals of the Gaussian PDF can be calculated using the Excel function NORM.DIST where

$$\int_{-\infty}^{T} g(\eta; y_m, u_m) d\eta \; = \; \mathrm{NORM.DIST}(T, y_m, u_m, \mathrm{TRUE})$$

Conformance probability for an *upper* limit is therefore

$$p_c = \mathrm{NORM.DIST}(T_U, y_m, u_m, \mathrm{TRUE}) \qquad \text{B.1}$$

Similarly, conformance probability for a *lower* limit is

$$p_c = 1 - \mathrm{NORM.DIST}(T_L, y_m, u_m, \mathrm{TRUE}) \qquad \text{B.2}$$

And conformance probability for a *two-sided* limit is therefore

$$p_c = \mathrm{NORM.DIST}(T_U, y_m, u_m, \mathrm{TRUE}) - \mathrm{NORM.DIST}(T_L, y_m, u_m, \mathrm{TRUE}) \qquad \text{B.3}$$

For example, suppose that in the case shown above $T_U = 1.96$, $y_m = 0$, $u_m = 1$ then

$$p_c = \int_{-\infty}^{T_U} g(\eta; y_m, u_m) d\eta \; = \; \mathrm{NORM.DIST}(1.96{,}0{,}1, \mathrm{TRUE})$$

i.e.

$$p_c = \mathrm{NORM.DIST}(1.96, 0, 1, \mathrm{TRUE}) = 0.975 = 97.5\%$$

The corresponding equations for calculating conformance probability for a t-distribution with ν degrees of freedom are:

Conformance probability for an ***upper*** limit

$$p_c = \mathrm{T.DIST}\left(\left(\frac{T_U - y_m}{u}\right), \nu, \mathrm{TRUE}\right) \tag{B.4}$$

Conformance probability for a ***lower*** limit

$$p_c = 1 - \mathrm{T.DIST}\left(\left(\frac{T_L - y_m}{u}\right), \nu, \mathrm{TRUE}\right) \tag{B.5}$$

And conformance probability for a ***two-sided*** limit

$$p_c = \mathrm{T.DIST}\left(\left(\frac{T_U - y_m}{u}\right), \nu, \mathrm{TRUE}\right) - \mathrm{T.DIST}\left(\left(\frac{T_L - y_m}{u}\right), \nu, \mathrm{TRUE}\right) \tag{B.6}$$

For example, suppose that in the case shown above $T_U = 1.96$, $y_m = 0$, $u_m = 1$ and $\nu = 3$ then

$$p_c = \mathrm{T.DIST}\left(\left(\frac{1.96 - 0}{1}\right), 3, \mathrm{TRUE}\right) = 0.928 = 92.8\%$$

With knowledge of the conformance probability, it becomes possible to evaluate the risk associated with a decision to accept or reject a result. For example, consider a decision based upon a measurement of some property of an item whose specification has a lower tolerance limit, T_L defining a single-sided tolerance interval C: $[T_L, \infty)$. When the measured value y_m is close to the limit value, a proportion of the PDF can be located both above and below the limit. Two scenarios are possible in this case:

a. The measured value is *within* the tolerance interval i.e. $y_m \geq T_L$

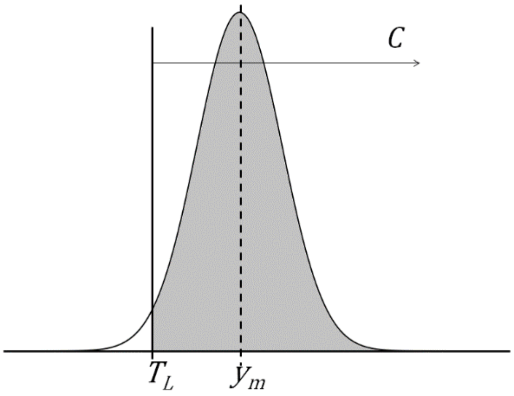


Figure 2: measured value within a tolerance interval that is defined by a single lower limit

The value of y_m suggests that the item *does* conform, however there are possible values for the measurand (unshaded region) that are *not* conforming. If a decision is taken that the item *is* conforming this (unshaded) area represents the probability of false acceptance (PFA).

b. The measured value is outside the tolerance interval i.e. $y_m < T_L$

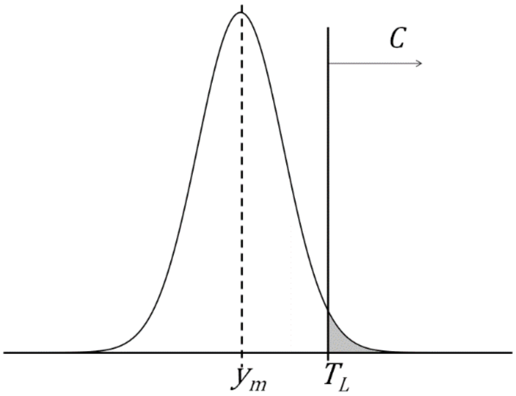


Figure 3: measured value outside a tolerance interval that is defined by a single lower limit

The value of y_m suggests that the item does *not* conform however there are possible values for the measurand that *are* conforming. If a decision is taken that the result is non-conforming this (shaded) region represents the probability of false rejection (PFR).

Note that in some situations the probabilities of false acceptance and false rejection are called the 'specific consumers risk' and 'specific producer's risk' – when an item has been falsely accepted as conforming it is the consumer that bears the cost, whereas when an item is falsely rejected it is the producer who bears the cost.

These examples illustrate the relationship between conformance probability and the associated 'specific' risks:

$PFA = 1 - p_c$ (applicable only when conformity has been accepted) B.7

$PFR = p_c$ (applicable only when conformity has been rejected) B.8

For example, suppose that in case a) $T_L = 0$, $y_m = +1.64$, $u_m = 1$ then

$$p_c = \int_{T_L}^{\infty} g(\eta; y_m, u_m) d\eta = 1 - \int_{-\infty}^{T_L} g(\eta; y_m, u_m) d\eta = 1 - \mathrm{NORM.DIST}(0{,}1.64{,}1, \mathrm{TRUE})$$

i.e. $p_c = 1 - \mathrm{NORM.DIST}(0, 1.64, 1, \mathrm{TRUE}) = 0.95 = 95\%$

If in this case a decision was made to 'Accept', the probability of false acceptance would be $PFA = 1 - p_c = 5\%$

because statistically speaking, 5% of the possible non-conforming values for the measurand could have resulted in the 'conforming' result y_m

Similarly, suppose that in case b) above we have $T_L = 0$, $y_m = -1.64$, $u_m = 1$ then

$$p_c = \int_{T_L}^{\infty} g(\eta; y_m, u_m) d\eta = 1 - \int_{-\infty}^{T_L} g(\eta; y_m, u_m) d\eta = 1 - \mathrm{NORM.DIST}(0, -1.64{,}1, \mathrm{TRUE})$$

i.e. $p_c = 1 - \mathrm{NORM.DIST}(0, -1.64, 1, \mathrm{TRUE}) = 0.05 = 5\%$

In this case if a decision to 'Reject' was made, the probability of false rejection would be $PFR = p_c = 5\%$

because statistically speaking, 5% of the possible conforming values for Y could have resulted in the 'non-conforming' result y_m

To restrict or minimise the risk of making incorrect decisions, constraints can be placed on the measured values that are accepted or rejected as conforming. These constraints define an acceptance interval A

The difference between the acceptance interval and the tolerance interval is the guard band w

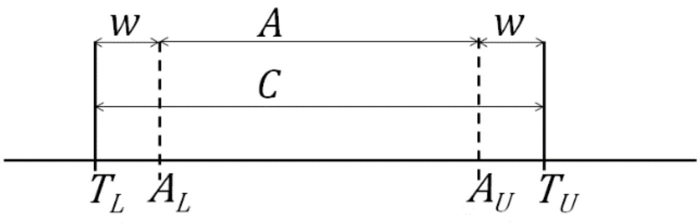


Figure 4: Tolerance interval C and 'stringent' acceptance interval A with associated guard bands, w

The choice of where to place the limits of the acceptance interval A_U and/or A_L either determines PFA_{max} or alternatively, the choice of PFA_{max} determines the acceptance limits. PFA_{max} is the largest value that PFA can have whilst the decision is to Accept, similarly, PFR_{max} is the largest value that PFR can have whilst decision is to Reject.

A common form of guard band is chosen to establish at least 95% confidence in the decision to accept a result[3] as conforming on the basis of a measured value y_m with associated standard uncertainty u_m i.e. the acceptance limits are chosen so that PFA_{max} = 5%. For a single-sided tolerance interval this corresponds to a guard band $w = 1.645\, u_m$. For the same coverage probability, the *same* single-sided guard band factor applies for a double-sided tolerance interval when only one side of the PDF significantly overlaps a tolerance limit. If significant overlap of both limits occurs the procedure outlined in Appendix C can be followed to establish a suitable guard band factor.

Note that in situations where the measurement potentially results in a different value of u_m each time the measurement is performed, such as usually occurs in calibration scenarios, it is likely to be necessary to evaluate p_c (and hence PFA or PFR) on a case-by-case basis. In such situations the interval corresponding to PFA (or PFR) varies on a case-by-case basis and an acceptance *limit* cannot be defined *a priori* (i.e. before the measurement is performed and u_m is evaluated).

Similarly, if a guard band is arbitrarily defined or is not defined in terms of u_m – it will be necessary to calculate p_c in order to report PFA (or PFR).

However, if the uncertainty is known to be fixed, as is often the case for measurement scenarios such as production testing or other scenarios in which the uncertainty is dominated by the process itself, then a fixed acceptance limit can be calculated that corresponds to PFA_{max}.

In practical situations the precise value of PFA may sometimes not be of interest for an individual result. Conformity can then be established with $PFA \leq PFA_{max}$ by requiring only that y_m is within the acceptance interval A .

[3] Similar choices can be made to set PFR_{max} in situations where the purpose of the decision process is to decide whether to reject a result.

Appendix C: Guard-band factor k_w

Single-sided specifications

For single sided specifications, the required size of a guard band can be determined as a multiple of the standard uncertainty, $w = k_w.u$ where guard band factor k_w is found for a Gaussian PDF by solving the equation

$$1 - PFA_{max} = \int_{-\infty}^{k_w} g(\eta; 0,1) d\eta$$

Using Excel Worksheet functions this is

$$k_w = \mathrm{NORM.S.INV}(1 - PFA_{max}) \quad \text{C.1}$$

Hence

$$A_L = T_L + k_w.u \quad \text{C.2}$$

$$A_U = T_U - k_w.u \quad \text{C.3}$$

Table of guard band factors for selected values of PFA_{max} for a single limit and a Gaussian PDF

PFA_{max} /%	k_w
0.1	3.0902
0.2275	2.8373
0.25	2.8070
0.455	2.6083
0.5	2.5758
1.0	2.3263
2.275	2.0000
2.5	1.9600
4.55	1.6901
5.0	1.6449
10.0	1.2816

For PDFs based upon a t-distribution the corresponding Excel function is T.INV i.e.

$$k_w = \mathrm{T.INV}(1 - PFA_{max}, \upsilon) \quad \text{C.4}$$

Double-sided specifications

The single-sided guard band factor can often be applied to establish guard bands for double-sided intervals.

A check can be performed to ensure that the guard band is consistent with the stated maximum risk of false acceptance. If not, a different guard band factor can then be calculated.

The process is as follows:

1. identify PFA_{max}, T_U, T_L, and u
2. first, calculate PFA for a double-sided specification (using B.7 with B.3 or B.6 as appropriate) with $y_m = (T_U + T_L)/2$
3. if $PFA > PFA_{max}$ then it is not possible to define an Acceptance Interval consistent with PFA_{max}
4. otherwise… calculate the single-sided guard band factor k_w using C.1 or C.4 as appropriate.
5. calculate $A_L = T_L + k_w.u$ (and for later use calculate $A_U = T_U - k_w.u$)
6. calculate PFA for a double-sided specification (using B.7 with B.3 or B.6 as appropriate) with $y_m = A_L$
7. if $PFA \approx PFA_{max}$ then the PDF does not extend 'significantly'[4] beyond *both* limits for a value at the limit of acceptance. The *single-sided* guard band factor (C.1 or C.4) is therefore suitable for a *double-sided* specification at the proposed PFA_{max}
8. <u>if instead</u> PFA is 'significantly' larger than PFA_{max} a different guard band factor is required. Precise calculation of the factor is not a straightforward procedure, and it is probably best obtained empirically… e.g. by progressively increasing k_w and repeating steps 5. and 6. until an acceptable interval is found, i.e. $PFA \approx PFA_{max}$

<u>For example</u>, suppose that we require $PFA_{max} = 0.050$ for a Gaussian PDF with

$T_L = -4$

$T_U = 4$

$u = 1$

In this case we find that $PFA = 0.0500$ when $y_m = (T_U + T_L)/2$

therefore, a guard band for $PFA_{max} = 0.050$ exists

the single-sided guard band factor is

$k_w = \text{NORM.S.INV}(1 - 0.05) = 1.64485$

$A_L = -4 + 1.64485 \times 1 = -2.35515$

$A_U = 4 - 1.64485 \times 1 = 2.35515$

Hence, for a double-sided specification, when $y_m = A_L$ we calculate

$PFA = 0.05000$

Since $PFA \approx PFA_{max}$ we can use the single-sided guard band factor for the double-sided specification.

[4] Deciding what is 'significant' depends upon the application and cannot be dictated in advance for all situations.

Suppose instead that $u = 2$. Following the same steps, we now find that

$A_L = -4 + 1.64485 \times 2 = -0.71030$

$A_U = 4 - 1.64485 \times 2 = 0.71030$

and, for the double-sided specification, when $y_m = A_L$ we calculate

$PFA = 0.05926$

which we may decide is 'significantly' larger than PFA_{max}

If we increase k_w by small increments and repeat the calculations for each new value, we find that when

$k_w = 1.79$, $PFA = 0.05028$, $A_L = -0.42$, $A_U = 0.42$

$k_w = 1.80$, $PFA = 0.04983$, $A_L = -0.40$, $A_U = 0.40$

$k_w = 1.796$, $PFA = 0.05001$, $A_L = -0.408$, $A_U = 0.408$

allowing us to select an appropriate guard band factor.

Appendix D: The problem with allowing decision rules that do not take account of measurement uncertainty

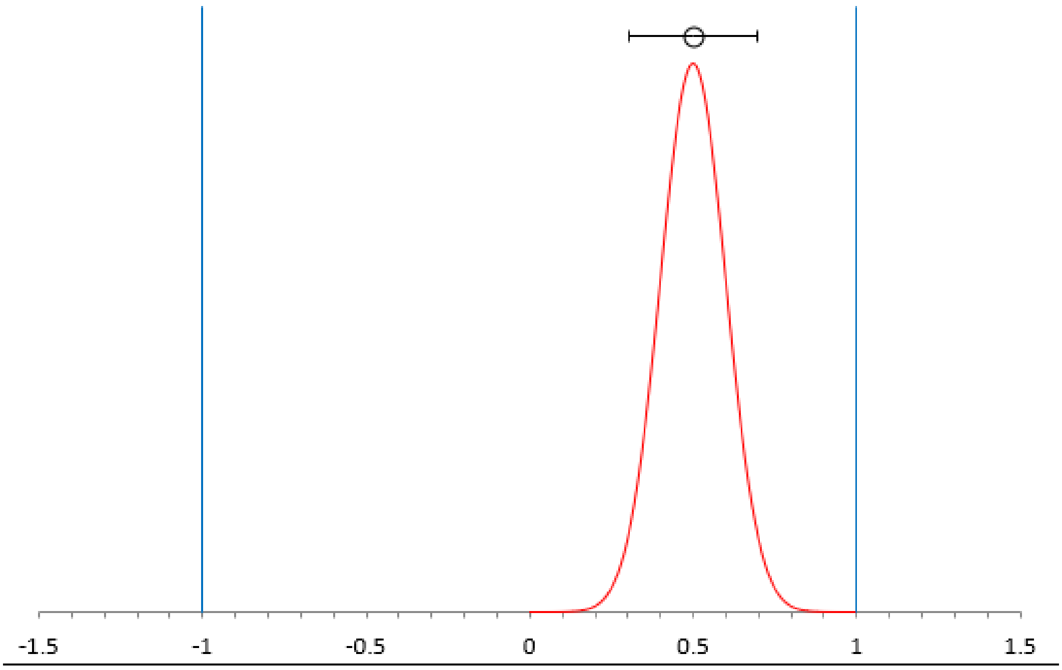
Conformity statements under ISO/IEC 17025:2017 require a Decision Rule (3.7) that takes account of measurement uncertainty. Some might argue that it is possible to ‘take account’ by ignoring it, if that is what the customer requests; however, this seems to require a rather contradictory belief that you can be ‘doing something’ by ‘not doing something’ (is it possible to ‘obey a red stop light’ by ‘not obeying a red stop light’?)

Despite the grammatical and logical inconsistency in this approach, others also argue that it is allowable because ‘the customer accepts the risk associated with ignoring uncertainty’. This too is a flawed argument as will be shown by a simple example…

Suppose that for some hypothetical reason Simple Acceptance with no account of measurement uncertainty *was* defined to be an acceptable Decision Rule i.e. PASS when the measured value is within the stated tolerance interval and uncertainty plays *no part* in the decision process…

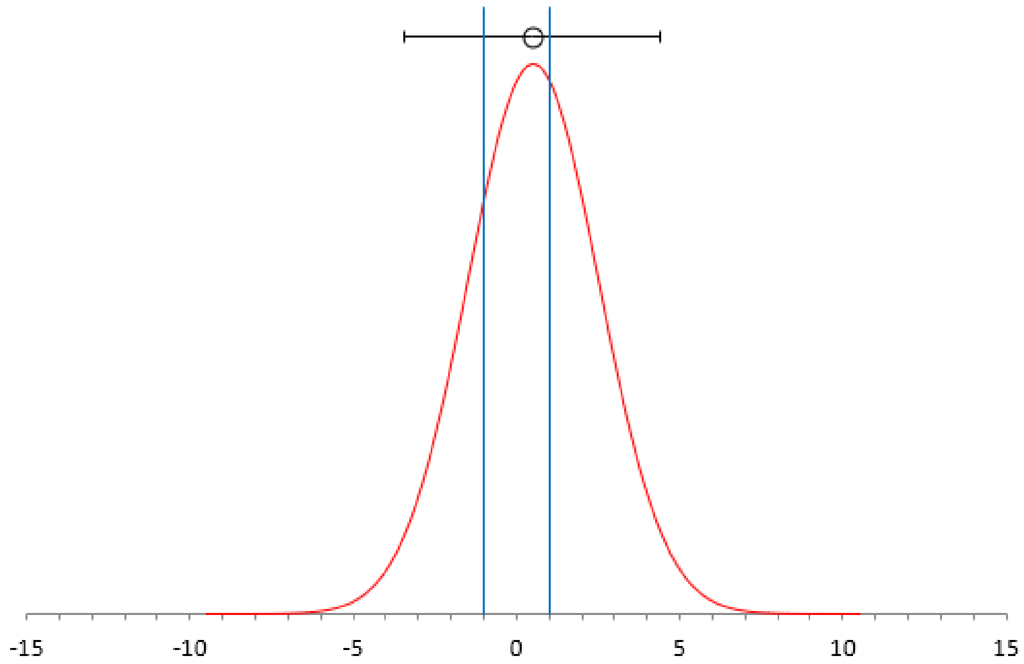
Suppose also that, for a particular measurement there is a tolerance of ±1 and the measured value equals 0.5

As the value is within the tolerance interval the result is therefore declared to be a PASS regardless of the measurement uncertainty

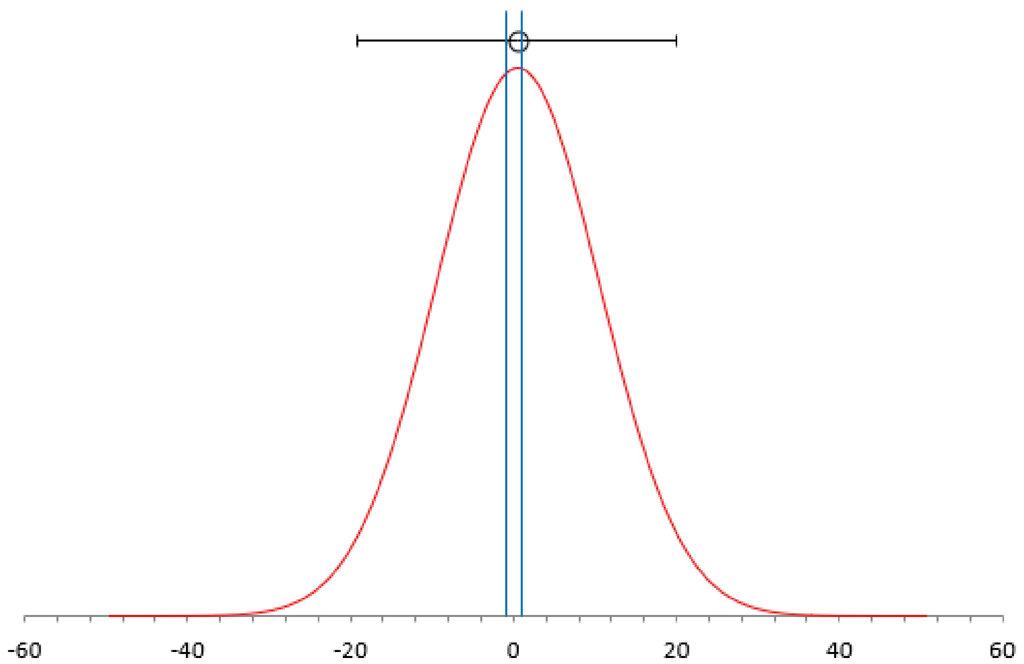


$u = 0.1$, $p_c = 100\%$: PASS according to Simple Acceptance rule with no account for u

In fact, *all of the following measurement scenarios* will result in a PASS according to this rule…



$u = 2$, $p_c = 37\%$: PASS according to Simple Acceptance rule with no account for u



$u = 10$, $p_c = 8\%$: PASS according to Simple Acceptance rule with no account for u

To reiterate, *all* of these scenarios (and an infinite number of others) are possible if there is "no account" taken of measurement uncertainty and the associated Risk will vary on a case-by-case basis.

It isn't therefore possible to 'accept the risk associated with ignoring uncertainty' as the risk is not only undefined, it is *undefinable when uncertainty is ignored*.

It *cannot* be argued in defence that "in practice this wouldn't be allowed to happen" and *at the same time* claim to "ignore uncertainty".

Suppose further, in this hypothetical situation, that a customer *did* understand that the risk was undefined and *still* wished to proceed, it begs the question 'for what legitimate purpose?' If (for some yet to be justified reason) such a Decision *were* to be allowed then, as in all cases, to avoid misrepresentation of the outcome the decision it would need to be accurately reported… for example:

> "Decision Rule: Simple Acceptance rule ignoring uncertainty, by which it is not possible to state any level of confidence or risk associated with the Decision"

It does not seem likely that this would be welcome, but to omit the final part of the sentence would misrepresent the basis for the Decision.

A further consequence of ignoring measurement uncertainty is that the outcome of such a conformity decision is not 'metrologically' traceable i.e. it *could not be used to provide traceability* for any subsequent measurement activity such as calibration, testing, inspection or certification.

It is not 'metrologically traceable' because it is not the result of an unbroken chain of measurements and associated uncertainties. In statistical terms it is not possible to establish a PDF for the measurand based upon a conformity statement using a rule that does not somehow, directly or indirectly, take account of measurement uncertainty.

Finally, it should also be noted that rules such as 'Simple Acceptance *ignoring* measurement uncertainty in the decision process but *reporting* measurement uncertainty together with the Decision outcome' are also *not* consistent with the ISO/IEC 17025:2017 definition of a *Decision Rule* - because uncertainty has not been involved in the decision process.

Reporting the uncertainty *post-decision* might allow risk to *subsequently* be evaluated, but it has not influenced the *earlier* decision to accept or otherwise – it therefore represents a situation where a decision is made regardless of risk.

The solution…

Often, in circumstances where 'the customer asks' the laboratory to 'ignore uncertainty' it is because they do not have sufficient understanding of uncertainty or of risk to realise what they are asking for. Usually however the customer actually does have some unarticulated belief about the appropriateness of the measurements - in other words there is some *implicit* idea of a point beyond which the uncertainty is too large.

Quantitatively establishing and applying that 'point' *takes measurement uncertainty into account*.

Simple Acceptance criteria can be therefore used as the basis for identifying the acceptance interval provided that it is used *together with* an identified constraint on the uncertainty, for example by agreeing an upper limit for measurement uncertainty or agreeing a limit to the measurement capability index.

Agreeing the limits for measurement uncertainty is a matter for review between the laboratory and their customer. The laboratory might for example point out that, being an accredited laboratory, they have *already* established values for the likely uncertainty of all key measurements...

To summarise:

A rule such as Simple Acceptance with no account for measurement uncertainty is ***not*** an appropriate Decision Rule under ISO/IEC 17025:2017.

- At best it would simply be technically worthless, having undefinable risk and providing no metrological traceability
- At worst it is misleading, using a laboratory's accreditation status to pass off a meaningless decision as something more credible than it is

Rules based upon Simple Acceptance criteria can be a part of a valid Decision Rule when used *together with* indirect accounting for measurement uncertainty. Under these conditions

- It provides traceability (a PDF can be established if required)
- There is a definable risk in the decision outcome

Reference documents

ISO/IEC 17025:2017 “General requirements for the competence of testing and calibration laboratories”

ILAC-G8:09/2019 “Guidelines on Decision Rules and Statements of Conformity”

ISO/IEC Guide 98-4:2012 (JCGM 106) “Uncertainty of measurement - Part 4: Role of measurement uncertainty in conformity assessment”

ISO/IEC Guide 98-3:2008 (JCGM 100) “Guide to the expression of uncertainty in measurement (GUM)”

ISO/IEC Guide 99:2007 (JCGM 200) “International vocabulary of metrology – Basic and general concepts and associated terms (VIM)”

ISO 10576-1:2003 “Statistical methods – Guidelines for the evaluation of conformity with specified requirements”

ISO 14253-1:2017 “Geometrical product specifications (GPS) - Inspection by measurement of workpieces and measuring equipment Part 1: Decision rules for verifying conformity or nonconformity with specifications (BS EN ISO 14253-1:2017)”

ISO 21748:2017 “Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation”

ISO 5725-2:2019 “Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method”

UKAS, M3003 “The Expression of Uncertainty and Confidence in Measurement (Edition 4, October 2019)”

“GUM-LPU uncertainty evaluation - importing measurement traceability from a conformity statement”

John Greenwood, Alen Bosnjakovic, Vedran Karahodzic, Paola Pedone, Fabrizio Manta and Maurice Cox

https://zenodo.org/record/4793471/files/EMUE_1_2_5.pdf?download=1