Using measurement uncertainty in decision-making and conformity assessment

L R Pendrill

SP Technical Research Institute of Sweden, Measurement Technology, Box 857, SE-50115 Borås (SE), Sweden
E-mail: leslie.pendrill@sp.se

Received 28 February 2014, revised 17 April 2014
Accepted for publication 25 April 2014
Published 11 July 2014

Abstract
Measurements often provide an objective basis for making decisions, perhaps when assessing whether a product conforms to requirements or whether one set of measurements differs significantly from another. There is increasing appreciation of the need to account for the role of measurement uncertainty when making decisions, so that a ‘fit-for-purpose’ level of measurement effort can be set prior to performing a given task. Better mutual understanding between the metrologist and those ordering such tasks about the significance and limitations of the measurements when making decisions of conformance will be especially useful. Decisions of conformity are, however, currently made in many important application areas, such as when addressing the grand challenges (energy, health, etc), without a clear and harmonized basis for sharing the risks that arise from measurement uncertainty between the consumer, supplier and third parties.

In reviewing, in this paper, the state of the art of the use of uncertainty evaluation in conformity assessment and decision-making, two aspects in particular—the handling of qualitative observations and of impact—are considered key to bringing more order to the present diverse rules of thumb of more or less arbitrary limits on measurement uncertainty and percentage risk in the field.

(i) Decisions of conformity can be made on a more or less quantitative basis—referred in statistical acceptance sampling as by ‘variable’ or by ‘attribute’ (i.e. go/no-go decisions)—depending on the resources available or indeed whether a full quantitative judgment is needed or not. There is, therefore, an intimate relation between decision-making, relating objects to each other in terms of comparative or merely qualitative concepts, and nominal and ordinal properties. (ii) Adding measures of impact, such as the costs of incorrect decisions, can give more objective and more readily appreciated bases for decisions for all parties concerned. Such costs are associated with a variety of consequences, such as unnecessary re-manufacturing by the supplier as well as various consequences for the customer, arising from incorrect measures of quantity, poor product performance and so on.

Keywords: uncertainty, decisions, risks, conformity

1. Introduction
Measurement is in most cases not an end in itself, but rather provides the means to make objective decisions, such as ‘do the new set of measurements differ from previous measurements?’ or ‘do measurements show that a product satisfies requirements?’ In attempting to answer questions such as these, the role of measurement uncertainty needs to be accounted for, particularly since a finite uncertainty will lead to risks of incorrect decisions.

There is increasing appreciation of the fact that better understanding of the role of measurement uncertainty in conformity assessment will

• aid the metrologist in setting a ‘fit-for-purpose’ level of measurement effort prior to performing a given task;
The role of measurement uncertainty in high-value manufacturing is reviewed in another paper in this special issue (Loftus and Giudice 2014). Decisions of conformity are currently made in many important application areas, such as environmental monitoring, the health sector and product safety testing, but without a clear and harmonized basis for sharing the risks that arise from measurement uncertainty between the consumer and the supplier.

The state of the art of uncertainty evaluation in decision-making and conformity assessment will be reviewed in this paper. Starting with published guides such as the recent ICGM 106:2012 guide, and current work in the EURAMET EMRP project NEW04 ‘Uncertainty’, new perspectives are being gained about the use of measurement uncertainty by extending analyses to multivariate, qualitative data, and the inclusion of measures of impact.

2. Essential steps in conformity assessment

In this review, how well the major recent guides and other publications about the role of measurement uncertainty in conformity assessment deal with the following essential steps for conformity assessment will be appraised:

(a) Define entity and its quality characteristics to be assessed for conformity with specified requirements (section 3).

(b) Set corresponding specifications on the measurement methods and their quality characteristics (such as maximum permissible uncertainty and minimum measurement capability) required by the entity assessment at hand (section 4).

(c) Produce test results by performing measurements of the quality characteristics together with expressions of measurement uncertainty (section 5).

(d) Decide if test results indicate that the entity and the measurements themselves are within specified requirements or not (section 6).

(e) Assess risks of incorrect decisions of conformity (sections 7 and 8).

(f) Final assessment of conformity of entity to specified requirements in terms of impact (sections 9 and 10).

3. Product (entity) specifications

The prime, overall motivation for conformity assessment is useful to have in mind from the start of any measurement, namely, that it

- provides confidence for the consumer that requirements placed on products and services are met;
- provides the producer and supplier with useful tools to ensure product quality; and
- is often essential for reasons of public interest, public health, safety and security, protection of the environment and the consumer and of fair trading (ISO 2013).

Conformity assessment, broadly defined, is any activity undertaken to determine, directly or indirectly, whether an entity (product, process, system, person or body) meets relevant standards or fulfils specified requirements. Decision-making about whether new measurements are consistent with earlier observations is included here as a special case of conformity assessment.

In evaluating product variations of the quality characteristic $\eta = Z$ in the ‘entity (or product) space’, measurements might be made on repeated items in a production process or by taking a sample of the population of items subject to conformity assessment. Products measured could be being manufactured in an industry or when subject to the wear and tear of use by the consumer. The corresponding probability density function (PDF), $f_{\text{entity}}(z)$, of the product characteristic will have a form determined ideally (in the absence of measurement or sampling uncertainty) by the variations in the inherent characteristic of product, process or system of prime interest in conformity assessment.

The established discipline of statistical quality control, including hypothesis testing on process parameters (with point and continuous estimators), is described extensively in the literature, for instance by Montgomery (1996). Tolerances on products can be set in terms of

- specification limits, $U_{\text{SL}}$ and $L_{\text{SL}}$, for the magnitude of a characteristic of any entity;
- for any entity, the maximum permissible (entity) error, MPE;
  - for a symmetric, two-sided tolerance interval: $\text{MPE} = (U_{\text{SL}} - L_{\text{SL}})/2$;
  - for a one-sided interval: $\text{MPE} = U_{\text{SL}} - \text{nominal}$, for instance, the interval between an upper specification limit and the nominal value of the characteristic.

As a principal requirement for conformity assessment in process control (Montgomery 1996), to ensure that product lies within specifications, limits on process capabilities are traditionally set in terms of minimum values of the process capability index, $C_p$, defined as

$$C_p = \frac{U_{\text{SL}} - L_{\text{SL}}}{N \cdot s_p} \quad (1)$$

where process variation characterizes with a standard deviation $s_p$ of estimated product variations and where $N = 6$ (corresponding to a coverage factor, $k = 3$ and 99% confidence) in the famous ‘six-sigma’ approach (Joglekar 2003) to statistical process control (SPC).

Measurement uncertainty has normally been assumed in these contexts to be negligibly small (Grubbs and Coon 1954).

4. Measurement specifications

Since neither the production nor measurement processes are perfect, there will always be some dispersion in the observed
product value either for repeated measurements of one item or for measurements of a series of items. Measurement uncertainty in a test result—an apparent product dispersion arising from limited measurement quality—can be a concern in conformity assessment by inspection since, if not accounted for, uncertainty can

- lead to incorrect estimates of the consequences of entity error;
- increase the risk of making incorrect decisions, such as failing a conforming entity or passing a non-conforming entity when the test result is close to a tolerance limit.

Practically in cases where measurement dispersion is of comparable size to actual product variation, it can be difficult to separate these (Kacker et al 1996, Rossi and Crenna 2006). This separation is, however, essential

- if one overestimates measurement dispersion; then actual dispersion in product values will not be detected and thus lead to poorer product quality. Additional costs will also be incurred if it is decided, on the basis of estimated poor measurement quality, to spend more on (unnecessarily) improving measurement resources.
- An underestimation of measurement dispersion will lead to unnecessary adjustment of the production process and thus to increased production costs together with poorer product quality, where spurious measurement dispersion is transferred to product dispersion.

4.1. Separating production and measurement errors—numerically

Accounting for risks such as those arising from measurement uncertainty and associated decision rules in conformity assessment are the main subject of this work. Despite the importance of this area, ‘there is no single method used to integrate the uncertainty of measurement into the decision-making process. The decision rules differ between products, fields of measurement, profession and countries’, quoting from the French standard on the use of uncertainty FD x07-022 (AFNOR 2004).

There are several ways of ensuring that limited measurement quality does not adversely affect conformity assessment. Risks and the consequences of incorrect decision-making in conformity assessment can be minimized with the following three steps:

(A) set limits on maximum permissible measurement uncertainties (equivalently, minimum measurement capability) and on maximum permissible consequence costs at the specification stage of any task (sections 5 and 8);
(B) agree on acceptable locations of the uncertainty interval with respect to a specification limit (sections 6 and 7);
(C) optimize measurement uncertainty proactively, ahead of a series of measurements, by designing experiments so that the sum of costs of measurement and of incorrect decisions of conformity is at a minimum (sections 9 and 10).

4.2. Separating production and measurement errors—conceptually

Some of the difficulty in separating production and measurement dispersion has to do with a lack of clarity in concepts, definitions and nomenclature which arises at the interface where two disciplines—metrology and conformity assessment—meet.

Two principally distinct, but closely related and easily confounded concepts coming from the two disciplines are, respectively,

- a ‘measurand’ is a quantity intended to be measured;
- a ‘quality characteristic’ is a quantity intended to be assessed.

This dichotomy is clearly illustrated in the concept diagram: figure B.16 of ISO 3534-2:2006 Applied Statistics, where the distinct pairs ‘quantity: measurement result’ and ‘characteristic: test result’ are juxtaposed.

Within metrology, Giordani and Mari (2012) have recently emphasized the conceptual distinction between measurement of a ‘general’ quantity—such as ‘length’—and of an ‘individual’ quantity—such as the ‘length of a specific object’, with reference to the JCGM 200:2012(E/F) VIM.

- Measurement of a ‘general’ quantity is much the domain of the physicist (Pendrill 2005) where relations (laws of Nature) amongst such quantities (which also give the corresponding relations amongst the measurement units associated with them) are fundamentally applicable, irrespective of particular objects (as for instance in Newtonian mechanics as applied to all bodies universally, from microscopic and cosmological scales). This ‘general’ perspective of measurement and quantities is clear for instance in the concept diagram for part of Clause 1 around ‘quantity’ of the VIM, where no explicit mention is made of either an entity or its quality characteristic. Indeed, throughout the metrology vocabulary VIM, there is no mention of ‘requirements’ in general, apart from special instances, such as ‘maximum permissible measurement error’ (VIM, section 4.26).
- In contrast, when dealing with conformity assessment, the focus is instead on the measurement (testing) of a specific object—i.e. of an ‘individual’ quantity, in particular singling out (usually a few) particular ‘quality’ characteristics of an object which are to be assessed with respect to specifications since they are judged to be essential in assuring quality of the product.

Because of the different points of departure of each discipline—‘measurement’ for the metrologist and ‘product’ for the conformity assessor—the role of measurement uncertainty in conformity assessment can be treated differently, as is evident from existing guides on the subject. Clear definitions of key conformity assessment concepts such as ‘entity’ and ‘quality characteristic’ can be found for instance in ISO 10576-1:2003 and their systematic use in the metrological literature can reduce confusion in the field. The entity subject to conformity assessment may be a single item, a collective sample of items or maybe not even a physical object but a
service. Irrespective of which entities are subject to conformity assessment, it is important to specify the assessment target as clearly as possible: ‘global’ conformity denotes the assessment of populations of typical entities, while ‘specific’ conformity assessment refers to inspection of single items or individuals, as defined by Rossi and Crenna (2006).

5. Measurement conformity assessment: limits on measurement capability factors

Confidence in the measurements performed in the conformity assessment of any entity (product, process, system, person or body) can be considered sufficiently important that the measurements themselves will be subject to the steps (a) to (f) above (section 2) as a kind of metrological conformity assessment ‘embedded’ in any product conformity assessment.

The corresponding specified requirements on measurement in that case are, respectively:

- limits on maximum permissible measurement uncertainty (or, equivalently, minimum measurement capability) when testing product;
- limits on maximum permissible error in the indication of the measurement equipment/system intended to be used in the measurements when testing product.

5.1. Maximum permissible uncertainty and minimum measurement capability

Variations associated with limited measurement quality, expressed in terms of a measurement uncertainty PDF, \( g_{\text{meas}}(x) \) of the quantity \( \xi = X \) in the ‘measurement space’, i.e. the measurand, may partially mask observations of actual entity quality characteristic dispersion with PDF, \( g_{\text{entity}}(c) \) introduced in section 3. To make clear the essential distinction between measurement variations and the quality characteristic variations that are the prime focus of conformity assessment, two different notation—\( X \) and \( Z \), respectively—have been deliberately chosen. As such, measurement variability is just one, and hopefully a relatively minor, source of uncertainty that needs to be accounted for when making decisions of conformity.

A first step to minimizing the effects of less than perfect measurement quality on conformity assessment is simply to set a limit proactively, before starting the measurements, on how large measurement uncertainty is allowed to be (Rukhin 2013). This limit is often expressed as a maximum permissible uncertainty or ‘target uncertainty’, MPU = \( 1/C_{m,\text{min}} \) in terms of a corresponding minimum measurement capability.

In a manner analogous to process control (section 3), a measurement capability index, \( C_m \), can be defined in terms of estimated measurement variations as

\[
C_m = \frac{U_{\text{SL}} - L_{\text{SL}}}{M \cdot u_m}
\]  

(2)

with standard measurement uncertainty \( u_m \) and typically \( M = 4 \) (corresponding to a coverage factor, \( k = 2 \) and 95% confidence). (Note that it is more usual in metrology to use a coverage factor of 2, while in SPC the corresponding coverage factor for process capability is often 3 (as in the six-sigma approach.).

In various sectors of conformity assessment, different limits on measurement capability have become established, with \( C_{m,\text{min}} \) ranging from typically 3 to 10. A common limit to ensure that measurement quality variations are small is \( u_m/\vartheta < 30\% \), as in measurement system analysis (MSA) in the automobile industry (AIAG 2002), for instance. Even in qualitative measurement, such as made on ordinal scales and using questionnaires, reference is made to minimum values of a reliability coefficient, given by

\[
R_\theta = \frac{\text{True variance}}{\text{Observed variance}} = \frac{\text{var}(\theta')}{\text{var}(\theta')} = \frac{\text{var}(\theta) - \text{var}(\vartheta)}{\text{var}(\theta)}
\]

where \( \theta = \theta' + \vartheta \) for an attribute \( \theta \) of an entity (object or item) with a value \( \theta' \) when the measurement error \( \vartheta \) is zero. In the literature (Linacre 2002), a recommended value of a reliability coefficient is 0.8, corresponding to a ‘separation’ of \( G = 2 \), or in other words, a measurement uncertainty \( \sqrt{\text{var}(\vartheta)} \) not larger than half the object standard deviation \( \sqrt{\text{var}(\theta)} \).

5.2. Maximum permissible (measurement) error

Much of conformity assessment in legal metrology is based (amongst other factors) on the testing of measurement instruments instead of the control of actual measurements of goods and utilities in society, as covered by the EU Directive Measuring Instrument Directive (MID)—both by type approval, initial and subsequent verification (Källgren and Pendrill 2006, EU Commission 2004). Typically, alongside more qualitative attribute requirements, such as inspection of correct instrument labelling and unbroken sealing of instruments, measurement specifications are also set by variable in terms of MPE, both for the main characteristic (e.g. indication of an electricity energy meter) as well as of any influence quantity (e.g. level of disturbing electromagnetic field, in EMC testing) to be tested through quantitative measurement.

Conformity assessment procedures in legal metrology can be regarded as a prototype for more general conformity assessment, for instance in the framework of EU Commission (2006) legislation.

A general question, as yet apparently unanswered in the literature, is the following: in the context of conformity assessment of an instrument (or measurement system), we have seen in this section how a maximum permissible measurement error, MPE, can be specified. In the context of product conformity assessment, a maximum permissible product error, MPE, is specified as discussed in section 3. The question is: what is the relation between these two MPEs?

6. Deciding if entity is within specified requirements

Having set limits—albeit so far somewhat arbitrarily—on (A) how large measurement uncertainty is allowed to be, as described in section 5, we now move to the next step, (B), aimed at minimizing risks and the consequences of incorrect decision-making in conformity assessment.
Unfortunately to date there is no consensus about this step (B) of agreeing on acceptable locations of the uncertainty interval with respect to a specification limit.

Typical scenarios are illustrated in figure 1 of the EURACHEM guide (2007), where the coverage interval of a test result has a significant overlap with a specification limit. Such an overlap can cause both ‘customer risk’ and ‘supplier risk’, as mentioned in the introduction to the French standard FD x07-022 (AFNOR 2004)—see further in section 7.

Rather strict decision rules are given for instance in ISO 10576-1:2003 which stipulate that in the first of a two-stage conformity test, the coverage interval (termed ‘uncertainty interval’ in that standard) must be wholly contained within a tolerance interval for an initial test result to indicate compliance and that the second stage shall be performed if and only if the uncertainty interval is inside the region of permissible values.

Another approach, followed for instance in the JCGM 106:2012 guide, is to define an acceptance interval of permissible values of the quality characteristic of the entity subject to conformity assessment that is narrower than the corresponding tolerance interval. The item is accepted as conforming if the measured value of the quality characteristic lies in an interval defined by acceptance limits \( (A_L; A_U) \), and rejected as non-conforming otherwise.

7. Risks of incorrect decisions of conformity—in percentage terms

A possible approach to removing the present arbitrariness and lack of consensus about (A) setting limits on measurement capability and (B) agreeing on acceptable locations of an uncertainty interval with respect to specification limits is to make the third step, namely (C) treating explicitly the risks of incorrect decisions of conformity arising from measurement uncertainty.

A number of recent publications (Fearn et al. 2002, Sommer and Kochsiek 2002, Forbes 2006, Rossi and Crenna 2006, Pendrill 2006, 2007, Pendrill and Källgren 2008, EURACHEM/CITAC 2007, JCGM 106:2012) have extended the ISO 10576-1 approach to include explicit consideration of risks, and develop general procedures for deciding conformity based on measurement results, recognizing the central role of probability distributions as expressions of uncertainty and incomplete information.

7.1. Percentage risk

An example of application of this kind of percentage risk approach to the role of measurement uncertainty in conformity assessment can be found in the construction industry (Hinrichs 2010).

The JCGM 106 guide (2012) for instance provides a number of plots (for example figure 17 in that reference). The JCGM 106 gives no details of what software is suitable for such calculations but for example programs such as Maple and MathCad can perform both symbolic and numerical evaluation of integrals such as in (4).

The EURACHEM/CITAC (2007) guide defines a specification zone in terms of acceptance and rejection zones through the introduction of guard bands at each specification limit. The size of a guard band—that is, the distance between a limit of the acceptance zone and the corresponding limit of the specification zone—is related to the value of test uncertainty and is chosen to meet the requirements of a particular decision rule. For instance, if the rule for deciding non-compliance is that the probability that the entity value (i.e. of the quality characteristic \( q \) = \( Z \)) in the ‘entity (or product) space’, described in section 3) lies above an upper specification limit, \( U_{SL} \), should be at least 95%, then the guard band size, \( g \), is set (as some multiple of the uncertainty) so that for an observed value of \( U_{SL} + g \), the probability that the entity value lies above the limit \( U_{SL} \) is 95%.

7.2. Sharing risks

The risk of making erroneous declarations of conformity can be ‘shared’ (ILAC 1996):
‘it may be appropriate for the user to make a judgment of compliance, based on whether the test result is within the specified limits with no account taken of the uncertainty. This is often referred to as shared risk since the end-user takes some of the risk that the product may not meet the specification after being tested with an agreed measurement method’.

The Geometrical Product Specification (GPS) standard ISO 14253-3:2011 emphasizes the importance of reaching agreement between customer and supplier about measurement uncertainty, preferably even as early as the pre-contract stage of a commission, usually in terms of an agreed target (or maximum permissible) measurement uncertainty.

Following actual measurements (that is, at the verification stage), normally it is the party providing the proof of conformance or non-conformance with a product specification or measurement equipment specification, i.e. the party making the measurements, who states the actual measurement uncertainty, according to GPS standard ISO/FDIS 14253-1:2013(E).

8. Comparative, ordinal and multinomial measurement and decision-making

There is an intimate relation between decision-making and relating objects to each other, encompassing not only quantitative properties but also comparative or even merely qualitative concepts, often involving human judgment, as will be discussed in this section.

Concepts used to describe objects can be classified into three general types: qualitative, comparative and quantitative, according to a standard view in the philosophy of science (Mari and Giordani 2012). Qualitative concepts allow the mere categorization of objects in classes, i.e. nominal properties; comparative concepts allow objects to be related to each other with respect to an overall order, i.e. ordinal properties, but any numbers assigned to these do not in general represent relations between the numbers represent relations between the objects.

8.1. Dichotomous case of decision-making: nominal properties and binomial classification

The simplest, dichotomous case of decision-making (section 7.1) can be regarded, in the presence of significant risks of incorrect decisions from measurement uncertainty, as an imperfect go/no-go classification of the quality characteristic of the entity being assessed with respect to a specification limit. When \( N \) repeated go/no-go trials are made, with \( d \) non-conforming entities, the estimated fraction, \( \hat{p} \), of non-conforming product \( (x = d/N) \), will follow a binomial distribution, for which the off-diagonal elements of the decision matrix (equation (3)) will be given by the well-known formula:

\[
\hat{P}_{j,k} = \frac{N!}{d!(N - d)!} \hat{p}^d (1 - \hat{p})^{N - d}.
\]

This dichotomous decision-making case can be regarded as an elementary example of a nominal classification into two categories—if one additionally denotes the two categories as ‘0’ and ‘1’, then we have a simple ordinal classification.

Although ‘qualitative’ analysis has been defined as the ‘Assessment of presence or absence of one or more analytes in sample due to its physical and chemical properties’ (Trullols et al 2004), the decision to classify as go/no-go is not exclusive to qualitative measurement (Pendrill 2011). Decisions of conformity can be made on a more or less quantitative basis—referred in statistical acceptance sampling as by ‘variable’ (i.e. with respect to a numerical relation (ISO 3951-1:2005)) or by ‘attribute’ (i.e. go/no-go decisions (ISO 2859:1999))—depending on the resources available or indeed whether a full quantitative judgment is needed or not.

An example of a quantitative analysis where go/no-go decisions are deemed adequate for purposes is found in legal metrology when ensuring the metrological performance of various types of measuring instrument (EU Commission 2004, Källgren et al 2006). Actual measurements of instrument errors for each meter are made quantitatively. But a sufficient measure of confidence in decisions of compliance is often provided for by specifying in the first case an attribute (i.e go/no-go) sampling plan (Montgomery 1996), such as in the MID directive (EU Commission 2004, 2006). On the one hand an acceptable quality level (AQL) of 1% (EU Commission 2004, 2006, Annexes F and F1), which is the poorest level of quality for the instrument manufacturer that the consumer would consider to be acceptable as a process average, in terms of the fraction of non-conforming product. Also specified in conformity assessment is a limiting quality level (LQL) of 7%, that is, the poorest level of quality that the consumer is willing to accept in an individual lot of instruments. For a typical case of electricity meters, with a sample of size \( n = 5000 \) entities drawn from a population of \( N = 1.5 \times 10^6 \), where \( d/N = 1.60\% \) are found on average to be rejectable, an estimated (statistical) sampling uncertainty, \( \sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} = 0.18\% \) (Pendrill 2006).

Other, less quantitative examples of decision-making include material hardness measurement; expert elicitation; questionnaires and sensory perception (e.g. of smell). Observed values in these cases—where not enough is known about the quantitative relation between measurement values and the quality characteristic of the entity subject to conformity assessment—are referred to an ordinal scale and any underlying variable, which might with more resources have provided a more quantitative basis for decisions, is termed ‘latent’.

8.2. Polytomous case of decision-making: ordinal properties and multinomial classification

While many of the guides on the role of measurement uncertainty in conformity assessment focus on the dichotomous case of decision-making, it is straightforward to generalize to multinomial classification into \( K \) categories on (at least) an ordinal scale, where the decision matrix, \( P \), equation (3) becomes a...
\[ K \times K \text{ matrix and} \]
\[ P(n_1, n_2, ..., n_K) = \frac{N}{n_1! n_2! ... n_K!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot ... \cdot p_K^{n_K} \]
where \( n_k \) entities are classified to be in category \( k \) prior to measurement. The probability, \( q_c \), of classifying the entity in a category \( c \) is related to (i) the accumulation of (unobserved) probabilities, \( p_k \), of the entity being in a number of categories \( k \) prior to measurement, and (ii) to the decision matrix, \( P \) (equation (3)), according to the expression: \( q_c = \sum_{k=1}^{K} p_k \cdot P_{k,c} \) (Bashkansky et al. 2007). This multinomial distribution has the following mean and variance, respectively: \( \bar{x}_c = N \cdot q_c; \text{var}(\bar{x}_c) = q_c \cdot (1 - q_c)/N \).

The measurement (standard) uncertainty in the estimated mean fraction of entities in each classification category \( c \) can be estimated as \( \sqrt{\text{var}(\bar{x}_c)} \).

In order to handle less quantitative observational data—where statistical tools are of limited applicability (Svensson 2001)—can be transformed to a quantitative scale, where the probability \( q_c \) is investigated as a function of an ‘explanatory’ variable, \( z_c \), such as the force applied to a hardness indenter. Helton et al. (2006) summarize a typical procedure when using expert elicitation (in the challenging prediction of risk in long-term nuclear waste storage, for instance): ‘as general guidance, it is best to avoid trying to obtain …[pdf] distributions by specifying the defining parameters (e.g. mean and standard deviation) for a particular distribution type. Rather, distributions can be defined by specifying selected quantiles (e.g., 0.0, 0.1, 0.25, …0.9, 1.0) of the corresponding cumulative distribution functions (CDFs), which should keep the individual supplying the information in closer contact with the original sources of information or insight than is the case when a particular named distribution is specified. Distributions from multiple experts can be aggregated by averaging’.

One approach is logistic regression which fits the log-odds to a linear function of explanatory variable \( z_c \), with two (or more) fitting parameters \( \theta \) and \( \beta \):
\[ \log \frac{q_c}{1 - q_c} = \theta - \beta \cdot z_c \]
(Theil 1970, McCullagh 1980) as an example of a generalized linear model applicable to non-Normal distributions. This not only deals with ordinal data but also allows a separation of linear model applicable to non-Normal distributions. This not is general a true distance metric—for example, it is not symmetric: \( D_{\text{KL}}(P, Q) \neq D_{\text{KL}}(Q, P) \)—its infinitesimal form, specifically its Hessian, is a metric tensor, the so-called Fisher information metric. Minimizing this distance corresponds to a maximum likelihood estimation of the quality characteristic, \( X \), of the entity subject to assessment. The associated covariance matrix \( \text{cov}(x)_{j,k} \) is the inverse of the information matrix, which has a \((j,k)\) element
\[ -E(\partial^2 D_{\text{KL}}(P \parallel Q)/(\partial \theta_j \cdot \partial \theta_k)) \] (Agresti 2013).

These measurement uncertainty estimates can be used when making weighted logistic regression fits as well as in assessing decision risks in subsequent conformity assessment.

9. Introducing costs and impact into risk assessment in conformity assessment by inspection

Traditional treatment of the risk of incorrect decision-making associated with measurement uncertainty, including concepts such as ‘shared risk’ and ‘guard-banding’ (Deaver 1998) in percentage terms (section 7), do not always readily relate to the principal aims of the decision-maker who often thinks in terms of impact and economy.

Ultimately when appropriate (or fit-for-purpose (Thompson and Fearn 1996)) levels of uncertainty and associated risks of incorrect decisions are to be set, reference will need to be made to measures of impact for various stakeholder groups. What appears to be statistically a certain sharing of risks (section 7.2) between consumer and supplier in purely percentage terms may be rather unfair when different economic consequences are taken into account (Pendrill 2006). The optimized uncertainty methodology (section 10) has demonstrated that traditional MPU ‘rule-of-thumb’ limits are often higher than the optimum uncertainty (Pendrill 2007) with correspondingly unnecessarily large consequence costs from incorrect decisions of conformity.

Choosing the tolerance limits and acceptance limits should be business or policy decisions that depend upon the consequences associated with deviations from intended product quality (Williams and Hawkins 1993, Montgomery 1996). Not only economics but other impacts such as livelihood can be at stake (King 2004).

9.1. Costs and economic risks in conformity assessment

Incorrect decisions in turn lead to a variety of consequences, with associated risks and costs: a supplier might be obliged, for instance to re-manufacture product unnecessarily. There are various consequences for the customer too, such as where incorrect decisions of conformity lead to incorrect measures of quantity (e.g. over-estimated or under-estimated entity values) and poor product performance. Incorrect decisions may even lead to litigation, where disputes in conformity assessment might end up in a legal process in court. Various cost models and how economics is introduced in risk assessment in conformity assessment are dealt with the remaining parts of this paper.
In figure 1 an overall picture of the classic decision matrix but including costs explicitly is given of sources of both profit and loss from the point of view of the supplier when assessing the conformity of a particular value of an important characteristic of the entity of interest.

Irrespective of the result of product (or entity) conformity, there will always be the costs of production and testing of product (at the centre of figure 1).

Then, for each specimen of product, the actual true value, $\mu$ (although unknowable exactly), of the characteristic will either conform or not conform depending on whether the value is inside or outside, respectively, of a specification limit ($U_{SL}$, upper specification limit in the current example). Correct decisions of conformity relate to both the profit made on selling product which has been correctly assessed to be conforming (top, left of figure 1), and the losses made on product correctly assessed to be non-conforming (bottom, right of figure 1).

A general formulation of the overall profit (Williams and Hawkins 1993) would be a sum of the various incomes and losses shown in figure 1, including (a) income from sales of passed, conforming product; (b) loss associated with customer risk (passed, non-conforming product); (c) cost of (all) manufactured product (exclusive test); (d) cost of testing (all) product; (e) loss associated with re-manufacturing with supplier risk (failed, conforming product); and (f) loss associated with re-testing with supplier risk (failed, conforming product).

Among a number of conceivable models of how test costs could vary with measurement uncertainty (Ramsey et al 2001), a common choice is to assume that the test cost depends inversely on the squared (standard) measurement uncertainty $u_m^2$, that is, $D/u_m^2$, where $D$ is the test cost at nominal test (standard) uncertainty $u_m$. As more effort is expended to reduce measurement uncertainty, the more it costs to measure.

9.2. Introducing cost into conformity assessment risks

In general, the impact of a wrong decision in conformity assessment is expressed as a risk $R$, defined as the probability $p$ of the wrong decision occurring multiplied by the cost $C$ of the consequences of the incorrect decision (AFNOR 2004):

$$R = p \cdot C. \quad (6)$$

9.2.1. Difficult estimation of cost and impact in conformity assessment

Economic costs for both measurement and consequence are sometimes difficult to estimate and not always the best measure of impact. Particularly, estimating costs as a consequence of incorrect decision-making can be difficult:

In an extreme case, deterioration of concrete might lead to having to demolish and reinstate the building at a cost estimated to be $7 million. (Fearn et al 2002)

The EURACHEM/CITAC guide (2007, section 6) claims that ‘the information needed to (choose a value of uncertainty which minimise(s) the costs of analysis plus the costs of the decisions) is very rarely available’. The French standard (AFNOR 2004) also writes, ‘In practice, difficulties in estimating this cost often result in evaluating the probability, which is also incorrectly referred to as the ‘risk’’. 
The counter-argument in this work in support of assessing risks in impact terms is to emphasize the many advantages of including costs and measures of impact, as summarised by Pendrill (2007): the decision-maker can arguably more readily relate to a cost than to a percentage risk; costs can be both positive and negative while percentages are always without a sign; even in difficult cases, it is better to attempt an economic analysis—however rough and ready—they assume that impact costs of incorrect decisions from measurement uncertainty are set arbitrarily to zero.

9.2.2. Cost and impact in measurement instrument conformity

In many areas of considerable impact in society, such as covered by legal metrology, costs can be objectively specified—even at the national economic level—and are often relatively simply modelled.

Costs associated with display errors for a particular type of instrument—say, petroleum fuel dispensers—are normally a linear function of instrument error, since the commodities (and such) themselves are normally charged in linear proportion to the quantity dispensed by each type of instruments. Secondly, for many utilities but also increasingly in the environmental area, prices of commodities and emissions are known. An additional advantage of costing risks rather than just in terms of percentages is that actual prices and costs can be included at any given moment. Finally, in making a financial analysis in national economic terms, there is a direct relation between taxes levied on goods and transactions and the costs of regulation in legal metrology. Examples of loss function implementation in legal metrology cover instrument categories such as

- electricity meters (Pendrill 2007);
- exhaust gas analysers (Källgren and Pendrill 2006);
- fuel meters (Pendrill and Källgren 2008).

10. Optimized uncertainty

In a typical conformity assessment situation, the needs and wishes of the consumer have to be balanced against the capabilities and promises of the supplier. In some cases, e.g., in legal metrology, a third party such as an independent testing laboratory may intervene to act as impartial arbiter in any dispute, for instance to represent not just one individual but a wider group, perhaps in a national economic context. Any one (or all) of these three main actors in conformity assessment may act as ‘decision-makers’.

10.1. Balancing test costs against consequence costs

An incorrect accept on inspection of a non-conforming object will lead to customer costs associated with out-of-tolerance product. Overall costs, consisting of a sum of testing costs, \( D \), and the costs, \( C \), associated with customer risk, can be calculated with the expression:

\[
E(\eta, \sigma) = D(\eta, \sigma) + C(\eta) = \frac{D}{\sigma^2} + \int_{\eta \notin R_{PV}} C(\eta) \cdot g(\eta | \hat{\eta}) \cdot d\eta
\]

with \( \hat{\eta} \in R_{PV} \), where \( R_{PV} \) denotes the region of permissible values and using equations (4) and (6)—an expression which can be applied to both specific and global conformity assessment.

Expression (7) is often evaluated in the literature (Forbes 2006) by setting \( C \) equal to zero inside \( R_{PV} \) and to a constant value outside the region of permissible values. Examples of explicitly including a cost function in the integrand when evaluating expression (7) for optimized uncertainty by variable can be found in Pendrill (2007) and references therein.

Equation (7) can be evaluated in two complementary ways, as summarized in figure 2:

- a range of quantity values of \( U_{SL} - h \cdot u_{m} \leq y_{m} \leq U_{SL} + h \cdot u_{m} \) for a given test uncertainty, \( u_{m} \), and ‘guard-band’ factor \( h \)—yielding an ‘operating cost characteristic’ analogous to the traditional, probability-based operating characteristic;
- a range of test uncertainties, \( u_{m} \), for a given quantity value \( y_{m} \leq U_{SL} \), the so-called ‘optimized uncertainty curve’.

These calculations are often made with reference to a single specification limit—\( L_{SL} \) or \( U_{SL} \), whatever the case may be. Since a maximum permissible measurement uncertainty is already set (step A—see discussion in section 5.1), it is often a good approximation to calculate with reference only to the nearest specification limit since the other limit would lie several multiples of uncertainty away.

Various cost models can be employed: linear models for metering in legal metrology, for instance; parabolic functions capturing varying customer satisfaction of market expectations, etc (Taguchi 1993). Studies of pre-packaged goods (Pendrill 2008) in legal metrology are a kind of prototype for a general treatment of (univariate, interval-scale) conformity assessment of any kind of product. A more recent example of this approach can be found in an application to geometrical product control in automobile industry (Pendrill 2010). The practice of guard-banding (Deaver 1994) can also be analysed in terms of costs, impact and optimized uncertainties (Pendrill 2009).

10.2. Global conformity assessment by variable and by attribute

Optimized uncertainty analysis can be carried out (Pendrill 2008) both by variable (ISO 3951:2005) and attribute (ISO 2859:1999).

Pendrill (2007) gives examples of nesting of product and measurement integrals to treat ‘global’ conformity assessment by variable, according to the following equation:

\[
E_{np} = \frac{D_{np}}{u_{measure}^2} + C_{np} \cdot \left[ \int_{\infty}^{U_{SL}} \frac{1}{\sqrt{2\pi} \cdot s_{instrument}} \cdot e^{-\frac{(y_{measure} - \hat{y})^2}{2s_{instrument}^2}} \cdot dy_{measure} \right. \\
\left. \times \int_{U_{SL}}^{\infty} \frac{1}{\sqrt{2\pi} \cdot u_{measure}} \cdot e^{-\frac{(y_{measure} - \hat{y})^2}{2u_{measure}^2}} \cdot dy_{measure} \right] \\
\]

(8)

for instance where the conformity assessment of electricity utility meters was modelled using linear cost models together...
with a plausible measurement distribution PDF \( g(x) \) and an observed distribution PDF, \( g(z) \), of display error for a number of household meters under national legal metrological control. Equation (8) is the model of consumer risk, where the subscript ‘np’ denotes the incorrect decision of conformity: non-conforming/pass. Such an analysis does not, as yet, take into account (i) possible variations of instrument error over different values of stimulus, for instance, over a range of electrical energy, or (ii) variations in frequency of use from instrument to instrument across a population of meters.

A corresponding conformity assessment by attribute would first calculate the aggregate consumer-related cost associated with a fraction non-conforming product with the expression

\[
C_{USL} = C_{np} \cdot \left[ \int_{-\infty}^{\infty} z \cdot \frac{1}{\sqrt{2\pi} \cdot \delta_{instrument}} \cdot e^{-\frac{(z - \mu_{instrument})^2}{2\delta_{instrument}^2}} \cdot dz \right]
\]

as an estimate of the costs associated with instruments exceeding the (upper) specification limit, USL, where \( C_{np} \) is the consequence cost by variable. In the example of household electricity meters, \( C_{USL} \) is of the order of 0.54 M€ nationally (an over-taxation since meters are slightly biased towards positive instrument errors) if all instruments in the country are accounted for.

Conformity assessment by attribute would be against a specification limit, \( SL_{p} \), set on the fraction non-conforming instruments acceptable nationally (e.g. 12%). Actual consumer risk can be expressed in percentage terms by the cumulative binomial distribution (equation (6)) beyond \( SL_{p} \) in terms of the sum

\[
\Phi_{\text{binomial}}(d, N, SL_{p}) = \sum_{p=SL_{p}}^{\infty} \frac{N!}{d!(N-d)!} p^d \cdot (1-p)^{(N-d)}
\]

for a sample of size \( N \) and actual observed number of non-conforming entities, \( d \).

In this case, an optimized sampling uncertainty can be deduced as that uncertainty at which overall costs, given as the sum of sampling costs and consequence costs according to the expression

\[
E_{USL,\text{attr}} = \frac{D_{np}}{u_{sample}^2} + C_{USL} \cdot \Phi_{\text{binomial}}(d, N, SL_{p})
\]

are minimized, where \( C_{USL} \) is given by equation (9).

Figure 3 illustrates how overall costs vary with the number, \( N \), of instruments sampled (assuming an infinite population) and how an optimized value of the sampling uncertainty \( \sigma_{\hat{p}} = u_{sample} = \sqrt{p(1-p)/N} \) along the \( x \)-axis (Pendrill 2006) can be identified in relation to traditional sampling planning limits AQL and LQL (Montgomery 1996) for the example of electricity meters. Actual costs are high, because the larger sample size (5000 meters) in the present example costs about 25 times as much as the around 1000 optimal sample size. Where the actual optimum sampling uncertainty (and sample size) lies will of course be determined both by actual sampling costs as well as the choice of specification limit (even that economically motivated) on the maximum permissible fraction non-conforming product.
This new optimized sampling uncertainty methodology extends traditional attribute sampling plans to include economic assessments of the costs of measuring, testing and sampling together with the costs of incorrect decision-making.

11. Conclusion and future work

The subject—the role of measurement uncertainty in conformity assessment—is new to many metrologists. We have, therefore, deliberately chosen in this paper to present not only a literature survey but also a discussion of concepts. It was emphasized that the study of the role of measurement uncertainty in conformity assessment addresses an area where two disciplines—conformity assessment and metrology—meet. Conceptual starting points may therefore differ and care has been expended in the text to follow standardized terminology as far as possible. Practically in cases where measurement dispersion is of comparable size to actual product variation, it can be difficult to perform the essential separation of these required in conformity assessment. Without that separation, there will always be significant risks of incorrect decisions of conformity. A starting point is to form a clear picture of different perspectives: in conformity assessment, the quality characteristic of an entity is the quantity to be assessed in relation to requirements while in metrology the measurand is the quantity intended to be measured. Existing guides, such as JCGM 106:2012, do not in our opinion give clear definitions of global risk, for instance.

The arbitrariness of many of the existing rules aimed at limiting decision risks—either rules of thumb on ratios of product to measurement dispersion or on percentage risk of incorrect decisions—may be resolved by economic consideration and there are examples of successful implementation in areas such as the legal metrological control of measurement instruments in a range of societal important sectors such as utility, commodity and environmental monitoring. Economic costs for both analysis and consequence are admittedly sometimes difficult to estimate and not always the best measure of impact. It is better, however, to attempt an economic analysis—however rough and ready—than to set arbitrarily impact costs of incorrect decisions from measurement uncertainty to zero. None of the major recent guides about the role of measurement uncertainty in conformity assessment reviewed here. New expressions for decision-making risks including costs have been presented in this recent work, including the operating cost characteristic curve as an extension of traditional statistical tools, with the addition of an economic decision-theory approach. Complementarity with the optimized uncertainty methodology is emphasized.

A summary is given in sections 9 and 10 of recent progress which significantly extends that found in the major recent guides about the role of measurement uncertainty in conformity assessment reviewed here. New mathematical and statistical approaches are required to address uncertainty evaluation in many modern metrology applications such as biochemistry, molecular biology and healthcare, which are not explicitly covered by existing GUM guidelines. This includes qualitative and subjective properties where formulation of metrological concepts is in its infancy (Fisher 1997, Pendrill 2011) compared with more traditional quantitative measurements. Common tools of statistics needed for evaluation of measurement uncertainty and in making decisions of conformity, which work readily for quantitative interval and ratio scales, are unfortunately not applicable (Svensson 2001) to the ordinal data typical of ‘human’ measurement (Pendrill et al 2010).

Acknowledgments

Thanks are due to several members of the NEW04 project for their constructive comments. The European Metrology Research Programme (EMRP, FP7 Art. 185) is jointly funded by the EMRP participating countries within EURAMET (www.euramet.org) and the European Union.

References

AFNOR 2004 Use of uncertainty in measurement: presentation of some examples and common practices French Standardisation FD x07-022
Agresti A 2013 Categorical Data Analysis 3rd edn (New York: Wiley) ISBN 978-0-470-46363-5
Automotive Industry Action Group (AIAG) 2002 Measurement Systems Analysis Reference Manual 3rd edn, Chrysler, Ford, General Motors Supplier Quality Requirements Task Force Bashkansky E, Dror S, Ravid R and Grabov P 2007 Effectiveness of a product quality classifier Quality Eng. 19 235–44
Deaver D 1994 Guardbanding with confidence Proc. NCSL Workshop & Symp. (Chicago, July–August 1994) pp 383–94
Deaver D 1998 Guardbanding and the world of ISO Guide 25: is there only one way? NCSL Workshop & Symposia; Nicholas M

Figure 3. Costs as a function of attribute sampling uncertainty for electricity meters.
Rossi G B and Crenna F 2006 A probabilistic approach to measurement-based decisions Measurement 39 101–19
Rukhin A L 2013 Assessing compatibility of two laboratories: formulations as a statistical hypothesis testing problem Metrologia 50 49–59
Sommer K-D and Kochsiek M 2002 Role of measurement uncertainty in deciding conformance in legal metrology OIML Bull. 43 19–24
Svensson E 2001 Guidelines to statistical evaluation of data from rating scales and questionnaires J. Rehab. Med. 33 47–8
Taguchi G 1993 Taguchi on Robust Technology (New York: ASME Press)
Theil H 1970 On the estimation of relationships involving qualitative variables Am. J. Sociol. 76 103–54
Thompson M and Fearn T 1996 What exactly is fitness for purpose in analytical measurement? Analyst 121 275–8
Trullols E, Ruisánchez I and Xavier Rius F 2004 Validation of qualitative analytical methods Trends Anal. Chem. 23 137
Williams E and Hawkins C 1993 The economics of guardband placement Proc. 24th IEEE Int. Test Conf. (Baltimore, MD)