Critical Test Appraisal

Uncertainty of Measurement

Prof. Dr. F. Vanstapel, MD PhD
Laboratoriumgeneeskunde
UZ-KULeuven

Teaching goals

Understand requirements of standards

Make sure that the exercise adds value

understand certainty

understand the diagnostic process

understand analytical errors

Find relevant information efficiently and use it effectively

Important steps in the life cycle of a method

- validation plan and method validation: analyse
  - measurement principle
  - robustness of assay design
- continuous learning:
  - analysis of internal quality control data
- use knowledge for method optimalization & risk containment

Uncertainty of Measurement
Requirements of the Standards

ISO 17025 § 5.4.6.
Testing laboratories shall have and shall apply procedures for estimating uncertainty of measurement.
In certain cases the nature of the test method may preclude rigorous, metrologically and statistically valid, calculation of uncertainty of measurement.

ISO 15189 § 5.6.2.
The laboratory shall determine the uncertainty of results, where relevant and possible. Uncertainty components, which are of importance, shall be taken into account.
Sources that contribute to uncertainty may include sampling, sample preparation, sample portion selection, calibrators, reference materials, input quantities, equipment used, environmental conditions, condition of the sample and changes of operator.

The lab shall have a procedure

The effort shall be commensurate

Preanalytical

Meterological Traceability

Uncertainty of Measurement
Requirements of the Standards

ISO 15189 § 5.6.3.

A programme for calibration of measuring systems and verification of trueness shall be designed and performed so as to ensure that results are traceable to SI units or by reference to a natural constant or other stated reference.

Where none of these are possible or relevant, other means for providing confidence in the results shall be applied, including but not limited to the following:

a) participation in a suitable programme of interlaboratory comparisons

b) use of suitable reference materials, certified to indicate the characterisation of the material

c) examination or calibration by another procedure

d) ratio or reciprocity-type measurements

e) mutual consent standards or methods which are clearly established, specified, characterized and mutually agreed upon by all parties concerned

f) documentation of statements regarding reagents, procedures or the examination system when traceability is provided by the supplier or manufacturer

Philosophy

1. Understanding Uncertainty

certainty
results from (structured) reasoning,
taking into account all known relevant elements (= state of knowledge).
Certainty is then the degree of belief (= probability)
that the facts that are found (= result of a measurement)
are relevant, so that an action is more likely to follow (= consequential).

Certainty ≈ Relevant (validated) Evidence
(U.S. Federal Rule of Evidence 401)

Our belief can be expressed as a probability:
L (conclusion | the facts that are found)
Our beliefs vary from weak to strong,
and rarely are absolute (0, 1).

1. Understanding Uncertainty

certainty results from (structured) reasoning,
taking into account all known relevant elements (= state of knowledge).
Certainty is then the degree of belief (= probability)
that the facts that are found (= result of a measurement)
are relevant, so that an **action is more likely to follow (= consequential)**.

The **certainty** that counts follows from
(structured) analysis of the state of knowledge, **in the head of the clinician**.

The fallacy of the diagnostic process:

P(fact | condition) ↔ L(condition | fact)

These are used interchangeably in common language, but they are not.

Test characteristics — Post-Test Probability — Bayes-theorema

also dependent on P(fact | condition absent), spectrum & prevalence

Failing to provide *info about test capability* is a source of uncertainty.
Failing to analyse all facts in a structured way is a source of (post-analytical) error.

2. Understanding Measurement

measurand: particular quantity subject to measurement

1. physicochemical phenomenon that is measured for the **analyte** that is studied
- report line : dimensions of measured phenomenon ... for analyte ...

2. state of the phenomenon, body, or substance under study
- report line : measured in matrix ..., at temperature ..., sampled … …
- reference values : under specified conditions (age, time of day, time post load …)

3. The **measurement** can change the quantity that is actually measured
e.*g*. potential difference between the terminals of a voltometric element
depends on the internal resistances of battery and voltmeter
e.g. the indicator participates in acid-base titrations
e.g. dilution and volume displacement effects in potentiometry

To the extent that such phenomena occur your report line has to specify
- the measurement principle
- the presence/or absence of a correction

Failing to define the measurand accurately is a source of (definitial) uncertainty.

2. Understanding Measurement

measurand: particular quantity *intended to be measured*

e.g. HbA1c
surrogate measurement for a surrogate parameter of multiple index-conditions

1. goodness of Glc-control
2. pathophysiological hypothesis: Glc as a driver of Glc-dependent effects
   - Glc- and insulin-desensitization & -irresponsiveness
     (hormonal & metabolic)
   - blood vessel- and nerve-damage
   - …

   (Brownlee, *Diabetes* **54**:1615-25, 2004)
3. indicator of prognosis

Failing to educate the user about the use of the results is a source of error.

2. Understanding Measurement

measurement
process to experimentally obtain an estimate of the magnitude of a **quantity**

1. the **measurement procedure** can be described and analyzed as a model
2. measurement presupposes a calibrated **measuring system**,
   preferably subsequently verified (iQC)

Often we limit ourselves to analysis of this component,
as we have direct access to it within the confinement of our labs.

This attitude neglects significant pre- and post-analytical steps of the total measurement process.

3. Understanding Error

precision and accuracy of experimental data

precision
reproducibility of results or the agreement between the numerical values for repeated measurements made under identical measurement conditions

1. always has a positive sign
2. measured as average deviation from the mean (or median), or as the spread or range, or as the SD on repeated measurements, …
3. can be expressed absolute or relative as fraction of the mean (or median)

trueness - accuracy
nearness to the *accepted* value (as the true value is generally not known)

1. can have a positive or negative sign
2. can be expressed absolute or relative as fraction of the mean (or median)

3. Understanding Error

determinate error *vs.* indeterminate error
bias or systematic error *vs*. random error

Call it by its name, but also :
Is method robust ?
Can you simplify procedure ?

error
difference between a measurement and the accepted value

determinate errors
1. arise from methodological flaws, and in principle can be accounted for
   - Instrumental errors: imperfect tools
   - Method errors: slowness of reactions, non-specific reactions, interferences, …
   - Personal errors: ignorance, carelessness, prejudices, physical limitations, …
2. have a definite value = **bias**, **systematic error**

indeterminate errors
1. arise from uncontrollable variations
2. produce scatter about a mean value = **random error**, **imprecision**

4. Understanding Uncertainty of Measurement

measurement uncertainty
(cited from info circulating in CLSI C51-P working group, unpublished communication)
parameter, associated with the result of a measurement, that characterizes the **dispersion** of the values that could **reasonably** be attributed to the measurand
(VIM93, ISO-15189 2003 3.17)
The parameter may be, for example, a standard deviation, (or a given multiple of it) or the half-width of an interval having a stated level of confidence.
(ISO 15195)

revised definition
(cited from info circulating in VIM working groups, unpublished communication)
… *based on relevant available information*
This dispersion is due to
definitional uncertainty of the **measurand**
random and **systematic** effects in the **measurement**

A restrictive definition of UM, will create little or no value.
By contrast, the **revised definition addresses key issues**

4. Understanding Uncertainty of Measurement

metrological traceability
property of a **measurement result** relating the result to a stated metrological reference through an unbroken chain of **calibrations of a measuring system** or comparisons

1. Each step in the chain contributes to the **measurement uncertainty** (extended VIM definition)

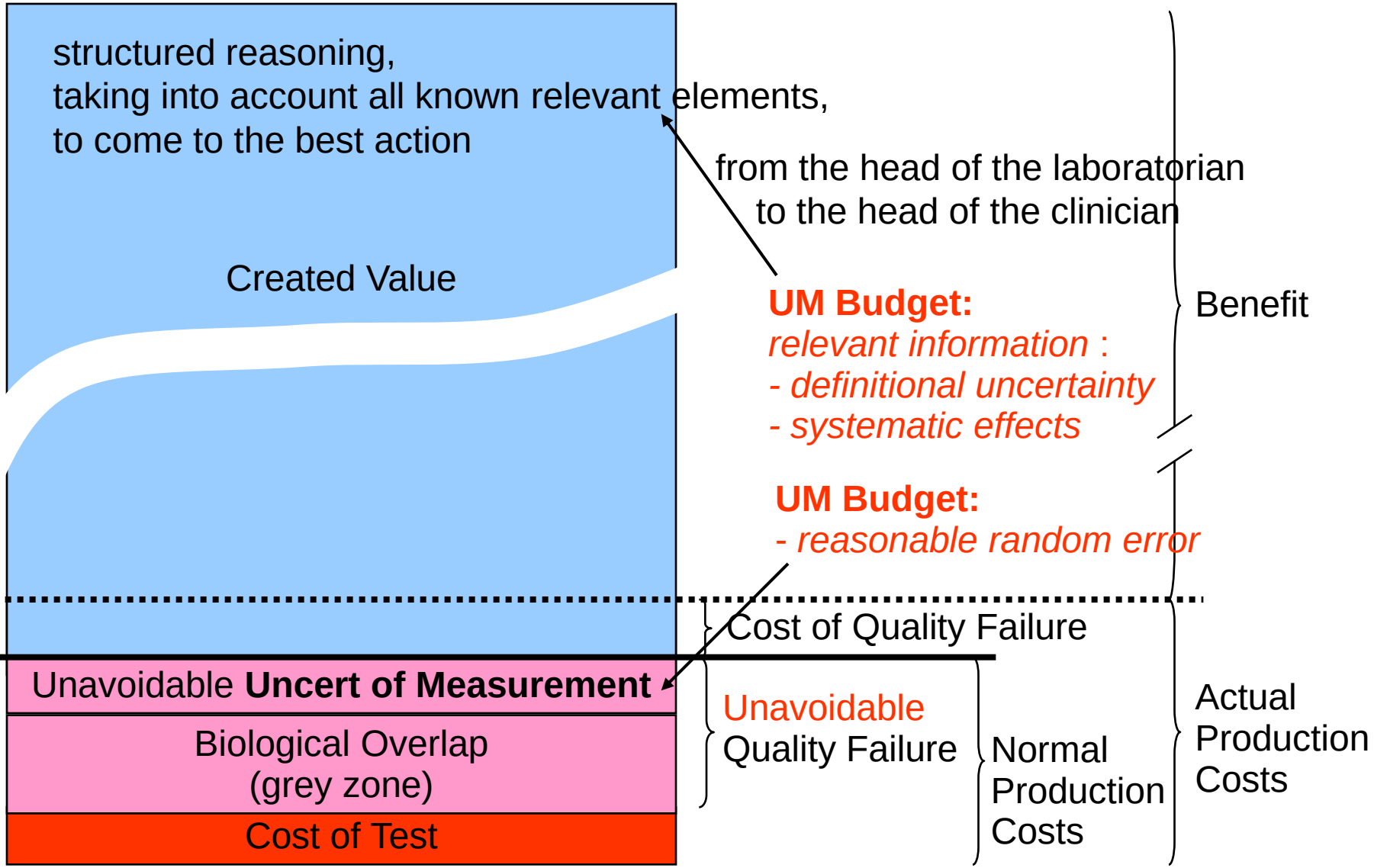
2. Results from a harmonized measurement system support the establishment of common reference intervals and decision limits, which allows for global application of clinical study findings.

3. Metrological traceability is facilitated by expressing results in common units, producing a measurement system that supports shared reference intervals and decision limits.

Here knowledge about UM creates value

Does analytical performance affect clinical utilization ?

Where can knowledge about UM create value ?



5. Understanding GUM (Guide to Uncertainty of Measurement)

Part of the Analytical Test Validation :

Systematic evaluation and appraisal of (analytical test) validity

(Analytical Test) Validity :

No measurement is possible without uncertainty

What are the components of that uncertainty ?

Size of that uncertainty ?

Importance of that uncertainty ?

fit for purpose = function of the intended use

Analysis:
Design for Quality

Proper starting & end point
of this think-experiment

5. Understanding GUM

Methodology

- Techniques for estimating dispersion and bias
  (in classical GUM, the latter is neglected)
- Specification of the measurement
- Identification of the sources (components) of uncertainty
- Quantification of these sources (in groups)
- Combination of quantified groups

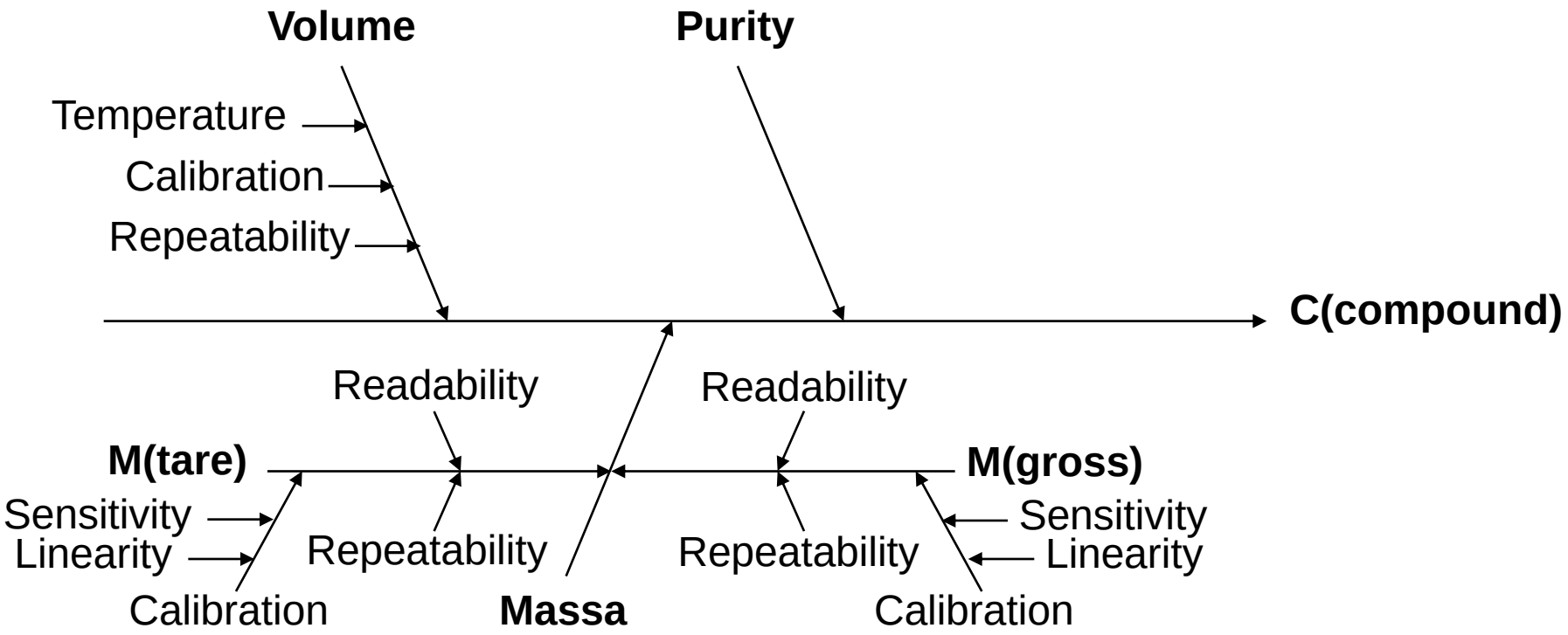
Sources of information

- data accumulated during validation and verification of a test
  - prior to its application in testing environment
    (implementation validation, use of certified reference materials, …)
  - during the life cycle of the method:
    - accumulated data of internal quality control (iQC)
    - inter laboratory studies, proficiency test schemes

5. Understanding GUM

Building models:

e.g. Calibrator formulation



The analysis resembles Fault Trees, treated in the Teaching Module on Risk

5. Understanding GUM

Quantifying uncertainty (mainly, reasonable dispersion)**:**

Type A:
by the statistical analysis of a series of repeat observations of the measurand obtained under specified measurement conditions.

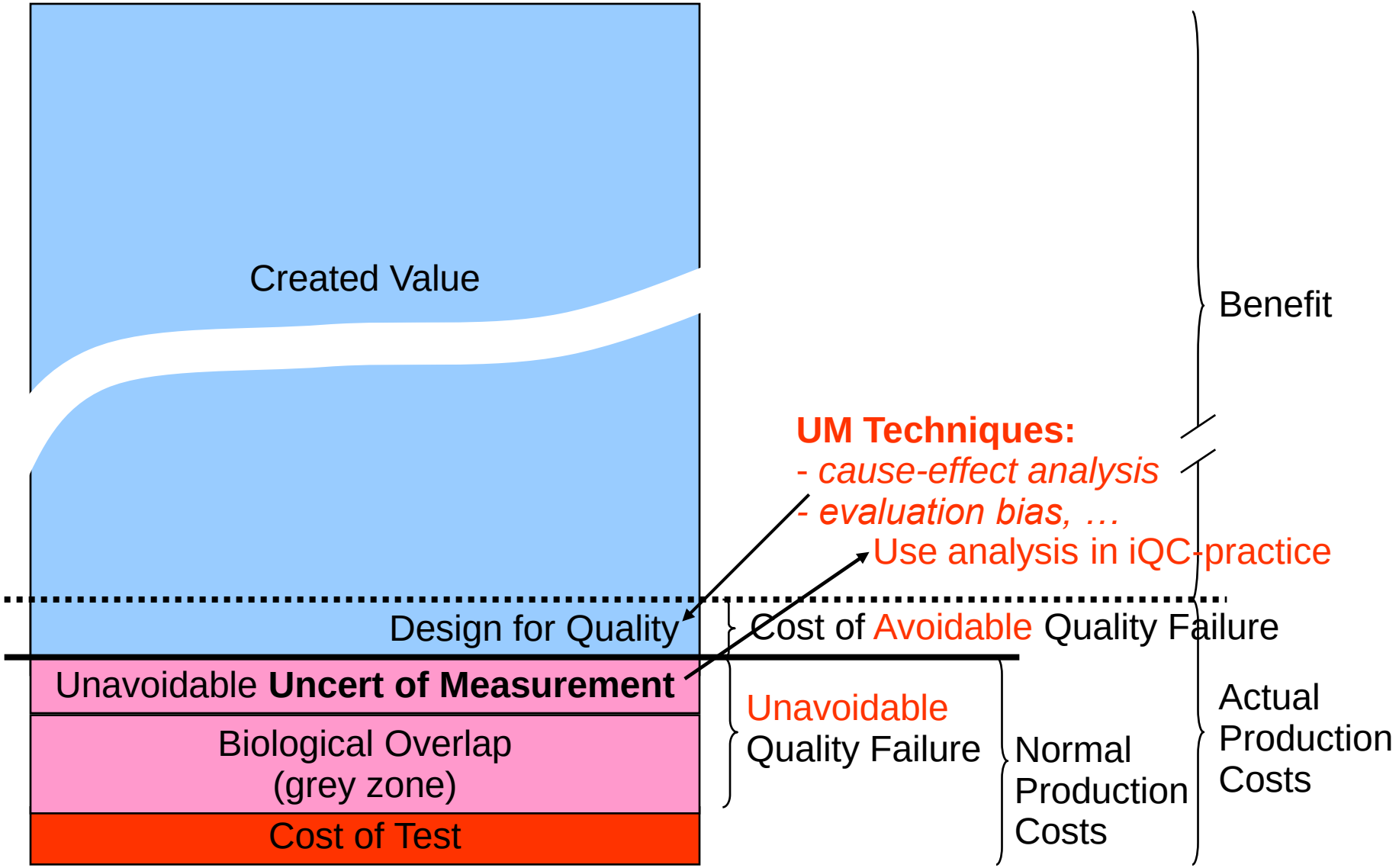
Type B:
based on available information
- previous measurement data
- manufacturer's specifications
- data provided in calibration certificates
- uncertainties assigned to reference materials

etc….

- Exploit your iQC-data
- Use UM-analysis
  to your advantage
  in iQC-practice

Does analytical performance affect clinical utilization ?

Where can knowledge about UM create value ?



Nuts and Bolts

General Relationship between the combined standard uncertainty and the uncertainty of the component quantities

First order Taylor Series approximation of the measurement process

$$u^2(y(x_1,x_2,\ldots)) = \sum_{i=1}^{N} (\delta y/\delta x_i)\, u_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\delta y/\delta x_i)\,(\delta y/\delta x_j)\, c_{i,j}$$

sensitivity

Covariance: $u_i u_j r_{ij}$

For independent parameters, the second term is absent.

Estimating SD from available information:

Gaussian distribution:
reading is within ± n mg with 95% confidence
SD = n / 1.96 mg

Uniform distribution:
Grade A volumetric flask certified within ± n ml
Extreme readings judged likely : model corresponds to a uniform distribution
SD = n / √3 ml

Triangular distribution:
Grade A volumetric flask certified within ± n ml
Experience learns extreme readings unlikely : model corresponds to a triangular distribution
SD = n / √6 ml

Propagation of determinate error: Sum or Difference

$$y = a + b - c \quad (1)$$
$$(y + \Delta y) = (a + \Delta a) + (b + \Delta b) - (c + \Delta c) \quad (2)$$
$$\Delta y = \Delta a + \Delta b - \Delta c \quad (2\text{-}1)$$

The absolute error of a sum is the algebraic sum of the absolute errors

Propagation of determinate error: Product

$$y = a\,b \quad (1)$$
$$(y + \Delta y) = (a + \Delta a)(b + \Delta b)$$
$$ab + a\,\Delta b + b\,\Delta a + \Delta a\,\Delta b \quad (2)$$
$$\Delta y = b\,\Delta a + a\,\Delta b + \Delta a\,\Delta b \quad (2 - 1)$$
$$\Delta y / y = \Delta a / a + \Delta b / b + \Delta a\,\Delta b / ab \quad (2 - 1) / 1$$
$$\approx \Delta a / a + \Delta b / b$$

The relative error of a product is the sum of the relative errors
and *mutatis mutandis*

$$y = a / b$$
$$\Delta y / y \approx \Delta a / a - \Delta b / b$$

Propagation of error: Exponential Calculations

$y = a^n$ (1)

with n (power or root), containing no uncertainty

$dy = n\, a^{n-1}\, da$ (2)

$dy / y = n\, (a^{n-1} / a^n)\, da$ (2 / 1)

$= n\, (da / a)$

The relative error of the power of a is exponent-times the relative error of a.

Propagation of error: Logaritms

$y = \text{Log } a = 0.434 \ln a$ (1)

$dy = 0.434\, da / a$ (2)

Propagation of error: Antilogaritms

$a = \text{antiLog } y$ (1)

$dy/0.434 = da / a$ (2)

The absolute error of the logaritm of a is proportional to the relative error in a, and conversely for the antilogaritm.

Accumulation of indeterminate error

Addition and subtraction:
The errors accumulate with a random algebraic sign.
A pythagorean rule applies to the absolute standard deviations

Multiplication and division:
The errors accumulate with a random algebraic sign.
A pythagorean rule applies to the relative standard deviations (CV's)

Exponential calculus:
Here the sign of the error is necessarily identical throughout the multipications, and thus the pythagorean rule does not apply.
The CV of a power is exponent-times the CV of the component quantity.

Significant Figure Convention

Sum

$$\begin{array}{r} 123.45 \\ +\quad 1.2345 \\ \hline 124.68 \end{array}$$

The last significant position of the result is determined by the last decimal position in the component quantity with the worst precision.

Product

$$\begin{array}{r} 3.33 \\ \hline \times\ 3.0001 \\ 10.0 \end{array}$$

The number of significant positions in the result is determined by the component quantity with the least number of significant figures.

Logaritm

$\log 3.00 * 10^{-3} = -2.523$

The chracteristic (to the left of the digital point) locates the digital point.
The mantissa (to the right of the digital point) has the same number of significant numbers as the original quantity.

Aleatory and epistemic uncertainties

Aleatory uncertainty : uncertainty induced by randomness

Epistemic uncertainty : uncertainty induced by imperfect knowledge about something that in principle is knowable

Basic example :

$y = y_0 + S_W\, p + S_B\, q$ (*e.g.* true mean ± within- and between-batch variance)

y_0 , S_W, S_B Epistemic variables
(imperfectly known expert estimates
for the true mean, and between and within variability)

p, q Aleatoric variables
(zero-mean unit-variance random variables)

Test-drive simulator: accumulation of internal quality control data 

Extended applications :

Any calculated parameter making use of estimated constants
When the confidence interval of such estimates is known, it can enter the model.

Internal Quality Control (iQC)

Statistical Process Control is treated in detail in a separate teaching module 

Use of iQC data

Eliminate a misconception about bias
In iQC follow up you see layers of data points,
jumping to new levels after a calibration event, new reagent lot, etc.
This is NOT BIAS BUT RANDOM ERROR
- you can not predict the sign of the change

ANOVA analysis of your iQC data : simplified UM model
This corresponds to a Taylor expansion
- indeterminate error evaluated directly
  (with abstraction of sensitivity)
- covariance = 0

Test-drive our simulator 

Use of iQC data

Take-home observations from running the simulator

The iQC long-term SD can approximate the combined UM, provided :
- a sufficient number of significant events are accumulated

An accurate estimate of the iQC target (central tendency & dispersion)
- depends on regular updates
- results in these benefits:
  - the false detection rate of the iQC-rule will decline
  - the true detection rate of the iQC-rule depends mainly on the power of the rule (but may also benefit)

ANOVA: identify principal components of variance &
- can method be made more robust ?
- can you simplify procedure ?

Use of iQC data

ISO 15189 § 5.6.2.

The laboratory shall determine the uncertainty of results, where relevant and possible. Uncertainty components, which are of importance, shall be taken into account.

Sources that contribute to uncertainty may include sampling, sample preparation, sample portion selection, calibrators, reference materials, input quantities, equipment used, environmental conditions, condition of the sample and changes of operator.

The lab shall have a procedure

Lab Procedure:

1. For quantitative tests, the UM is available on demand.
At relevant levels, the combined UM is approximated by
the long-term iQC imprecision times a coverage factor (1.95).
2. The consultative function of the lab addresses other elements of UM (definitional uncertainty, systematic and spurious errors).

New Developments

Probabilistic Model *vs* Risk Model

Classical MU : Probabilistic Model
- construct model for all steps in the analytical process
- estimate / assign indeterminate error for each step
- compound the error-propagation
- limited to quantitative results
- reported as a confidence interval

estimate of total dispersion · coverage factor

Extended MU : Risk Models 
- inventory of significant quality failures
  - insufficient analytical quality
  - insufficient specificity
    - analytical non-specificity
    - biological / physiological interfering responses
- estimate extent and frequency of such (spurious) events
- applicable to quantitative and non-quantitative results
- reported as an inventory of pitfalls (= body of knowledge)

Reasonable Dispersion *vs* Total Error

Classical MU :

Data model $X = X_{accepted} + e_i * random(0,1)$

Reasonable variability

Total Error :

Data model $X = [X_{true} + \Delta_i] + e_i * random(0,1)$

Epistemic variables

Bias

Fallacy of the total error (bias) model :

While bias does occur,
as soon as it is known,
it no longer exists (except for its epistemic uncertainty),
and as long as it is not known,
it can only enter the calculus as a hypothetical.

Literature

NIST reference on Constants, Units and Uncertainty
NIST-TN 1297 (1994)

The expression of uncertainty in quantitative measurement
EAL-G23 (1996)

Quantifying Uncertainty in Analytical Measurement
EURACHEM / CITA Guide CG4 (QUAM 2000.1)

Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025
ILAC G17 (2002)

Guide de Métrologie á l'usage des laboratoires d'analyses de biologie médicale
(Collège Français de Méterologie)

Uncertainty of Measurement in Quantitative Medical Testing
(Australasian Association of Clinical Biochemeists, 2004)

Internet resources

- http://physics.nist.gov/cuu/index.html
- http://www.measurementuncertainty.org/mu/guide/index.html