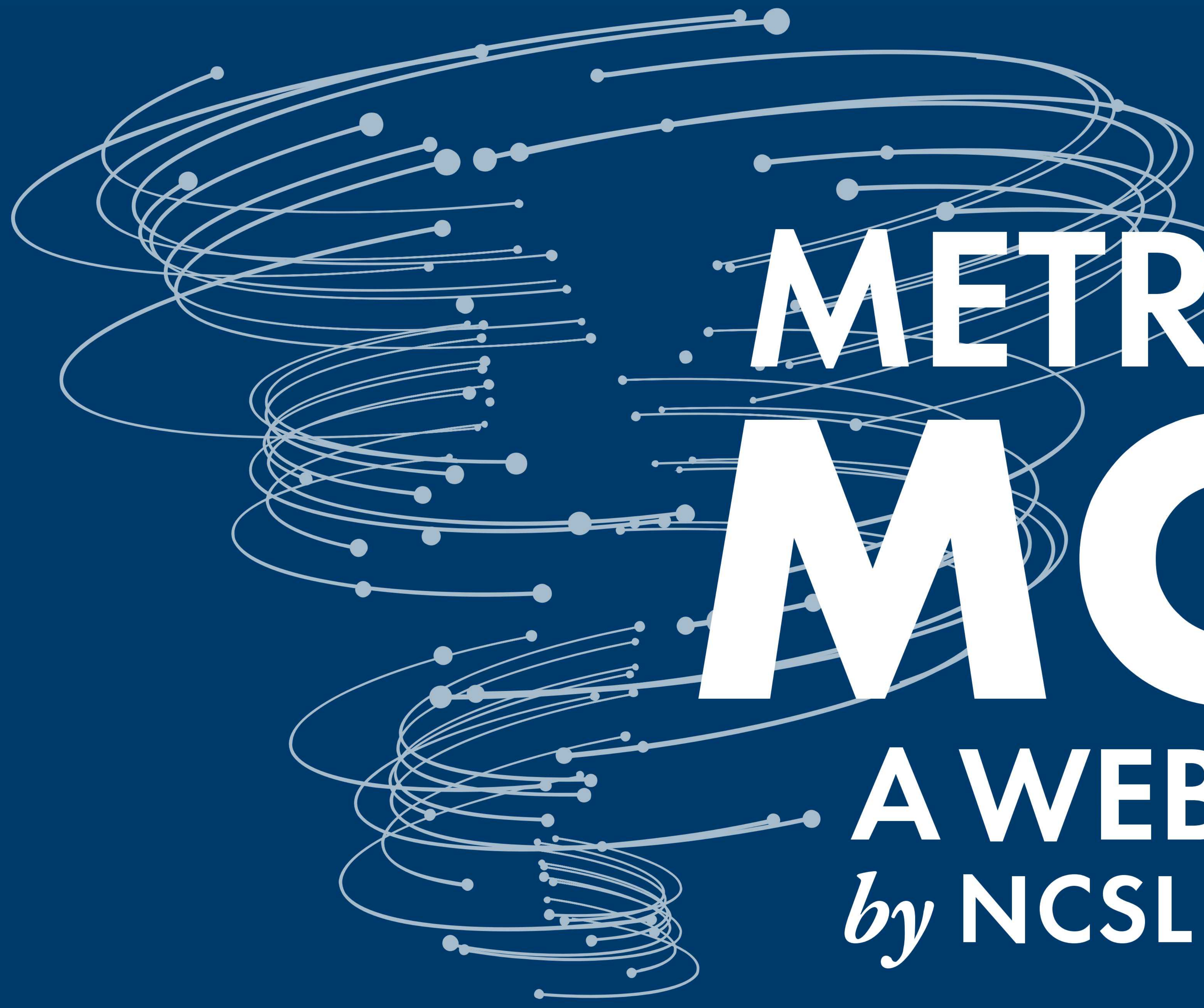
SAND2020-5632PE



METROLOGY *in* MOTION

A WEBINAR SERIES *by* NCSL INTERNATIONAL

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Uncertainty for Technicians

Part 1: What is this uncertainty thing, anyway?

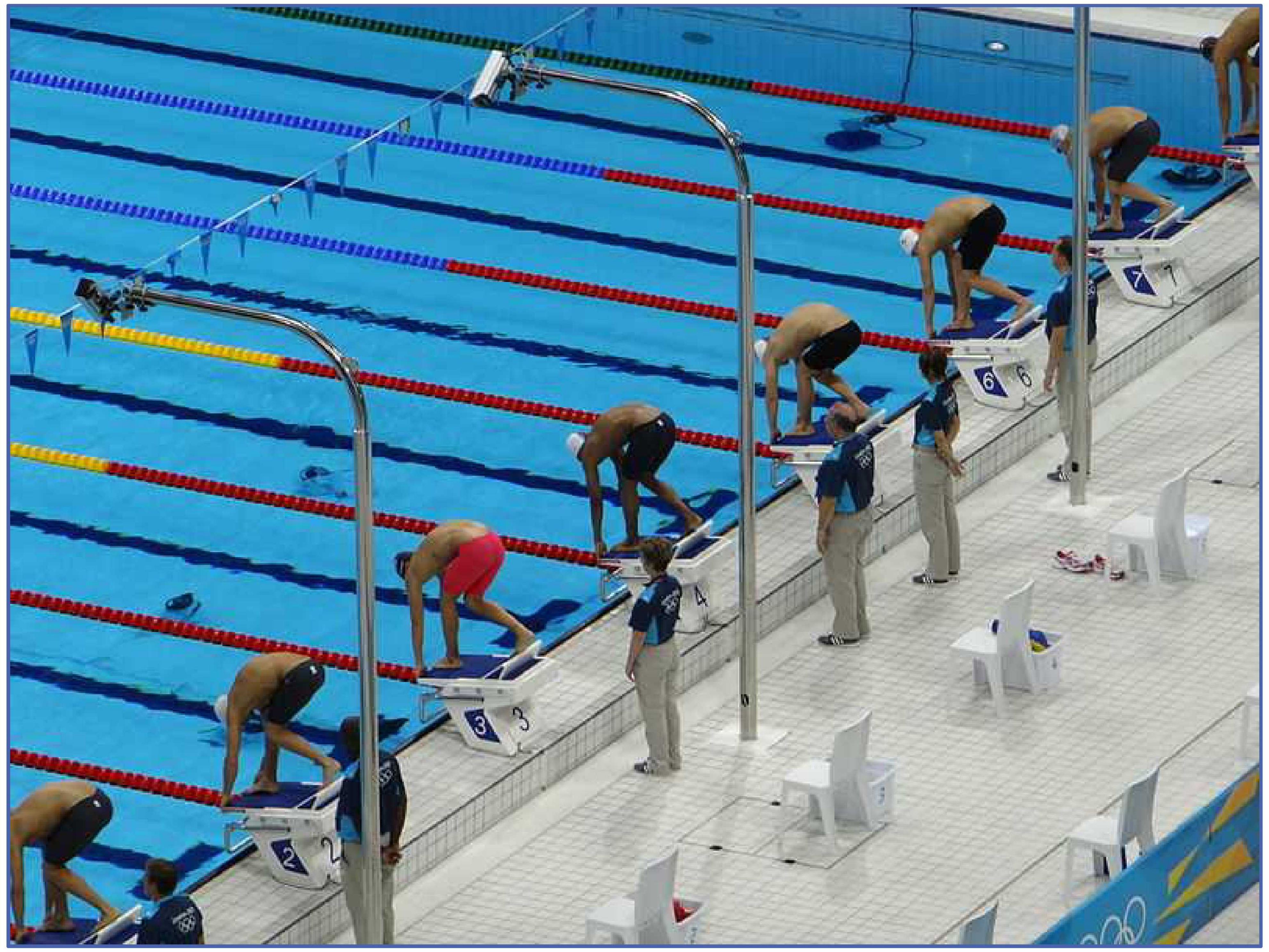
Collin J. Delker

Primary Standards Lab,
Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

METROLOGY *in* MOTION
A WEBINAR SERIES *by* NCSL INTERNATIONAL

Why do we care about uncertainty?

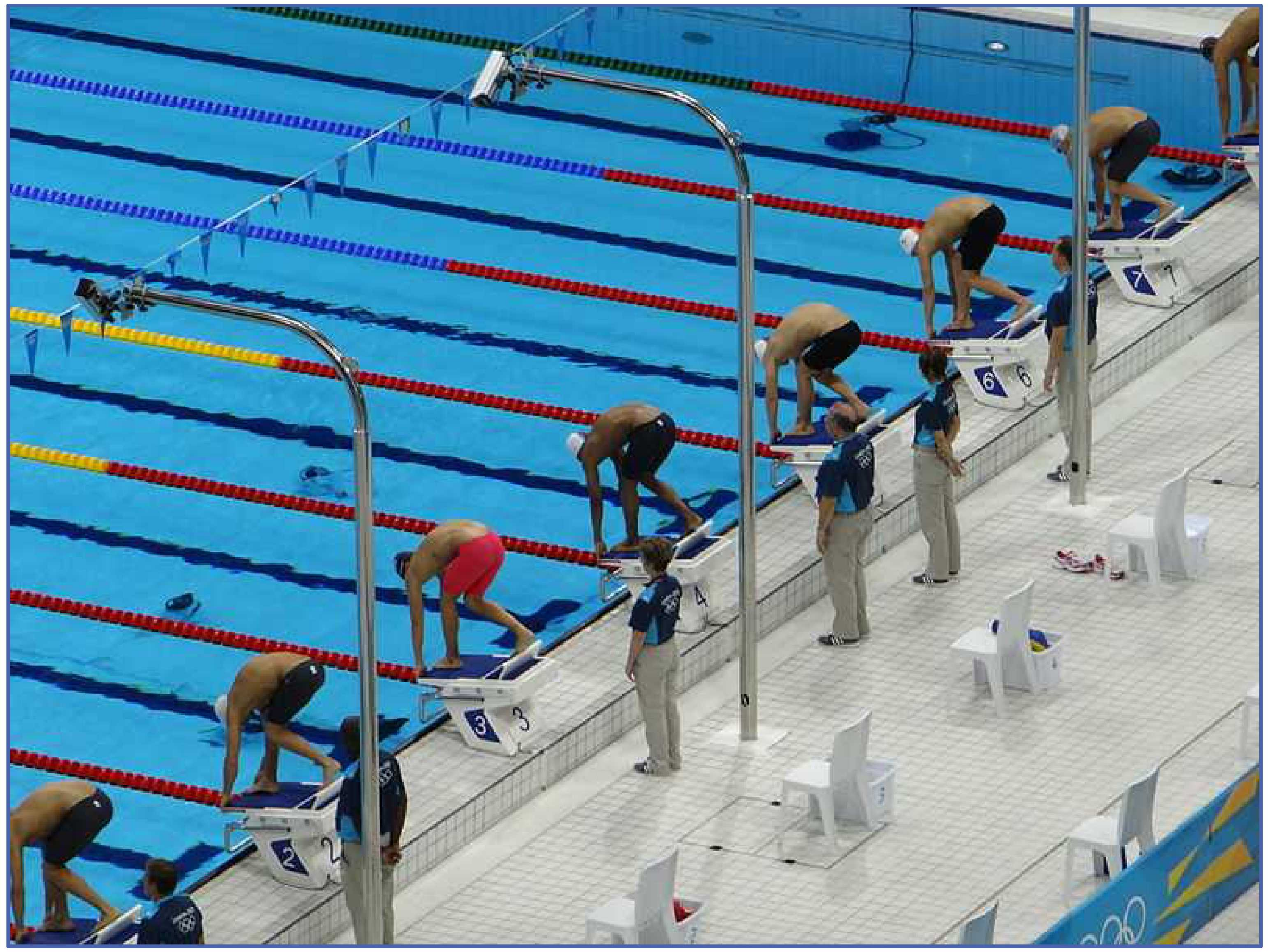


Olympic Swimming Pools

- 3 cm tolerance in length between lanes
- $t = d/v$
- 50 m ± 1.5 cm lane at 2.4 m/s → ± 0.006 s

- Results of 1972 games overturned because of this uncertainty
- Most Olympic events are timed with 0.001 s resolution, but swimming must be timed to 0.01 s because of this uncertainty.

Why do we care about uncertainty?



Swimming facility regulations have tolerances on:

- Dimensions
- Temperature
- Salt content of water
- Water flow/turnover rate
- Illumination
- PA system loudness and frequency response
- Vibrational frequencies of diving boards

Why do we care about uncertainty?

New Mexico Philharmonic, nmphil.org

Violin Precision

- Frequency of vibrating string: $f_2 = \frac{f_1 L_1}{L_2}$
- Smallest perceptible frequency change:
  5 “cents” ~ C_5 = 523 Hz ± 1.5 Hz
- Open string length: 32 cm (440 Hz, A_4)

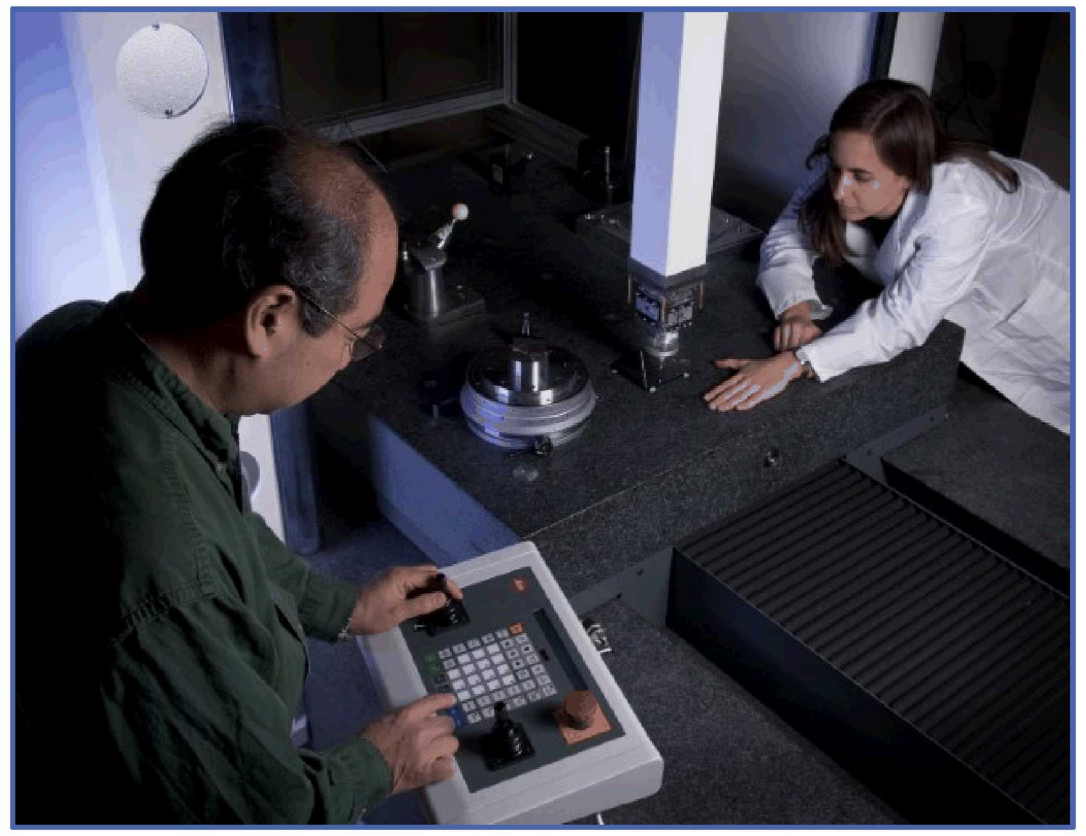
- → String must be pressed within ± 0.5 mm for in-tune pitch

Metrology at Sandia National Laboratories

Sometimes the stakes are higher than just our entertainment…

https://www.youtube.com/watch?v=-9GYcThATBg

Sandia Primary Standards Laboratory

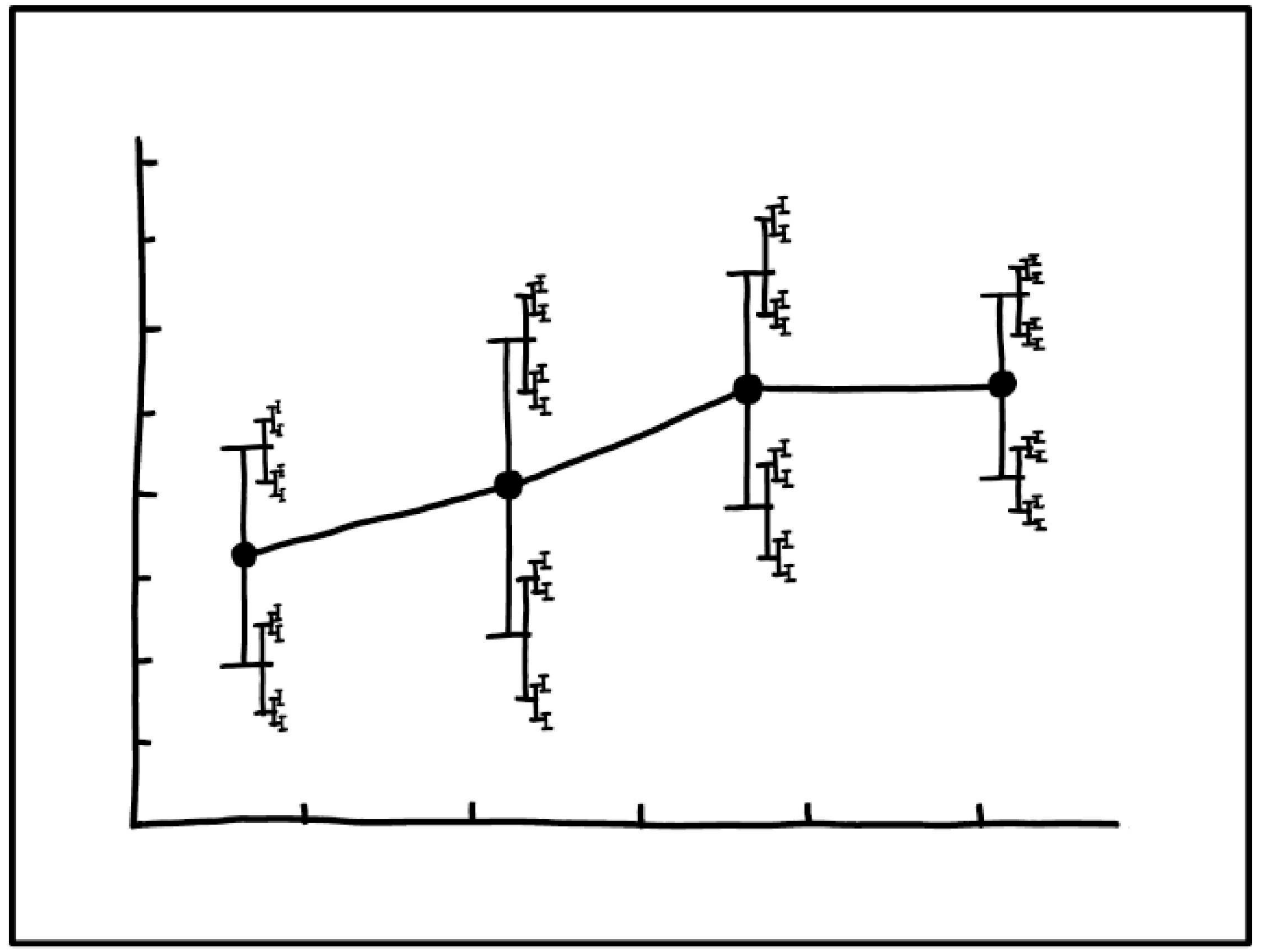


Poll

- What type of measurements to you typically make?
  - Electrical
  - Dimensional
  - Mass, Force
  - Pressure, Vacuum
  - Temperature
  - Radiation
  - Other

Part 1 Outline

- The calibration cycle
- Traceability
- Metrology and uncertainty vocabulary
  - The old-school “Error Approach”
  - The modern “Uncertainty Approach”
- Type A and Type B evaluations of uncertainty
- Uncertainty Sources
- Uncertainty Budgets



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.

Parts 2 and 3 Outline

- Part 2: Determining Type A, Type B, and combined uncertainties
- Part 3: Indirect measurements (GUM method), measurement risk, and Test Uncertainty Ratios (TURs)

The Calibration Cycle

"Everyone who uses the results of experiments must sooner or later ask 'How reliable are these results?' … the honest experimenter must provide the reader with some measure of the reliability of the results."

— S. J. Kline and F. A. McClintock, Describing Uncertainties in Single Sample Experiments

Calibration vs. Tolerance Testing

Calibration: The comparison of a device under test to a known reference standard

- Certificate reports the measured value and uncertainty

Tolerance Test: A pass-or-fail determination based on whether a measurement falls within required limits

- Certificate reports a pass-or-fail status and TUR
- Typically requires TUR of at least 4 (stay tuned for Part 3!)

Adjustment: Modification of the device's properties to ensure performance within tolerance

- Often performed as part of a tolerance test or calibration, but not always

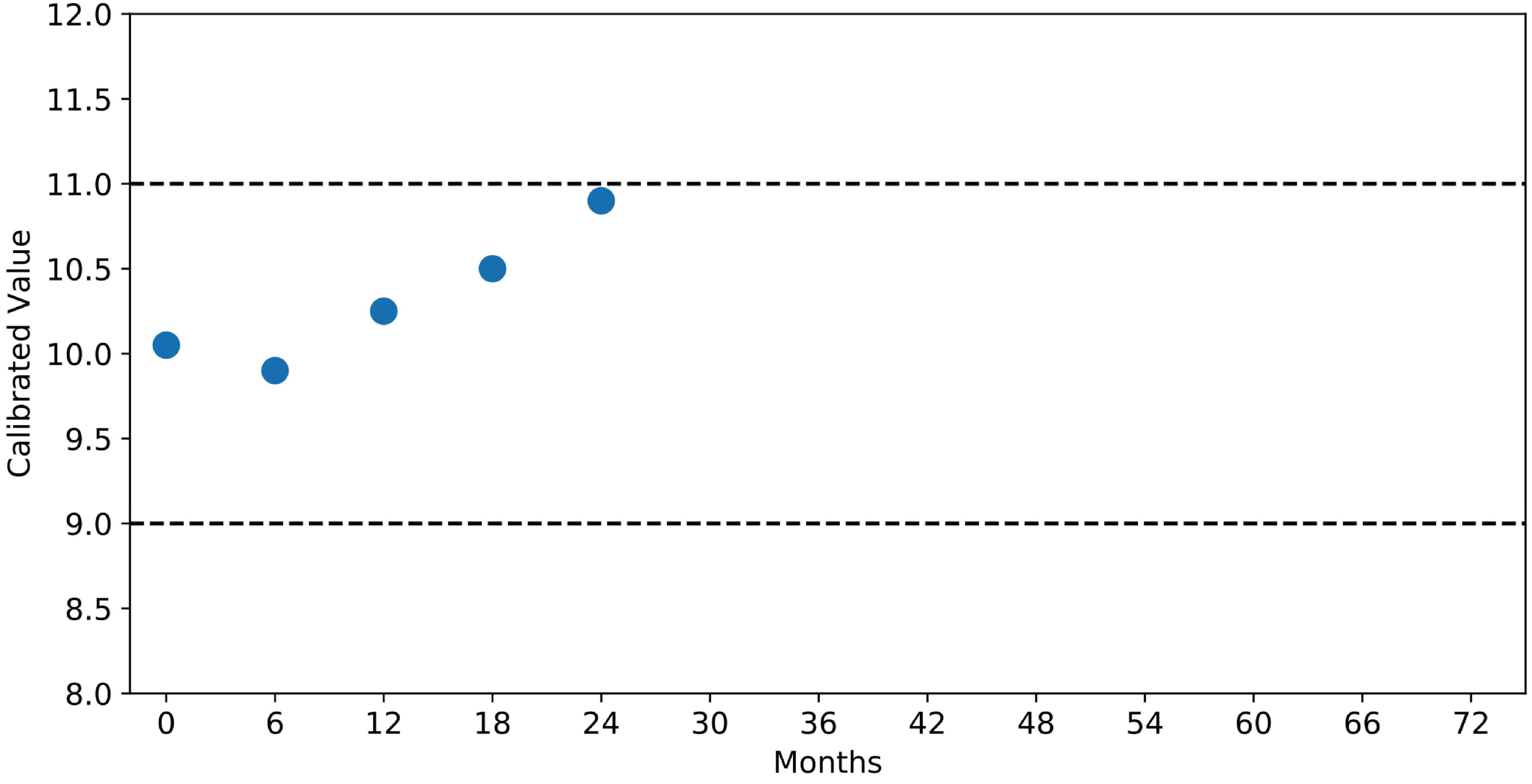
Calibration ≠ Adjustment

You can calibrate a dog, but adjusting one would be a bad idea.

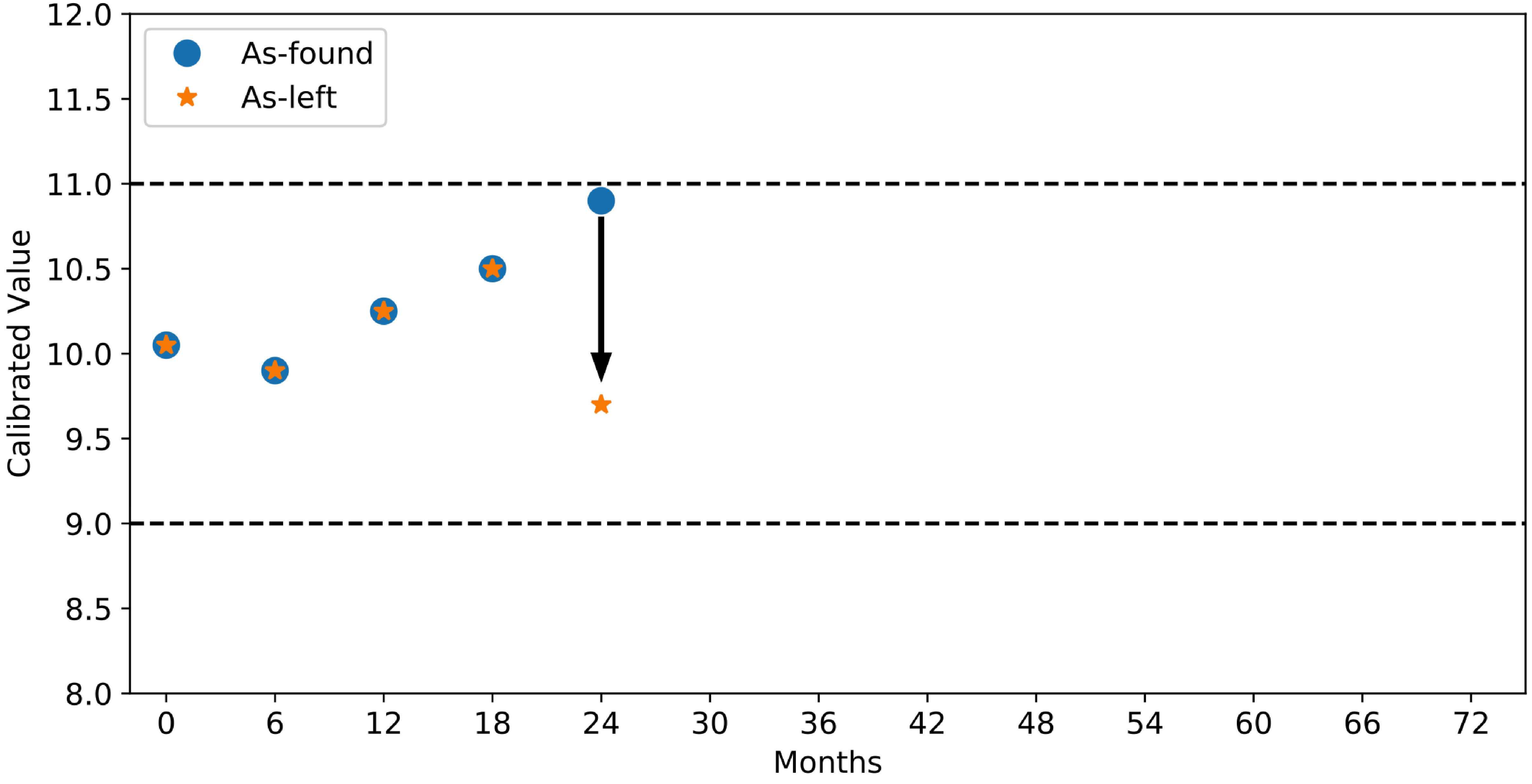
Ymnes/CC0-1.0

Credit: Andrew Mackrory, SNL

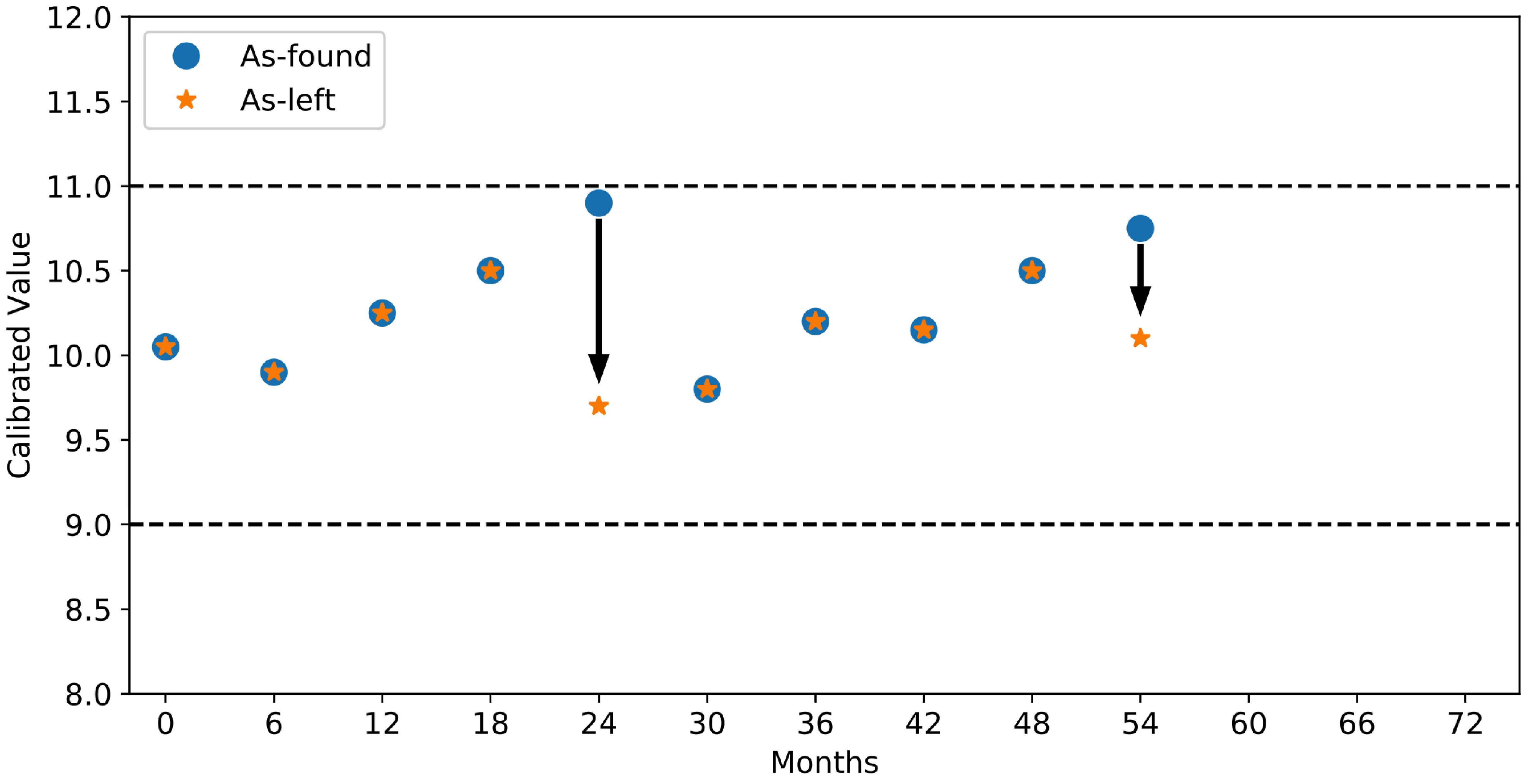
The Calibration Cycle



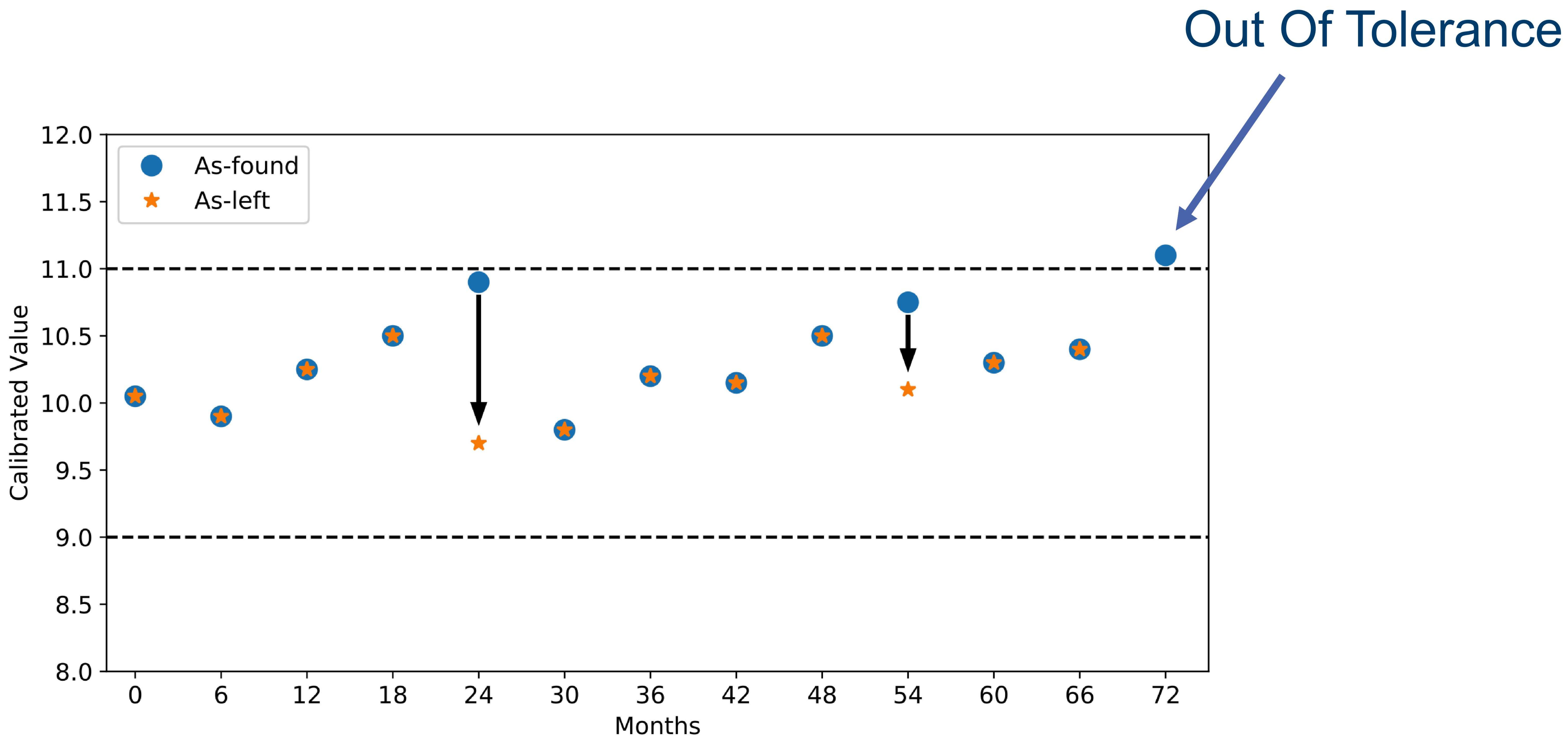
The Calibration Cycle



The Calibration Cycle

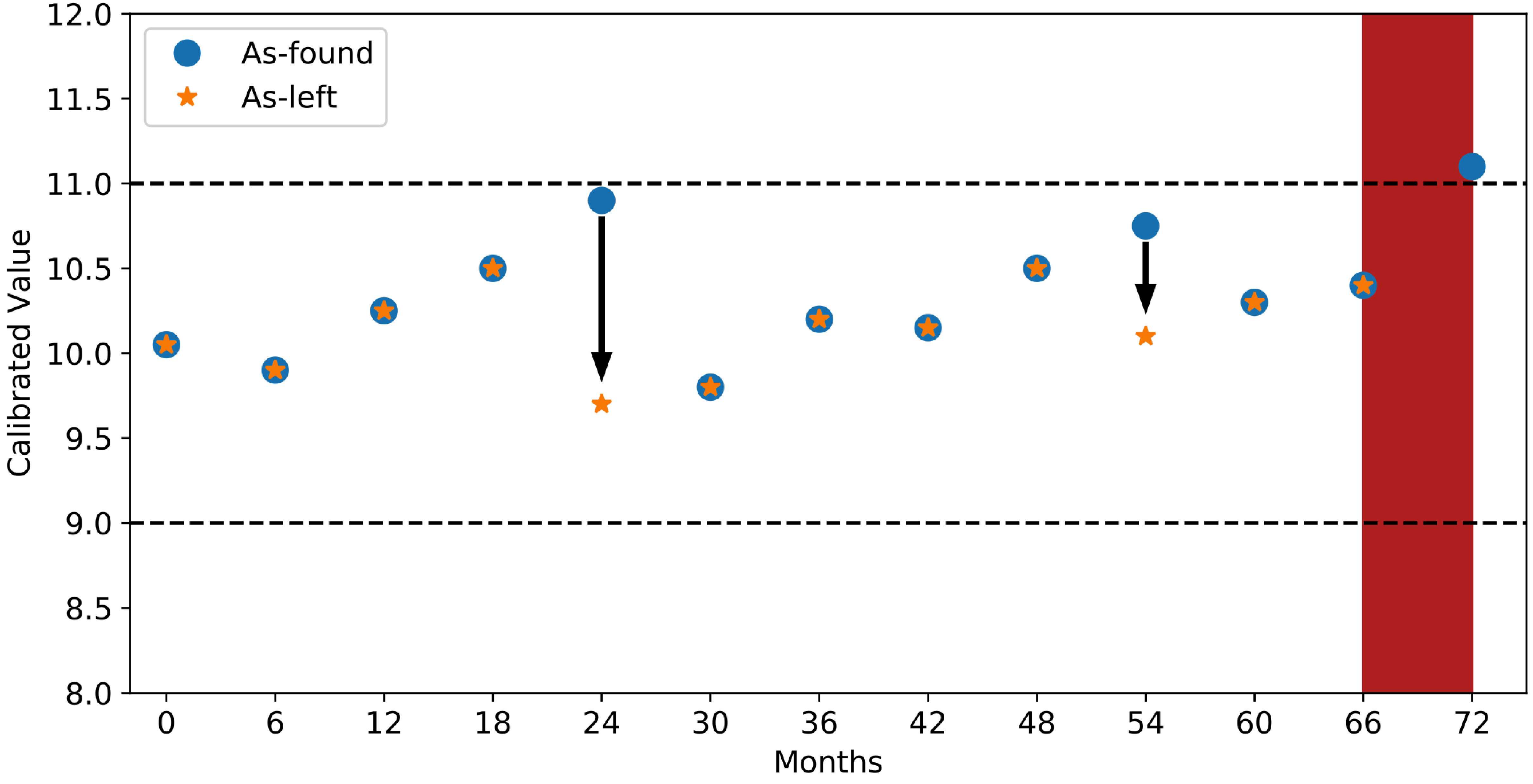


The Calibration Cycle



The Calibration Cycle

All measurements made using the equipment during this time are suspect.



Traceability

- A **traceable measurement** result is connected to the SI base units through a documented, unbroken chain of calibrations, each contributing to total uncertainty.

Mass Traceability Chain Before 2019

IPK Kilogram prototype

BIPM 1 kg working standard

NIST primary 1 kg standard

NIST working mass set

$1.99

Supermarket scale

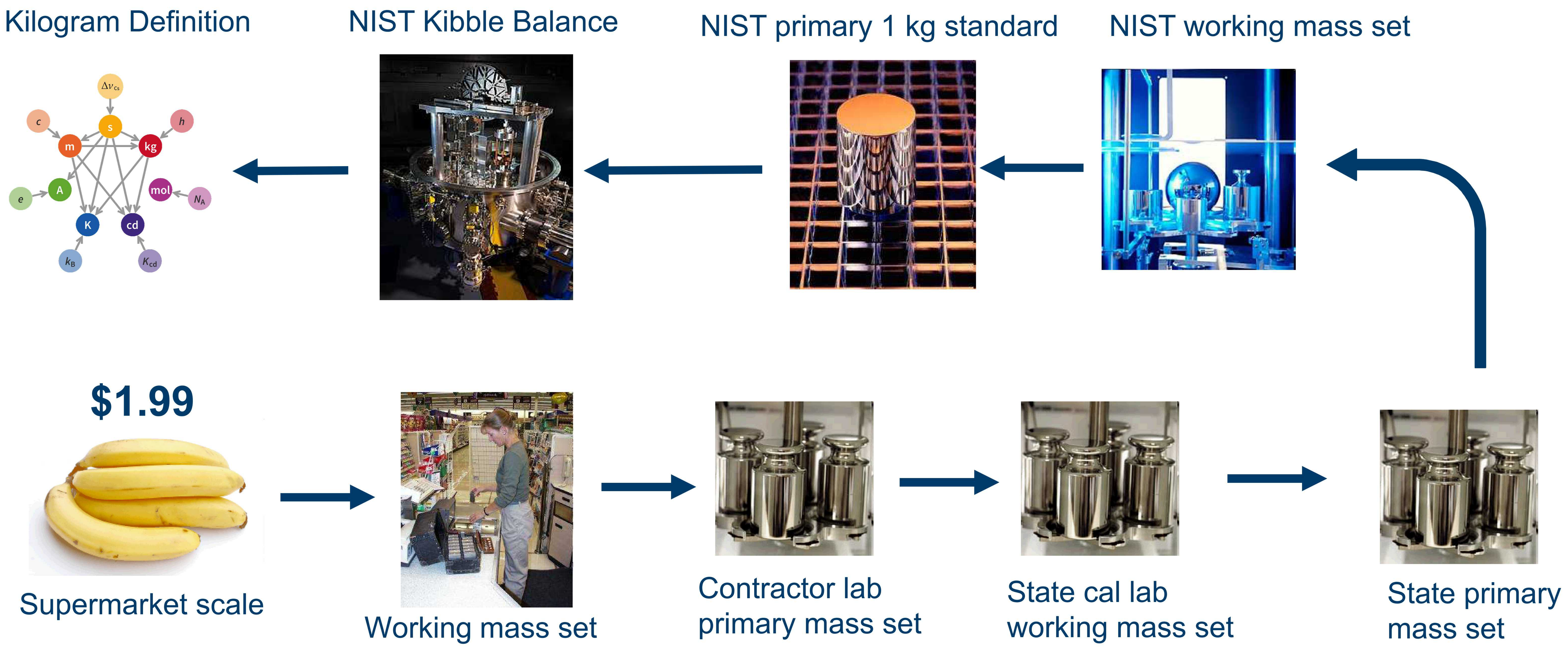
Working mass set

Contractor lab primary mass set

State cal lab working mass set

State primary mass set

Mass Traceability Chain After 2019*



* See NCSLI Webinar “Metrology and the Updated SI!” by Pat Abbott for why this isn’t the case just yet.

Uncertainty Terminology

"The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement."

— Guide to the Expression of Uncertainty in Measurement (GUM) 3.4.8

See also: JCGM 200:2012, the "VIM"

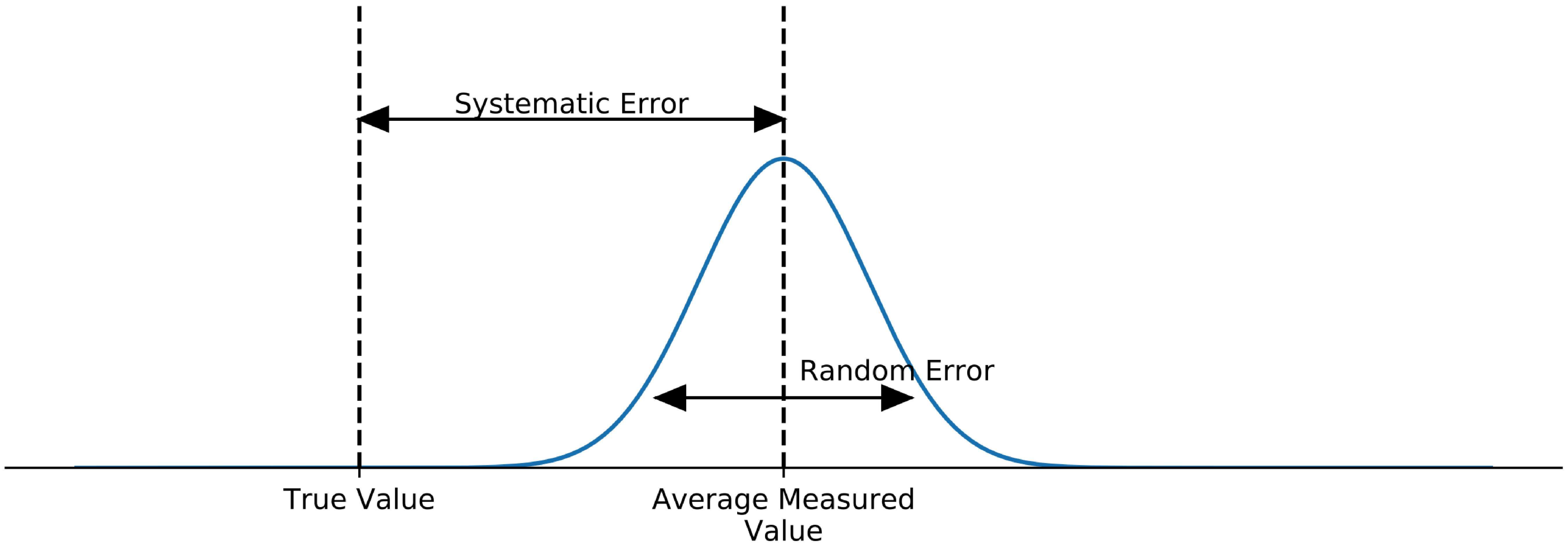
The “Error Approach”

Y = True Value + Measurement Error

= True Value + Systematic Error + Random Error

- Y: Measured Value
- Systematic Error: Component of error that is predictable or constant
- Random Error: Component of error that is unpredictable

The “Error Approach”



- “True Value” can never be known, so exact Measurement Error is not quantifiable.
- Error Approach **is not the preferred terminology.**

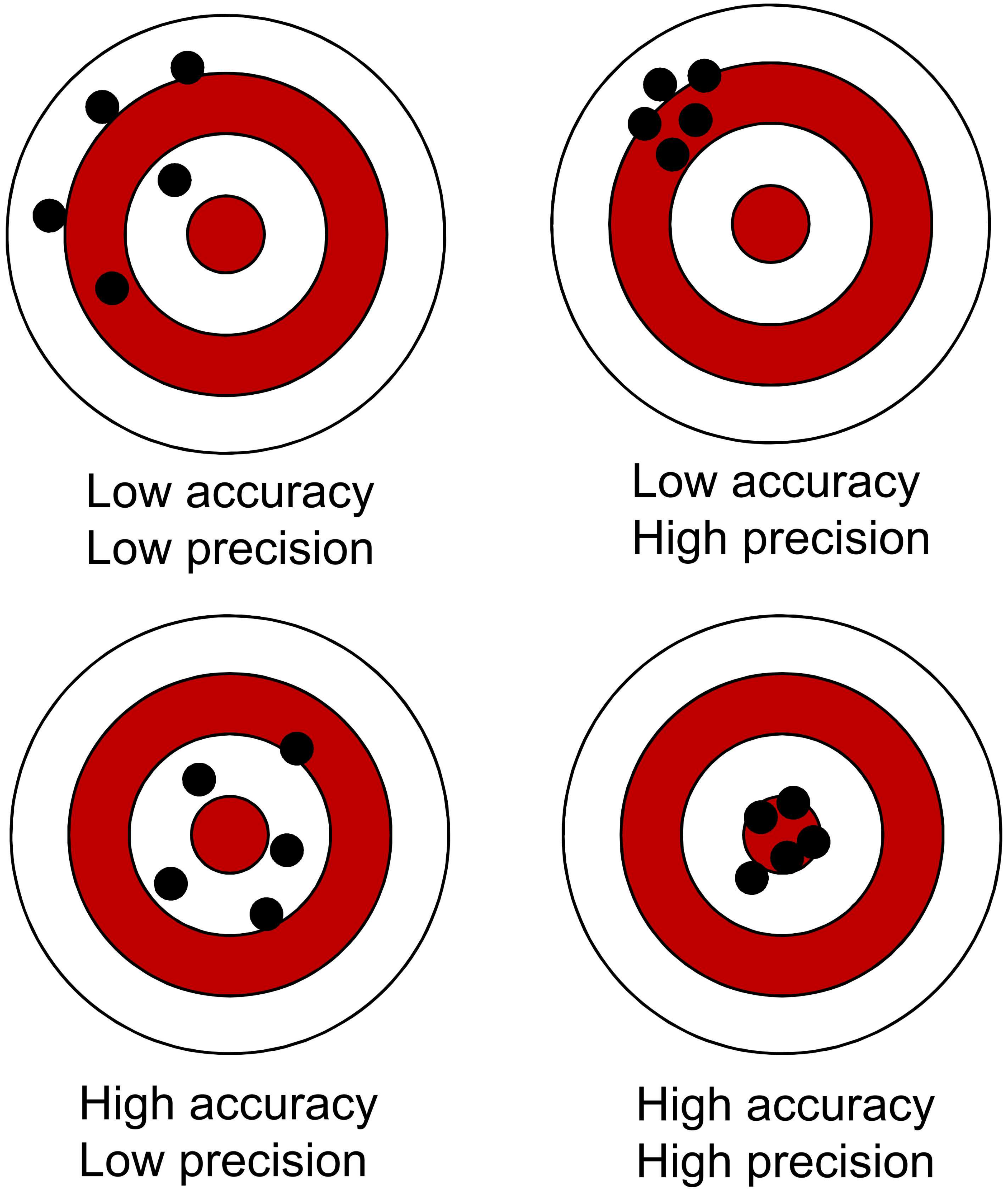
- Still encountered, so presented here as a reference

Accuracy and Precision

- Accuracy and Precision are related to measurement **repeatability**.

- **Accuracy**: Closeness of repeated results to the true value.
- **Precision**: Closeness of repeated results to each other.

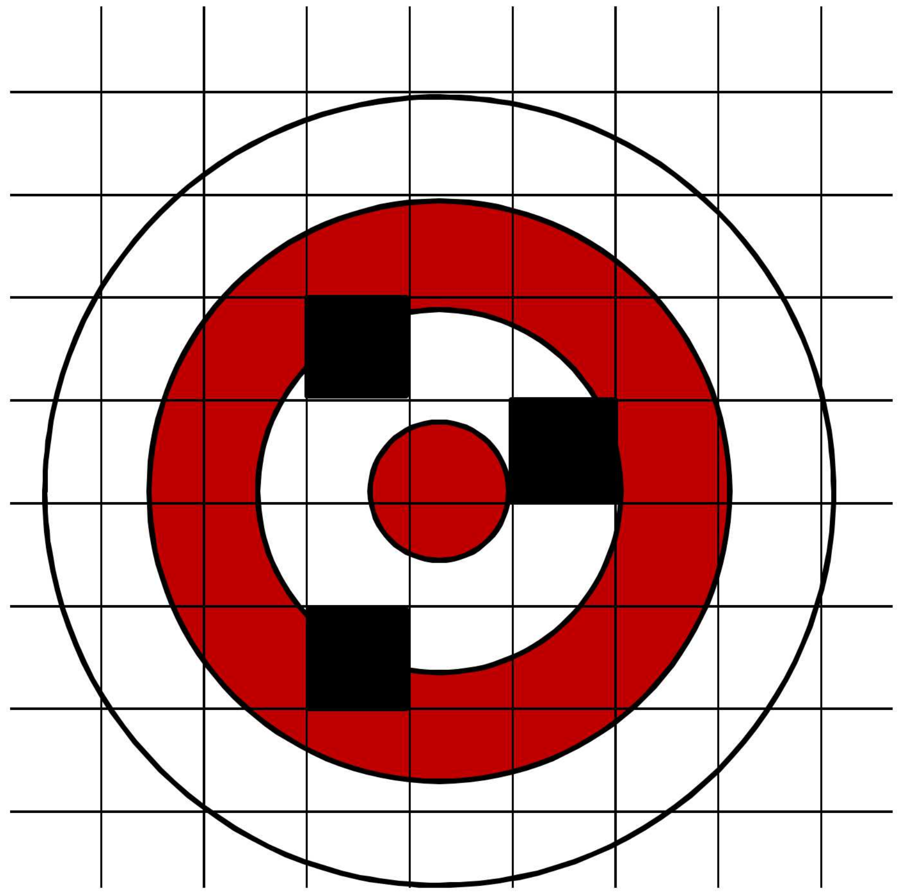
- In metrology, these terms are useful but are typically used in a **qualitative** sense.



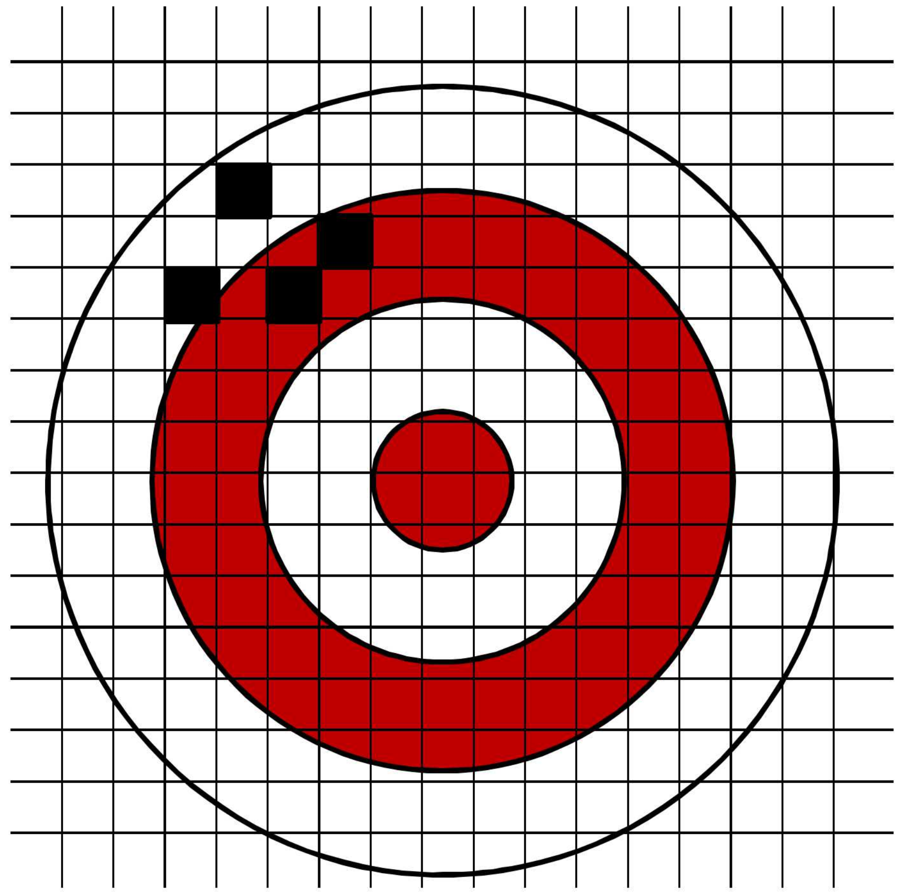
Resolution ≠ Accuracy ≠ Uncertainty

Resolution: Smallest change that can be distinguished by a measurement instrument.

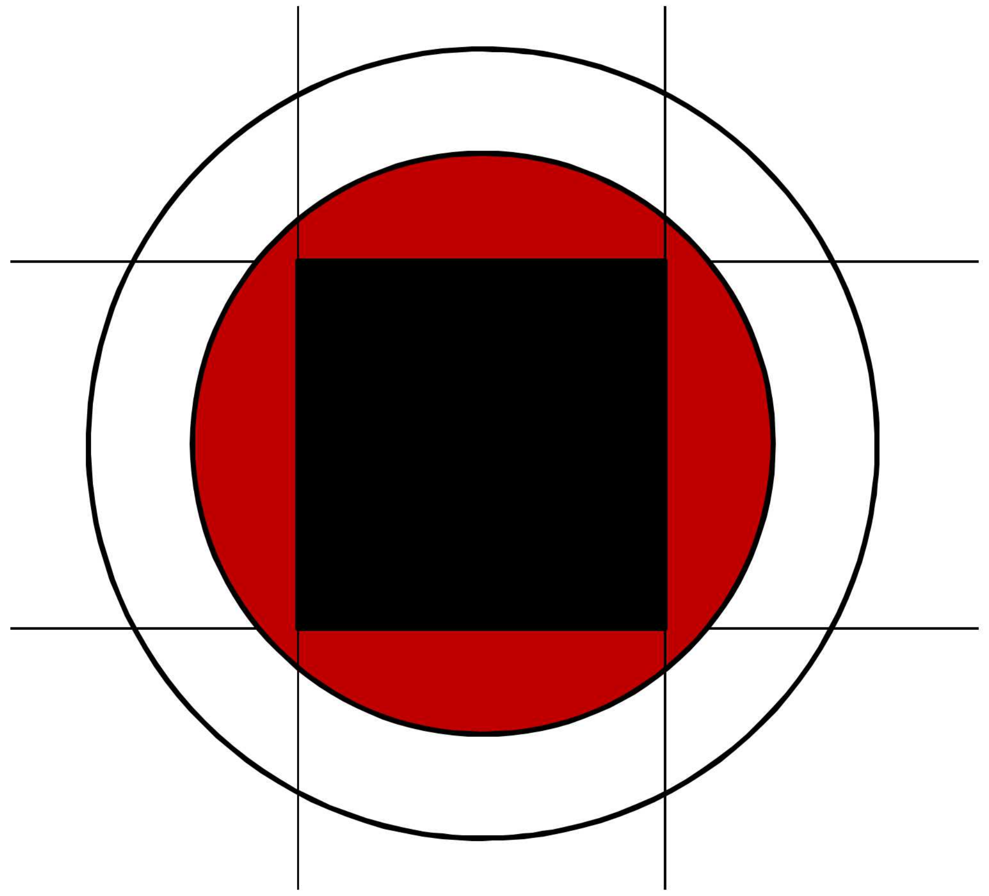
Resolution may contribute to uncertainty, but is not the same thing.



Low resolution



High resolution



Terrible resolution

The "Uncertainty Approach"

Y ± Measurement Uncertainty

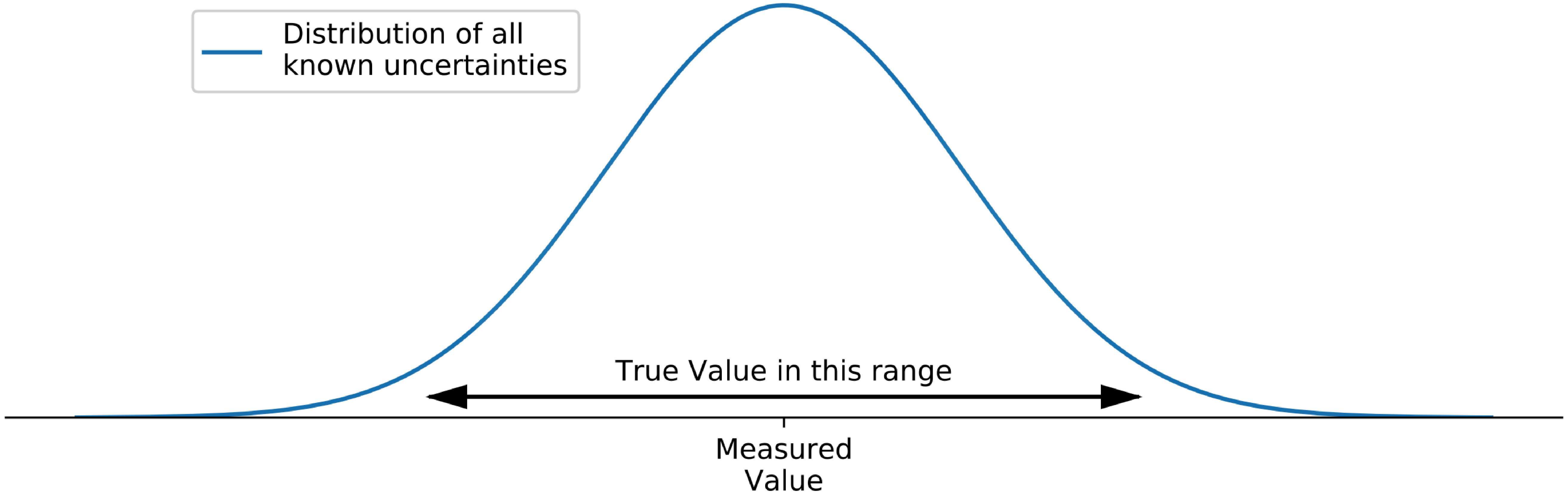
- Y is the best estimate of the true value.
- Uncertainty is a **quantifiable range of probable values** for the measurand based on all available information.
- The preferred terminology of uncertainty and metrology.

- Measurement Uncertainty is the combination all known uncertainties of the measurement, expressed at a specific level of confidence (usually 95%).

The "Uncertainty Approach"

The true value can never be known, only estimated.

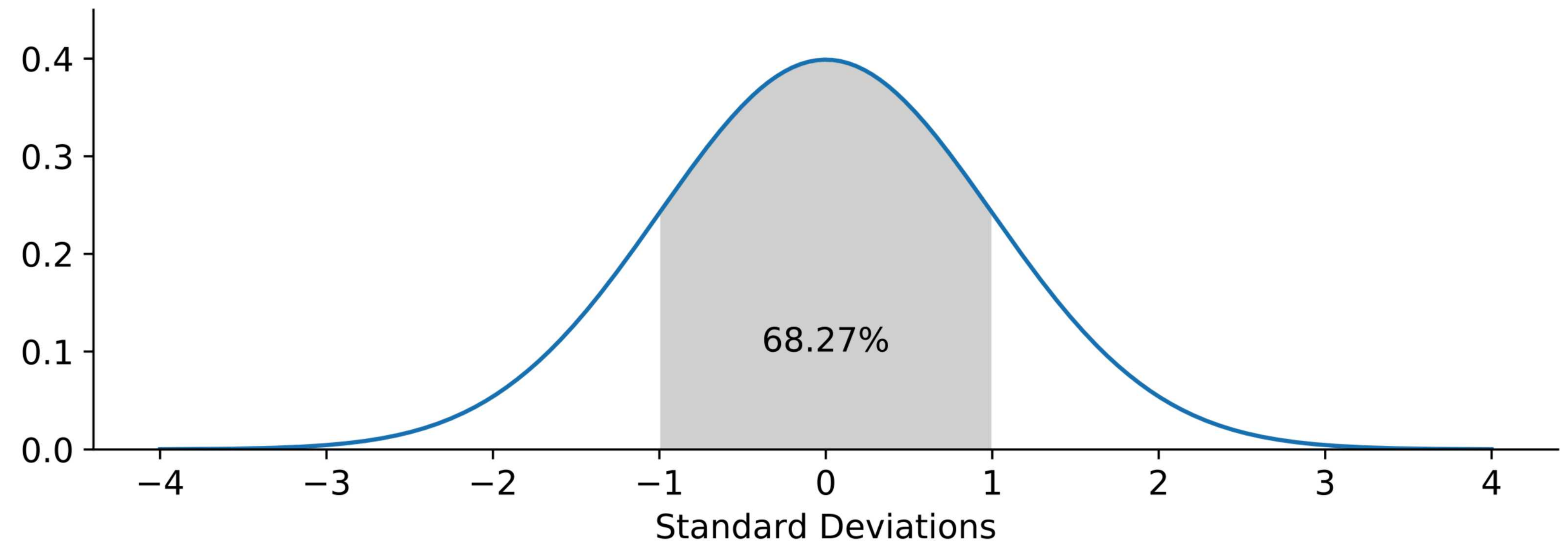
Uncertainty describes where the true value is likely to be.



Expanded Uncertainty

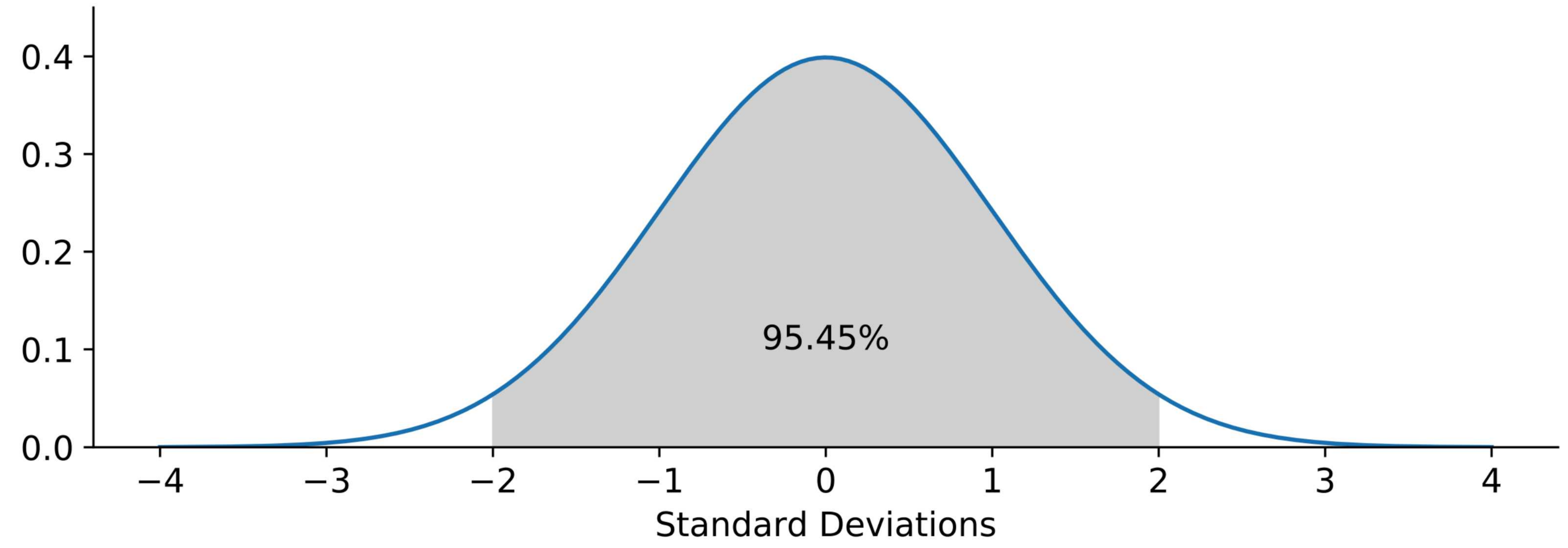
Standard Uncertainty: Uncertainty as one standard deviation (68% confident)

- Lower case u



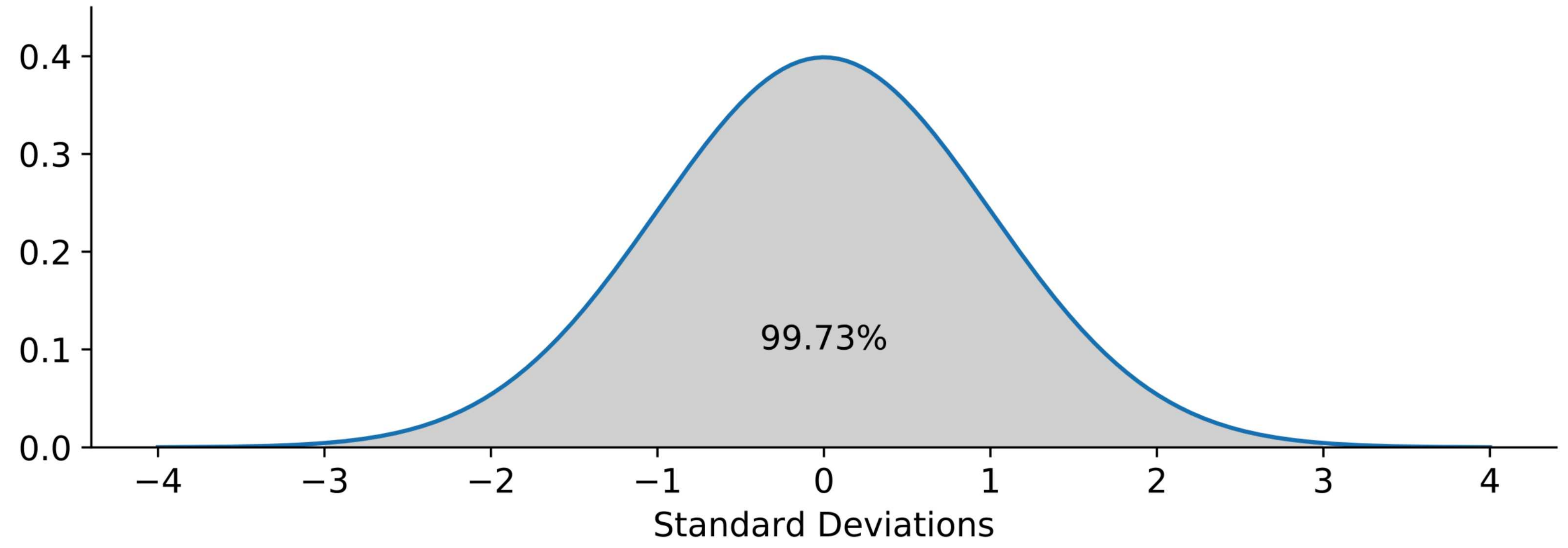
Expanded Uncertainty: Uncertainty expressed at a higher level of confidence, usually 95%

- Upper case U, often with confidence as superscript U^{95}



Coverage factor: Multiplier used to expand the standard uncertainty

- k (usually around 2)



Type A and Type B Uncertainty Evaluation

Type A: Uncertainty derived from statistical analysis of current test data

Type B: Uncertainty derived from other sources

This categorization applies to **how the uncertainty was determined**, not the physical source of the uncertainty!

Type A and Type B Uncertainty Evaluation

Type A

- Spread of multiple measurements
- Repeatability, reproducibility
- Check Standard Statistics
- Historical data analysis
- Design of Experiments studies

Type B

- Calibration certificates
- Manufacturer's specifications
- Reference data from handbooks
- Known environmental fluctuations
- Other information

My Type A might be your Type B. In the end, it usually doesn't matter anyway:

$$U = k\sqrt{typeA^2 + typeB^2}$$

Repeatability and Reproducibility

Repeatability: Variation under the **same** operator, measuring system, and procedure over a short period of time.

Reproducibility: Variation due to **different** operators, measuring systems, and/or procedures, often over longer periods of time.

Physical Sources of Uncertainty

- Calibration and standards
- Procedures
- Environmental factors
- Operator
- Equipment selection
- Choice of measurement model
- Other random variation

Any of these could be Type A or Type B

Uncertainty Budget

Uncertainty	Standard Uncertainty μΩ/Ω	Description
Reference standard	0.01	Time-of-test uncertainty as calibrated by NIST
Drift uncertainty	0.006	Statistical analysis of drift on the reference standard
Resistance bridge uncertainty	0.013	1Ω Specifications for MI-6010C bridge
Resistance bridge linearity	0.0029	1Ω Specifications for MI-6010C bridge
Resistance bridge resolution	0.00029	1Ω Specifications for MI-6010C bridge
Temperature fluctuation	0.0021	Includes oil bath stability and temperature coefficient of standards
Pressure fluctuation	0.0017	From pressure coefficients of standards and typical atmospheric changes during test
Repeatability uncertainty	0.00862	Pooled standard deviation of 200 measurements
Combined uncertainty	0.0196	RSS of above components (k = 1)

Credit: PSL DC Lab

What to do with all these uncertainties?

- Make sure they're expressed as standard uncertainties
- Then Root-Sum-Square them together!

$$u_c = \sqrt{\sum_{i=1}^{N_A} u_{A_i}^2 + \sum_{i=1}^{N_B} u_{B_i}^2}$$

Brainstorming

- Brainstorm some sources of uncertainty when calibrating a car’s speedometer by comparing against the time between mile markers using a stopwatch.

Poll: What's the biggest source of uncertainty?

- Accuracy of stopwatch
- Accuracy of mile marker placement
- Human error in starting/stopping the watch
- Assumption that vehicle travels with constant velocity
- Curvature of road
- Weather conditions
- Procedure – single mile marker or average of multiple?
- Lunar gravitational pull on vehicle

Review

- **Calibration**: Comparison to a known reference standard. Results in statement of uncertainty.
- **Adjustment**: Modification of device parameters to meet requirements.
- **Tolerance Test**: Determination of whether a measurement falls within some limits.
- **Traceability**: Unbroken chain of calibrations connecting a measurement to the SI
- **Accuracy**: Closeness of repeated measurements to the true value.
- **Precision**: Closeness of repeated measurements to each other.
- **Uncertainty**: Range of values where the true value is likely to be.
- **Standard Uncertainty**: Uncertainty in terms of one standard deviation (68% confident).
- **Expanded Uncertainty**: Uncertainty at a higher (usually 95%) level of confidence.
- **Type A Evaluation of Uncertainty**: Derived from statistical analysis of current test data.
- **Type B Evaluation of Uncertainty**: Derived from other sources.

Next Time…

How do you actually calculate Type A and Type B Uncertainties?

METROLOGY *in* MOTION

A WEBINAR SERIES *by* NCSL INTERNATIONAL

Uncertainty for Technicians

Part 2: Putting some certainty in your uncertainty budgets

Collin J. Delker

Primary Standards Lab,
Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

METROLOGY *in* MOTION
A WEBINAR SERIES *by* NCSL INTERNATIONAL

Part 2 Outline

- Type A Uncertainty Evaluation
    - Repeated measurements
    - Pooled variance and repeatability
    - Reproducibility
- Type B Uncertainty Evaluation
    - Probability distributions
    - Normal, Uniform, Others
- Combined Uncertainty
- Expanded Uncertainty

Poll – SI Trivia

Which of these has NOT been used as the basis of the SI meter:

- The length of a Platinum-Iridium bar (IPM)
- The length of a pendulum with a specific period
- The distance from the equator to the poles
- The length traveled by light during a specific amount of time
- A specific number of wavelengths of radiation emitted by Krypton-86

Type A and Type B Uncertainty Evaluation

- **Type A**: Uncertainty derived from statistical analysis of current test data
- **Type B**: Uncertainty derived from other sources

- This categorization applies to **how** the uncertainty was determined, not the **physical source** of the uncertainty!

Type A and Type B Uncertainty Evaluation

Type A

- Spread of multiple measurements
- Repeatability, reproducibility
- Check Standard Statistics
- Historical data analysis
- Design of Experiments studies

Type B

- Calibration certificates
- Manufacturer's specifications
- Reference data from handbooks
- Known environmental fluctuations
- Other information

"Type A" and "Type B" describe **how** the uncertainty was evaluated

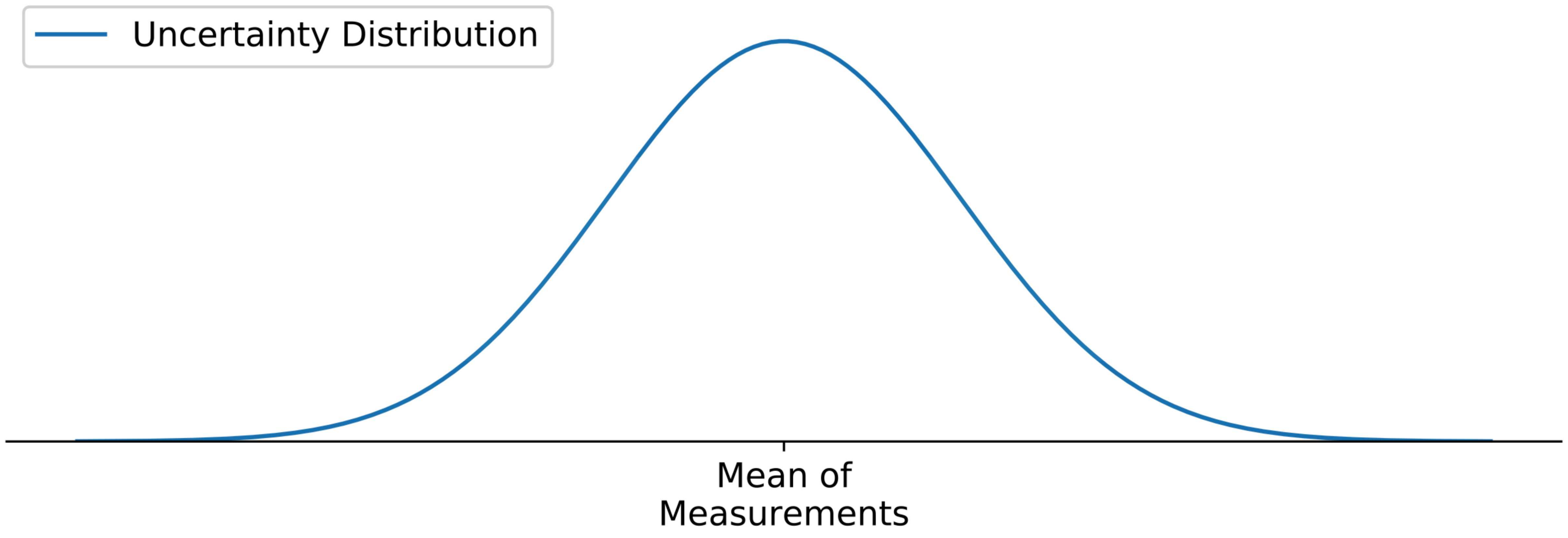
Type A Uncertainty

Evaluated using statistics on current measurement data

"Ideally, we would like to repeat all measurements enough times using enough observers and enough diverse instruments so that the reliability of the results could be assured by the use of statistics."

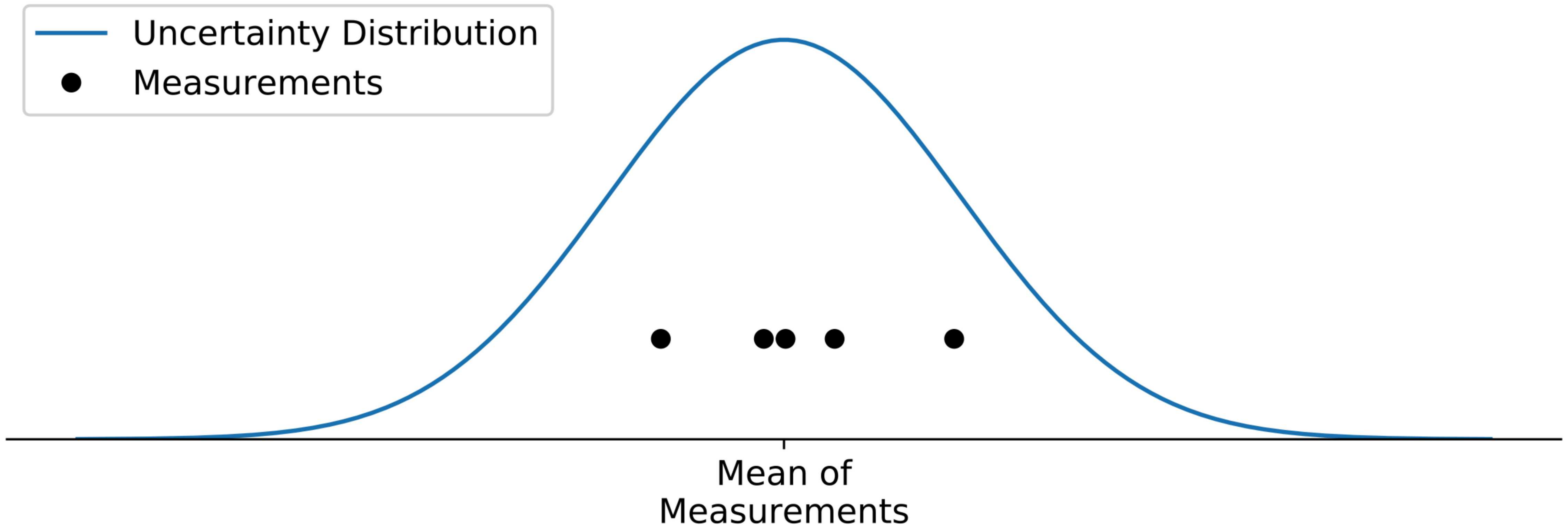
— S. J. Kline and F. A. McClintock, Describing Uncertainties in Single Sample Experiments

Repeated Measurements



- Repeated measurements fall within the bell curve

Repeated Measurements



- Repeatability of measurements fall within the bell curve
- Take N independent measurements under the same conditions to capture the repeatability uncertainty.

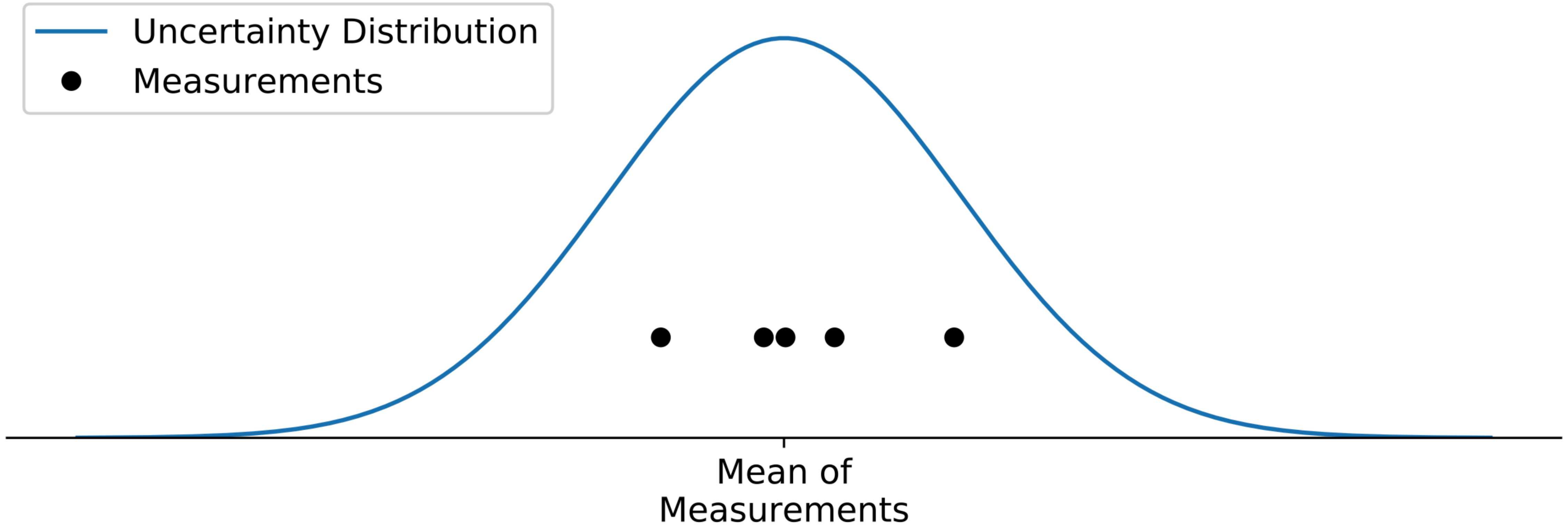
- This attempts to estimate the actual distribution with a **sample** of measurements.

Some Statistics Tools

- For multiple measurements $x_1,\ x_2, x_3, \ldots x_N$:
- Mean: $\bar{x} = \frac{1}{N}\sum x_i$
- Standard Deviation: $s(x) = \sqrt{\frac{1}{N-1}\sum(x_i - \bar{x})^2}$
  - Note: Use standard deviation of the **Sample** (STDEV.S in Excel)
- Variance: $s(x)^2$

Repeated Measurements

- How well do we know the MEAN value of this distribution based on these measurements?



- Standard Deviation **of the Mean**: $s(x)/\sqrt{N}$ (sometimes “standard error of the mean”)
- Type A uncertainty: $u_A(x) = s(x)/\sqrt{N}$
- Degrees of Freedom: N - 1

Example: Measurement of Lava Temperature

USGS/Public Domain

10 repeated measurements using thermocouple

Measurement	Temperature °C
1	1130.60
2	1210.64
3	1211.25
4	1047.04
5	1135.67
6	1189.19
7	1237.59
8	1164.33
9	1225.51
10	1165.79

Example: Measurement of Lava Temperature

- Mean: 1171.76 °C
- Standard Deviation: 57.01°C
- Standard Error of the Mean = 57.01 / √10 = 18.03 °C

- Type A Standard Uncertainty:
  - $u_A(T) = 18.03$ °C

10 repeated measurements using thermocouple

Measurement	Temperature °C
1	1130.60
2	1210.64
3	1211.25
4	1047.04
5	1135.67
6	1189.19
7	1237.59
8	1164.33
9	1225.51
10	1165.79

Multiday Repeated Measurements

M number of days

N repeated measurements

	Day 1	Day 2	Day 3	Day 4	Day 5
#1	0.96	.99	1.04	1.03	1.05
#2	1.03	0.96	0.96	0.98	0.99
#3	1.01	0.98	0.96	0.97	1.04
#4	0.98	1.01	0.99	0.99	1.01
#5	1.00	0.97	1.02	0.99	1.01
Mean	0.996	0.982	0.994	0.992	1.02
Variance	0.00073	0.00037	0.00128	0.00052	0.00060
Std. Dev	0.0270	0.0192	0.0358	0.0228	0.0245

1Ω Resistor calibrated 5 times on 5 different days

Credit: Josh Reynolds, PSL DC Lab

Repeatability and Reproducibility

Repeatability: Variation under the same operator, measuring system, and procedure over a **short period of time**.

- "within-day variation"

Reproducibility: Variation due to different operators, measuring systems, and/or procedures, often **over longer periods of time**.

- "between-day variation"

Repeatability – within day variation

Repeatability: Pooled standard deviation (average of the daily variances)

- s_i: Standard deviation of day i
- M: number of days
- $s_r = \sqrt{\frac{\sum s_i^2}{M}}$

Or if different number of measurements each day:

- n_i: Number of measurements on day i
- $s_r = \sqrt{\frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)}}$

Resistor Example: Repeatability

	Day 1	Day 2	Day 3	Day 4	Day 5
#1	0.96	.99	1.04	1.03	1.05
#2	1.03	0.96	0.96	0.98	0.99
#3	1.01	0.98	0.96	0.97	1.04
#4	0.98	1.01	0.99	0.99	1.01
#5	1.00	0.97	1.02	0.99	1.01
Mean	0.996	0.982	0.994	0.992	1.02
Variance	0.00073	0.00037	0.00128	0.00052	0.00060
Std. Dev	0.0270	0.0192	0.0358	0.0228	0.0245

1Ω Resistor calibrated 5 times on 5 different days

N = # repeats = 5
M = # days = 5

Repeatability:
Average the daily variances

$$s_r = \sqrt{\sum \frac{s_i^2}{M}} = \sqrt{\frac{0.00073+0.00037+0.00128+0.00052+0.00060}{5}}$$

$$= 0.026$$

Reproducibility – between day variation

Reproducibility: Standard deviation of the daily means

- μ_i: Mean of measurements made on day i
- $\bar{\mu}$: Grand Mean: Mean of all measurements
- M: Number of days
- $s_R = \sqrt{\frac{1}{M-1}\sum(\mu_i - \bar{\mu})^2}$

Resistor Example: Reproducibility

	Day 1	Day 2	Day 3	Day 4	Day 5
#1	0.96	.99	1.04	1.03	1.05
#2	1.03	0.96	0.96	0.98	0.99
#3	1.01	0.98	0.96	0.97	1.04
#4	0.98	1.01	0.99	0.99	1.01
#5	1.00	0.97	1.02	0.99	1.01
Mean	0.996	0.982	0.994	0.992	1.02
Variance	0.00073	0.00037	0.00128	0.00052	0.00060
Std. Dev	0.0270	0.0192	0.0358	0.0228	0.0245

1Ω Resistor calibrated 5 times on 5 different days

N = # repeats = 5
M = # days = 5

Reproducibility:
Standard Deviation of the daily means

$$\bar{\mu} = \frac{1}{M}\sum \mu_i = \frac{0.996 + 0.982 + 0.994 + 0.992 + 1.02}{5} = 0.997$$

$$s_R = \sqrt{\frac{1}{M-1}\sum (\mu_i - \bar{\mu})^2} = 0.014$$

Repeatability and Reproducibility in an Uncertainty Budget

Repeatability and Reproducibility as Type A uncertainties:

Repeatability: $u_{A_{repeat}} = \frac{s_r}{\sqrt{NM}}$

Reproducibility: $u_{A_{repr}} = \frac{s_R}{\sqrt{M}}$

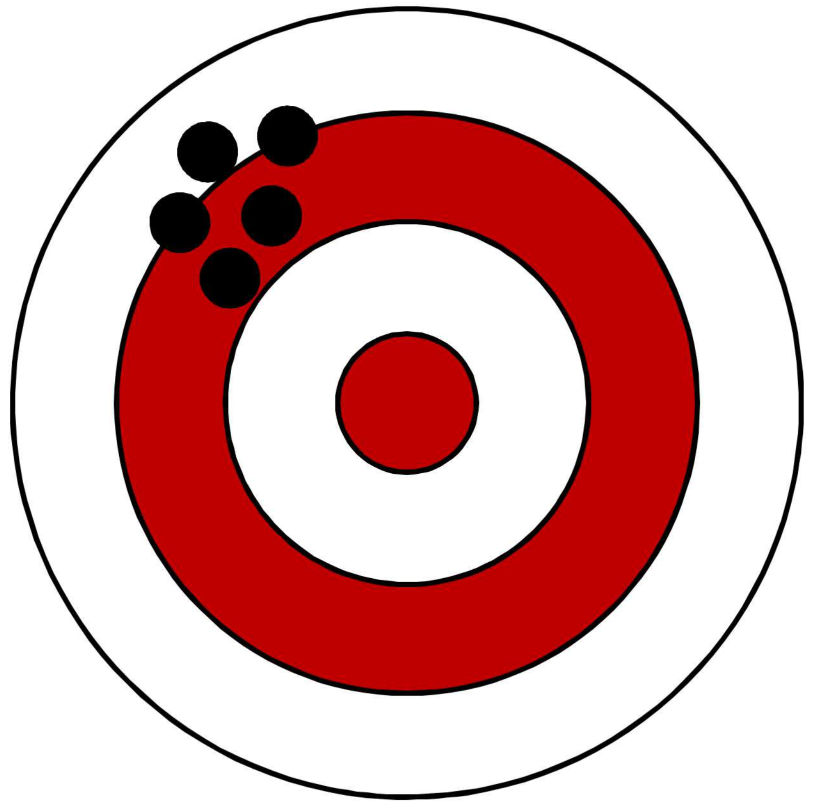
- N = Number of repeated measurements each day
- M = Number of Days

Infinite Measurements ≠ Zero Uncertainty

- $u_A(x) = \frac{s_r}{\sqrt{N}}$

- If I take a bazillion measurements, I'll have no uncertainty, right?

- **WRONG.** Type A uncertainty may approach zero, but there's always other Type B uncertainties!



Type B Uncertainty

Evaluated using other methods

"It is inevitable that the statements of reliability will be based in part on estimates. This must be true since by definition statistics cannot be applied to all of the errors."

— S. J. Kline and F. A. McClintock, Describing Uncertainties in Single Sample Experiments

Type B Uncertainty

- Uncertainty derived from other sources
  - Calibration certificates
  - Manufacturer’s specifications
  - Equipment resolution
  - Finite digital precision/rounding
  - Reference data from handbooks
  - Known environmental fluctuations
  - Other information

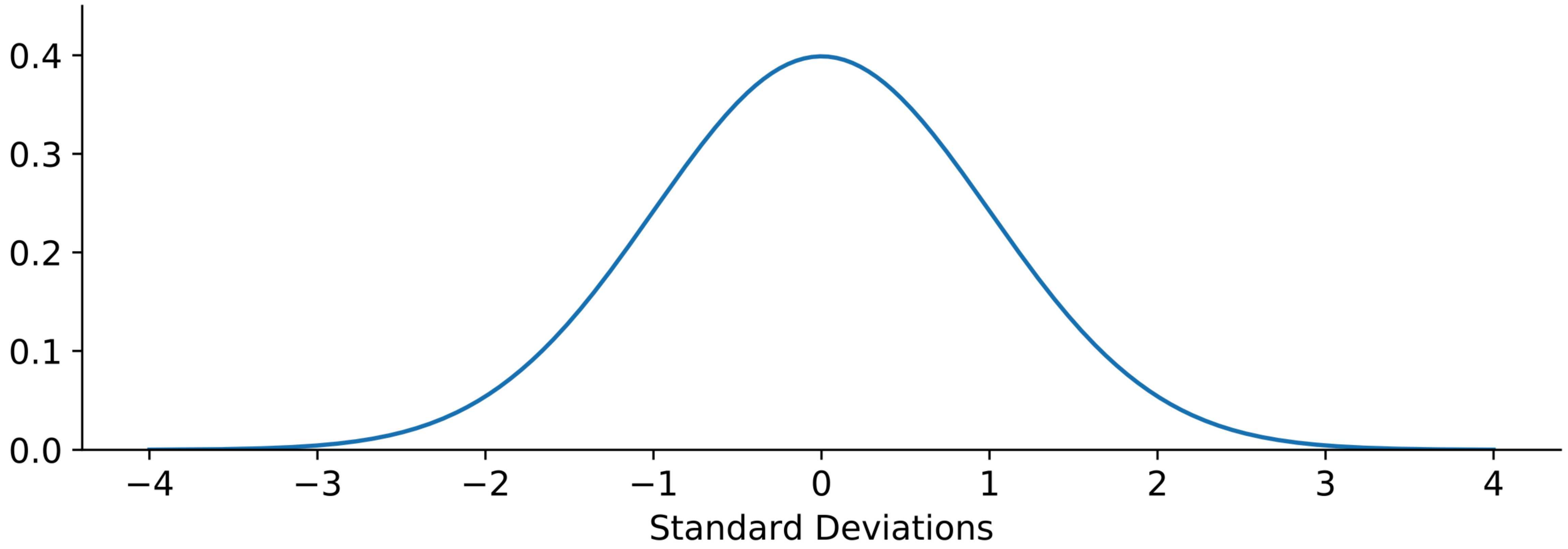
Type B Uncertainty: Probability Distributions

- To combine multiple uncertainty components, they all must be "standardized" – expressed in terms of standard deviations.

- Each uncertainty can be assigned a probability distribution depending on what is known. This distribution is used to standardize the uncertainty value.

- Choose a probability distribution based on what you know about the uncertainty.

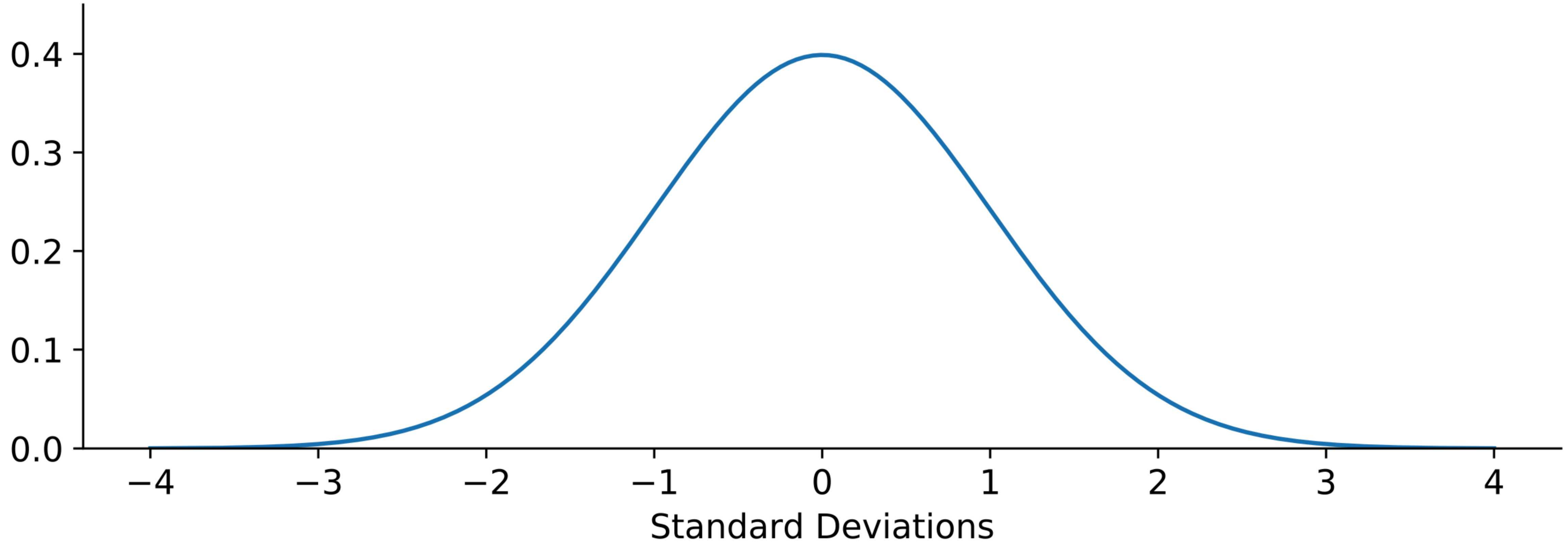
Normal Distribution



Normal Distribution

- Use a normal distribution if you are given:
  - Uncertainty in terms of standard deviations ("… given at the three standard deviation level")
  - Uncertainty with a k value ("… the uncertainty is ± 0.1Ω (k=2)")

- Standard Uncertainty $u = \frac{U}{k}$

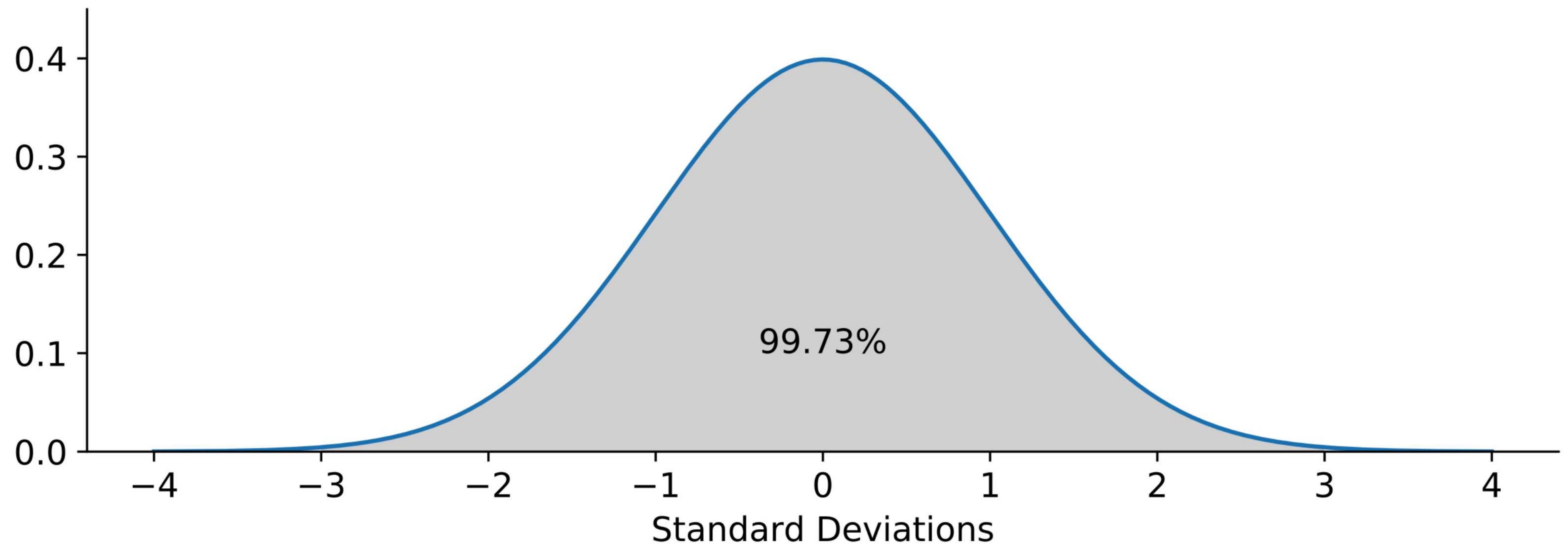


Normal Distribution

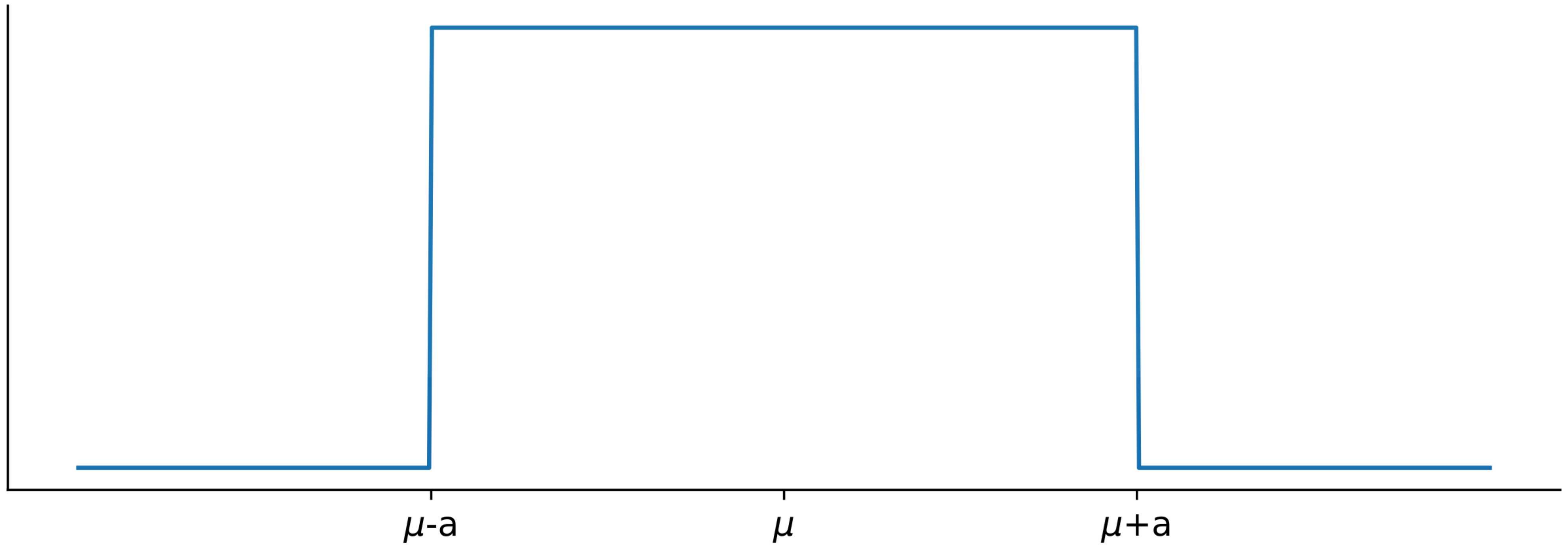
- Use a normal distribution if you are given:
  - Uncertainty with a confidence as a percent ("… the uncertainty is ± 0.1Ω to 95% confidence")

- Look up k value given the confidence.
- Standard Uncertainty $u = \frac{U}{k}$

Confidence	k
99.73%	3
99.0%	2.58
95.45%	2
95%	1.96
90%	1.64
68.27%	1

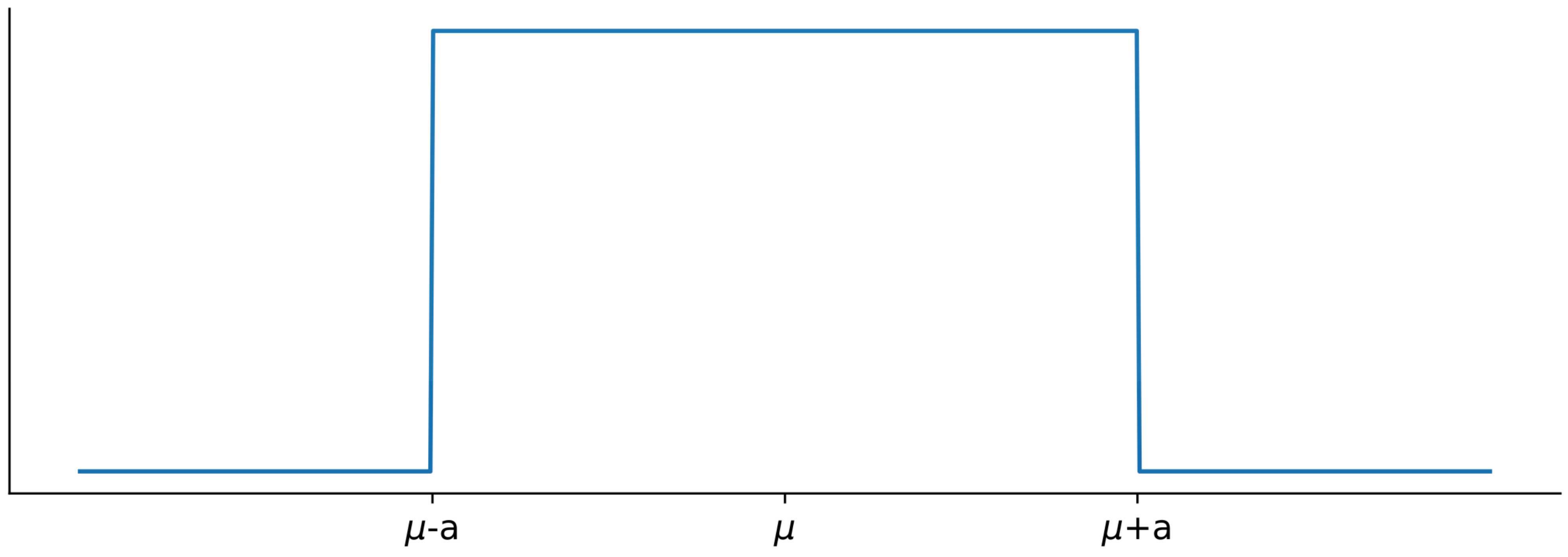


Uniform (Rectangular) Distribution



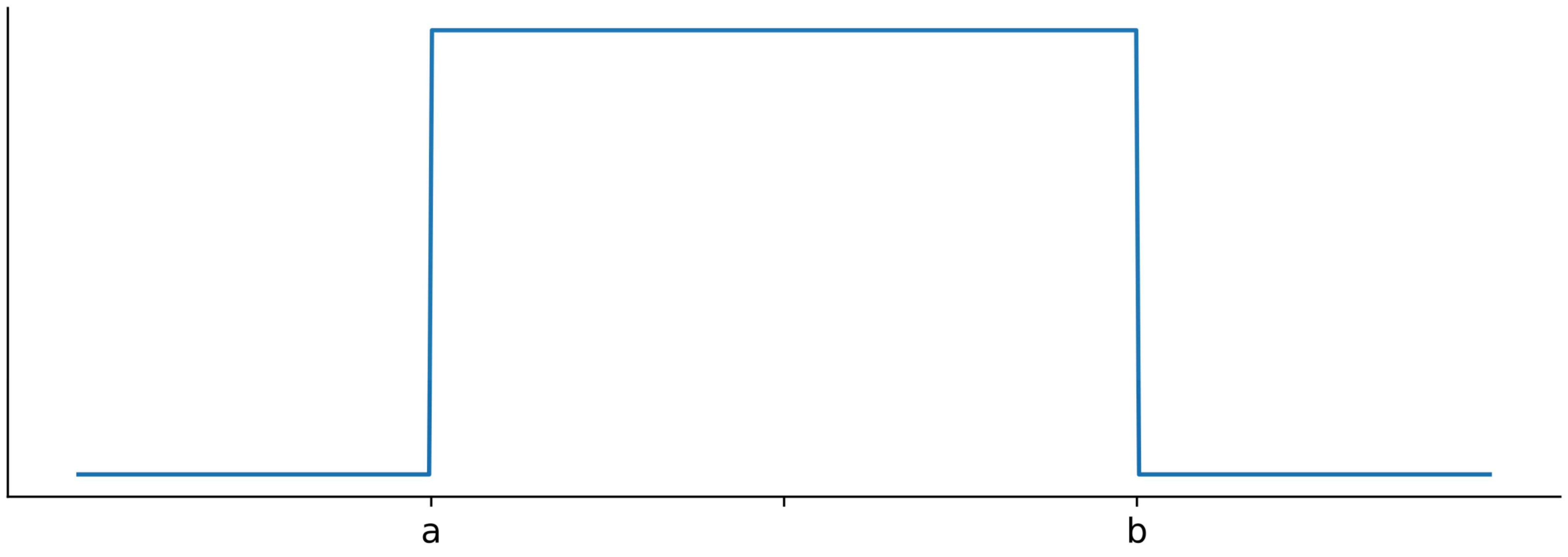
Uniform (Rectangular) Distribution

- Use a uniform distribution if you are given:
  - ± Tolerance (a) with no other information ("…the accuracy is ±1%")
- Standard uncertainty $u = \frac{a}{\sqrt{3}}$



Uniform (Rectangular) Distribution

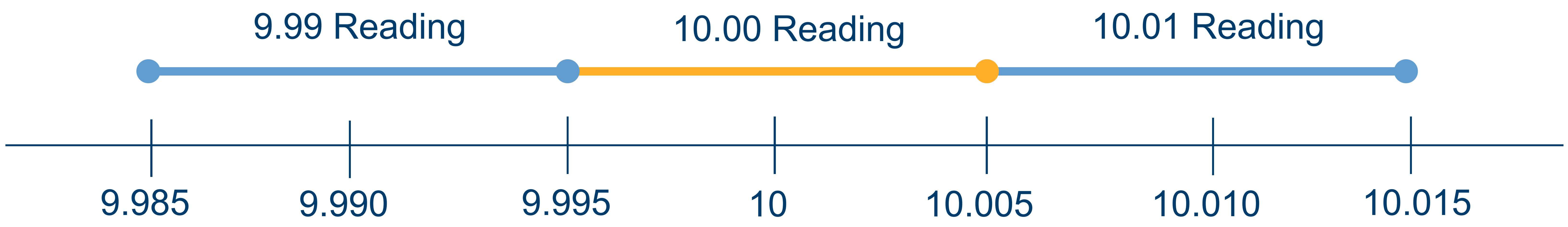
- Use a uniform distribution if:
  - value must be between a and b, with no other information
- Standard uncertainty $u = \frac{b-a}{\sqrt{12}}$



Equipment Resolution

- Use a uniform distribution if you are given equipment resolution, with least significant digit r
  - This is uniform distribution with half-width $a = r/2$

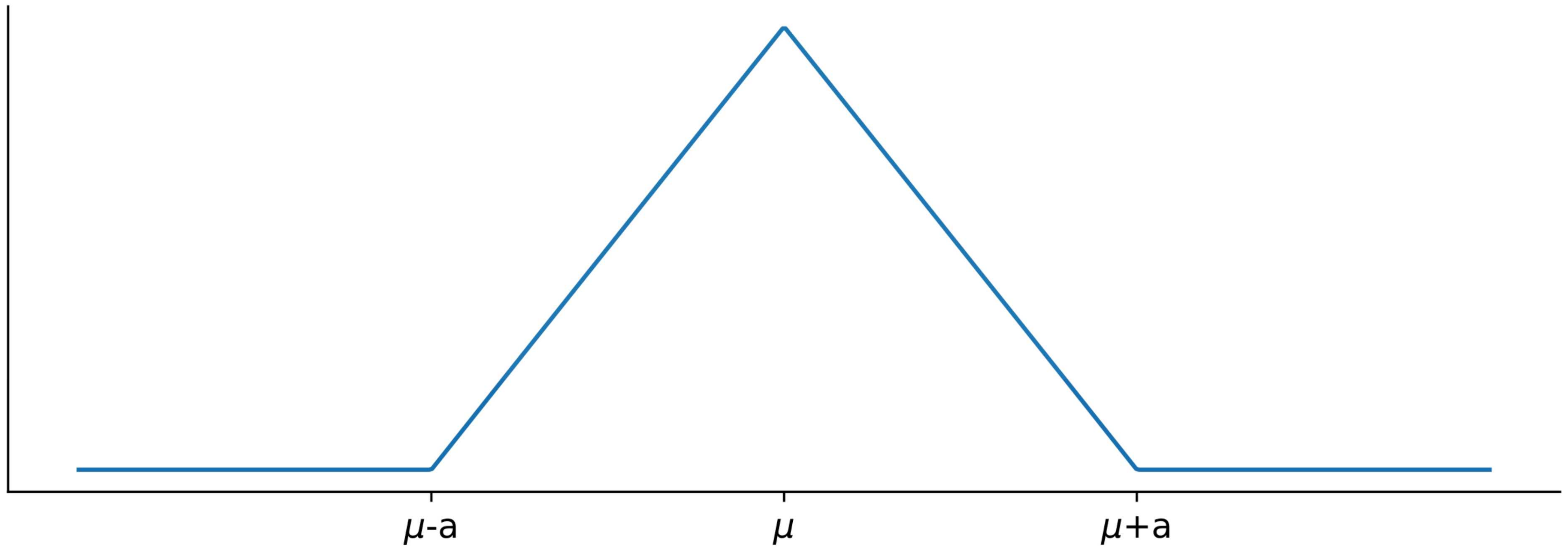
- Standard uncertainty $u = \frac{a}{\sqrt{3}} = \frac{r}{2\sqrt{3}}$



Triangular Distribution

- Use a triangular distribution if you are given:
  - ± Tolerance, but more likely to be near center
- Standard uncertainty $u = \frac{a}{\sqrt{6}}$

- Sometimes used for analog readouts (make sure procedure specifies how/whether to interpolate between markings)



A Few Other Distributions

- **Arcsine/U-Shaped:** Use when value fluctuates sinusoidally
- **Trapezoidal:** Combination of two uniforms; bounded but more likely in center
- **Exponential**: Value must be positive
- **Poisson**: Use when counting things
- **Student T**: Estimating the mean and standard deviation using N (small) samples

Example – Lava Temperature

The instrument / thermocouple combination was tested by comparison with a Standard Platinum Resistance Thermometer (SPRT). The (SPRT) is traceable to the National Institute of Standards and Technology (NIST). The type T temperature calibration was performed on the International Temperature Scale of 1990 (ITS-90), using the standard ITS-90 reference functions for thermocouples as defined in NIST Monograph 175. The instrument / thermocouple combination was found to be within the following tolerance:

TC TYPE	RANGE	Tolerance
N	1000°C to 1300 °C	± 75 °C

Limitation: This report is valid only for the instrument / thermocouple combination tested and over the range listed above.

Tolerance with no other information given. Assume uniform distribution.

$$\square_B(\square) = \frac{75}{\sqrt{3}} = 43.3\ °\mathrm{C}$$

Example – Resolution

- Thermocouple Reader

- Least significant digit $r = 0.01\ ℃$
- Uniform Distribution, $a = r/2 = 0.005℃$
- Type B Standard Uncertainty =

$$u_B = \frac{a}{\sqrt{3}} = 0.0029\ ℃$$

Combined Uncertainty

What to do with all these uncertainty components?

- Once all uncertainties are standardized using distributions in the previous section, they can be added in quadrature (RSS method)

$$u_c = \sqrt{\sum_{i=1}^{N_A} u_{A_i}^2 + \sum_{i=1}^{N_B} u_{B_i}^2}$$

- Take the Square **R**oot of the **S**um of the **S**quares of the uncertainty components!

- u_c is the combined standard uncertainty (“c” for combined, lowercase “u” for standard).

Lava Temperature Example

$$u_A(T) = 18.03\ \text{°C}$$

$$u_B(T) = 43.3\ \text{°C}$$

$$u_{B_{res}}(T) = 0.0029\ \text{°C}$$

$$u_c(T) = \sqrt{u_A^2(T) + u_B^2(T) + {u_B}_{res}^2(T)}$$

$$= \sqrt{18.03^2 + 43.3^2 + 0.0029^2} = 46.90\ \text{°C}$$

Expanded Uncertainty

- A confidence level of 68% isn't very reassuring. A higher **level of confidence** is desired.

- Multiply u_c by a **coverage factor** k to get **expanded uncertainty** U.

- k depends on **degrees of freedom** – the number of independent pieces of information used to determine the standard deviation.

Degrees of Freedom

Degrees of freedom – the number of independent pieces of information used to determine the standard deviation

- Type A degrees of freedom: Typically $N - 1$
- Degrees of freedom for Type B can be harder to determine, depending on the source.
- Most often, assign $\nu = \infty$ to Type B uncertainties if there is no other knowledge.
  - But be careful if $u_B \gg u_A$. Refer to GUM for guidance.

Effective Degrees of Freedom

- Each uncertainty component has its own degrees of freedom. Combine using **Welch-Satterthwaite** formula:

$$\nu_{eff} = \frac{u_c^4(y)}{\sum_{i=1}^{N_A} \frac{u_{A_i}^4(y)}{\nu_{A_i}} + \sum_{i=1}^{N_B} \frac{u_{B_i}^4(y)}{\nu_{B_i}}}$$

Lava Temperature Example

$$\nu_{eff} = \frac{u_c^4(y)}{\sum_{i=1}^{N_A} \frac{u_{A_i}^4(y)}{\nu_{A_i}} + \sum_{i=1}^{N_B} \frac{u_{B_i}^4(y)}{\nu_{B_i}}}$$

$$= \frac{46.90^4}{\frac{18.03^4}{9} + \frac{43.3^4}{\infty}} = 412$$

Student T Table

- Look up *k* from this table using the column for the desired level of confidence.
- It gives a slightly higher *k* to account for the limited information.

- Or use Excel:
  - =T.INV.2T(1-p, v)

Degrees of Freedom *v*	Fraction p in Percent					
	68.27	90	95	95.45	99	99.73
1	1.840	6.310	12.710	13.970	63.660	235.8
2	1.320	2.920	4.300	4.530	9.920	19.21
3	1.200	2.350	3.180	3.310	5.840	9.220
4	1.140	2.130	2.780	2.870	4.600	6.620
5	1.110	2.020	2.570	2.650	4.030	5.510
6	1.090	1.940	2.450	2.520	3.710	4.900
7	1.080	1.890	2.360	2.430	3.500	4.530
8	1.070	1.860	2.310	2.370	3.360	4.280
9	1.060	1.830	2.260	2.320	3.250	4.090
10	1.050	1.810	2.230	2.280	3.170	3.960
11	1.050	1.800	2.200	2.250	3.110	3.850
12	1.040	1.780	2.180	2.230	3.050	3.760
13	1.040	1.770	2.160	2.210	3.010	3.690
14	1.040	1.760	2.140	2.200	2.980	3.640
15	1.030	1.750	2.130	2.180	2.950	3.590
16	1.030	1.750	2.120	2.170	2.920	3.540
17	1.030	1.740	2.110	2.160	2.900	3.510
18	1.030	1.730	2.100	2.150	2.880	3.480
19	1.030	1.730	2.090	2.140	2.860	3.450
20	1.030	1.720	2.090	2.130	2.850	3.420
25	1.020	1.710	2.060	2.110	2.790	3.330
30	1.010	1.700	2.040	2.090	2.750	3.270
35	1.010	1.700	2.030	2.070	2.720	3.230
40	1.010	1.680	2.020	2.060	2.700	3.200
45	1.010	1.680	2.010	2.060	2.690	3.180
50	1.010	1.680	2.010	2.050	2.680	3.160
100	1.005	1.660	1.984	2.025	2.626	3.077
∞	1.000	1.645	1.960	2.000	2.576	3.000

Student T

William Sealy Gosset

Annals of Eugenics/Public Domain

Head Brewer, Guinness

Sami Keinänen/CC-BY-SA 2.0

Lava Temperature Example

- $\nu_{eff} = \dfrac{u_c^4(y)}{\dfrac{u_A^4(y)}{\nu_A} + \dfrac{u_B^4(y)}{\nu_B}} = \dfrac{46.90^4}{\dfrac{18.03^4}{9} + \dfrac{43.3^4}{\infty}} = 412$

- k = 2.01 (95.45% with $\nu_{eff} = 412$)

Final Expanded Uncertainty

- $U = k \cdot u_c$
- $U = 2.01 \cdot 46.9\ ℃ = 93.8\ ℃$

- Final answer should include average measured value, expanded uncertainty, **and** level of confidence.
- Typically round uncertainty to 2 significant figures.

$$\boldsymbol{T = (1172 \pm 94)\ ℃, \qquad (95.45\%\ \text{confidence})}$$

Summary

- Type A: standard deviation of the mean of N measurements
  - $u_A = s(x)/\sqrt{N}$
- Type B: choose appropriate distribution and standardize
- RSS to combine all uncertainty components
  - $u_c = \sqrt{u_A^2 + u_B^2}$
- Calculate Degrees of Freedom using W-S formula
- Determine coverage factor using Student-T Table
- Expand uncertainty to desired level of confidence
  - $U = k \cdot u_c$

Review of Key Terms

- **Type A Evaluation of Uncertainty**: Derived from statistical analysis of current test data
- **Repeatability**: Measurement variation under the same conditions in a short time period
- **Reproducibility**: Measurement variation under differing conditions or long time periods
- **Type B Evaluation of Uncertainty**: Derived from other sources
- **Normal distribution:** Use when uncertainty is given with a confidence or *k* value
- **Uniform (or rectangular) distribution**: Use when uncertainty is bounded with no other information
- **Root-Sum-Square (RSS):** Use to combine multiple uncertainty components
- **Degrees of Freedom**: Number of independent pieces of information in determining a value
- **Coverage Factor**: Multiplier to expand uncertainty to desired level of confidence

Next Time…

- How do you know this measurement is good enough?

METROLOGY *in* MOTION

A WEBINAR SERIES

by NCSL INTERNATIONAL

Uncertainty for Technicians

Part 3: I have my uncertainty. Now what?

Collin J. Delker

Primary Standards Lab,
Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

METROLOGY in MOTION
A WEBINAR SERIES by NCSL INTERNATIONAL

Poll

- Do you typically perform:
    - Calibrations on measurement equipment
    - Tolerance tests on measurement equipment
    - Measurements on components to verify they meet specifications
    - Measurements in an R&D setting
    - Other

Part 3 Outline

- Review of uncertainty evaluation
- Uncertainty in indirect measurements
    - GUM equation
    - Monte Carlo approach
- Measurement Decision Risk
    - Specific Risk and Global Risk
- Test Uncertainty Ratio and the 4:1 “rule”
- Guardbanding
- Decision Rules

Review of Uncertainty Evaluation: Deflategate

AFC Championship January 2015

- New England Patriots accused of deflating footballs below the regulation pressure.

- NFL Regulation: Football must be inflated to 12.5 to 13.5 “pounds”
    - Assuming they mean psig, this is equivalent to 13 psig ± 0.5 psig

Exponent Engineering: The Effect of Various Environmental and Physical Factors on the Measured Internal Pressure of NFL Footballs, 2015.

Deflategate: Type A Uncertainty

- At halftime, each football was measured with two gauges. Results from one ball:
  - Gauge 1: 11.50 psig
  - Gauge 2: 11.80 psig

- N = 2 measurements
- Average: 11.65 psig
- Standard Deviation: 0.212 psig
- Standard Uncertainty: 0.212 / √2 = 0.15 psig
- Degrees of Freedom: 1

Exponent Engineering, 2015

Deflategate: Type B Uncertainty

- Resolution: $r = 0.05$ psig
  - Uniform distribution, $a = \frac{r}{2} = 0.025$
  - $u_{Bres} = \frac{0.025}{\sqrt{3}} = 0.014$
- Equipment Spec: ± 1% full scale = ± 0.20 psig
  - Uniform distribution, a = 0.2
  - $u_{Bspec} = \frac{0.2}{\sqrt{3}} = 0.12$
  - Degrees of freedom: 2 (estimated using GUM G.4.2)

Exponent Engineering, 2015

Deflategate Uncertainty

- Standard Uncertainty:
  - $u_c = \sqrt{u_A^2 + u_{Bspec}^2 + u_{Bres}^2} = \sqrt{0.15^2 + 0.12^2 + 0.014^2} = 0.19 \text{ psig}$
- Degrees of Freedom:
  - $\nu_{eff} = \dfrac{u_c^4}{\frac{u_A^4}{\nu_A} + \frac{u_{Bspec}^4}{\nu_{Bspec}} + \frac{u_{Bres}^4}{\nu_{Bres}}} = \dfrac{0.19^4}{\frac{0.15^4}{2} + \frac{0.12^4}{2} + \frac{0.014^4}{\infty}} = 2.2 \rightarrow 2$
  - $k = 4.3$ for 95% confidence from t-table lookup
- Expanded Uncertainty:
  - $U = k \cdot u_c = 0.82 \text{ psig}$

- Final Value: 11.65 psig ± 0.82 psig (95% confidence)
- NFL Requirement: 13.0 psig ± 0.50 psig

Indirect Measurements

"It is necessary to compute the uncertainty in the results from the estimates of uncertainty in the measurands… Propagation of uncertainty is not a trivial exercise…"

— S. J. Kline, The Purposes of Uncertainty Analysis

Direct vs. Indirect Measurement

- **Direct Measurement**: The measuring instrument measures the quantity directly.
- **Indirect Measurement**: Several measurements are made and combined through some mathematical model.

- **Measurement Model**: Equation used to calculate what you want to know.
  - Area = Length × Width
  - Resistance = Voltage/Current
  - Density = Mass/Volume = Mass/(Length × Width × Height)
  - Etc.

- Can’t just RSS the individual measurement uncertainties; they’re not all equally weighted.

Combining Uncertainty in Indirect Measurements

- First, find total uncertainty in each individual measurement $u(x_i)$
- Use methods from previous session to combine all known Type A and Type B uncertainties

Example:

- $R = \frac{V}{I}$
- Nominal value of voltage source: V
- Average of repeated current measurements: I
- Total uncertainty in voltage source: $u(V)$
- Total uncertainty in current measurement: $u(I)$

Combining Uncertainty in Indirect Measurements

- First, find total uncertainty in each individual measurement $u(x_i)$
- Use methods from previous session to combine all known Type A and Type B uncertainties

Example:

- $R = \frac{V}{I}$
- Nominal value of voltage source: $V = 3.30\ \mathrm{V}$
- Average of repeated current measurements: $I = 10.0\ \mathrm{mA}$
- Total uncertainty in voltage source: $u(V) = 0.0016\ \mathrm{V}$
- Total uncertainty in current measurement: $u(I) = 0.005\ \mathrm{mA}$

Option 1: GUM Equation

- Calculate the combined uncertainty using the GUM equation

$$u_c(y) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)}$$

- Looks intimidating, but it's really just RSS with extra weighting factors (called the sensitivity coefficients) for each term.

Option 1: GUM Equation

$$u_c(y) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)}$$

- Resistance: $R = V / I$

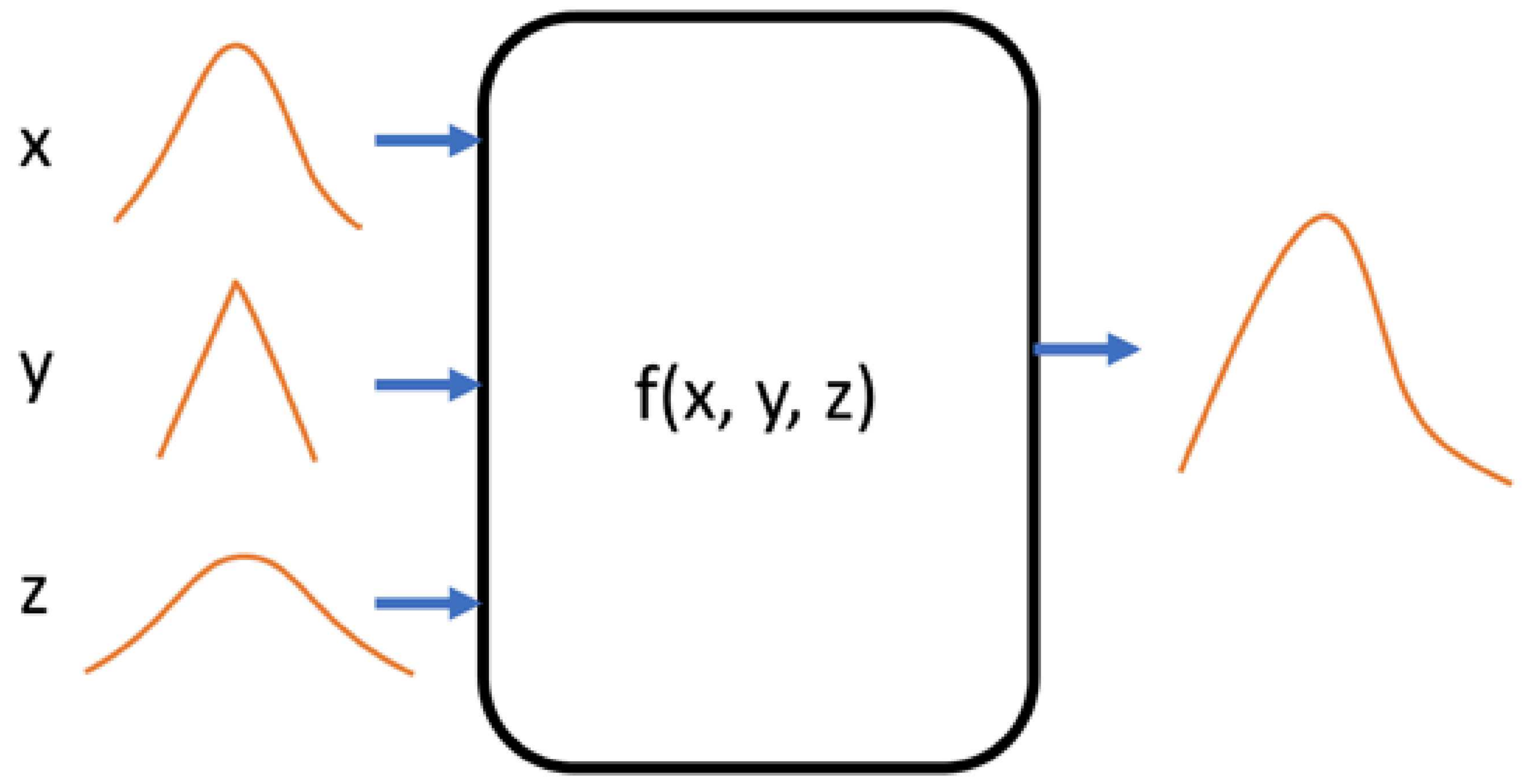
- Sensitivity coefficients (note conversion to base units):
  - $\frac{\partial R}{\partial I} = \frac{-V}{I^2} = -\frac{3.30}{0.01^2}, = -33000$
  - $\frac{\partial R}{\partial V} = \frac{1}{I} = \frac{1}{.01} = 100$

$V = 3.30 \text{ V}$
$I = 10.0 \text{ mA}$
$u(V) = 0.0016 \text{ V}$
$u(I) = 0.005 \text{ mA}$

- Combined standard uncertainty:
  - $u_c(R) = \sqrt{\left(\frac{-V}{I^2}\right)^2 u^2(I) + \left(\frac{1}{I}\right)^2 u^2(V)} = \sqrt{(33000^2)(0.000005^2) + (100^2)(0.0016^2)} = 0.23\ \Omega$

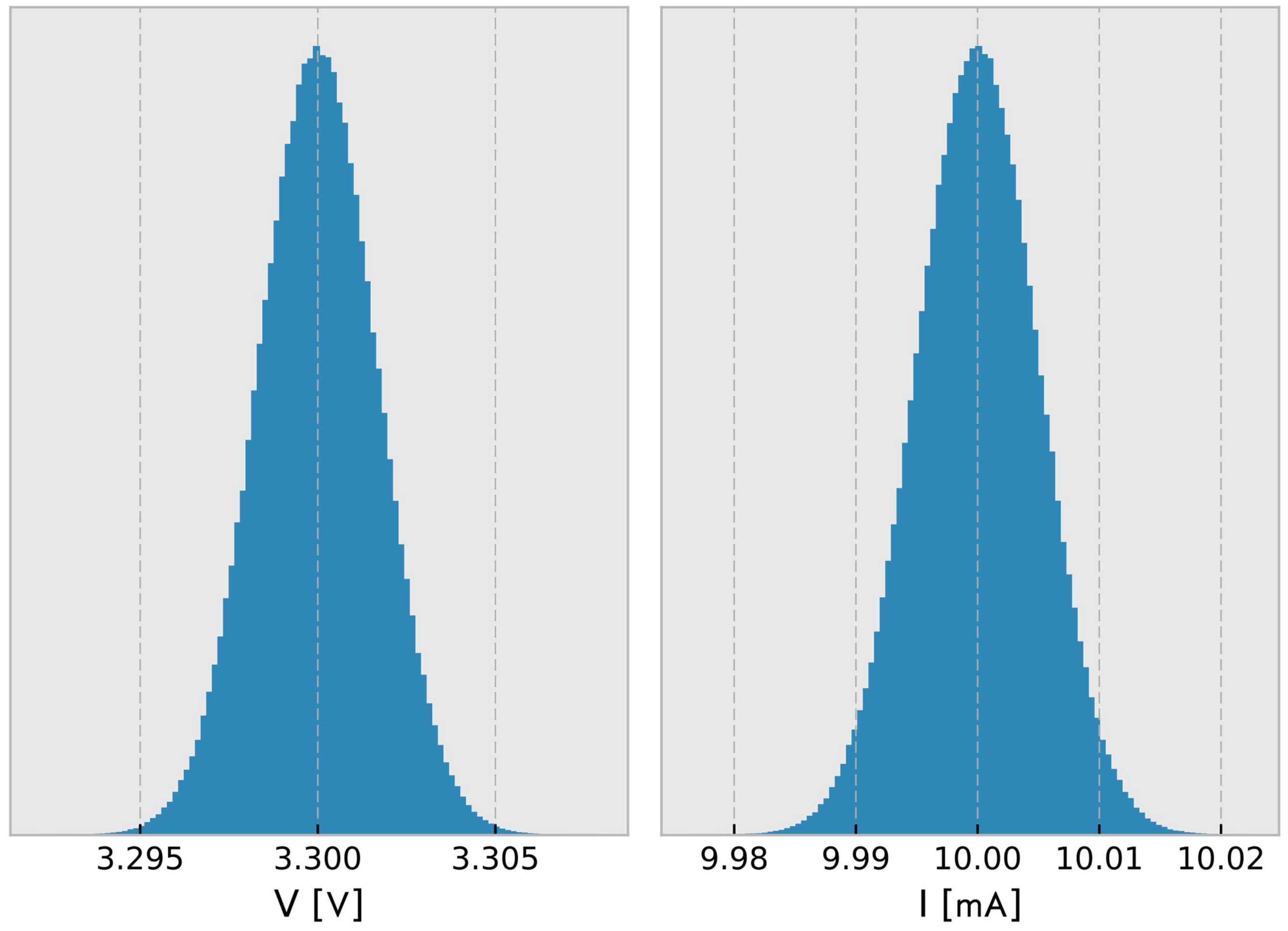
Option 2: Monte Carlo Approach

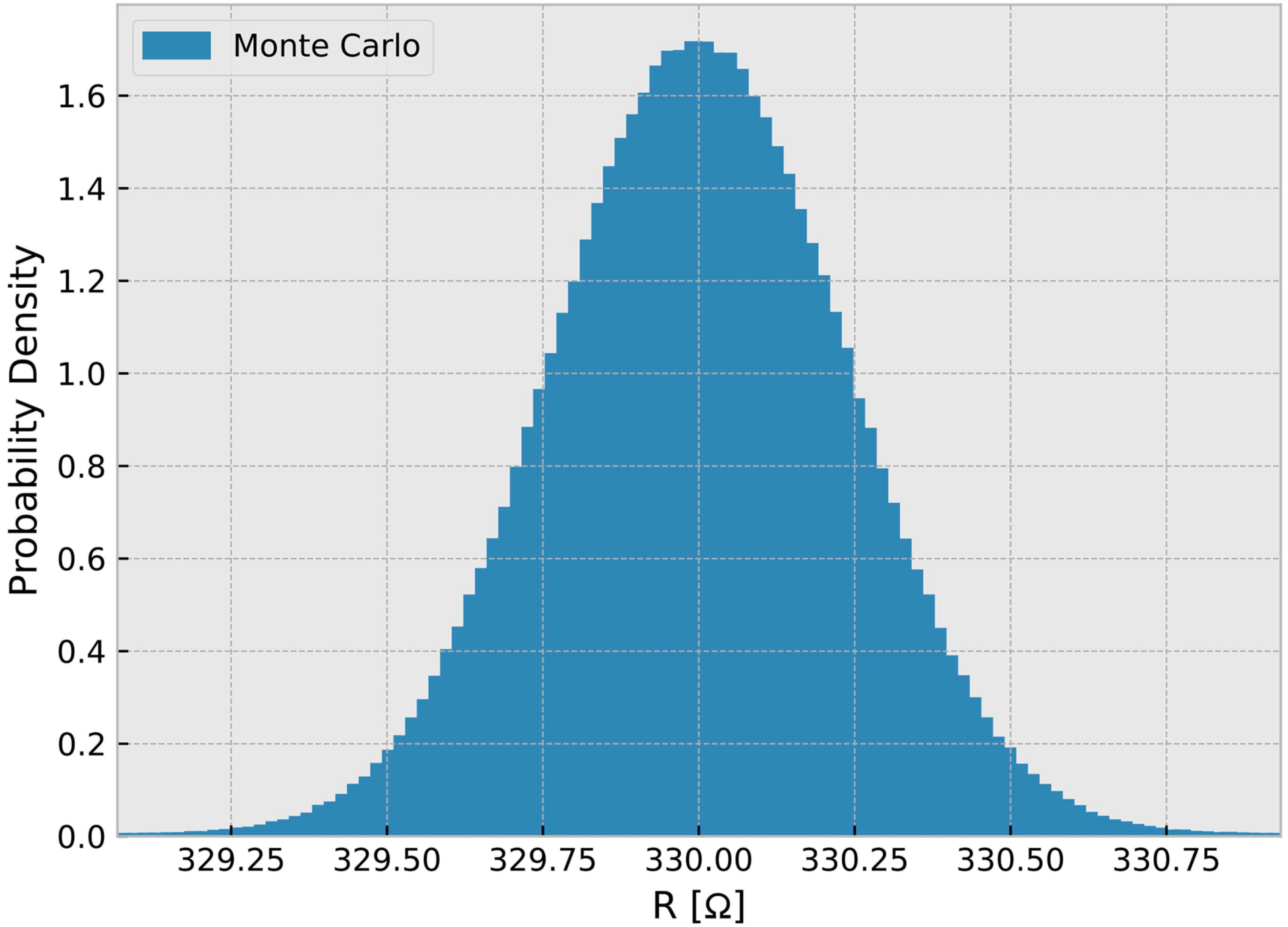
- Generate random samples of all inputs and combine using measurement model.
- Resulting distribution describes uncertainty in output.



Option 2: Monte Carlo Approach

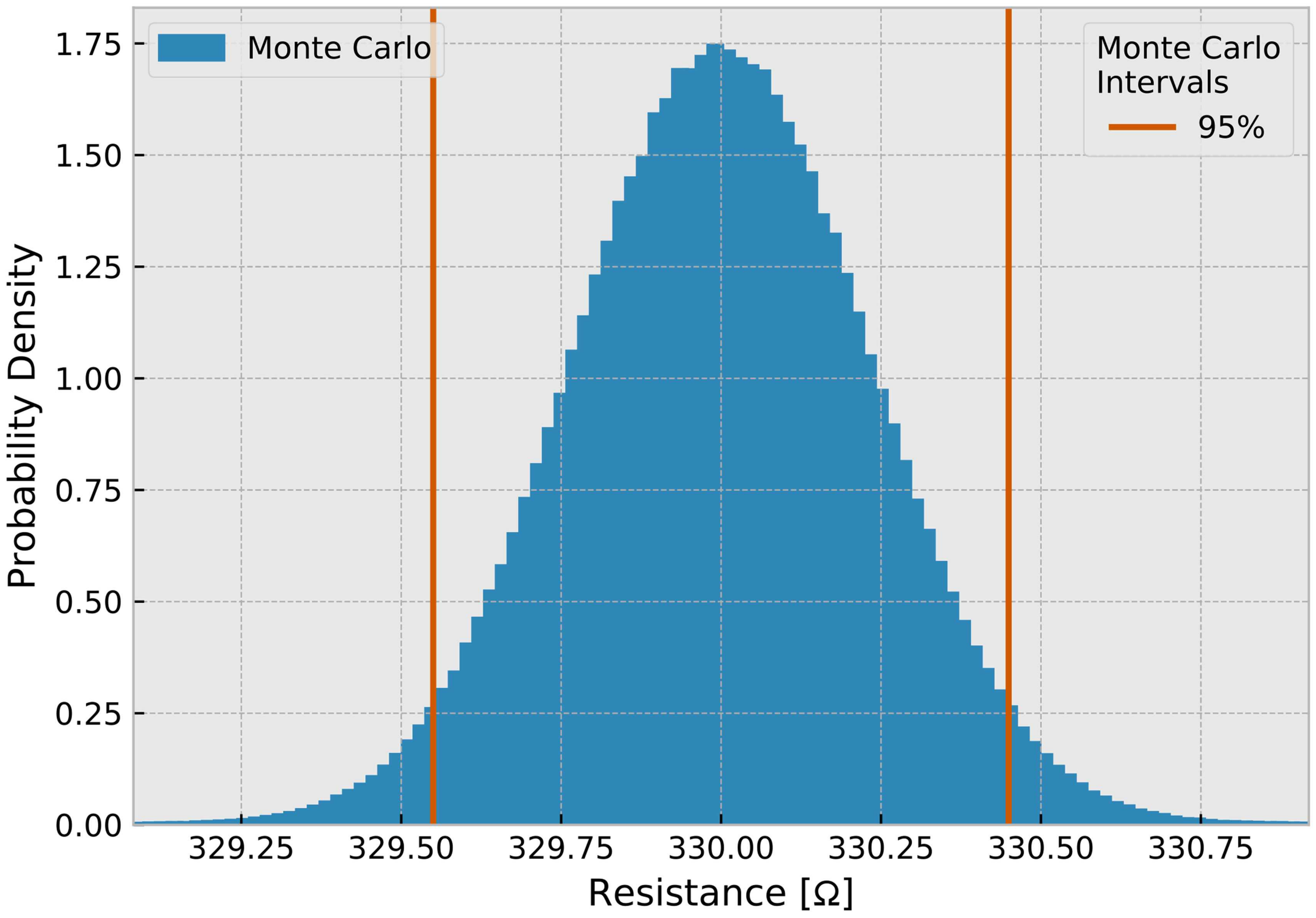
- Generate random samples of all inputs and combine using measurement model.
- Resulting distribution describes uncertainty in output.





Option 2: Monte Carlo Approach

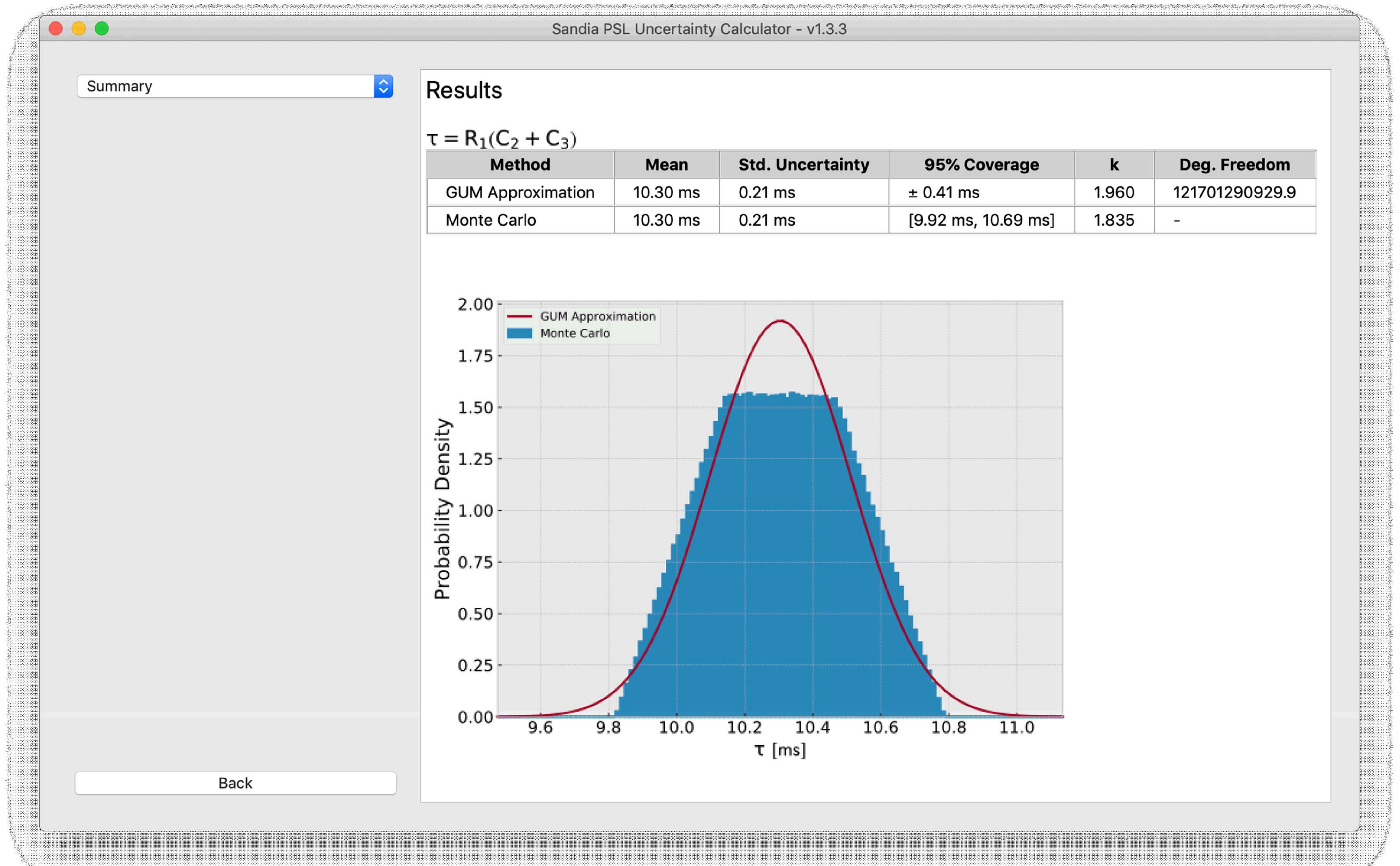
- Use percentiles to find 95% coverage range



Indirect Measurements in Practice

- Nobody wants to do calculus, and Excel is terrible at Monte Carlo (without macros, at least).
- Typically, other software is used, such as:
  - Sandia Uncertainty Calculator (**https://sandiapsl.github.io**)
  - NIST Uncertainty Machine (**https://uncertainty.nist.gov**)

Covered in depth by workshops at NCSLI Tech Exchange, tutorial sessions at Workshop & Symposium



Sandia Uncertainty Calculator

Measurement Decision Risk

"The more critical the decision, the more critical the data. The more critical the data, the more critical the measurement."

— NASA Publication 1342, 1994

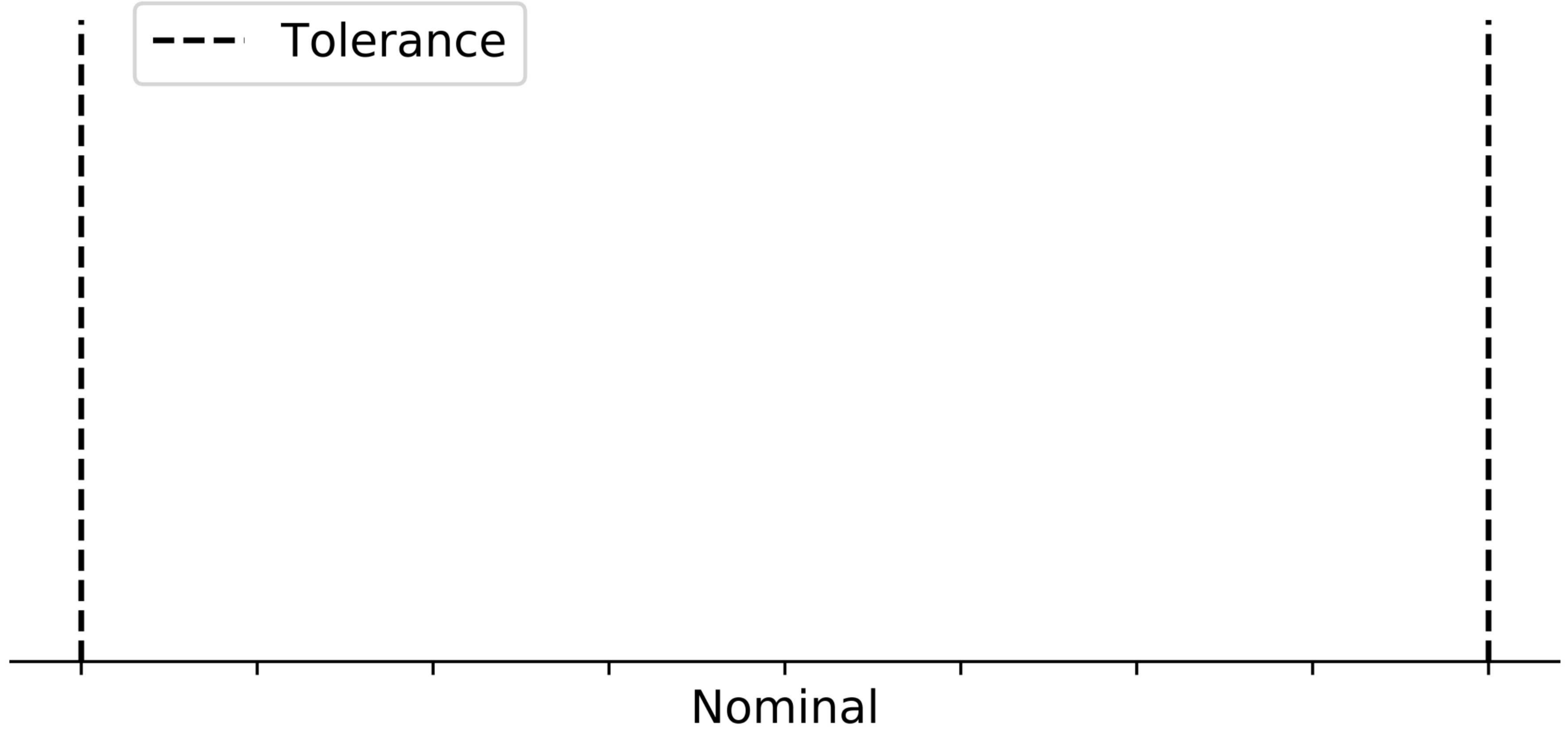
Types of Measurement Decisions

- Can we use this football in the Super Bowl?
- Should the aircraft take off based on the indicated air speed?
- Does the widget we manufactured meet specifications?
- Should we allocate more funding to this R&D project based on initial results?
- Is the power supply operating in tolerance?

- Measurement is used to verify that a **specification** or **requirement** is met.

Measurement Risk

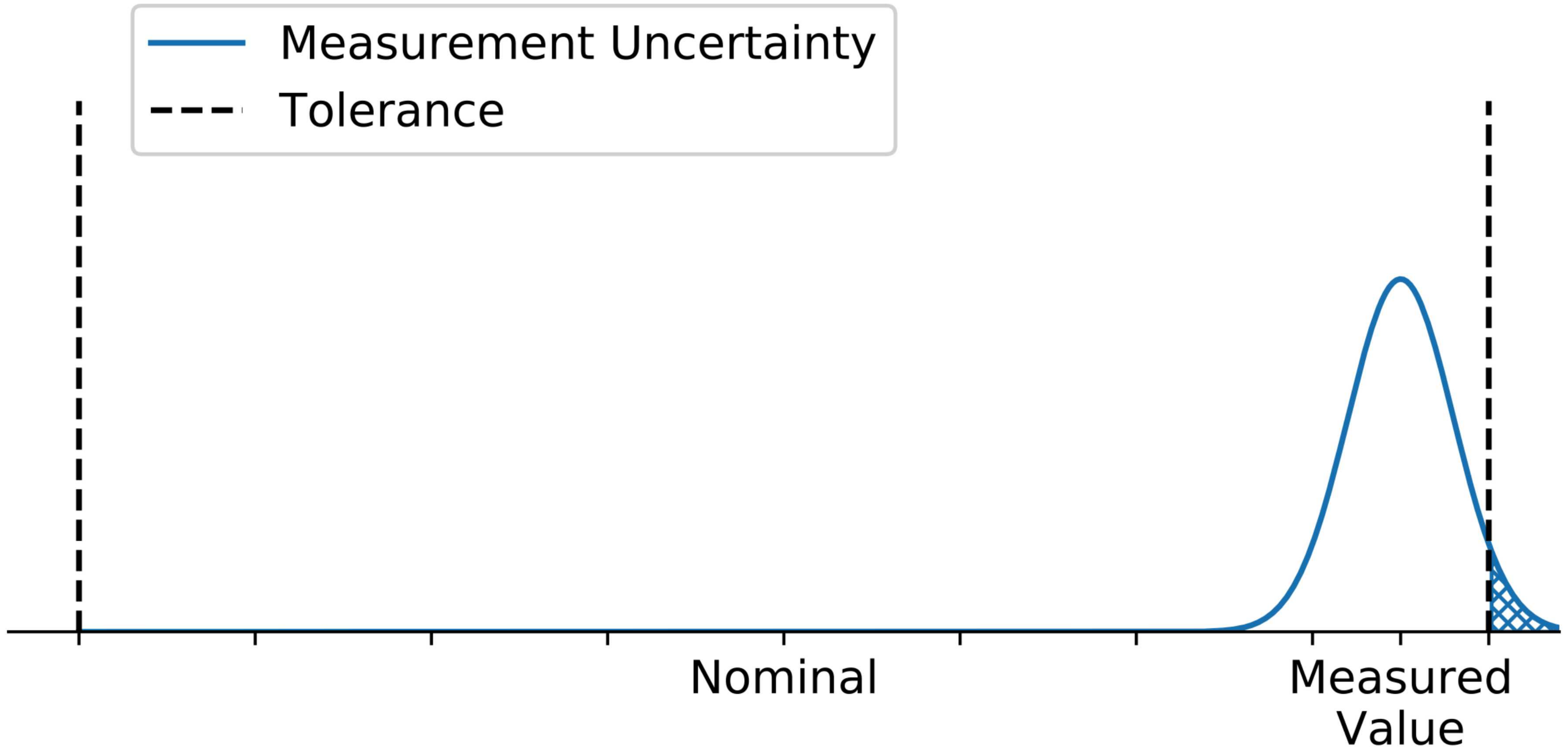
A specification must be between the dashed lines:



The tolerance can be for a measurement instrument undergoing calibration *or* engineering specifications on manufactured product

Measurement Risk

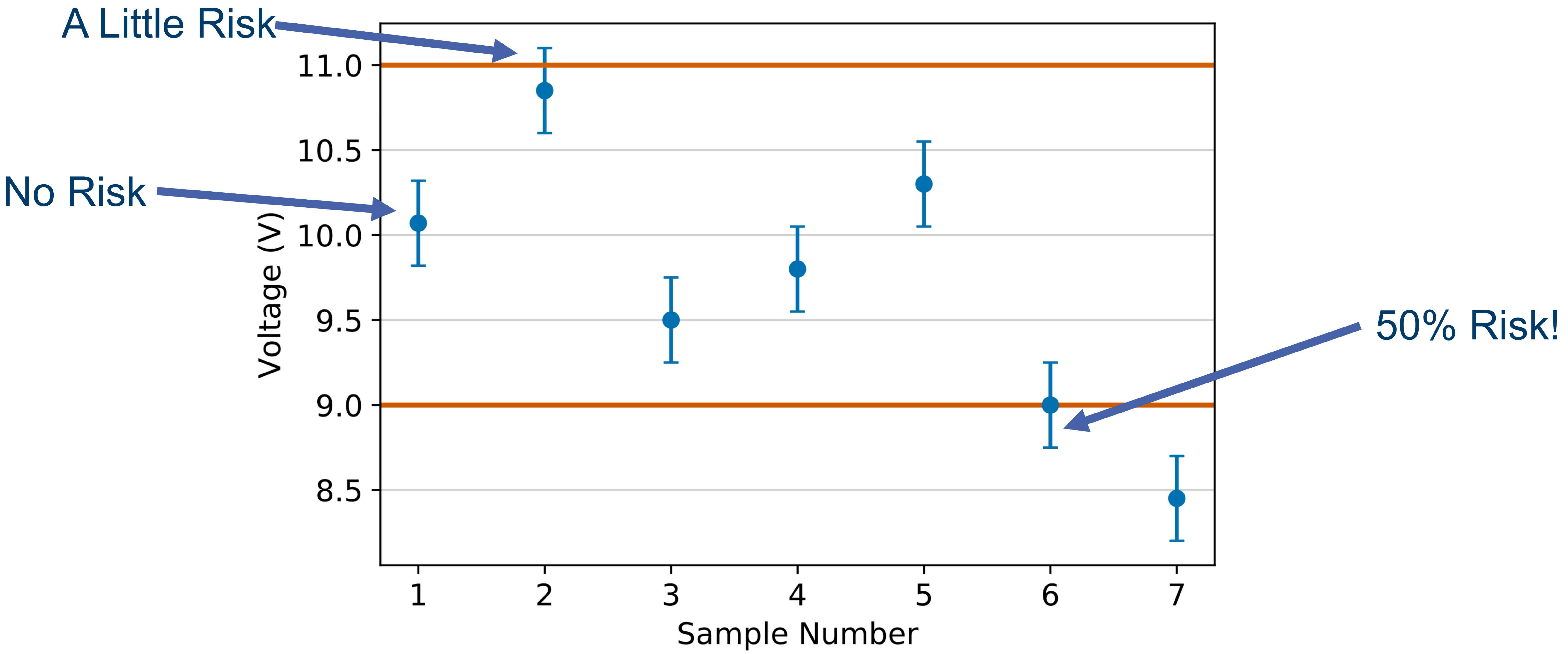
A measurement is made to check the requirement. What are the odds that the pass/fail decision is incorrect?



Remember, the true value lies somewhere within the normal distribution. The cross-hatched area represents the probability that the true value is actually out of spec, even though the measured value is in spec (False Accept).

Specific Risk

Specific risk: Probability that a decision based on a **specific** measurement is incorrect.

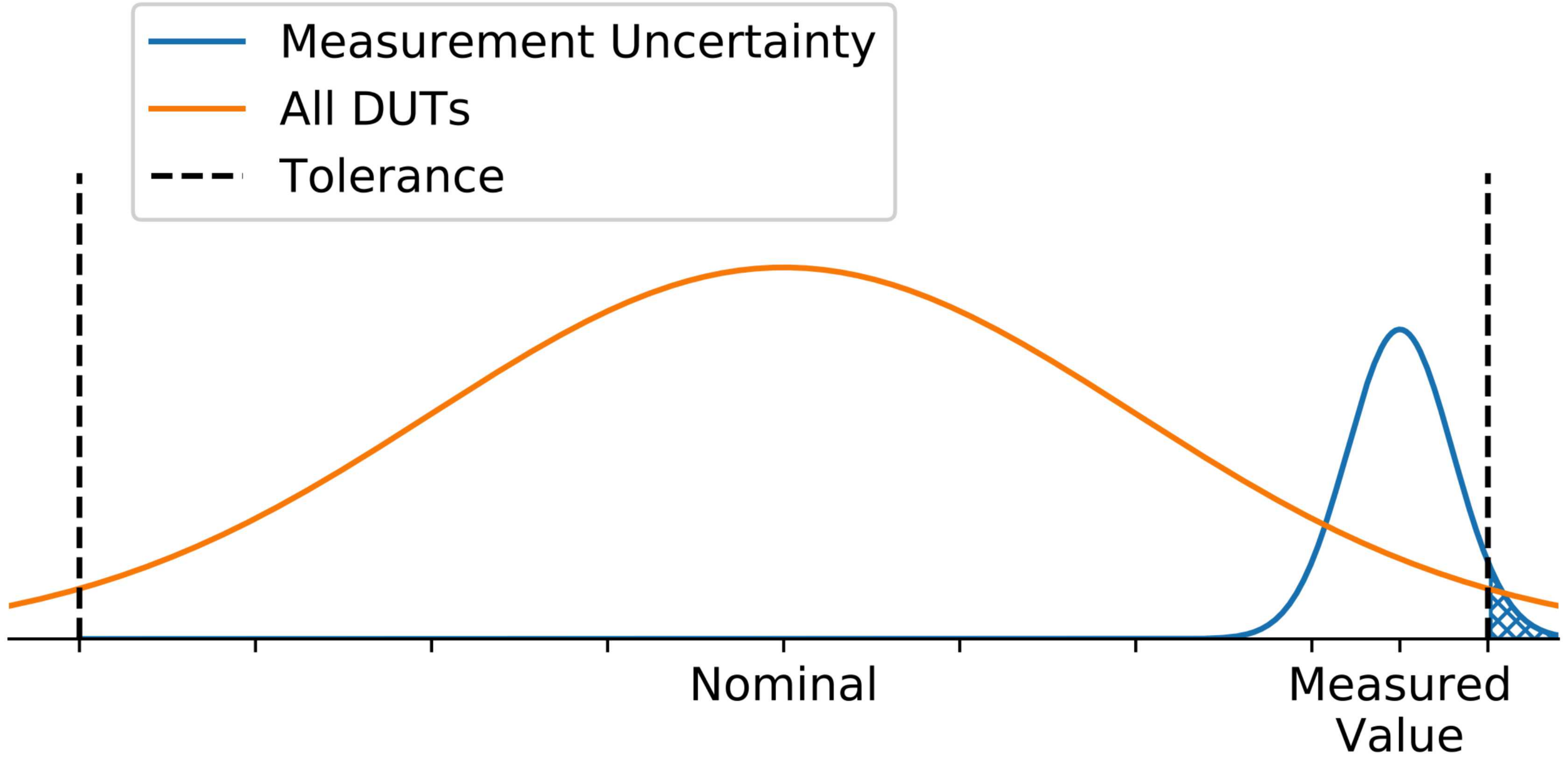


Also called *Conditional Risk*

Global Risk

Global Risk: Probability of making an incorrect decision on *any* part in the population. Also called *Consumer's Risk* or *Unconditional Probability of False Accept.*

- Combines knowledge of measurement uncertainty with knowledge/historical data of all possible parts.



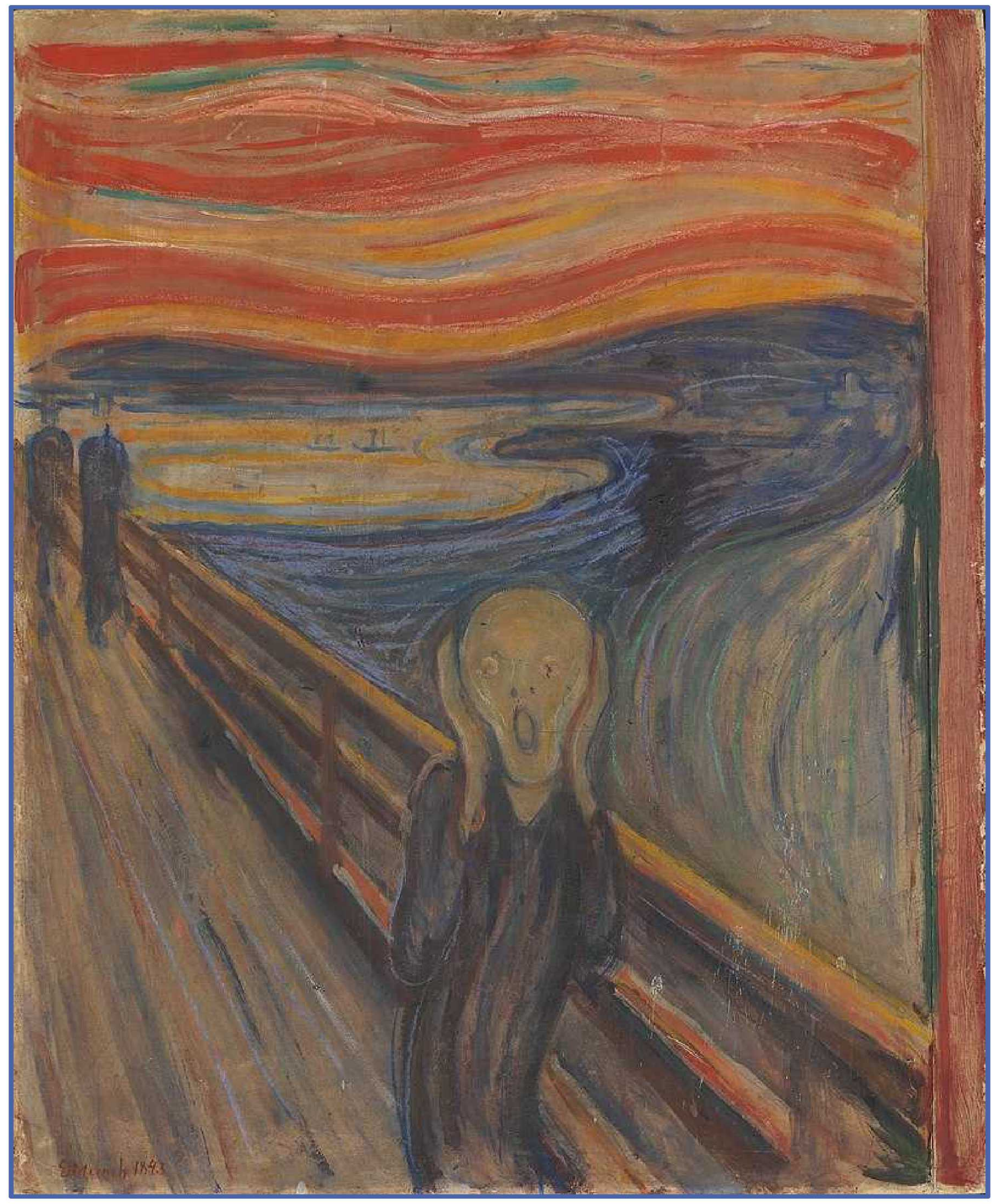
- Can be hard to get this data in a calibration setting – usually estimated using In-Tolerance-Probability: itp = (# devices calibrated within tolerance)/(total number of devices calibrated)

Global Risk

Global Risk: Probability of making an incorrect decision on *any* part in the population. Also called *Consumer's Risk* or *Unconditional Probability of False Accept.*

- Combines knowledge of measurement uncertainty with knowledge/historical data of all possible parts.

$$PFA = \int_{-\infty}^{SL_L} \left(\int_{SL_L}^{SL_U} \frac{1}{\sigma_m \sqrt{2\pi}} e^{-\frac{1}{2\sigma_m^2}(y-t)^2} dy \right) \frac{1}{\sigma_p \sqrt{2\pi}} e^{-\frac{1}{2\sigma_p^2}\left(t-\mu_p\right)^2} dt$$
$$+ \int_{SL_U}^{\infty} \left(\int_{SL_L}^{SL_U} \frac{1}{\sigma_m \sqrt{2\pi}} e^{-\frac{1}{2\sigma_m^2}(y-t)^2} dy \right) \frac{1}{\sigma_p \sqrt{2\pi}} e^{-\frac{1}{2\sigma_p^2}\left(t-\mu_p\right)^2} dt.$$


Edvard Munch/Public Domain

Test Uncertainty Ratio

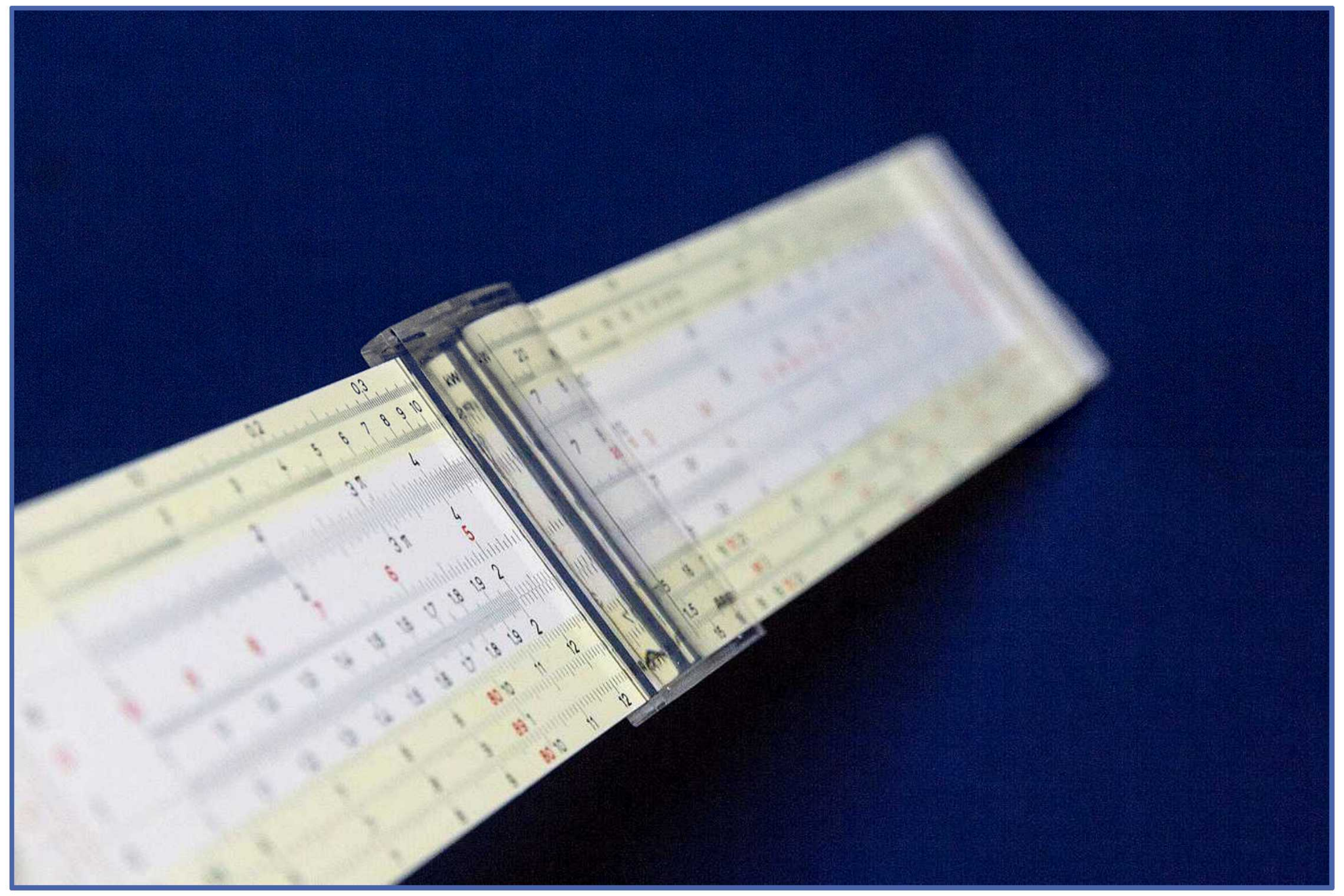
- **Test Uncertainty Ratio (TUR):** A simple metric for evaluating the level of risk in a pass/fail decision based on a requirement

TUR = (± Tolerance)/(± Measurement Uncertainty @ 95%)

- Generally accepted rule: If the TUR is ≥ 4, the measurement is adequate.

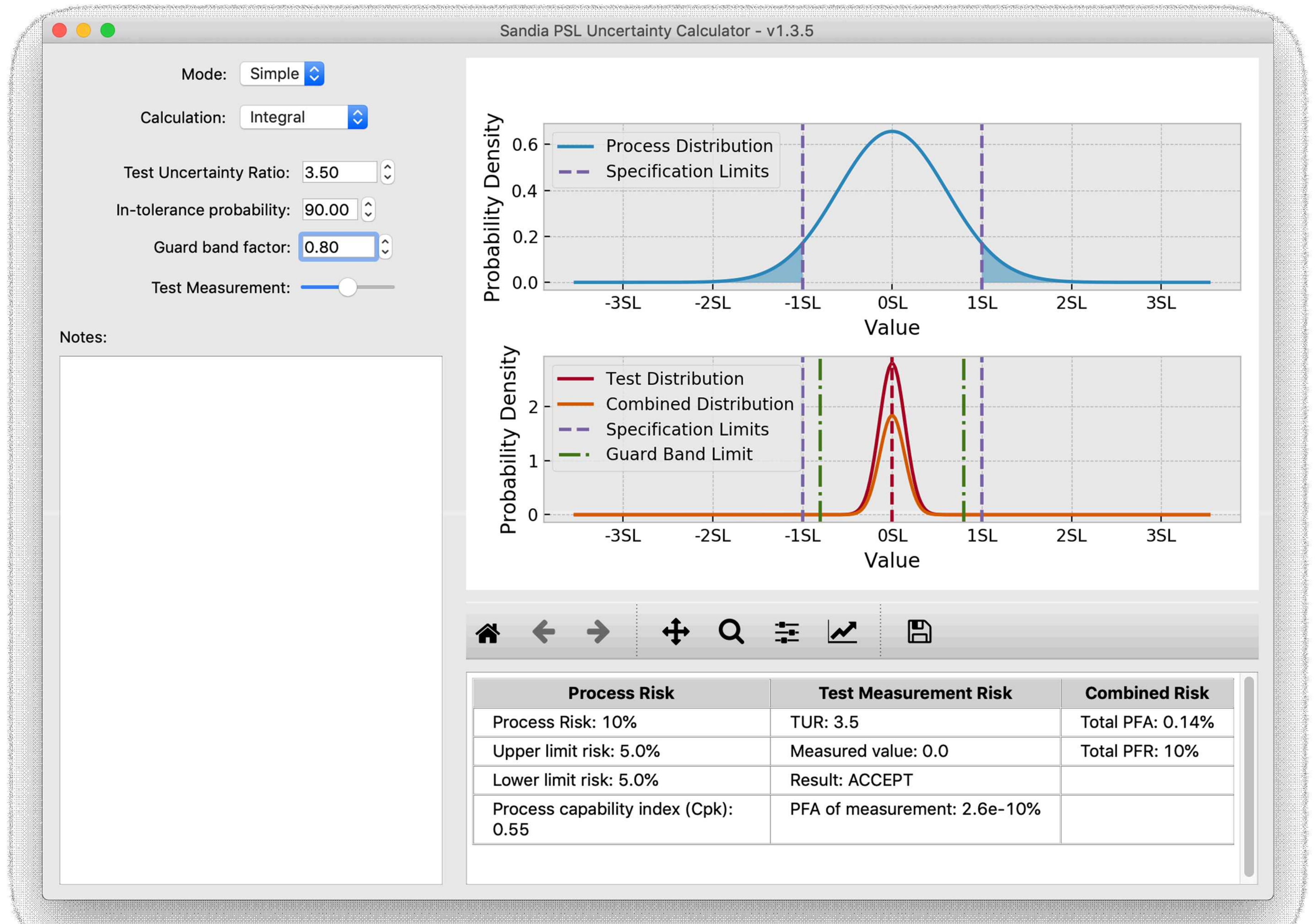
Risk Calculation Tools

- TUR as a quick metric for simple risk evaluation



Jan1959/CC BY-SA 4.0

- Software for full false accept false reject evaluation



Sandia Uncertainty Calculator https://sandiapsl.github.io

Test Uncertainty Ratio

TUR = (± Tolerance)/(± Measurement Uncertainty @ 95%)

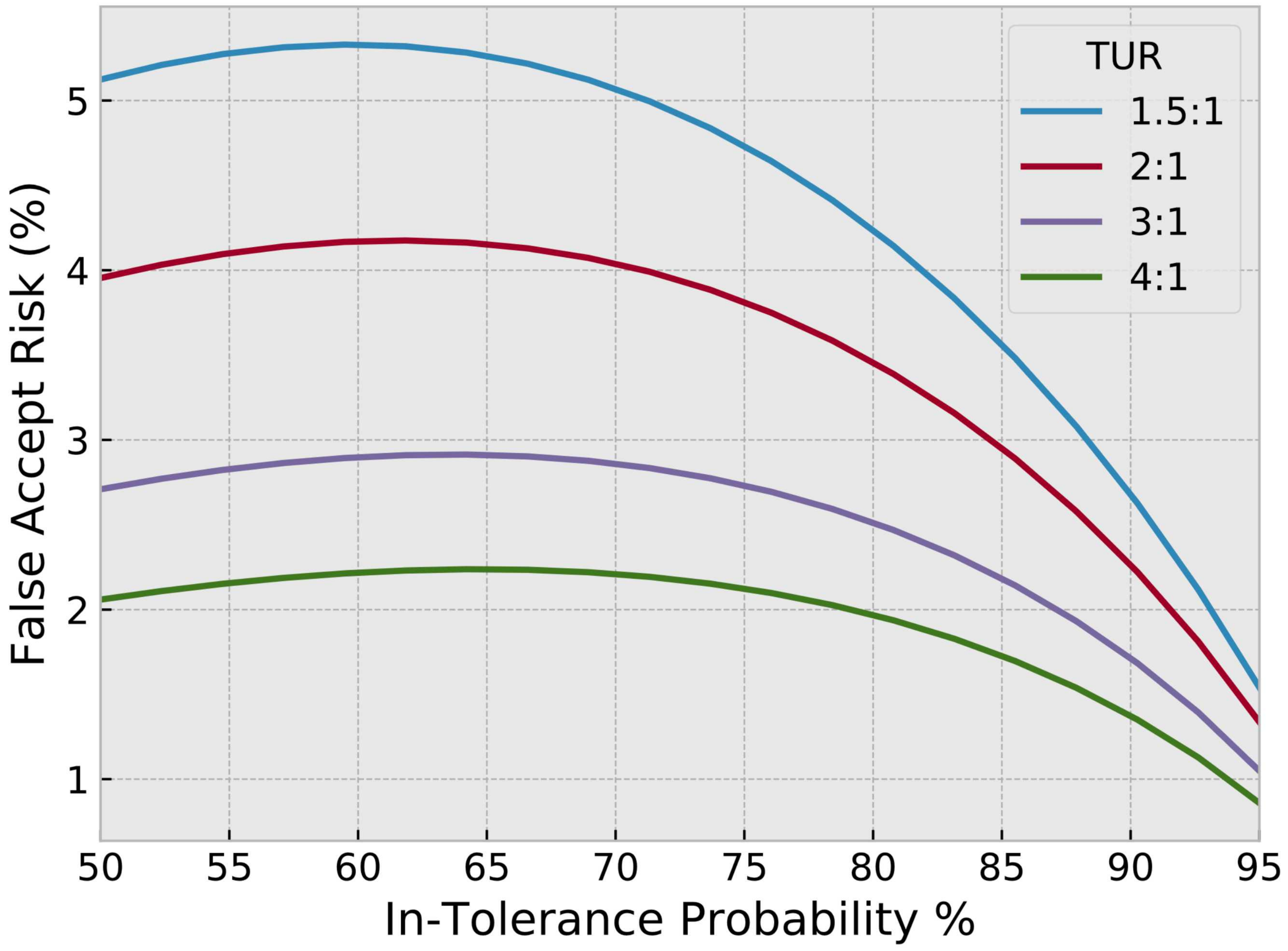
- Example:
  - Calibrating a voltage source with specification 1000 V ± 3 V
  - Measurement uncertainty = ± 0.25 V, 95% confidence
  - TUR = 3/0.25 = 12

Test Uncertainty Ratio

TUR = (± Tolerance)/(± Measurement Uncertainty @ 95%)

- Also occasionally encountered:
  - **Test Accuracy Ratio (TAR):** Only includes manufacturer specification in denominator.
  - **Measurement capability index:** GUM decided there's too many confusing definitions of TUR, so they added yet another one (same as the above definition of TUR).

Why 4:1?



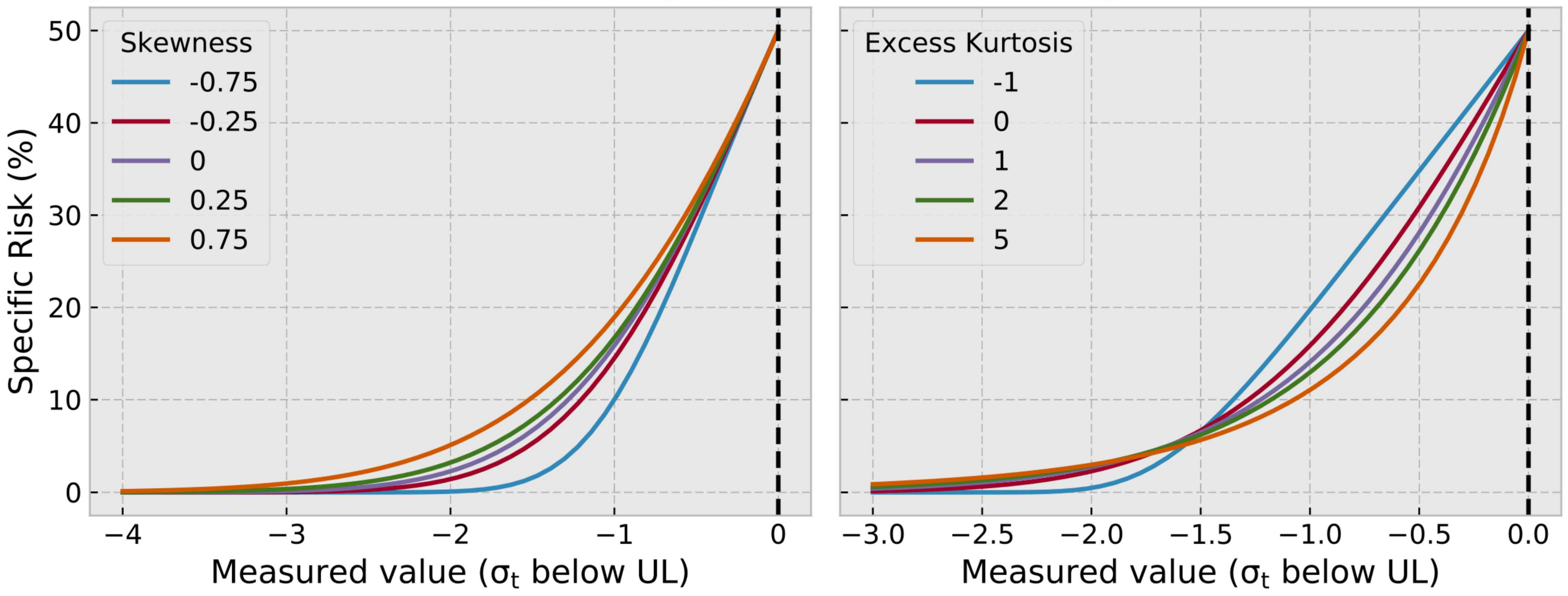
- ANSI/NCSL Z540.3: “incorrect acceptance decisions … shall not exceed 2% … where not practicable to estimate this probability, TUR shall be equal to or greater than 4.”
- 4:1 TUR guarantees False Accept risk is less than 2% for in-tolerance probability ≥ 80%.

When is using the 4:1 rule inappropriate?

Some assumptions of the 4:1 rule:

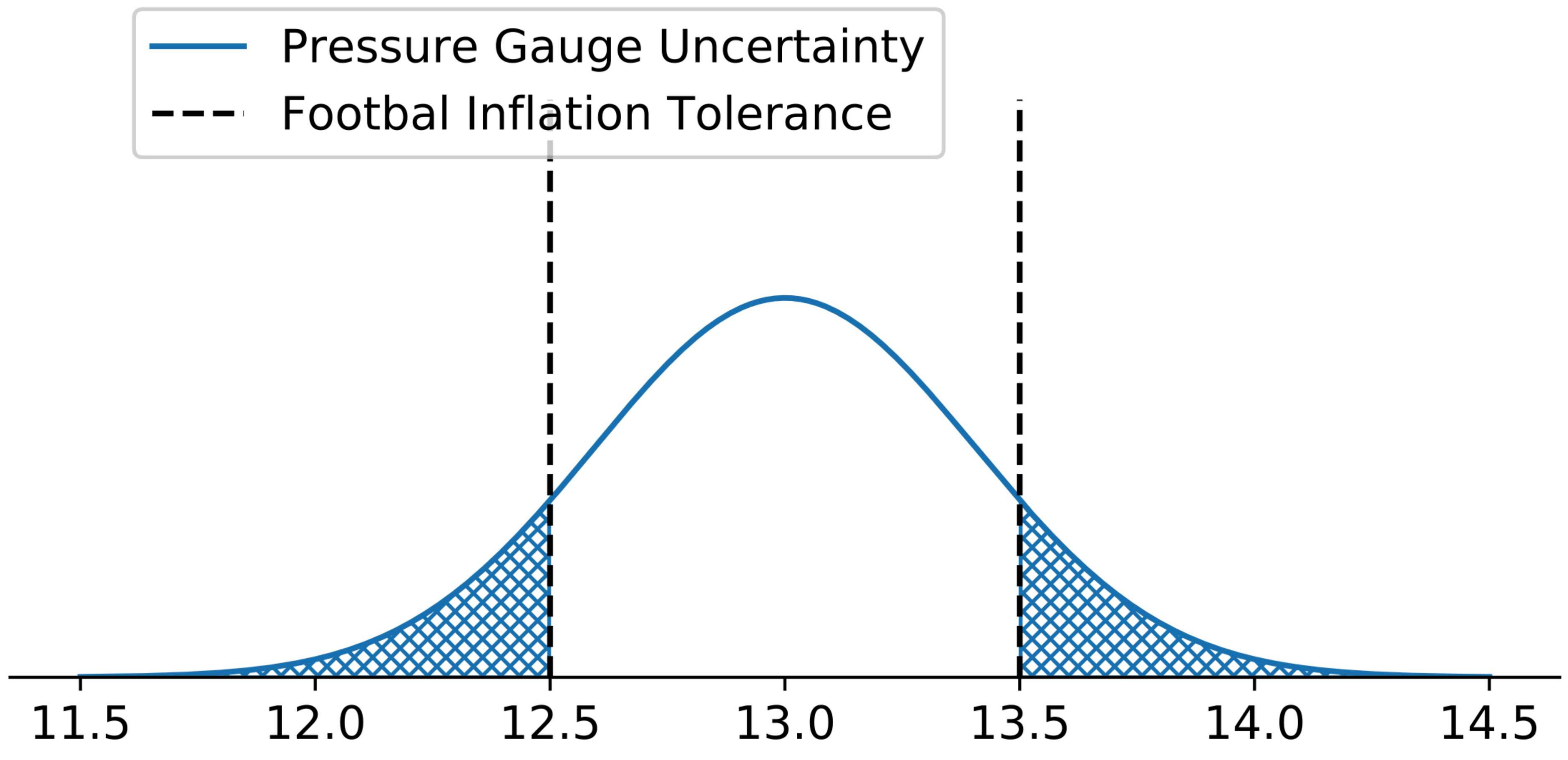
- Measurement uncertainty is normal and unbiased
- Average of all products being tested is normal and centered between tolerance limits
- There is sufficient data to characterize both probability distributions
- 2% False Accept rate is acceptable for the given application

Stay tuned for NCSLI Workshop and Symposium 2020 “Evaluating Risk in an Abnormal World”:



Deflategate TUR

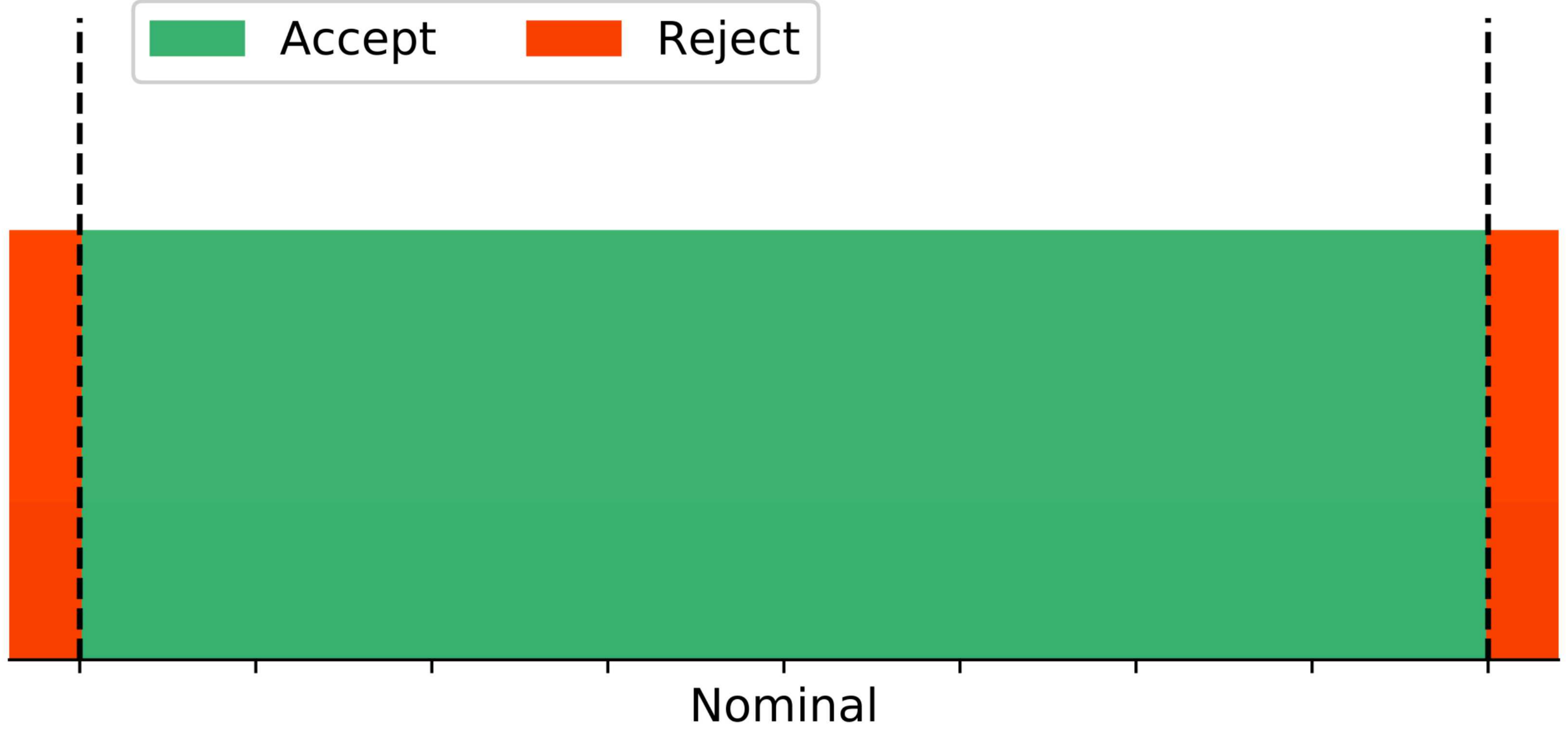
- NFL football pressure requirement: 13.0 psig ± 0.5 psig
- Estimated uncertainty in gauges used by referees: ± 0.81 psig @ 95% confidence
- TUR = 0.5/0.81 = 0.6 << 4 → **Gauges were not adequate!**

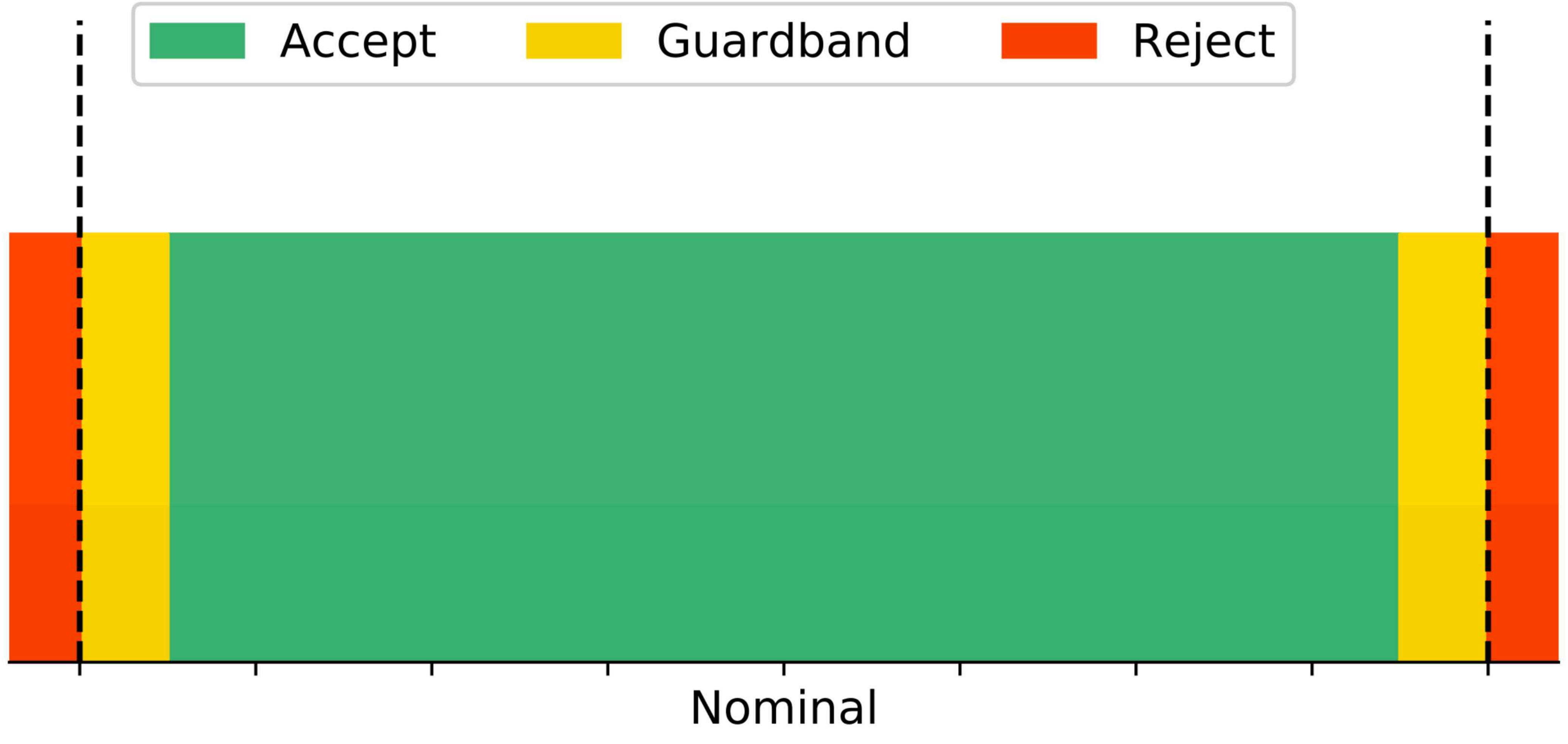


Specific false accept risk at 13.0 psig: ~21%

Guardbanding

- If TUR < 4, use a guardband to reduce the risk of an incorrect decision.
- Parts measuring in the green are accepted.
- Parts measuring in the yellow guardband are rejected, even if the measurement value indicates it is within specifications.





Guardbanding

- No standardized method for computing a guardband, but RSS is most common:

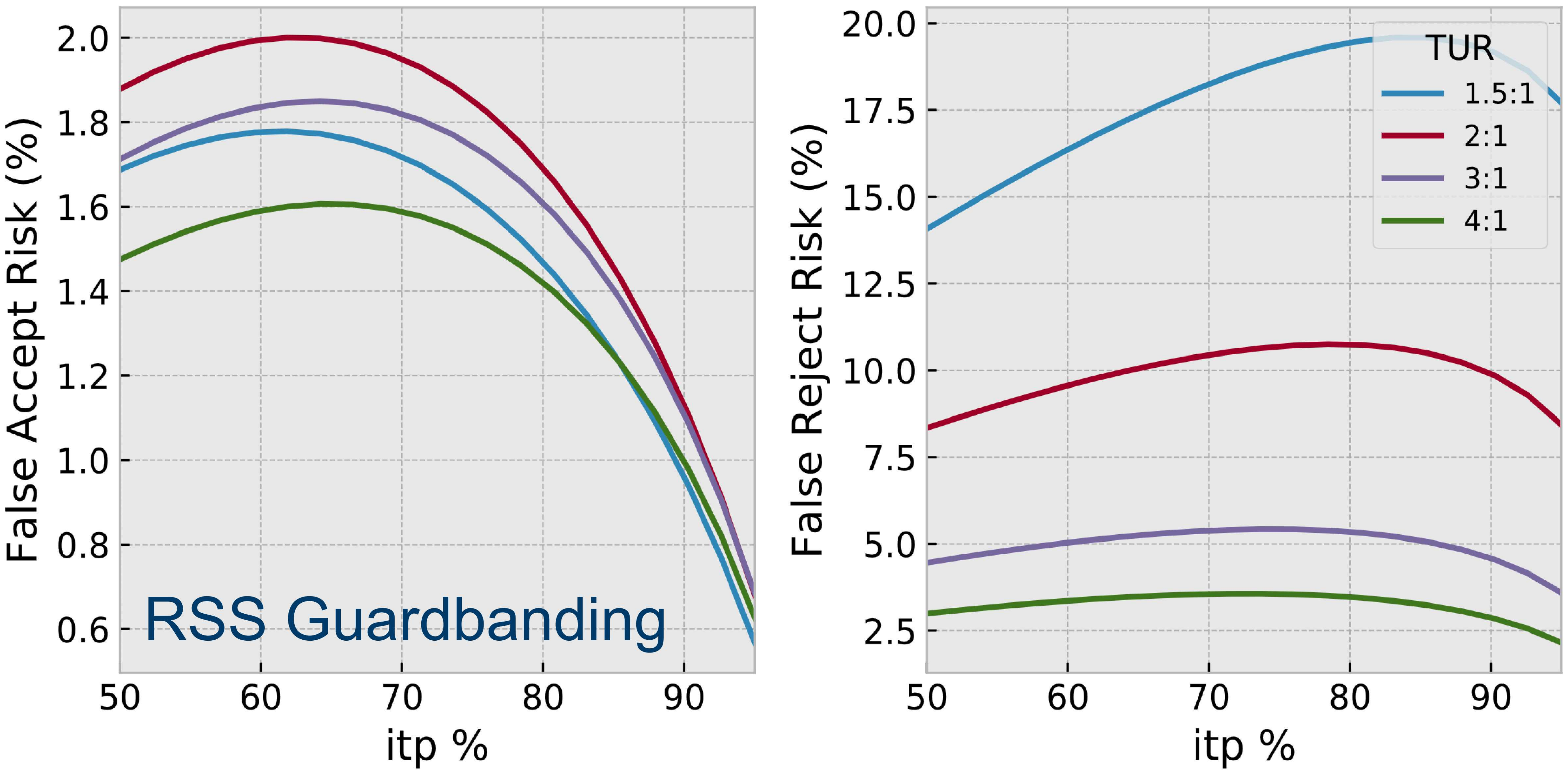
- RSS Method: Guardband Factor (GBF) $= \sqrt{1 - \frac{1}{TUR^2}}$

- Acceptance Limit = Tolerance × GBF

- Example:
  - Calibrating a voltage source with requirement is ± 3 V
  - TUR = 2
  - GBF = $\sqrt{1 - \frac{1}{2^2}} = 0.87$
  - Acceptance Limits = 3 × 0.87 = ± 2.60 V

Guardbanded Risk

- With RSS guardband, **False Accept will always be less than 2%.**
- However, False Reject probability shoots way up!
- Generally, TUR > 1.5 is required, even with guardbanding.



Decision Rules

- **Simple Acceptance:** Accept anything that measures within the limits. Use when TUR ≥ 4.

- **Guarded Acceptance:** Accept anything within reduced guardbanded limits. Use when 1.5 < TUR < 4.

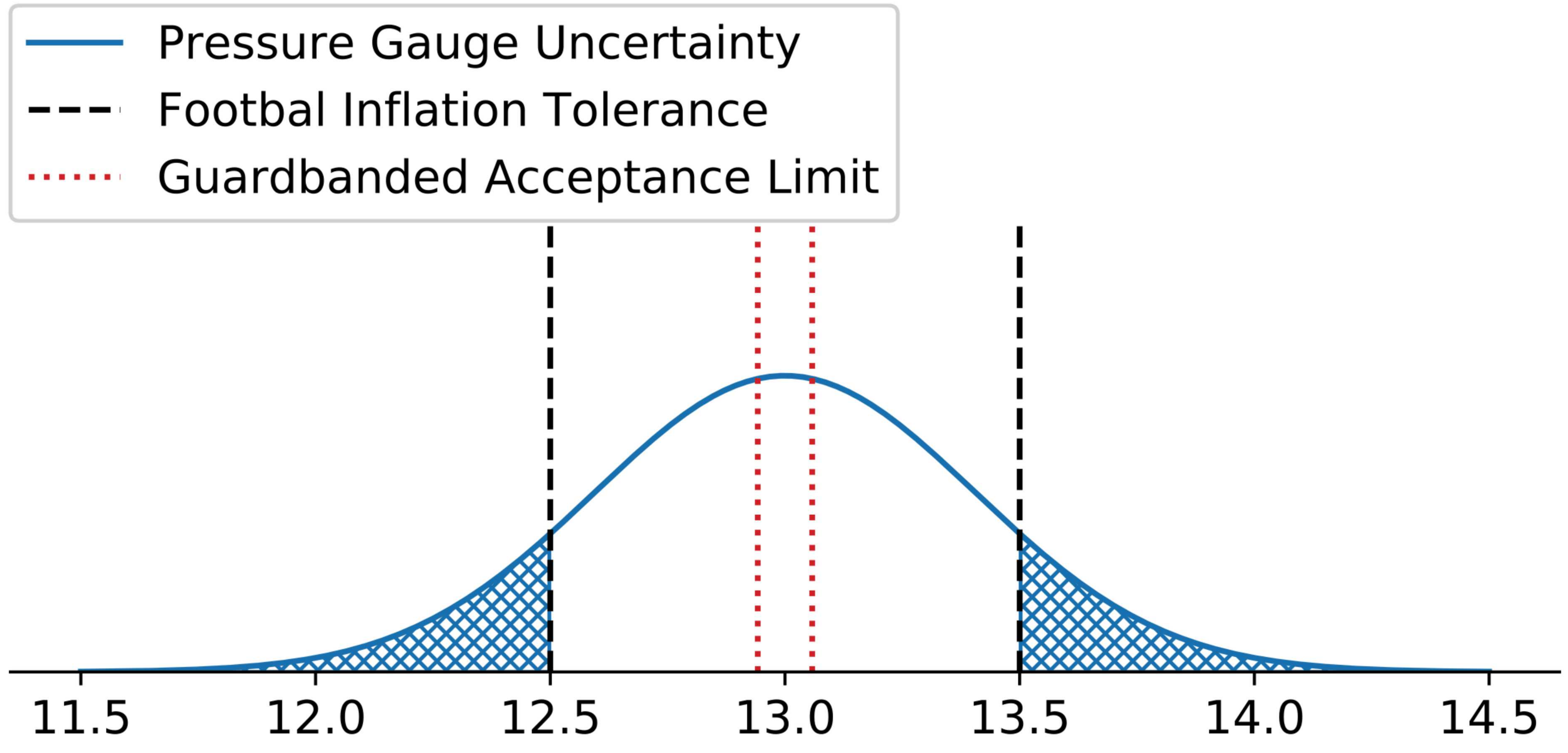
- **Indeterminate Acceptance:** Accept anything within reduced guardbanded limits. Reject anything outside of limits. Anything measuring in the guardbanded range is “indeterminate”.

Deflategate Guardbanding

- Football TUR = 0.5/0.81 = 0.6
- GBF = $\sqrt{1-\frac{1}{TUR^2}} = \sqrt{1-\frac{1}{0.6^2}} = \sqrt{-1.78}$ → Wait, that doesn't work...
- Generally, if TUR < 1.5 a measurement cannot be guardbanded.

Deflategate Guardbanding

- TUR = 0.5/0.81 = 0.6
- "Method 6" of NCSL Handbook for Z540.3:
  - GBF = 0.05



False accept rate: ~0.8%
False reject rate: ~81%

Measurement Assurance Plans

A spreadsheet to keep track of *all* measurements required to validate a product (or calibrate an instrument), the equipment used to measure the requirement, and the associated TURs.

What quantity are you measuring?		How accurate do the measurements need to be?			How will you ensure your equipment is capable of making this measurement?			
Quantity Measured	Requirement Number	Value or Range of Values Measured	Specification Limits	Guardbanded Acceptance Limit	Equipment Used	Measurement Uncertainty	TUR	Notes
Voltage	N/A	60 µV	± 6 µV	± 5.7 µV	Keysight 3458A	± 2 µV	3	GB caclulated using RSS method.

Review of Key Terms

- **Direct Measurement:** Value is read directly off a measuring instrument.
- **Indirect Measurement:** Value is obtained using a formula with multiple measurements.
- **GUM Equation:** An analytical method for combining uncertainties of an indirect measurement.
- **Monte Carlo Approach:** A numerical method for combining uncertainties of an indirect measurement.
- **Specific Risk:** Probability of incorrect decision given a particular measurement result.
- **Global Risk or Consumer's Risk:** Probability of incorrect decision on *any* measurement result on equivalent products.
- **TUR:** Ratio between product tolerance and measurement uncertainty.
- **4:1 Rule:** A measurement with TUR ≥ 4 is suitable to make a decision based on a tolerance.
- **Guardbanding**: Reduction in product tolerance to reduce global and specific risk.
- **Decision Rule:** Statement of how pass or fail decisions are made.

Acknowledgements

- Sandia Primary Standards Lab
  - Eric Forrest
  - Josh Reynolds
  - Raegan Johnson
  - Andrew Mackrory
  - Ricky Sandoval
  - Kyle Hoff
  - Aaron Meyrick
  - Tim Kypta
  - Tim Vanderburg
  - Wyatt Hodges
  - Faith Tinnin
  - Roger Burton
  - Meaghan Carpenter
- Sandia Statistical Sciences Department
  - Steve Crowder
  - Dan Campbell
  - Nevin Martin

References

- GUM, GUM supplements, and VIM, available at: http://www.bipm.org/en/publications/guides/
- NIST Technical Note 1297 (A publication from NIST that gives an overview of the GUM approach), available at: https://dx.doi.org/10.6028/NIST.TN.1297
- NIST Technical Note 1900 (Lots of uncertainty calculation examples), available at: https://doi.org/10.6028/NIST.TN.1900
- SI unit definitions are in the SI Brochure, available at: http://www.bipm.org/en/si/si_brochure/
- Kimothi, S.K., Uncertainty of Measurements, ASQ Quality Press, 2002.
- Dieck, R. H., Measurement Uncertainty methods and applications,
- Rabinovich, Semyon G., Measurement Errors and Uncertainties, American Institute of Physics, NY. (more mathematical)
- Sandia Uncertainty Calculator (https://sandiapsl.github.io)
- NIST Uncertainty Machine (https://uncertainty.nist.gov)
- Crowder, S., Delker, C., Forrest, E., Martin, N., Introduction to Statistics in Metrology. Springer (in review 2020).

METROLOGY *in* MOTION

A WEBINAR SERIES

by NCSL INTERNATIONAL