[PREPRINT] Exact methodology for (bio)assay validation based on product specification in line with USP <1033>

Davor Josipovic, Independent consultant, Belgium
(davor.josipovic@rovad.be)

Keywords: USP <1033>, ICH Q2(R2), analytical procedure validation, accuracy and precision, total analytical error, bioassay validation, assay validation, potency assay validation, analytical target profile, validation methodology

Abstract

Current validation methodologies -- whether based on accuracy and precision or total analytical error (TAE) and risk -- fall short when the analytical target profile (ATP) is defined directly in terms of product specifications, as is the case in the USP <1033> guideline for biological assay validation. In this paper, we critically examine the limitations of these existing methodologies and introduce a statistically rigorous and exact methodology tailored to the ATP formulation. While this novel methodology is demonstrated in the context of potency assay validation in line with the USP <1033> guideline, it is broadly applicable to other analytical procedures governed by ICH Q2(R2), with only minor adaptations. A freely accessible online application has also been developed to facilitate discussion and adoption of the novel methodology in practice.

Introduction

The principal guideline for validation of analytical procedures within the pharmaceutical industry is ICH Q2(R2) [1] which outlines two major methodologies for validation of analytical procedures: one based on accuracy and precision, and other based on total analytical error (TAE) and risk.

The USP <1033> [2] guideline adheres to these two general methodologies and applies them to bioassays, particularly the potency assays. Before these validation methodologies can be applied however, validation acceptance criteria need to be defined. USP <1033> [2] states that “when there is an existing product specification, acceptance criteria can be justified on the basis of the risk that measurements may fall outside of the product specification”. This requirement is usually formalized in the Analytical Target Profile (ATP)

of the procedure. Then the guideline proposes how to derive these acceptance criteria for accuracy and precision or TAE and risk, such that they comply with the ATP.

In this paper we show that both of these methodologies fall short in capturing the essence of the ATP when it is defined in terms of product specifications, and that to correctly validate the procedure a novel validation methodology is required.

The paper is structured as follows. In the first section the general assumptions are made explicit in a measurement model that serves as the foundation for later inference. In subsequent section the ATP is defined and formalized based on the measurement model. In the next section accuracy and precision methodology is assessed, showing how it fails to meet the ATP requirements. A similar conclusion is drawn in the section on TAE-based validation. Then the exact validation methodology is introduced, and it is shown how it fully satisfies the ATP requirements and how it resolves the shortcomings of the prior two methodologies. The final discussion section addresses important but peripheral topics not covered in the main body of the paper.

The assumptions

To properly address the research question, it is essential to begin with clear definitions and explicit assumptions. Central to this is the measurement model, which formally describes how the analytical procedure measures relative potency (RP).

The measurement model

A measurement model that is consistent with USP <1033> [2] definitions and formulas is the following:

$$X \coloneqq \mu_X \mathrm{BW} = \mu_T \frac{\mu_X}{\mu_T} \mathrm{BW} = \mu_T \tau_X \mathrm{BW} \tag{1}$$

where X is the measured value, μ_T the true (target) relative potency, $\tau_X = \mu_X/\mu_T$ the trueness factor (or multiplicative systematic error), and $B \sim \mathrm{logN}(0, \sigma_B^2)$ and $W \sim \mathrm{logN}(0, \sigma_W^2)$ the independent lognormal random variables with σ_B^2 the between and σ_W^2 the within run variability of the procedure.

Concretely, this means that every time one measures relative potency, the measured value X can be decomposed into the true potency value of the measurand, and the systematic bias and the random realizations of between and within variance components of the procedure. Note that the terms “measurement”, measured “value”, “measurand”, “true value” and “systematic error” are borrowed from GUM [3] in their original definitions.

Furthermore, knowing that the product of two independent lognormal random variables is lognormal, one can also deduce that:

$$X \sim \log\mathrm{N}(\ln\mu_T + \ln\tau_X, \sigma_{IP}^2)$$

where for convenience $\sigma_{IP}^2 \coloneqq \sigma_B^2 + \sigma_W^2$.

Lastly, for plotting purposes, X is often represented as relative error (%) through the following transformation:

$$(X/\mu_T - 1) \cdot 100$$

Measuring production samples

When the potency of a manufactured product is measured with this procedure, then the true value (μ_T) in the measurement model is the true (albeit unknown) relative potency of that manufactured product. Assuming the production process produces products with $\mu_P P \sim \log\mathrm{N}(\ln\mu_P, \sigma_P^2)$ relative potency, we can substitute μ_T with $\mu_p P$ in Eq. 1. Our measured value X becomes then the measured value of one realization of this production process:

$$M \coloneqq \mu_P \tau_X PBW \sim \log\mathrm{N}(\ln\mu_P + \ln\tau_X, \sigma_{IP}^2 + \sigma_P^2) \qquad (2)$$

Now $\mathrm{Var}[\ln M] = \sigma_M^2 \coloneqq \sigma_{IP}^2 + \sigma_P^2$, which means that the observed variability of the logarithm of our measured potency value is composed of the production process variability and analytical procedure variability, just as stated in USP <1033> [2], but now deduced formally from a measurement model which can be used for inference.

The measured value refers to the output of a single implementation of the procedure on a test article, ideally from one run and one replicate. In contrast, the reportable value (RV) [4] extends this concept by being a function of a set of measured values, hence $RV \coloneqq f(\{M_{nk}\})$ where n is the run and k the replicate index. The function f is usually -- but not necessarily -- the geometric average.

Global versus local performance parameters

Lastly make note of an important distinction between the parameters τ_X and σ_{IP}^2, and $\tau_X(\mu_T)$ and $\sigma_{IP}^2(\mu_T)$ respectively. The former are called global, and the later are called local because they depend on (i.e. are a function of) the true value. Local parameters are more general and can be freely substituted in place of the global ones in the measurement model above. The distinction between the two is important because during validation these parameters are estimated in function of the true value (and thus *depend* on the true value), while during acceptance criteria determination in the USP <1033> [2] guideline,

global parameters are used. This simplification results in the “curse of global limits" and will be discussed in due time.

Analytical Target Profile (ATP)

Before initiating validation, an Analytical Target Profile (ATP) must be defined. According to ICH Q14 [5] the ATP is a statement that defines “the performance characteristics describing the intended purpose and the anticipated performance criteria of an analytical measurement".

In USP <1033> and its latest draft [2,6] the ATP is defined based on product (release) specification (e.g. [0.70; 1.43] RP), an assumed production geometric variability ($\sigma_P = 0.048$ RP) and the requirement that whenever the relative potency of a manufactured product is reported (i.e. $f(\{M_{nk}\})$), the probability that the reportable value is out of product specification is not more than 1 %. Our use of these concrete values is for the sole purpose of making the exposition more comprehensive and illustrative (cf. the discussion section). In practice the % can be much lower, because falling out of specification means facing hard time justifying the result, potentially resulting in incidents and CAPAs -- even when the underlying product is compliant. While production variation cannot be directly controlled, one can define validation criteria for the analytical procedure to ensure failures due to measurement variability remain acceptably low. Hence the above requirement results in the following ATP example where for the sake of simplicity we base it on the most elementary form of the reportable value:

[ATP] The procedure must be able to quantify relative potency (RP) in a range from 0.5 to 2.0 RP such that, under an assumed lognormal manufacturing distribution geometrically centered on 1.0 RP with a geometric standard deviation of 0.048 RP, the expected probability of reportable values (one run, one replicate) falling outside [0.70; 1.43] RP is less than 1%.

The general form of this ATP can be formalized as:

$$\mathrm{P}(LSL < f(\{M_{nk}\}) < USL) \geq 1 - \omega \tag{3}$$

where LSL and USL denote the lower and upper product specification limits, ω the maximum allowable risk of measuring outside these limits and $f(\{M_{nk}\})$ the reportable value. The ATP example can then be formally stated as:

$$\mathrm{P}(0.70 < M < 1.43) \geq 1 - 0.01 \tag{4}$$

or in layman's terms: The probability to measure manufactured products (M) outside of product specification should not exceed 1 %.

Accuracy and precision validation

The accuracy and precision definitions are taken from USP <1033> [2] and rewritten in function of the parameters of the measurement model (Eq. 1). For accuracy this results in:

$$RB \coloneqq \tau_X - 1$$

And for precision:

$$IP_{GCV} \coloneqq e^{\sigma_{IP}} - 1$$

Note that to compute the RB, one needs to assume a true or known relative potency value (μ_T, cf. Eq. 1) which USP <1033> calls the “target potency” [2] or “expected potency” [6]. If the true value is not (assumed) known, then there is no way to estimate RB. USP <1033> [2] suggests that for validation purposes target potencies can be constructed “by dilution of the standard material or [based on] a test sample with known potency”. For IP USP <1033> [2] prefers to use the Kirkwood’s [7] GCV formula instead of the RSD. Preference of one formula over the other is a matter of taste in this context as they are both transformations of σ_{IP}^2 and used merely for reporting but carrying no deeper meaning or inferential value.

The classical way to validate procedures is by setting global acceptance limits on accuracy (RB) and precision (IP). To determine them, original USP <1033> [2] uses the process capability index, more specifically the $\hat{C}_{pm}$ variant, which is sometimes called the Taguchi capability index. The latest USP <1033> [6] draft uses an exact solution. It is called “exact” in this paper because the solution is derived directly from Eq. 4 after substitution of process parameters from the ATP. Both methods result in solutions as global (RB, IP) pairs that satisfy the ATP. These solutions are depicted in Fig. 1 as “Cpm” and “Exact”, with only distinction being that “Cpm" is more conservative, imposing unnecessarily strict limits on RB.

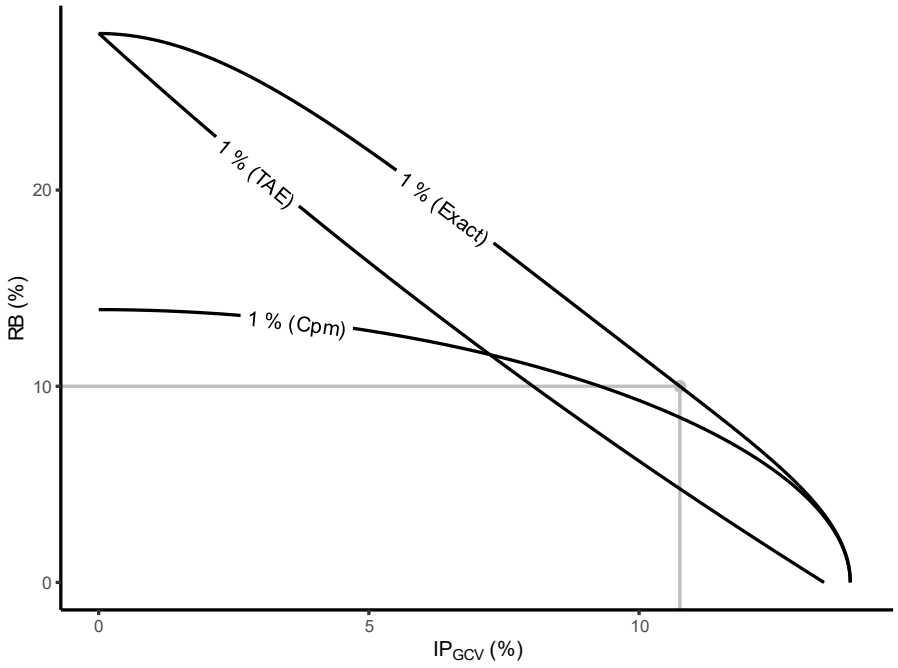

Fig. 1: Procedure performance depicted in function of global RB and IP. Curves show the global (RB, IP) pairs that correspond to procedure performance that would result in 1 % out-of-specification measurements under the ATP requirements, computed using $\hat{C}_{pm}$, TAE and the Exact way. The grey rectangle covers all procedures having $RB \leq 10$ % and $IP_{GCV} \leq 10.75$ %.

It is important to note that the ATP doesn't correspond to a single global (RB, IP) limit: rather a multitude of limits are possible. Concretely this means that *any* procedure with a set of global (RB, IP) parameters that is located below the "Exact" curve, satisfies the ATP. Unfortunately, in practice only one global (RB, IP) pair is taken from the "Exact" (or even worse the "Cpm") curve and is then used as the acceptance limit for the procedure during validation. For example, $\mathrm{RB} \leq 10$ % and $\mathrm{IP_{GCV}} \leq 10.75$ %. This pair is depicted by the grey dot in Fig. 1 and all procedures falling within the grey rectangle are deemed acceptable, even though many valid procedures just outside this rectangle (but still below the "Exact" curve) would also satisfy the ATP, but in practice would be rejected.

A further limitation of this validation methodology is what we call the "curse of global limits". The acceptance limits are imposed globally across the full working range of the procedure. Yet, as shown in Table I, a procedure's actual performance is highly dependent on the true value (μ_T) that is being measured, meaning the procedure's characteristics are local, not global. Imposing global limits can cause issues: For example, the ATP clearly defines the assumed production process under which in our given example it is extremely unlikely to produce products close to 2 RP. Yet by setting global (RB, IP) acceptance limits one is imposing equal quality standards on the procedure's performance when measuring at 1 RP as at 2 RP true value. This excessively strict requirement is not warranted by the

ATP and will be solved in the "exact validation" section where the novel validation methodology is presented.

Table I: The locality of the procedure's performance. At each "local" level, the performance of the procedure is expressed in terms of RB and IP. The values are taken from the USP <1033> [2] example.

Level (RP)	$\boldsymbol{RB}$ (%)	$\boldsymbol{IP_{GCV}}$ (%)
0.50	3.23	6.81
0.71	0.06	7.29
1.00	4.97	8.46
1.41	2.91	6.28
2.00	9.72	7.24

Total analytical error (TAE) validation

In the previous section on accuracy and precision we established that the Analytical Target Profile (ATP) cannot be satisfied by a single global (RB, IP) acceptance limit. This naturally leads us to consider whether the Total Analytical Error (TAE) methodology fares any better.

At its core, the TAE methodology aims to define acceptance limits around the true measured value such that the procedure's measurements fall within these bounds with a high probability. Formally, this is expressed as:

$$\mathrm{P}(LAL < X < UAL) \geq 1 - \rho \tag{5}$$

where LAL and UAL represent the lower and upper acceptance limits for the procedure's total analytical error (hence the acronym TAE), and ρ the maximal risk of observing a result outside this range.

However, even in this initial form, the problem is ill-posed, as Eq. 5 allows for infinitely many valid (LAL, UAL, ρ) solutions. To arrive at a unique solution, one typically imposes two additional constraints:

1. $\ln LAL$ and $\ln UAL$ are forced to be symmetric around $\ln \mu_T$. TAE acceptance limit (TAL) becomes then: $TAL \coloneqq [\exp(\ln \mu_T \pm \ln AL)]$.
2. Fixing ρ to an arbitrary but intelligible value (i.e. 5 %).

For the reader concerned about "arbitrary" ρ, note that ρ and AL influence each other: larger ρ implies a narrower AL and vice versa, but the pair (AL, ρ), whatever the initially chosen ρ may be, is a unique solution under Eq. 5.

Next, consider the scenario where the analytical procedure is used to measure manufactured products. In this case, the true value μ_T in X becomes the random variable $\mu_P P$, which once substituted in Eq. 5 results in the same form as Eq. 3. This shows that if

the production process is perfect (i.e. if $\mu_P P$ always resolves to a fixed RP and hence has no variability), then TAL and risk (ρ) have the same meaning as respectively product specification limits (L/USL) and probability to measure outside of specification (ω), and can be interchanged. This is the approach taken in the latest USP <1033> draft [6], which is a gross oversimplification as in practice a production process always has some variability and a translation from product specifications to procedure acceptance limits is required. The translation can be formally stated as follows:

$$\underset{\ln AL}{\operatorname{argmax}}\, \mathrm{P}(|\ln X - \ln \mu_T| < \ln AL) \geq 1 - 0.05 \Rightarrow \mathrm{P}(0.70 < M < 1.43) \geq 1 - 0.01$$

which in layman's terms says that one is seeking the widest possible TAL for the procedure's total error, such that it guarantees ATP compliance when applied to manufactured products.

The resulting unique solution is $\ln AL = 0.247\ \ln \mathrm{RP}$ and $\rho = 5\ \%$. This solution is depicted as the "TAE" curve using the corresponding (RB, IP) pairs in Fig. 1. The "TAE" curve shows that any procedure that has a total error within this AL interval at 5 % risk necessarily also complies to the ATP (i.e. the "TAE" curve is always below or at the "Exact" curve). Showing that the procedure is within the TAL during validation then often results in making an Analytical Error Profile plot as depicted in Fig. 2 from which one can see that the expected 95 % total error interval of the procedure (purple lines) is well within the TAL (brown lines).

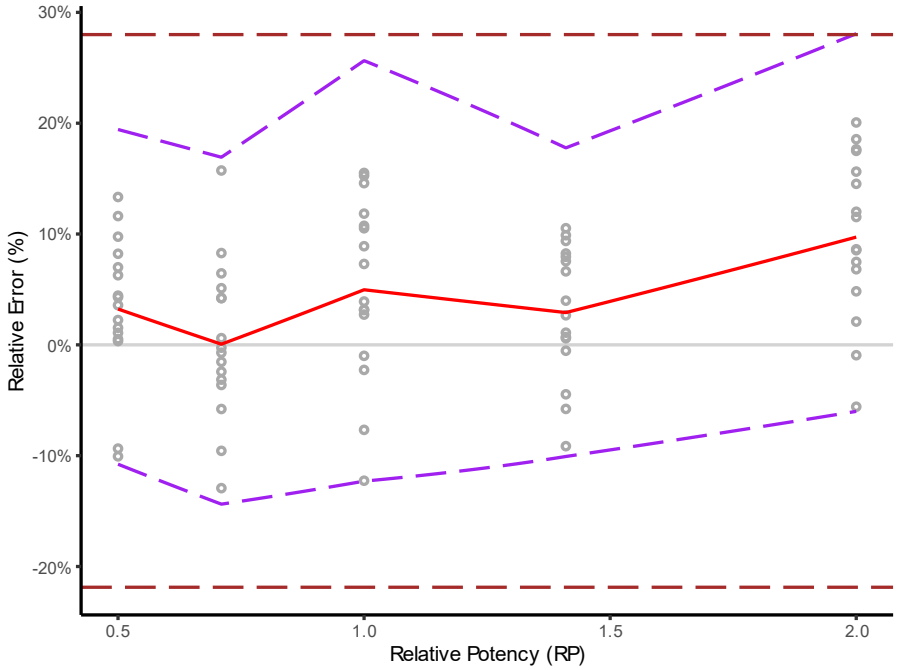

Fig. 2: Total Analytical Error profile of the procedure. The experimental measurements (grey circles) are plotted as relative error (%) in function of true relative potency (RP). The red curve is the expected relative bias (RB) and the purple dotted

interval is the expected 95 % total error interval of the procedure's measured values X. The brown dotted horizontal lines are the TAE acceptance limits (TAL).

Compared to accuracy and precision validation methodology, the TAE methodology is more inclusive: It validates any procedure below the "TAE" curve shown in Fig. 1, which already covers more area than the rectangular regions defined by fixed (RB, IP) limits from the "Exact" or "Cpm" curves. Yet, limitations remain. The "TAE" methodology still excludes many valid procedures that are in between the "TAE" and the "Exact" curve hence does not fully comply to the ATP requirement. What is more is that the TAE methodology is also affected by the same "curse of global limits" which is attributed to the accuracy and precision methodology. That is, a single global TAL is applied enforcing equal quality standards across the entire working range.

Exact validation

In previous sections we showed that both accuracy and precision, and TAE and risk methodologies are inadequate to correctly capture the ATP requirements. Both methodologies are susceptible to rejecting valid methods and are affected by the "curse of global limits". In this section we propose a methodology for validation that doesn't try to translate the ATP but uses it directly to validate the procedure. This direct approach avoids the pitfalls of earlier methods and ensures full alignment with local procedure's performance and the ATP by design.

At the core of the methodology is a direct evaluation of the procedure's performance with respect to the ATP, as formalized in Eq. 3. The locality of procedure's performance is modelled with functions $\tau_X(\mu_T)$ and $\sigma_{IP}^2(\mu_T)$ which are formed by linear interpolation of the estimated parameters shown in Table I. Other interpolation techniques can also be used. The final solution for the probability to fall out of specification is based on numerical optimization that results in highly accurate estimates of at least 7 significant digits.

While the computation of the exact solution is relatively straightforward, presenting the results in an accessible and actionable way is more challenging. The visualization shown in Fig. 3 represents a practical compromise between complexity and intelligibility, and effectively communicates how the procedure performs relative to the ATP.

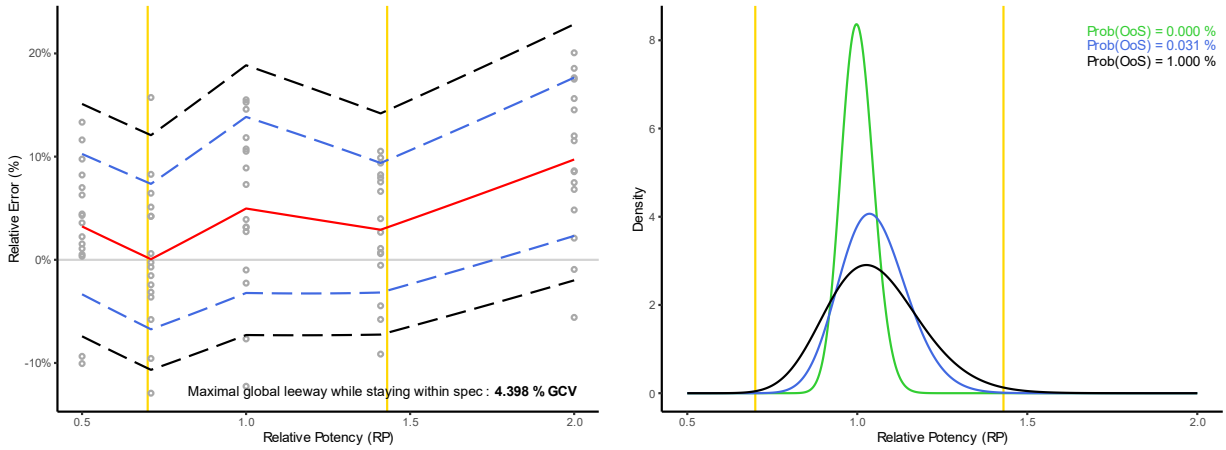

Fig. 3: On the left plot the experimental measurements (grey circles) are plotted as relative error (%) in function of true relative potency (RP). The red curve is the expected relative bias (RB) and the blue dotted interval is the added intermediate precision (IP). (These estimates are taken from Table I) One can state roughly that the blue dotted interval covers about 68 % of the measurements. The density curves in the right plot reflect the performance in routine. The green density represents the RP of the production process. The blue density tells us what will happen when we measure these products with our procedure (i.e. based on the procedure's performance summarized with the blue dotted lines). One can see that the measurements would remain well within the boundaries of the product specification (i.e. the yellow bars). The black density has exactly 99 % of its area within the yellow bars, which then translates to the black dotted acceptance interval of the procedure as the maximal addition of global IP to the current performance (blue dotted lines) while still meeting the ATP requirements. Hence the difference between black and blue dotted lines can be interpreted as the maximal global IP that the procedure can incur while still remaining within the ATP specification. The results shown have been validated by simulation for correctness.

The novelties of this exact validation methodology are:

1. *Direct validation against the ATP*: The methodology avoids translation of the ATP into auxiliary acceptance criteria, eliminating the risk of false rejections due to translational errors. The methodology is exact in the sense that it is validating directly against the ATP.
2. *Integration of process knowledge*: The methodology considers performance of the procedure less important in regions where the production process is less likely to produce products (and vice versa). This is consistent with the (assumed) knowledge about the production process embedded in the ATP.
3. *Recognition of local procedure behavior*: Unlike with global acceptance limits, this methodology acknowledges that a procedure may perform well in some parts of the range and less well in others. This local flexibility allows for compensation: strong performance in critical regions can offset weaker performance elsewhere, provided the overall probability of falling out of specification remains within the ATP bounds.

To make this methodology more tangible, Fig. 3 is made interactive in an online demo application [8]. Users can adjust key components -- such as the product specifications, the manufacturing distribution, or the procedure's performance characteristics (globally or

locally) -- and receive immediate feedback on the procedure's suitability in routine use. This makes it possible to explore trade-offs and optimize the procedure in terms of development effort versus routine impact.

Discussion

While the focus of this paper's ATP example has been on validating procedures based on the measured value (i.e. one run and one replicate), extending it to a more complex reportable value and checking its impact based on different formats is straightforward.

The production process in the ATP has been (assumed) known and -- for the sake of simplicity -- set to $\log\mathrm{N}(\mu_P = 1, \sigma_P^2 = 0.048^2)$ as calculated in USP <1033> draft [6] example. But within this novel validation framework it could be defined in various ways. For example, the production process variability could be defined as a "proportion of the overall manufacturing variance" [6] (i.e. as proportion of $\mathrm{Var}[\ln M]$). Note also that in the same USP <1033> draft [6] the production process variability is derived from the reportable value which is based on an (arbitrary) chosen (RB, IP) pair from the "Exact" curve. This is a dangerous practice as production process variability should not depend on an arbitrary chosen IP nor on the format of the reportable value, but rather on the σ_{IP}^2 estimate from the procedure, if anything.

The estimates in Fig. 3 are based on point-estimates of procedure's performance parameters and don't account for uncertainty due to the limited number of samples in the validation design. To account for this, the ATP could equally well be defined in terms of tolerance limits, e.g. to have 95 % confidence in the validation result.

We deliberately excluded certain aspects of validation, such as dilutional linearity and range determination, which are adequately addressed in USP <1033> [2]. Their treatment remains unchanged under the proposed validation approach.

We also haven't assessed the procedure's ability to detect defective (i.e., truly out-of-specification) manufactured products simply because this is not within the scope of the validation process as defined in USP <1033> [2]. Nonetheless, evaluating this capability is important from a quality control and risk management perspective. The exact validation methodology presented in this paper enables such an assessment by leveraging the production process knowledge embedded in the ATP as well as the estimated procedure performance. Specifically, one can calculate the probability of correctly identifying products that are truly out of specification, providing quantitative insight into the diagnostic sensitivity of the procedure. This capability has broader implications: these detection probabilities can inform process control strategies, support risk-based decision-

making, and align with broader quality-by-design principles advocated in ICH guidelines. As such, we advocate for incorporating this form of analysis alongside traditional validation elements to provide a more complete picture of a procedure's fitness for purpose.

Conclusion

This paper has demonstrated that validation methodologies based on accuracy and precision, and TAE and risk, fall short when the ATP is defined directly in terms of product specifications, as recommended by USP <1033> [2,6]. Rather than attempting to translate such specifications into surrogate metrics like accuracy, precision and TAE, a more robust approach is to align the validation methodology directly with the ATP itself. To this end, we have proposed a novel, statistically exact validation methodology that avoids the limitations of traditional approaches and ensures full compliance with the ATP. The methodology is also made freely available through an interactive online application at https://apps.rovad.be/usp-1033/. We hope this contribution will stimulate further dialogue on modernizing analytical procedure validation in line with current regulatory expectations.

References

1. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) ICH Q2(R2) guideline on validation of analytical procedures. 2023. https://www.ich.org/page/quality-guidelines
2. United States Pharmacopeia (USP). General Chapter <1033> Biological assay validation. 2010. https://doi.org/10.31003/USPNF_M912_01_01
3. International Organization for Standardization (ISO). Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) (ISO/IEC Guide 98-3). 2008. https://www.iso.org/standard/50461.html
4. Schofield T, van den Heuvel E, Weitzel J, Lansky D, Borman P. Distinguishing the analytical method from the analytical procedure to support the USP analytical procedure life cycle paradigm. Pharmacopeial Forum. 2019;45.
5. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH Q14 guideline on analytical procedure development. 2023. https://www.ich.org/page/quality-guidelines
6. United States Pharmacopeia (USP) Draft document: General Chapter <1033> Biological assay validation. 2024 Sep 30. https://doi.org/10.31003/USPNF_M912_102010201_01

7. Kirkwood TBL. Geometric means and measures of dispersion. Biometrics. 1979;35(4):908-9. http://www.jstor.org/stable/2530139
8. Josipovic D. Online tool for bioassay validation based on product specification. 2025. https://apps.rovad.be/usp-1033/

Declarations

Conflict of Interest: The author declares no conflict of interest.