Apples and Oranges,- case studies on similarity, comparability and equivalence regarding potency determination

CASSS Bioassays 2024

Hermann Beck,

Analyt. Dev. Bioassays

F. Hoffmann-La Roche Ltd

In potency determination (analytics in general), we are always comparing

- Sample A
- Method A
- Analytical result
- Method condition A
  - Ready-to-Use
  - Cell-line A
  - Robot
  - …..
- Old Ref.Standard
- ….
- A

VS.

- Sample B
- Method B
- Specification
- Method condition B
  - Cells in Culture
  - Cell-line B
  - Manual
  - ….
- New Ref.Standard
- ….
- B

In potency determination (analytics in general), we are always comparing

- Sample A
- **Method A**
- Analytical result
- **Method condition A**
  - **Ready-to-Use**
  - **Cell-line A**
  - Robot
  - …..
- **Old Ref.Standard**
  - ….
  - A

VS.

- Sample B
- **Method B**
- Specification
- **Method condition B**
  - **Cells in Culture**
  - **Cell-line B**
  - Manual
  - ….
- **New Ref.Standard**
- ….
- B

Comparison as basis for informed decisions

VS.

Our approach: **how we compare depends on**

the actual question to be answered (goal of the comparison)

and the type of objects to be compared

→ ***focus more on science or statistic***

Key-question: what is *similar enough, what is a meaningful difference*?

how much blur is acceptable?

what can happen?

In potency determination (analytics in general), we are always comparing

- Sample A
- **Method A**
- Analytical result
- Method condition A
  - Ready-to-Use
  - Cell-line A
  - Robot
  - …..
- Old Ref.Standard
  - ….
  - A

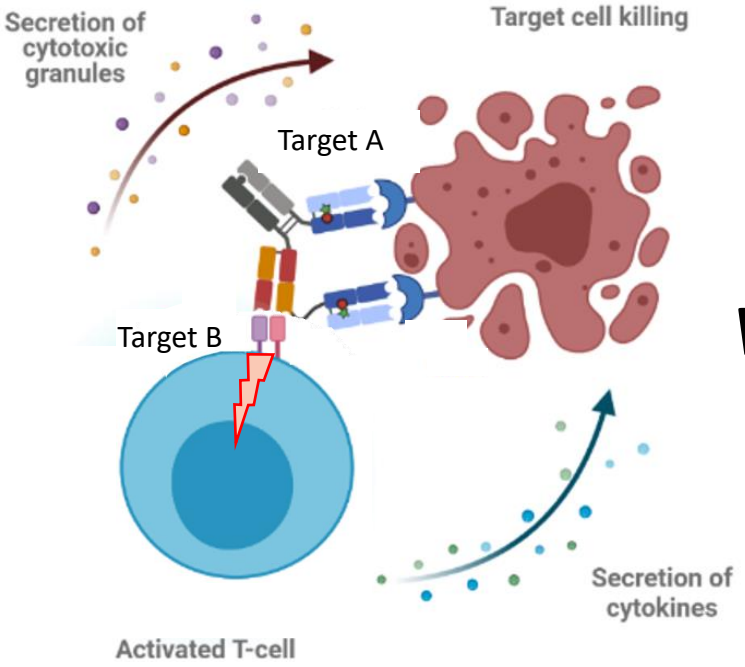
VS.

- Sample B
- **Method B**
- Specification
- Method condition B
  - Cells in Culture
  - Cell-line B
  - Manual
  - ….
- New Ref.Standard
- ….
- B

Comparing different types of assays

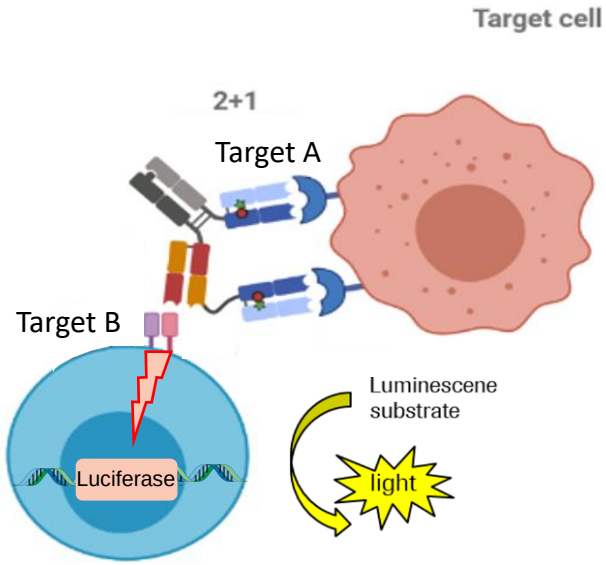
Bispecific Mab (2+1-format), MoA = target-cell killing & T-cell activation

Cell-killing assay with primary cells

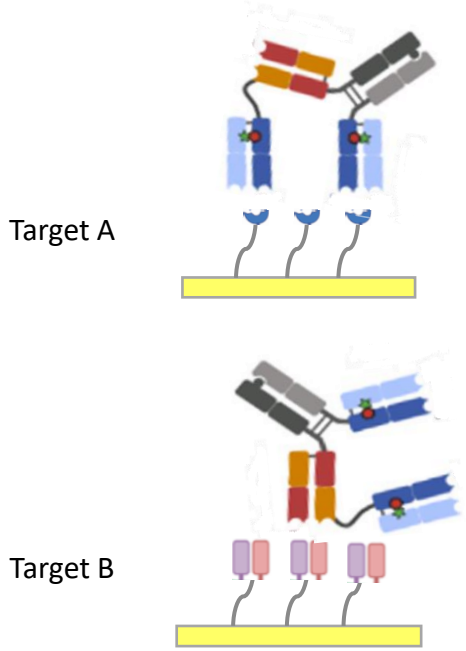


VS.

T-cell activation ReporterGene Assay with cell-lines



Target-binding by SPR



relation to clinical effect/
MoA-reflectiveness

technical QC suitability

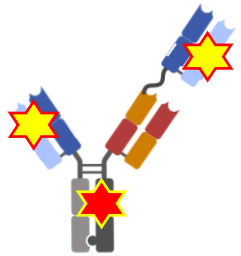
comparing different types of assays

Example: **Assays based on primary cells vs. cell line-based ReporterGene Assay (RGA)**

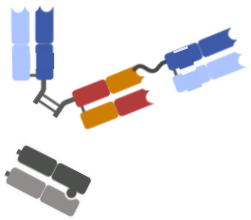
The actual question: is the cell-line based assay MoA-reflective, i.e. is what it displays relevant for the patient?

Approach: test sample panel in both assays and compare

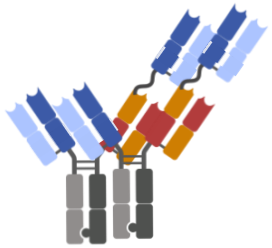
Which samples? With spiking-samples (50%..150% RS) you just learn about method capabilities
→ use «real» (stab, CQA,…) samples, as they are more relevant for the patient



Oxidations, deamidations ,….

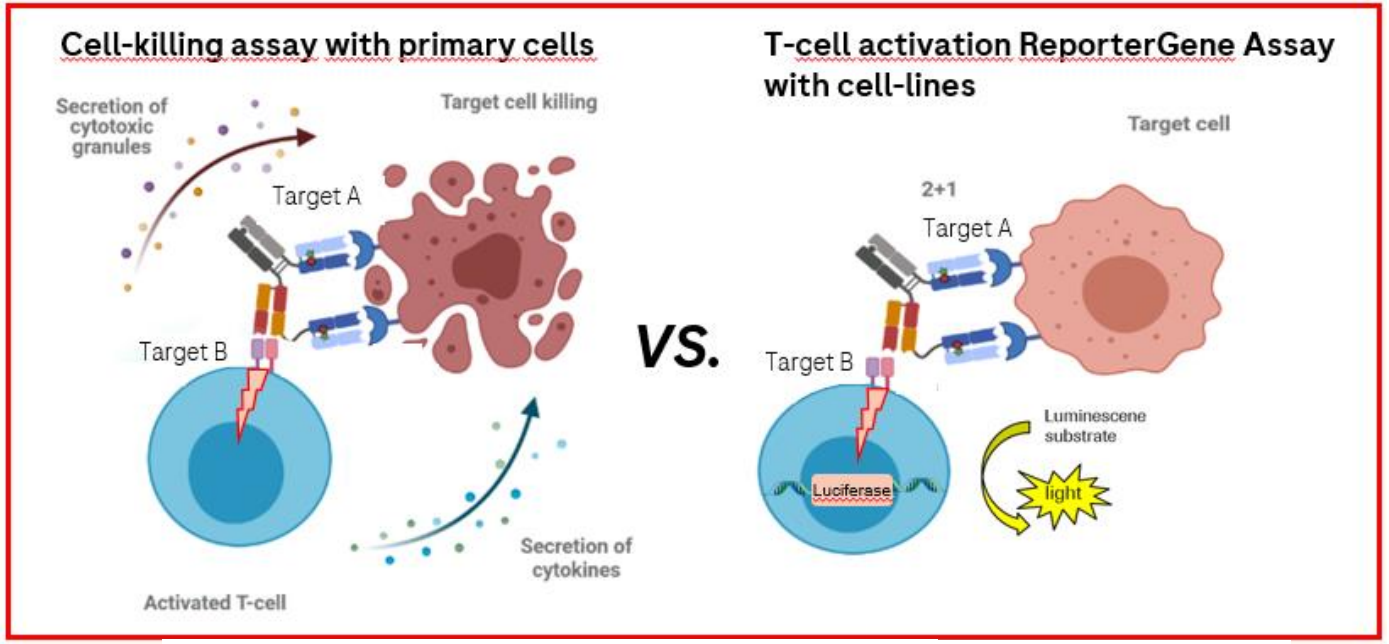


Fragments



Aggregates

…..



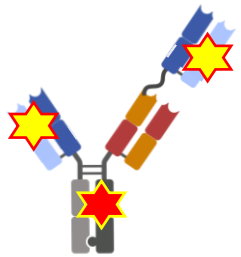
comparing different types of assays

Example: **Assays based on primary cells vs. cell line-based ReporterGene Assay (RGA)**

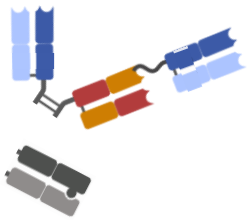
The actual question: is the cell-line based assay MoA-reflective, i.e. is what it displays relevant for the patient?

Approach: test sample panel in both assays and compare

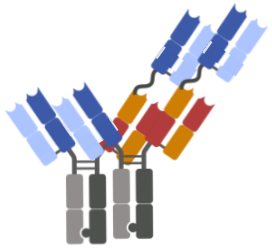
Which samples? With spiking-samples (50%..150% RS) you just learn about method capabilities
→ use «real» (stab, CQA,…) samples, as they are more relevant for the patient



Oxidations, deamidations ,….



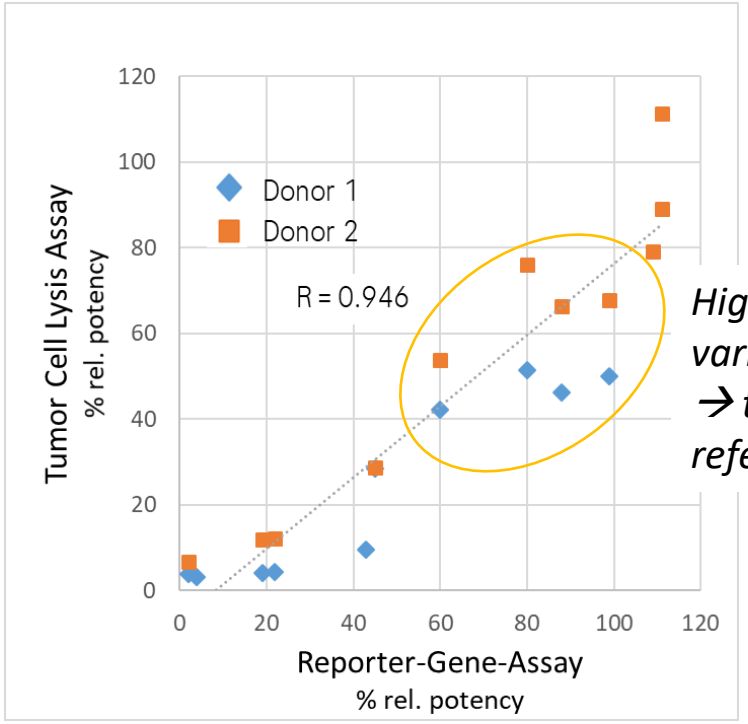
Fragments



Aggregates

….

Correlation of cell-killing assay with primary cells and RG-Assay with 2 cell lines



High donor-to-donor variability
→ *there is no absolute reference*

Conclusion: primary cell- and ReporterGene potency assay results correlate in tendency,- statistical correlation is not meaningfull.
The RG-assay is MoA-reflective

In potency determination (analytics in general), we are always comparing

- Sample A
- Method A
- Analytical result
- **Method condition A**
  - **Ready-to-Use**
  - **Cell-line A**
  - Robot
  - …..
- Old Ref.Standard
- ….
- A

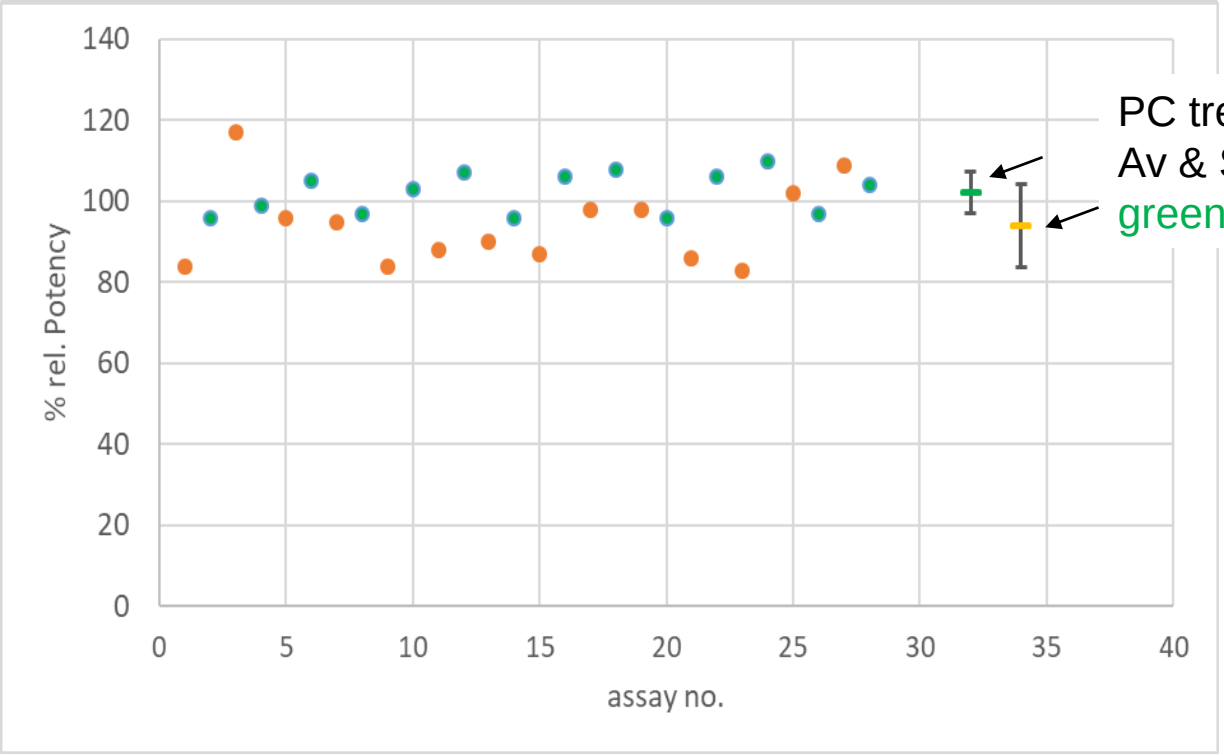
VS.

- Sample B
- Method B
- Specification
- **Method condition B**
  - **Cells in Culture**
  - **Cell-line B**
  - Manual
  - ….
- New Ref.Standard
- ….
- B

Comparing different assay *conditions*

The actual question: *can both conditions be used alternatively, i.e. are they «like-for-like»?*

Assay performed under condition A & B (hypothetical)



PC trending data:
Av & SD of conditions
green and orange

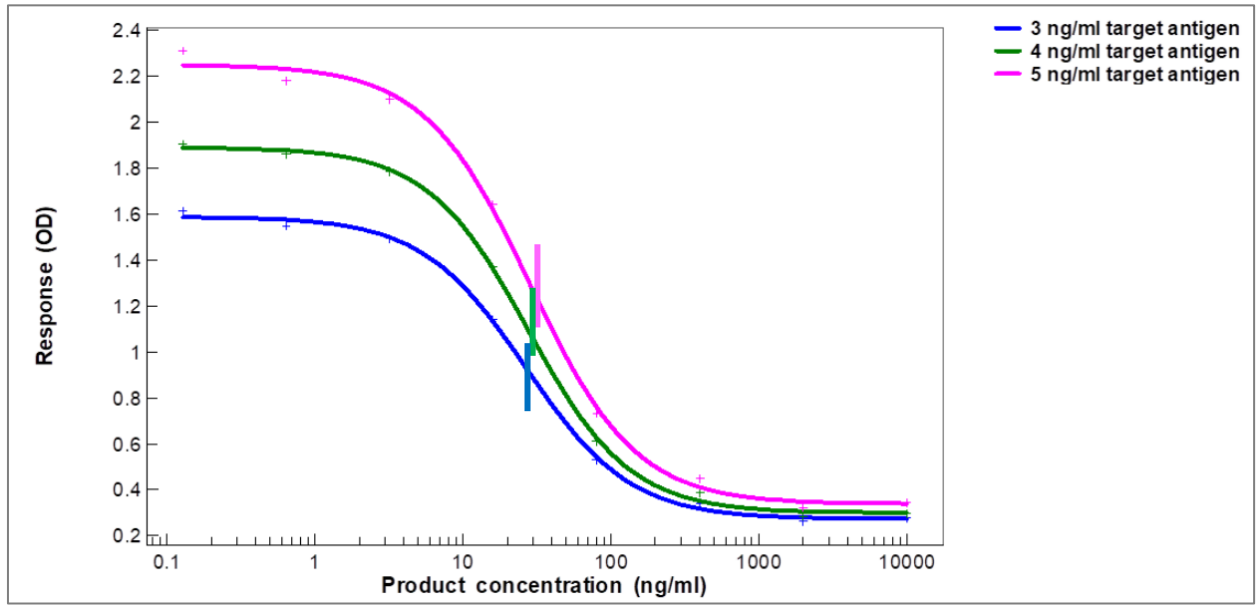
Is the difference between the two different operating conditions acceptable or not?

→ *Are they like-for-like or define 2 different assays?*

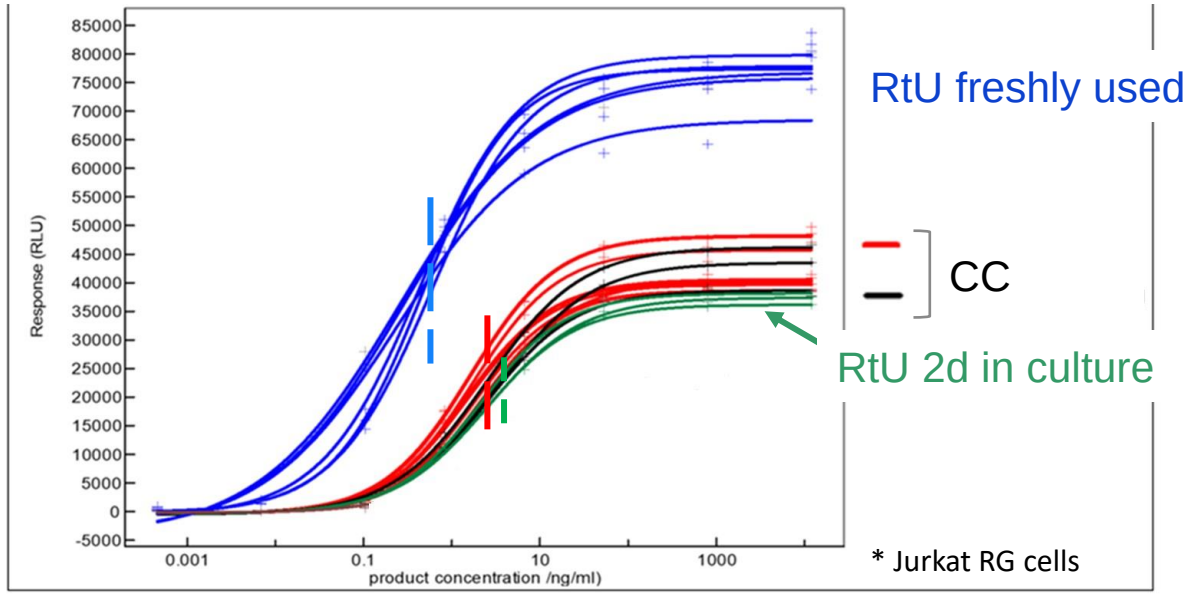
Comparing different assay *conditions*

Estimating potential criticality

ELISA: different target concentrations



Ready-to-Use (RtU) vs. Cells in Culture (CC) *

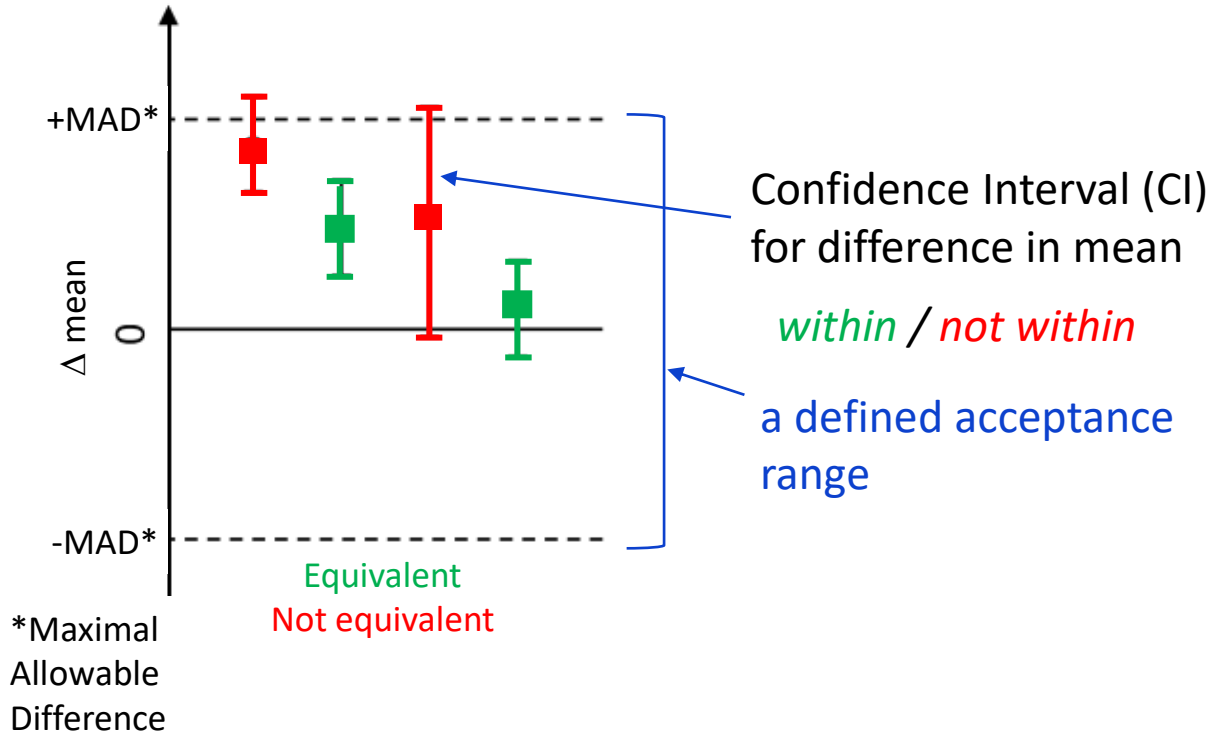


Difference in OD-range and signal discrimination only
→ can both be controlled by acceptance criteria
➔ No deeper comparative evaluation performed

Strong differences in DRCs indicate likelyhood of differences in accuracy and precision
➔ Statistical equivalence evaluation recommended

Evaluation of equivalence

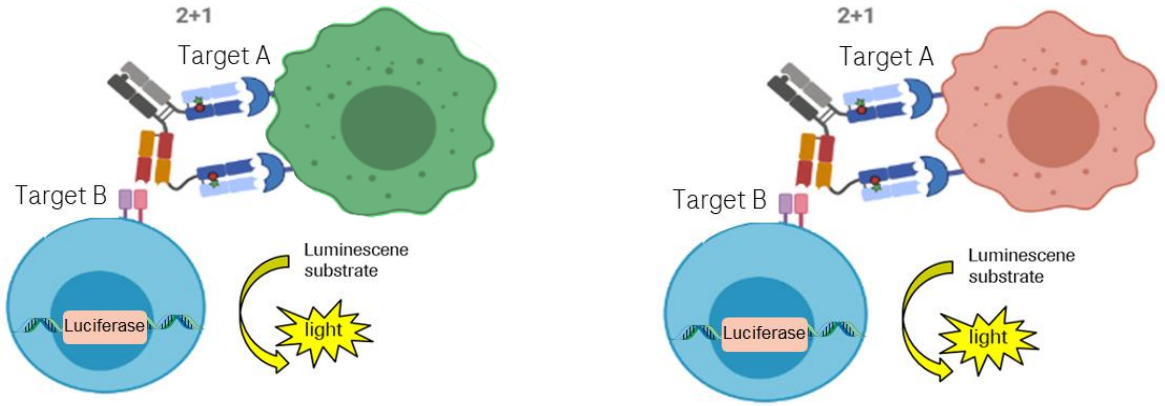
TOST = gold-standard for evaluation of equivalence. TOST assesses whether the mean difference between two groups and the corresponding CI lies within pre-defined MAD.*



Comparing different assay *conditions*

Example:

Target cell-line A *vs.* Target cell-line B

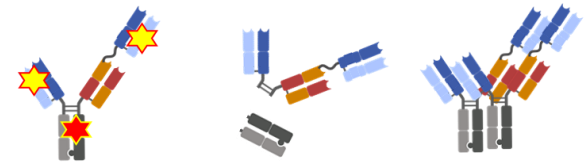


The actual question: *can both cell-lines be used alternatively in the assay*

Evaluation: *2-tiered approach*

1) do the 2 cells display «real samples» differently?

→ Analysis of «real» samples *

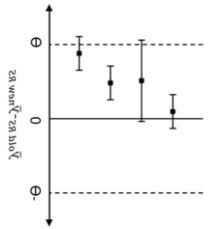


*Stab., CQA,…-samples

2) do the 2 cells deliver equivalent method capabilities?

→ Analysis of spiking samples*

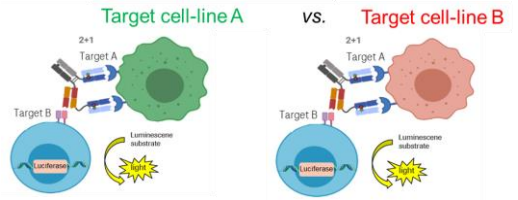
→ TOST & comparison of variabilities



*RS at different concentration levels

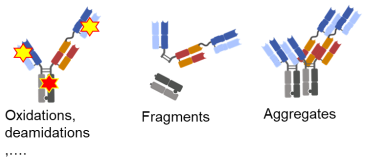
Comparing different assay *conditions*

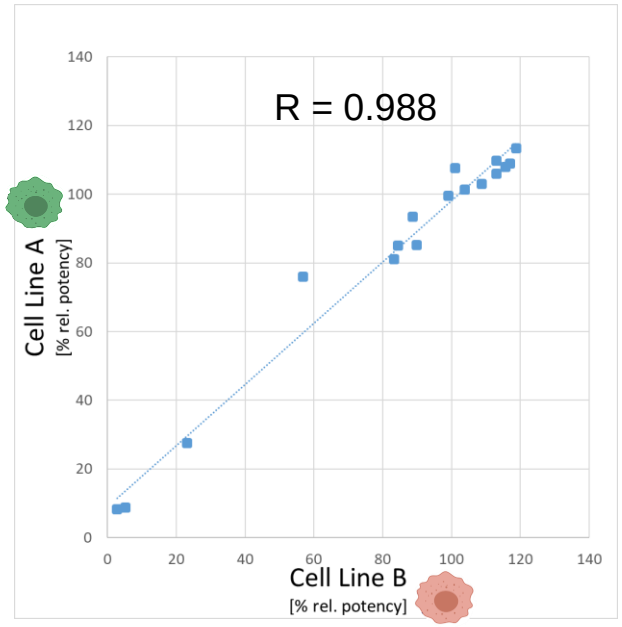
Example: **Target cell-line A vs. Target cell-line B**



2-tired approach

1) do the 2 cells display «real samples» differently?

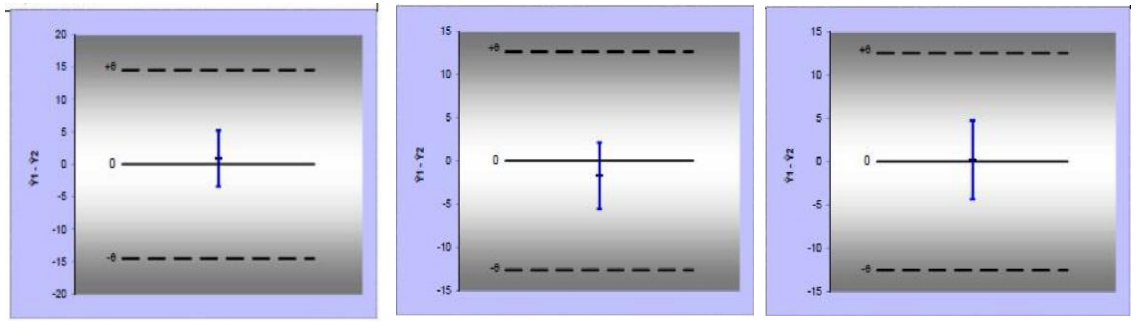




2) do the 2 cells deliver equivalent method capabilities?

→ Analysis of spiking samples* (2 x 18), TOST

50%-level 100%-level 150%-level



Spiking level	50%	100%	150%
Difference in RSD between cell-line A and B results	1.4%	0.9%	1.6%

Conclusion: assay performance is highly comparable/equivalent; both cells can be used in the assay like-for-like

In potency determination (analytics in general), we are always comparing

- Sample A
- Method A
- Analytical result
- Method condition A
- Ready-to-Use
- Cell-line A
- Robot
- …..
- **Old Ref.Standard**
- ….
- A

VS.

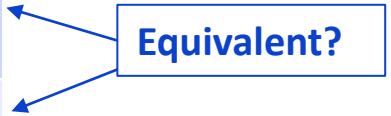
- Sample B
- Method B
- Specification
- Method condition B
- Cells in Culture
- Cell-line B
- Manual
- ….
- **New Ref.Standard**
- ….
- B

Comparing different *materials*

Example: 2 ReferenceStandard (RS) for the same product

Case-study
the «old» RS produced from technical, GLP-Tox-material to be superseded by the «new» RS produced from v0.2-material

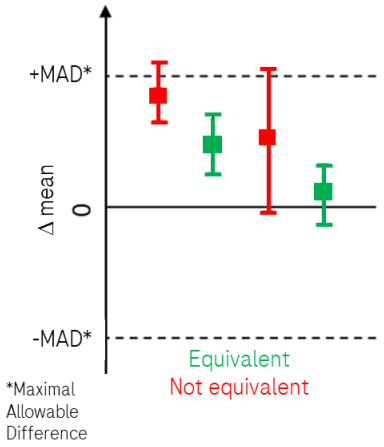
	Material/Process	Potency
«old» RS	GLP-Tox, technical material	100%, per definition* * as it was the 1st RS in the project
«new»RS* * 1st to be commercialized RS	v0.2-process	tbd, per qualification



Comparison of old and new ReferenceStandards for clinical phases

Statistical 2-step Approach

1) Test on equivalence



equivalence ⇨ old and new RS have same potency

⇩ non-equivalence

2) Determination of new RS potency

CI95 (n)

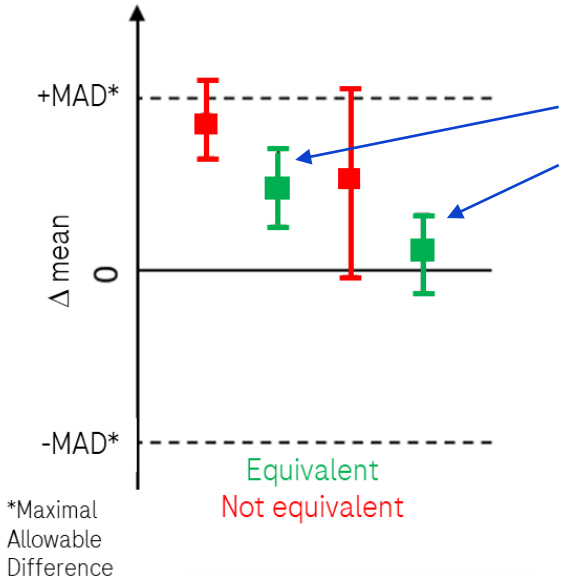
considering specification range & method variability

Example: ReferenceStandard (RS) for clinical phases vs. 1st to be commercialized RS

1) Test on equivalence of old and new RS

Approach

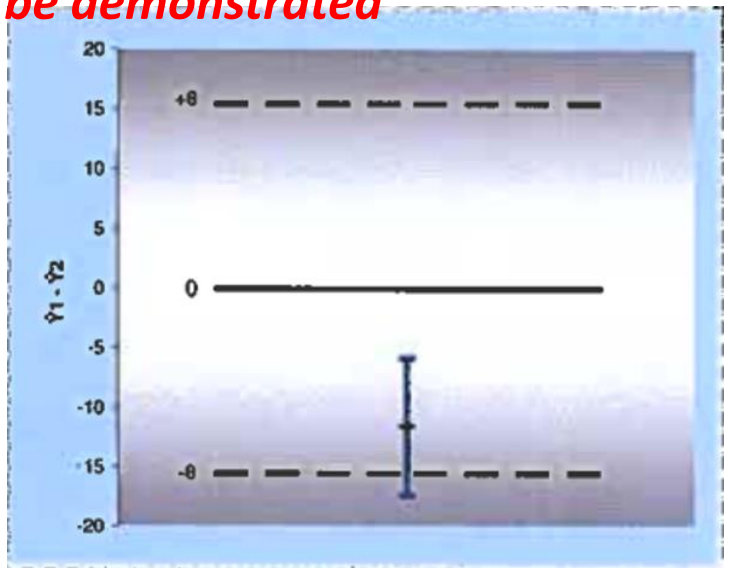
- Old and new RS were analyzed side-by-side, i.e. on same assay-plates ← *Number of required n (results) calculated in advance*
- Outlier test, CL 90%
- TOST ← *CI 90%, MAD = 2SD (from PC-trending)*



Difference in mean with corresponding 90% CI is within acceptance range
→ same potency value as old RS can be assigned to new RS

Case-study: the «old» RS produced from technical, GLP-Tox-material to be superseded by the «new» RS produced from v0.2-material

Result from TOST: e***quivalence between old and new RS could not be demonstrated***



Side-by-side testing, n = 2 x 21
TOST:
CI 90%, equiv.limit: 2SD (from PC-trending)

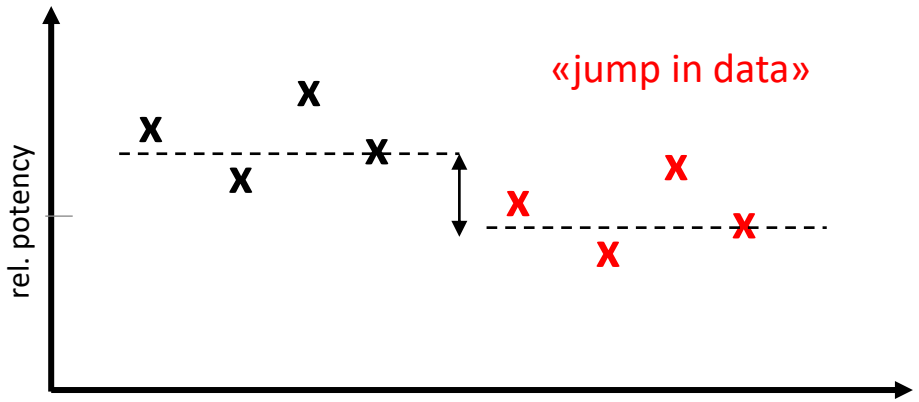
confidence interval is not completely contained within the range of MAD

Surprise! Physchem-data didn't let expect relevant differences in potency, in advance

Old and new RS with different potencies
Consequences

both RS have their own potencies → material analyzed against both RS has diffferent potencies,- relative to reference

Example: new RS has a higher potency than old RS



X: analyzed against «old» RS-potency
X analyzed agains «new» RS-potency

This may challenge stab,-data evaluation;
could indicate differences between samples where in reality there are none;
….

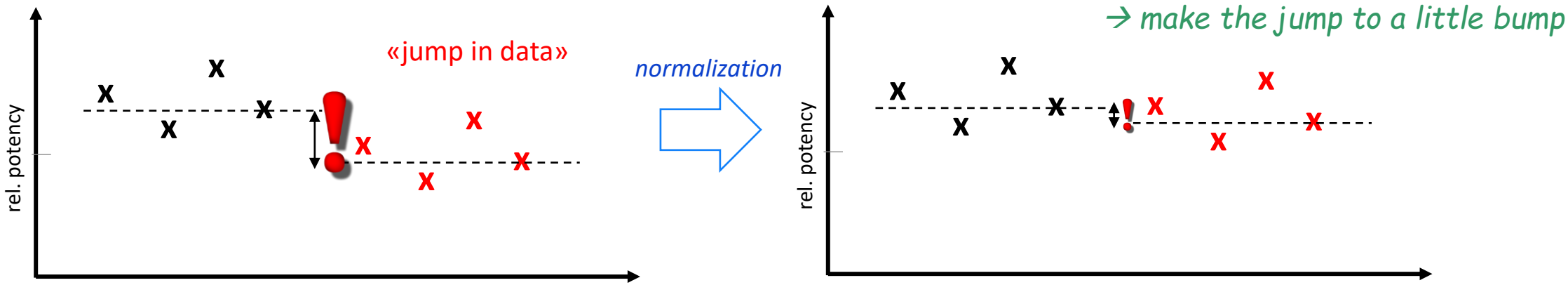
To avoid the «jump in (batch-release-, stab.-,…)data», both RS potencies have to be connected by a factor and data need to be normalized to one of the two.

To neutralize the «jump in data» as much as possible, the potency of the new RS should be known as good as possible .

Old and new RS with different potencies
Consequences

both RS have their own potencies → material analyzed against both RS has diffferent potencies,- relative to reference

Example: new RS has a higher potency than old RS



X: analyzed against «old» RS-potency
X analyzed agains «new» RS-potency

To avoid the «jump in (batch-release-, stab.-,…)data», both RS potencies have to be connected by a factor and data need to be normalized to one of the two.

To neutralize the «jump in data» as much as ~~possible~~ necessary, the potency of the new RS should be known ~~as good as possible~~ good-enough

Approach for ReferenceStandard (RS) for clinical phases vs. 1st to be commercialized RS

1) Test on equivalence of old and new RS

failed

2) Determination of the potency of the new RS

By a statistical approach to calculate the potency of the RS with a targeted accuracy

- Define a CI95 for the new RS potency considering the (anticipated) specification range and method variability
- **Calculate number of measurements n** that are required to reach the targeted CI95

CI95 (n) ~ expected value ± CI-factor x method precision

potency interval which contains the true RS potency value in 90 out of 100 cases

Case-study:
the «old» RS produced from technical, GLP-Tox-material to be superseded by the «new» RS produced from v0.2-material

Back to our example

The potency of the new RS was determined as average of 40 measurements

	Material/Process	Potency
«old» RS	GLP-Tox, technical material	100%, per definition*
«new» RS	v0.2-process	**108%**, per qualification

* as it was the 1st RS in the project

→ **many small differences**
in Σ LMWs, Σ HMWs, Acidic Peaks, Basic Peaks… **added-up to the** observed **difference in potencies between old and new material**

Discussed alternative approach
for 1st to be commercialized RS

Old and new RS were analyzed side-by-side, i.e. on same assay-plates, with fixed n *(20)*

Difference of potencies new vs old RS is

< 5% → *"similar enough"*
Old and new RS considered equi-potent

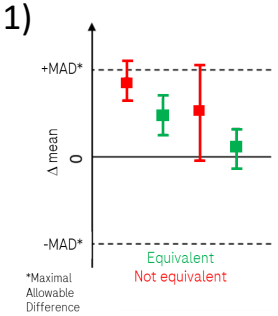
> 5% → *meaningful different*
Normalization factor calculated based on the n = 20

Comparison of old and new ReferenceStandards for clinical phases
Comparing the 2 approaches

Statistical Approach

- 2-step

1)



failed

2)

CI95 (n)

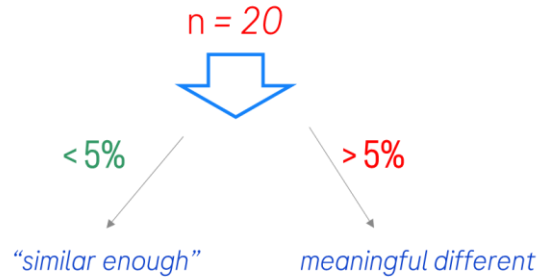
statistical equivalence testing

if required, the potency of the new RS is determined with a *high degree* of accuracy

- Case-dependent
- Consideres
  - specification range
  - method variability

Simplified Approach

- 1-step



the potency of the new RS is determined with accuracy considered *sufficient*

- Standardized (one-fits-all)
- Radical simplification
  → easy to apply and explain

Summary

- To compare A with B is our daily business
- ***Our approach:* how we compare depends on**
  - …on the type of objects to be compared
  - …on the question to be answered

  → defines the samples to be analyzed

  → defines if focus is more on science or statistics, or a combination of both
- The question of what is *similar enough*/what is a *relevant difference* is the key. And estimating this is the real difficulty.

  → *How much blur is acceptable? What can happen?*

Acknowledgements

Mentors, Peers & Contributors

- Adelheid Rohde
- Alex Briguet
- Alokta Chakrabarti
- Christel Richter
- Florian Cymer
- Johannes Dietrich
- Ngoc Nguyen
- Nico Bartel
- Philipp Metzger
- Silke Werz
- Till Koenig
- Yukina Umemoto

Doing now what patients need next