Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Statistical Considerations in Setting Acceptance Criteria

1. Reference intervals (and outlier tests)
2. Tolerance intervals
3. Process performance indices, 6 sigmas, …
4. Summary

Disclaimer

The views expressed here are those of the author and may not necessary reflect those of the German Regulatory Authorities

Reference Intervals

- Mean ± 2 standard deviations (Mean ± 3 s)
- Easy to implement
- Data should be symmetrically (ideal: normally) distribu
- Otherwise: Transformation of the data
- Reliable, unbiased estimation of mean and standard deviation necessary

Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

1-2-3-σ-Rule

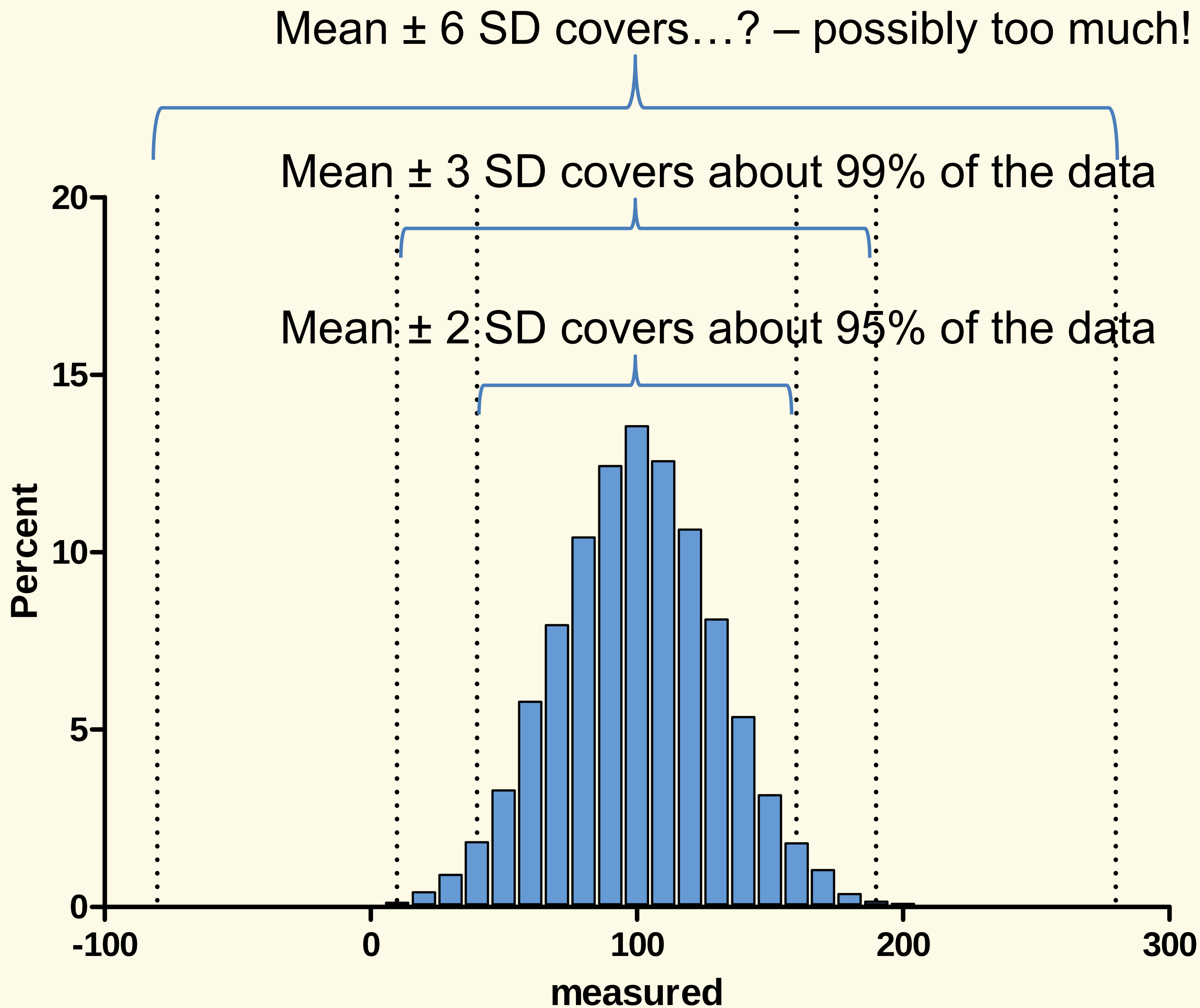
Reference Intervals include a certain amount of the data (assumption: normal distribution)

Interval	Probability	
Mean ± …	Within	Outside
… 1 s	0.683	0.317
… 2 s	0.954	0.046
… 3 s	0.998	0.003
… 1.64 s	0.900	0.100
… 1.96 s	0.950	0.050
… 2.58 s	0.990	0.010

95% of the data can be found in the interval [Mean - 2 s, Mean + 2s]

Or: With 95% probability a value lies in this interval

1. Reference Intervals

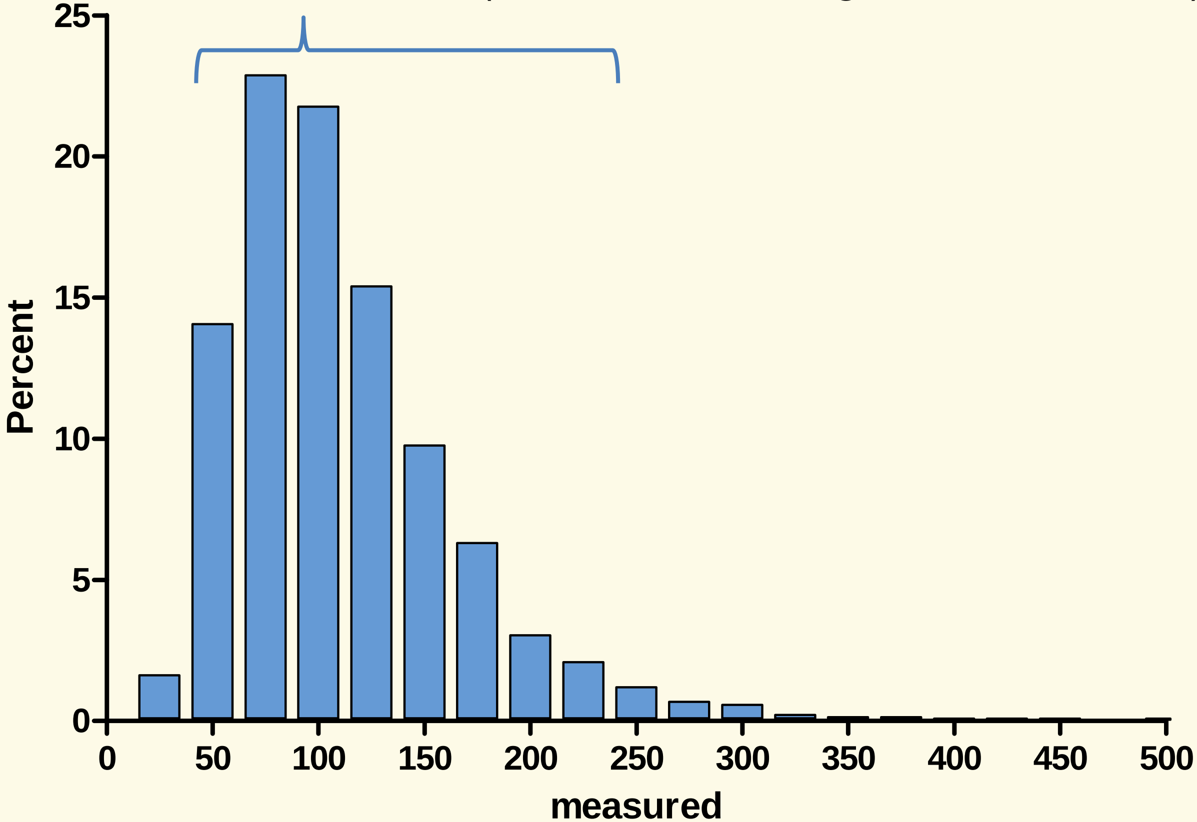


Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Skewed distributions / non-normal distributed data:

Transformation of data (→ skewed specification)

Specification
39.8 – 251.2 (Mean ± 2s, of logarithmised data)

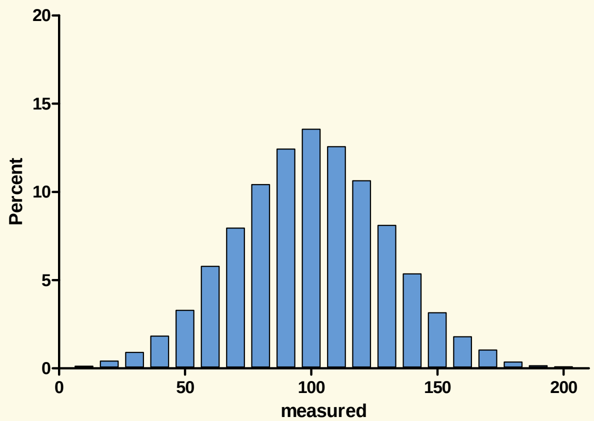


Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

How much data needed to set up specifications?

Simulated with true population parameters: μ = 100, σ = 30

N=10,000

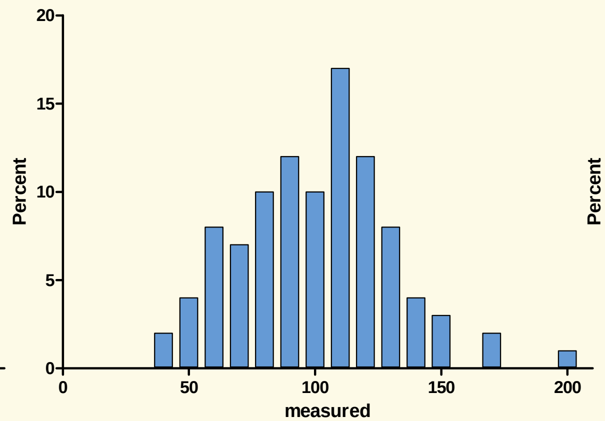


95%-Confidence Interval

$\bar{x}$ = 99.9 [99.3 – 100.5]

s = 30.0 [29.6 – 30.4]

N=100

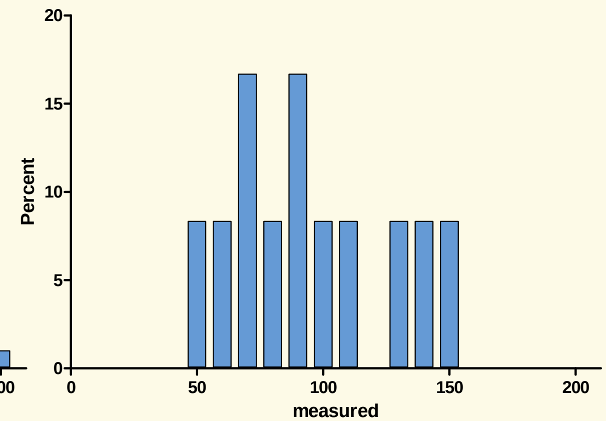


95%-Confidence Interval

$\bar{x}$ = 100.9 [95.0 – 106.8]

s = 29.9 [26.3 – 34.8]

N=12



95%-Confidence Interval

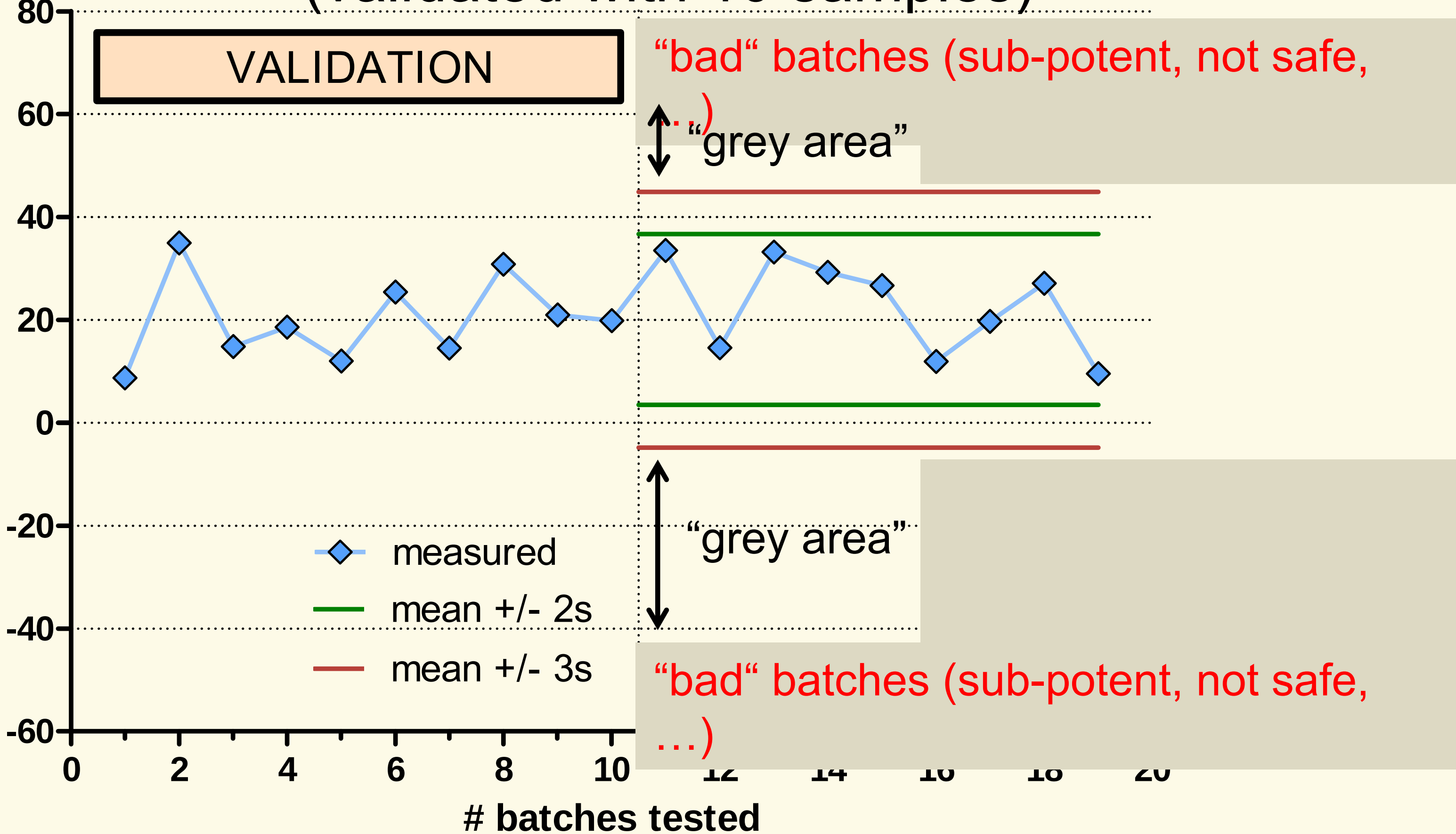
$\bar{x}$ = 96.2 [76.6 – 115.8]

s = 30.9 [21.9 – 52.5]

Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Example Reference Interval I

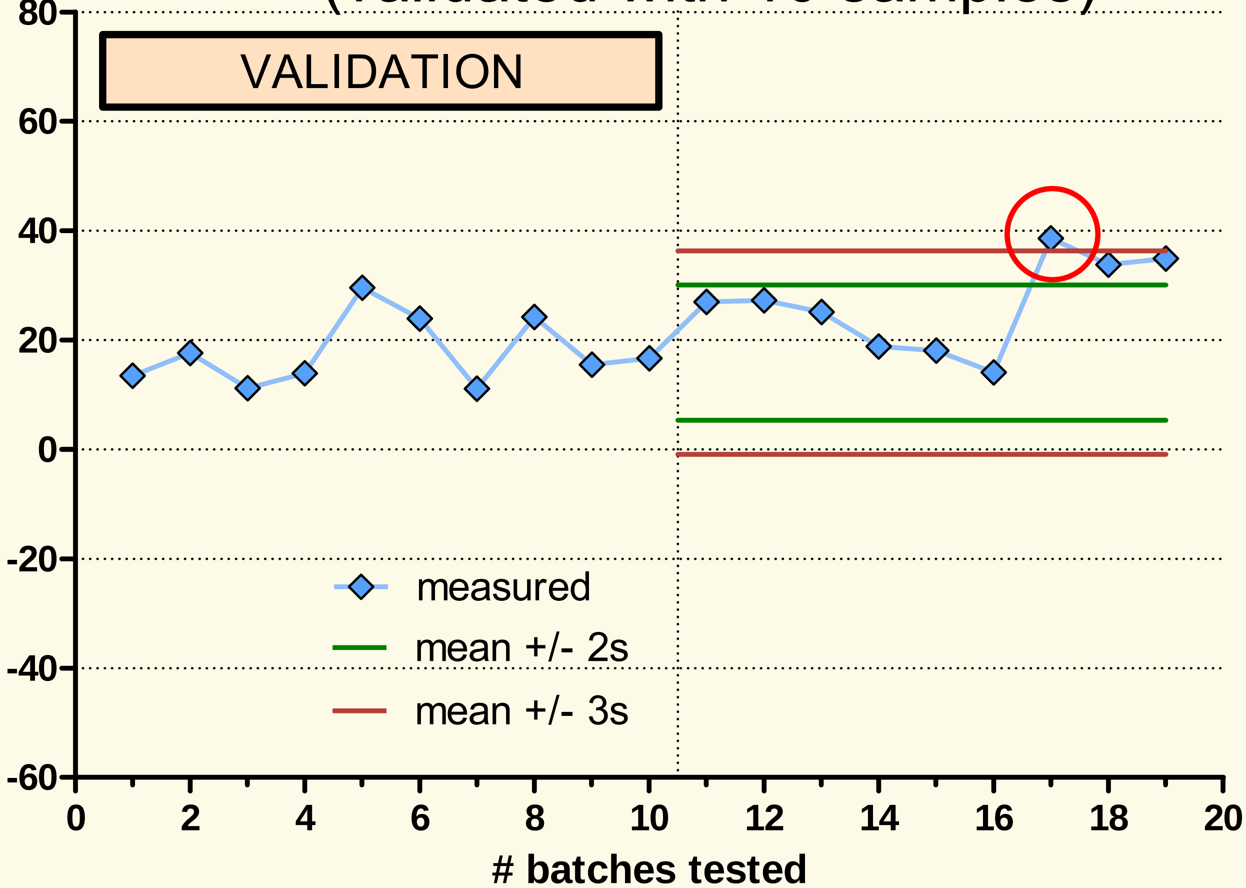
Mean +/-2s as "warning" limits, Mean +/-3s as "intervention" limits (validated with 10 samples)



Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Example Reference Interval II

Mean +/-2s as "warning" limits, Mean +/-3s as "intervention" limits (validated with 10 samples)



Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

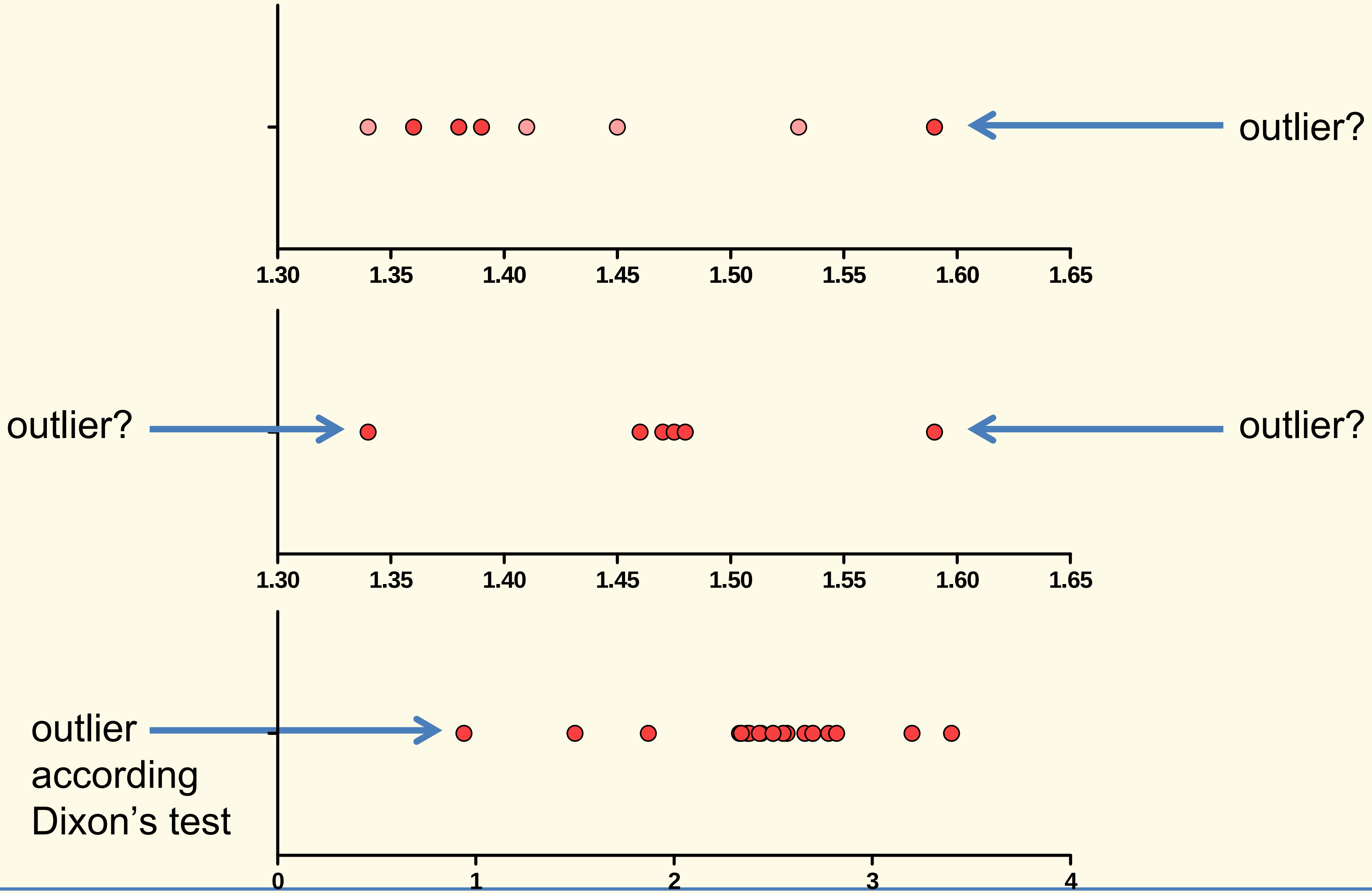
Outliers and outlier tests

Is it an outlier? – Or does it belong to the population?

It depends on how much information we have…

Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

There are several outlier tests… but be careful using them!



Tolerance Intervals

Intervals that cover percentiles of the population with a certain probability

Non-parametric TI – percentiles:

- Example: [$p_{0.05}$ – $p_{0.95}$] might serve as TI for the mean 90% of the population

- Example: Smallest – largest observation [$y_{(1)}$ – $y_{(n)}$] might serve as TI for whole population
- BUT: To cover actually 90% of the population with [$y_{(1)}$ – $y_{(n)}$], N=19 measurements are necessary
- AND: To cover actually 90% of the population with [$y_{(1)}$ – $y_{(n)}$] with 95% probability, N=46 measurements are necessary

2. Tolerance Intervals

Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

TI (2–sided) for normal distributed data:

$$\bar{x} \pm k_{n;\gamma} s$$

with
$\bar{x}$ as mean of validation samples,
s as standard deviation,
n as number of samples tested,
γ as proportion of population covered, and

$$k_{n;\gamma} = \sqrt{\frac{n+1}{n} t_{n-1;\frac{1+\gamma}{2}}}$$

(t: quantile of t distribution)

This TI covers γ (%) of the population

Guttman (1970), Rasch (1996

2. Tolerance Intervals

TI (2–sided) for normal distributed data:

$\bar{x} \pm k_2 s$

with
$\bar{x}$ as mean of validation samples,
s as standard deviation,
n as number of samples tested,
γ as proportion of population covered,
p as probability for γ% coverage (tolerance level), and

$$k_2 = \sqrt{(n-1)\left(1+\frac{1}{n}\right) z^2{}_{(1-p)/2} \Big/ \chi^2{}_{\gamma;n-1}}$$

(z: quantile of normal distribution and
χ^2: quantile of χ^2-square distribution)

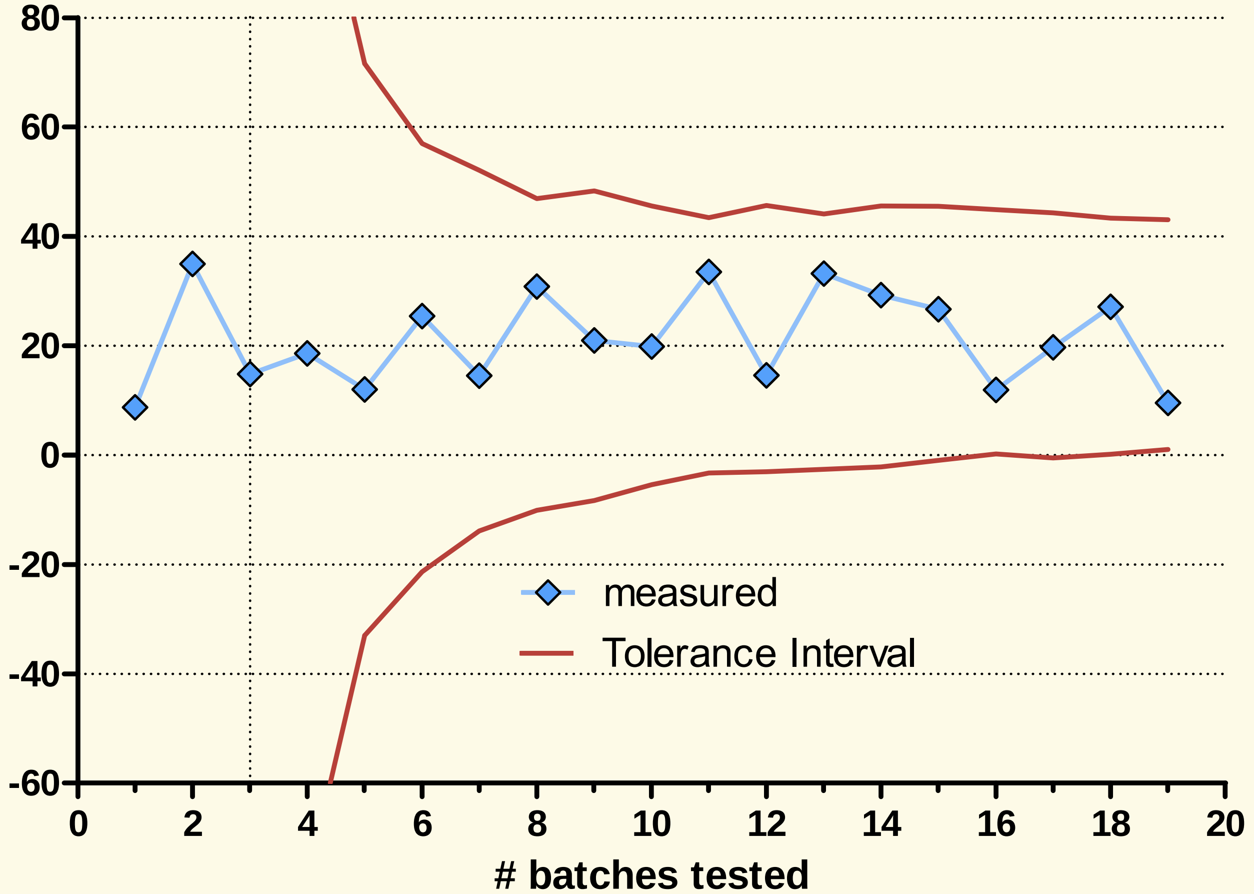
This TI covers γ (%) of the population <u>with probability p (%)</u>

Howe (1969)

Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Dynamic Tolerance Intervals

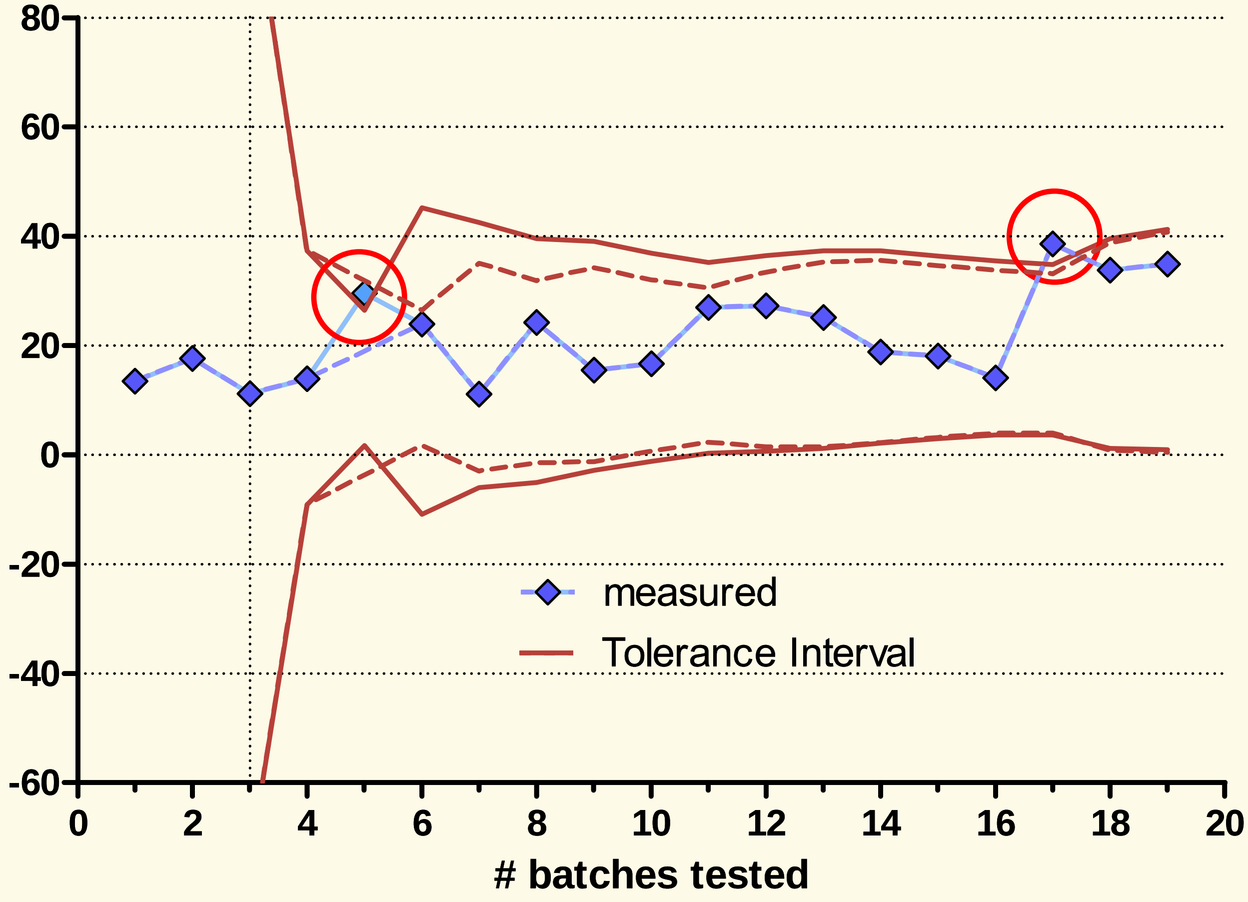
(γ=0.95, p=0.95; repeatedly calculated for n≥2 observations)



Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Dynamic Tolerance Intervals

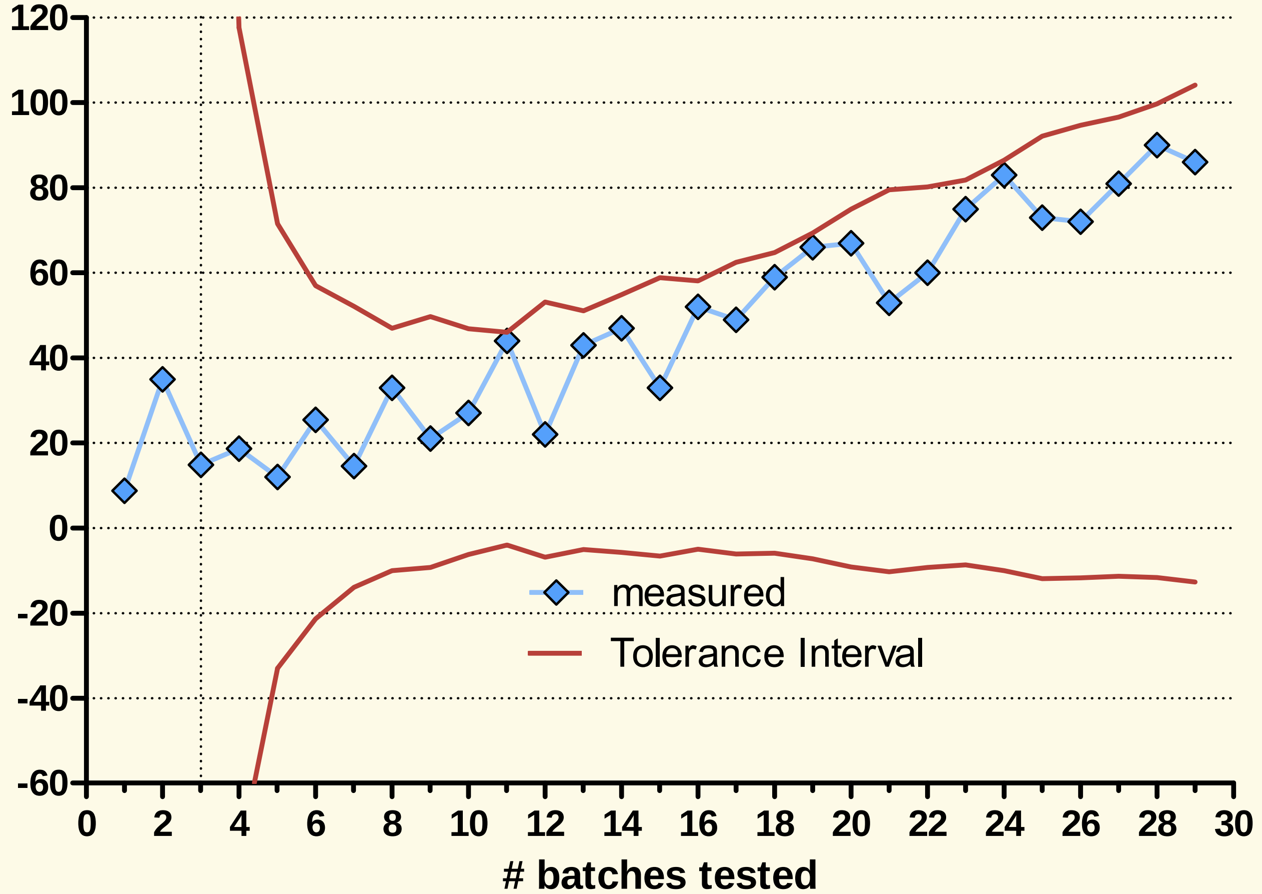
What about early OOS results? May belong to population – or may be OOS!



Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Dynamic Tolerance Intervals

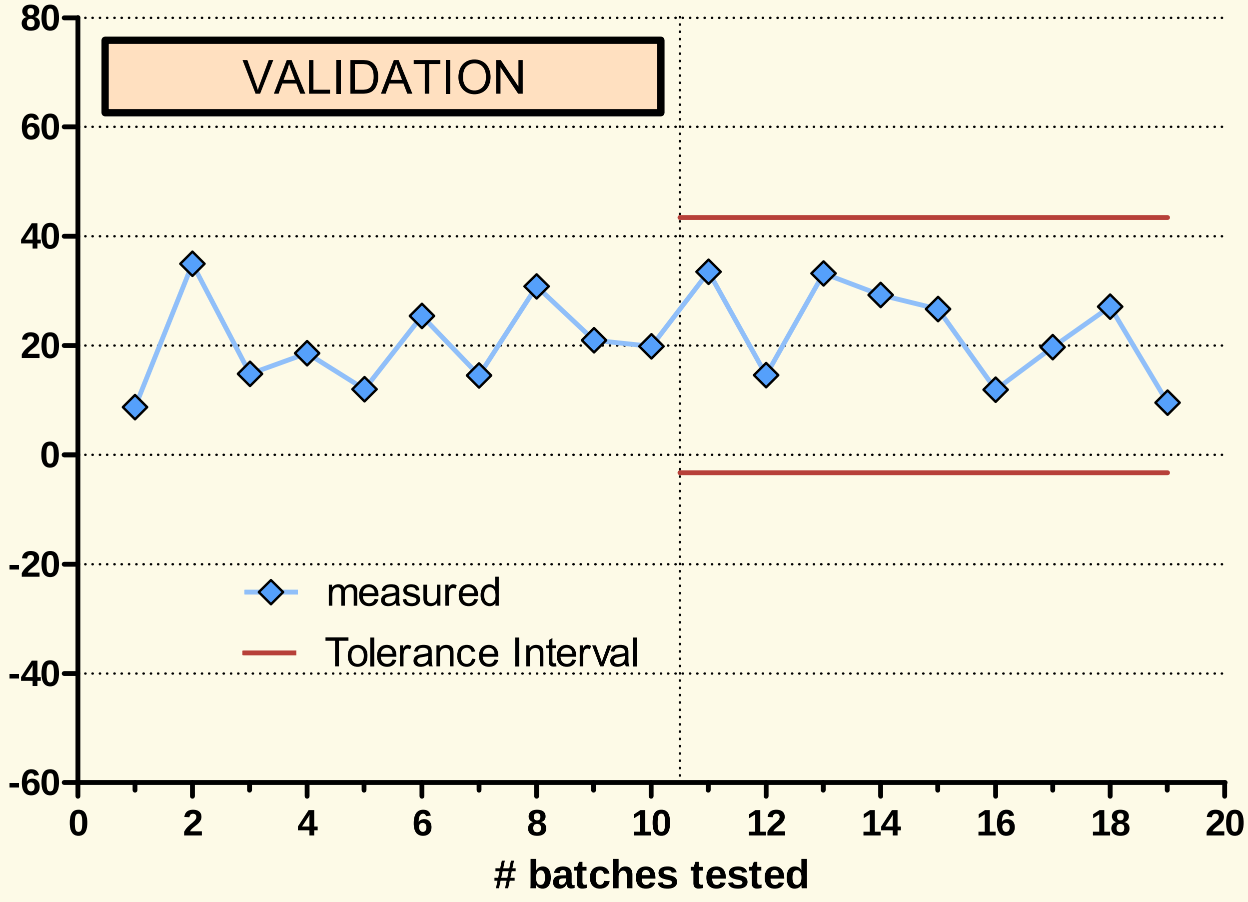
TI widens, when data shows a trend – may lead to undesired effects



Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

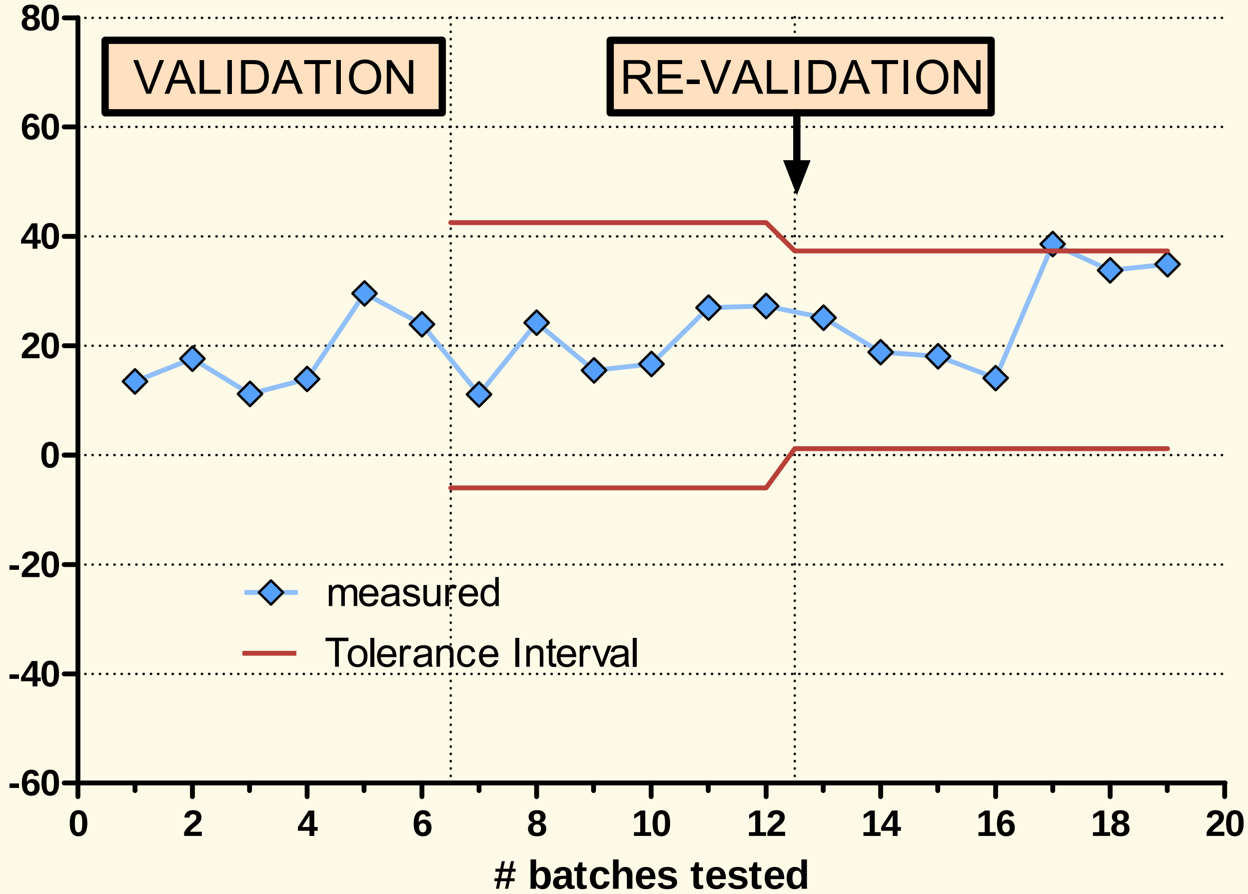
Fixed Tolerance Intervals

With a sufficient amount of validation data the TI will be more reliable



Paul-Ehrlich-Institut
Federal Institute for
Vaccines and Biomedicines

Two-Step Approach



Process performance indices, 6 sigmas, …

- PPI: Similar to reference limits (mean ± 3SD x Ppu)
- Advantage to be product specific
- BUT: Wouldn't such limits be to wide (and could include e.g. OOS batches)?
- At least: Limits must exclude range, where sub-potency, non-safety, … starts
- Thus: 6 sigmas are not recommended (for normal or similar distributed data these would include anything)

Summary

Crucial: Reliable estimation of the specification limits
-Too narrow intervals could result in too many re-tests and / or falsely rejected batches
-Too wide intervals could lead to falsely released batches (sub-potent, safety concerns, …)

- Specification based on what we have observed so far (might be few), thus future results may represent what we missed in validation
- Samples used for validation were too homogeneous (too narrow specifications) – bad luck?
- Samples used for validation had high variability (new processes, untrained personnel, too wide specifications)
- Validation performed with few samples, repeatedly tested (should be avoided!)

Summary II

- For a complete new product it will be difficult to set up reliable specification limits with only 3 batches/tests
- risk of miss-specification might be high, especially if variability of parameter of interest is expected to be high; a re-validation should be planned (n=8-12 batches) → 2-step-approach

- No universal tool available – parameter / product dependent
- Aim to obtain high sensitivity to detect critical batches (sub-potent, safety risks, …) and
- To obtain high specificity in order to avoid false negative results and to limit unnecessary re-tests

Thank you for your attention
☺