Statistical opportunities related to ICH Q2(R2) and Q14

TIM SCHOFIELD

IABS 10TH ANNUAL STATISTICS WORKSHOP

ROCKVILLE, MD, NOVEMBER 14, 2024

Outline

- Uses of intervals
- Validation design considerations
- Reportable result replication strategy
- Combining accuracy and precision
- Additional considerations
- References

Uses of intervals intervals
Q2(R2) – recommended data

- ***Accuracy*** *should be reported as the mean percent recovery of a known added amount of analyte in the sample or as the difference between the mean and the accepted true value, together with an appropriate* ***100(1-α) percent confidence interval*** *(or justified alternative statistical interval).* ***The observed interval should be compatible with the corresponding accuracy acceptance criteria****, unless otherwise justified.* [1]
- ***The standard deviation, relative standard deviation*** *(coefficient of variation), and an appropriate* ***100(1-α) percent confidence interval*** *(or justified alternative statistical interval) should be reported.* ***The observed interval should be compatible with the corresponding precision acceptance criteria****, unless otherwise justified.* [1]
- ***Combined accuracy and precision*** *can be evaluated by use of a* ***prediction interval, a tolerance interval, or a confidence interval****.* [1]

Uses of intervals
Q14 – bioassay example

- Framed as an **analytical target profile** (ATP)

Performance Characteristics	*Acceptance Criteria*
Accuracy	*The **95% confidence interval of the slope** of the fitted regression line between theoretical and measured potency **falls within a range of 0.8 to 1.25*** *The **upper and lower 90% confidence interval** for the relative bias calculated **at each potency level** is **not more than 20%***
Precision	*Upper **95% confidence interval for the average intermediate precision across levels** across the reportable range (95% CI % geometric coefficient of variation) is **not more than 20%***

Linearity:
Should be 90% CI to conform to the *equivalence* requirement in USP <1033>

Should be 0.80 to 1.20 to conform to USP <1033>

Trueness:
Should allow across levels

Should be -20% to 25% to conform with USP <1033>

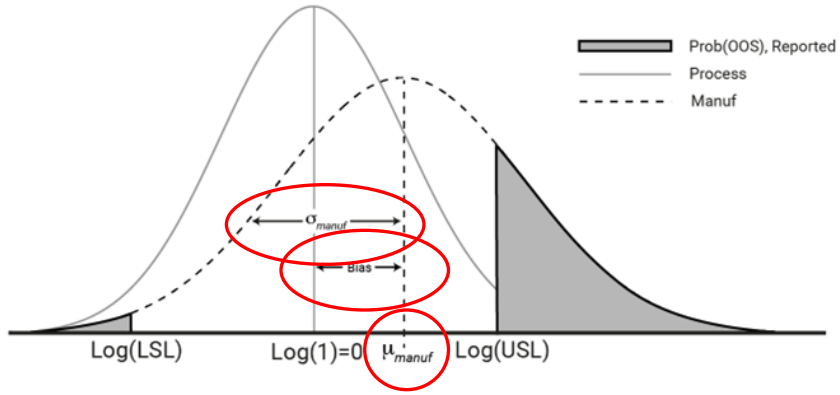
Intermediate Precision.:
Noninferiority requirement

Statistical opportunities to support use of confidence intervals

- The uses of confidence intervals to address ***equivalence and noninferiority*** require a focus on ***acceptance criteria*** and ***study design***
- This is in conflict with ICH recommendations
  - ***Study design***:
    - ***Recommended Data***: *Accuracy should be assessed using an appropriate number of determinations and concentration levels covering the reportable range (e.g., **3 concentrations/3 replicates** each of the full analytical procedure).* [1]
    - Same recommendation for repeatability; silent on intermediate precision
      - 3-concentrations is insufficient to assess linearity
      - 3-replicates of the procedure may not support confidence intervals

Statistical opportunities to support use of confidence intervals (cont.)

- ***Acceptance criteria***:
  - The Q14 bioassay example uses "industry standards" for acceptance criteria
    - ***Linearity***: confidence interval on slope within ***0.8 to 1.2*** (not 1.25)
    - **Relative accuracy**: confidence interval on relative bias ***<20%***
    - ***Precision***: upper confidence interval on %GCV ***<20%***
  - USP <1033> uses ***Prob(OOS)*** to establish acceptance criteria on relative bias and intermediate precision
  - ***Cpm*** dropped from consideration



- $\boldsymbol{\mu_{Manuf} = Bias}$ when spec centered on 0
- $\boldsymbol{\sigma^2_{Manuf} = \sigma^2_{Process} + \sigma^2_{RR}}$ with a restriction on percentage due to σ^2_{RR}

$$Prob(OOS) = \Phi\left(\frac{\log(LSL) - \mu_{Manuf}}{\sqrt{\sigma^2_{Manuf}}}\right) + \left[1 - \Phi\left(\frac{\log(USL) - \mu_{Manuf}}{\sqrt{\sigma^2_{Manuf}}}\right)\right]$$

Validation design considerations

- *The experimental design of the validation study should reflect the **number of replicates used in routine analysis** to generate a reportable result. **If justified**, it may be acceptable to perform some **validation tests using a different number of replicates** …* [1]
- *It is acknowledged that information about multiple performance characteristics may be derived **from the same dataset**.*
  - Versus ***separate repeatability*** and ***intermediate precision*** studies
- This acknowledges that you can validate the ***method*** rather than the ***procedure*** [3]
  - ***Method*** = performance of a run to get a ***measurement***
  - ***Procedure*** = replication strategy to get a ***reportable result***
- ***Justification:*** to ***support confidence intervals*** and to incorporate ***assay factors rather than replicates***

Validation design considerations (cont.)

- At least ***5 concentration/potency levels*** – to assess linearity
- Integrating ***ruggedness factors*** (e.g., using factorial design)
- ***Sample size***:
  - ***Accuracy*** (using an equivalence approach): smallest n such that $n \geq \frac{\left(t_{\alpha:n-1}+t_{\beta/2:n-1}\right)^2 \cdot \sigma^2}{Bias^2}$, where σ is the expected precision and $Bias$ is related to the ATP requirement on relative accuracy
  - ***Precision*** (using noninferiority approach): smallest n to satisfy $\frac{\sigma}{TMU} \leq \sqrt{\frac{\chi^2_{\alpha:n-1}}{\chi^2_{1-\beta:n-1}}}$, where σ is the expected precision and TMU is the target measurement uncertainty (ATP requirement on precision)

Reportable result replication strategy

- *… or to **adjust the number of replicates in the analytical procedure** based on data generated during validation.* [1]
  - Combinations of number of runs (n) and number of replicates within runs (k)

Combinations (Requirement σ < 5.0%)

Reps (k)	Runs (n)			
	1	2	3	6
1	7.2%	5.1%	**4.1%**	2.9%
2	6.4%	**4.5%**	3.6%	2.6%
3	6.0%	4.2%	3.4%	2.4%
6	5.7%	4.0%	3.3%	2.3%

Conf. Bound (Requirement σ < 5.0%) [5]

Reps (k)	Runs (n)			
	1	2	3	6
1	11.8%	8.2%	6.7%	**4.7%**
2	10.7%	7.5%	6.0%	4.2%
3	10.3%	7.2%	5.8%	4.1%
6	9.9%	6.9%	5.6%	3.9%

  - Note 1: Lab might operate with n = 6 runs until there's enough information to reduce this
  - Note 2: Groups of successive runs are typically correlated, and effectively replicates

Combining accuracy and precision

- ***Combined Approaches for Accuracy and Precision****: An alternative to separate evaluation of accuracy and precision is to consider their total impact by assessing against a combined performance criterion.* [1]

$$\Pr(-\lambda < Y - \tau < \lambda) \geq P, or \text{ [6]}$$
$$\Pr(-\lambda + \tau < Y < \lambda + \tau) \geq P$$

τ = *reference value* (e.g. 1000 mg/gm), λ = *acceptable limit* (e.g., 2%), yielding the range 1000·(0.98) = ***980*** to 1000·(1.02) = ***1020***

- Two interpretations:
  - ***Probability*** that the next reportable result falls into the range $(-\lambda + \tau)$ to $(\lambda + \tau)$ is $\geq P -$ ***prediction interval***
  - **Proportion** of all future reportable results falling into the range $(-\lambda + \tau)$ to $(\lambda + \tau)$ is $\geq P -$ ***tolerance interval***

Combining accuracy and precision (cont.)

- ***Advantages*** (relative to individual parameters):
  - Validation success is not dependent upon the choice of selected acceptance criteria on accuracy and precision
  - Can be used as the primary basis of design and analysis, or to resolve a validation failure
- ***Disadvantages***:
  - Design is not as straightforward as for individual parameters
    - Accuracy by equivalence
    - Precision by noninferiority
  - The study size can be large if the study goal is to ensure the proportion of passing results (by tolerance interval) versus the probability of the next result passing (by prediction interval)

Additional considerations

- *Platforms*
  - *An analytical procedure that is suitable to test quality attributes of different products without significant change to its operational conditions, system suitability, and reporting structure*. [1,2]
  - It may be required to "verify" that the method performs as expected
  - Opportunity for ***Bayesian analysis***
- *Product variability*
  - Manufacturing variability is the sum of product and assay variances
  - Need data or a heuristic (e.g., bioassay is 80% of total variability) to isolate the impact of the bioassay on Prob(OOS)
- *Linearity*
  - The ***linearity slope*** is associated with bias ($b \neq 1$) in ***comparing reportable results*** (e.g., stability)
  - There is currently no bases of acceptance criteria or design (sample size) for assessing linearity (note: linearity criterion in Q14)

References

1. ICH Q2(R2) *Validation of Analytical Procedures,* Mar 2024
2. ICH Q14 *Analytical Procedure Development,* Mar 2024
3. USP <1033> *Biological Assay Validation, USP* Pharmacopeial Forum (PF), 50(6), Nov 2024-Jan 2025
4. *Distinguishing the Analytical Method from the analytical Procedure to Support the USP Analytical Procedure Life Cycle Paradigm*, USP Pharmacopeial Forum (PF), 45(6), Nov-Dec 2019
5. Burdick, R.K., Borror, C.M., Montgomery, D.C. (2005), Design and Analysis of Gauge R&R Studies, SIAM, 143-148
6. USP <1210> *Statistical Tools for Procedure Validation,* USP NF, May 2018

Thank you!