IPAC-RS Comments on "USP <1210> Statistical Tools for Procedure Validation"

[Pharm Forum 40(5); In-Process Revision]

USP Chapter <1210> is a non-mandatory chapter, proposed as a companion chapter to "Validation of Compendial Procedure <1225>". USP Chapter <1210> introduces new statistical methods, such as two one-sided test (TOST) of statistical equivalence and corrected Akaike Information Criterion (AICc), which can be utilized during analytical method validation; however, it does not exclude the use of more traditional methodologies such as signal-to-noise ratio for the determination of limit of detection and quantitation. Overall, the document reads as a compilation of statistical equations and tools, with statistical interpretations of a number of validation parameters that are different from the current common practice. Thus, the new chapter represents a departure from current practice within the pharmaceutical industry. The USP authors greatly increase the complexity of method validation while providing little to no rationale for doing so. The recommendations appear to be purely statistical and devoid of rationale grounded in practice.

In general, IPAC-RS appreciates the initiative to introduce this chapter. IPAC-RS agrees with the principle of using statistics to support the validation process and enhance understanding of the properties of analytical methods. However, the risk in producing this USP document in such a prescriptive way is that more problems are created than are necessarily solved.

Historically, the purpose of USP chapters has been to provide guidance on a topic and to set public standards, rather than to "discuss" a topic. Therefore the chapter should be re-written to help guide analysts, or else be published as a discussion piece in an academic research journal outside of the USP.

The major concern is that the specific statistical methods discussed in this document will ultimately be *required* in practice. For example, when FDA Guidances for Industry reference USP chapters, they make no distinction between "below 1000" and "above 1000" chapters,[1] and thereby even USP's "informational chapters" could take on a role of a default standard. This could set a concerning precedent around chapters such as <1210>, which propose statistical approaches dictating what methods to apply rather than allowing companies to do what is most appropriate in a given situation. The text of the chapter should strongly and clearly stress that the described statistical methods are only suggestions or possibilities, but should not be construed as prescriptive, USP-recommended or required approaches.

[1] See, for example, FDA CDER/CBER Draft Guidance for Industry Analytical Procedures and Methods Validation for Drugs and Biologics, February 2014, especially pages 8, 11 at http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm386366.pdf

The chapter's intended use and intended audience are also unclear. More thought (and clarification in writing as part of the chapter) should be given to such questions as who this chapter is aimed at, and what the reader is expected to do with the information.

Another concern is that some methods are based on newly published papers instead of well-established statistical methods. This naturally poses two risks:

- a proposed statistical method could be found to have errors later on, as more research is done in that area,
- a proposed statistical method could be "not fit-for-purpose".

IPAC-RS therefore recommends that much further work be done before the chapter could be used beneficially and unambiguously.

Examples of issues that need to be resolved are listed below.

1. Are the listed statistical methods the only ones allowed by USP? Even though the chapter's number (above 1000) designates it as "informational", there is a well-based apprehension that compliance officers, inspectors, and others would read it more prescriptively. Some language therefore is needed to emphasize the "optional" nature of the advice given in the chapter, and perhaps some guidance on selection and justification of alternative approaches.

2. There could be significant compliance issues with such prescriptive guidance. For example, would it be possible to have a statistical program (e.g. SAS/R) available that supports methods in this USP document, as many of the computations envisioned by this chapter are not currently built into any software?

3. The chapter would have substantial implications with respect to the level of statistical ability required of an analyst to implement the described methods. How would this be resolved? Do the USP authors foresee large training programs, bespoke "black box" software, or increased expert statistical resource requirements? There is a potential risk that because of the level of complexity, validation will be seen as a "check box" exercise rather than individuals developing insights and understanding of the underlying principles and requirements.

4. The chapter does not provide any clear guidance on what to do under what circumstances. Without full clarity about what is, and what is not, expected in a certain situation, the amount of work associated with validating an analytical method could significantly increase when following this chapter.

5. Some of the advice in this chapter would appear to imply increases in the amount of experimentation during validation and substantially more experimental investigation during the pre-validation stage, e.g., to assess suitability of poolability, of trueness, etc. Has USP tried to assess the full impact of the requirements in the new chapter? Such an assessment would seem a responsible first step before USP's adopting a chapter.

6. As a specific example of the above concern: Chapter <1210> states that intermediate precision can be justified by a single experiment; however, it is noted that more detailed

pre-validation data are required to determine the parameters for the single intermediate precision experiment. There is a concern that for complex analytical procedures, such as Cascade Impactor testing (which is also labor- and resource-intensive), this approach, compared to the conventional method, will significantly increase the number of experiments required in order to complete the validation. What will be the unintended consequences?

7. It is unclear whether the expectation is to apply the proposed statistical procedures to all instances of method validation or whether it is acceptable to use a more pragmatic "risk based" approach, applying the methods from this chapter only to high-risk methods. Furthermore, a company typically would not validate compendial methods so the procedures of this chapter should only be applied to new methods.

8. Must the methods listed in the chapter be applied even when the underlying statistical assumptions do not apply?

9. Must a company collect data following the design structures given in the document even if others may be preferable?

10. The document makes no attempt to address how to set appropriate acceptance criteria. This is almost always the most difficult part of validation; and without it, the whole validation process is undermined. A more useful guidance document would be one on how to reasonably set acceptance criteria for TOSTs for accuracy, precision and linearity, since under most circumstances scientists have difficulty setting acceptance criteria. The guidance should also address the selection of acceptance criteria for use with prediction intervals and tolerance intervals as proposed in Section 3.3.2.

11. The document gives no specific limits for what would be considered an acceptable differential bias between two data sets, or how to calculate this from historical data.

12. Does the extended discussion of a specific experimental approach – such as simultaneous investigation of bias and intermediate precision – imply that separate investigations are inappropriate?

13. It is not clear how the tools proposed in the chapter could be used in the product development phase when there is limited process experience and limited real sample information.

14. The chapter should stress that AIC measures the *relative* quality of a statistical model. One model could have a better AIC than another but still be unfit as a calibration tool.

15. There are many different approaches within statistical modeling, for example bootstrapping techniques. A systematic literature review could identify a couple of methods suitable for modeling calibration relationship. The chapter would benefit from adding these methods as alternatives to the suggested TOST method.

16. It would be interesting to have some idea about the suggested TOST method's robustness when the assumptions of normally distributed response values and constant standard deviations for different levels are weak.

17. The notation is illegible in several places, for example in formulas 20-22 on page 25 of the pdf version from http://www.usppf.com

Specific Page/Line Comments:

Page, Line or Section	**Original Text**	**Proposed Change**	**Justification of Proposed Change**
Pages 1-2 Briefing and Introduction		Align the concept of ***validation*** with the lifecycle approach.	The term *validation* is used for the entire lifecycle of activities in the USP Stimuli article on 'Lifecycle Management of Analytical Procedures'
page 4, lines 13-15	*"Finally, although some of the statistical methods may appear new, **~~they~~** are currently used as standard practice in many industries outside of the pharmaceutical industry."*	Change to: "Finally, although some of the statistical methods may appear new, **most of them** are currently used as standard practice in many industries outside of the pharmaceutical industry."	The statement is true for the methods presented in section 3 (accuracy and precision). However the TOST approach for modeling the calibration relationship seems relatively new, which is indicated by an unpublished reference.

Page, Line or Section	Original Text	Proposed Change	Justification of Proposed Change
page 7, line 20	*"The normality assumption can be investigated using the Shapiro–Wilk's test (4) or by visual inspection of normal quantile plots. If data appear to be non-normal, possible remedies include data transformation, change of range end-point, **~~increased sample size~~**, or confidence intervals based on more appropriate probability models."*	Delete "increased sample size"	Increased sample size may simplify the judgment whether a normality assumption is valid. But a larger sample would not automatically lead to normality.
Page 9. Section 3.1	Chapter <1210> introduces NSD (normalized standard deviation) as an alternative statistical approach to % RSD for determining accuracy and precision. The normalizing constant, used in the calculation of the NSD, can be set as the % label claim, according to the chapter.	Further thought and guidance should be given regarding the criteria to be used to identify the normalizing constant. It should also be considered that %RSD (%CV) is a very well established measure of variability and the introduction of NSD may lead to confusion.	At lower concentrations, if the measured values were slightly below the label claim, this would result in an artificially lower NSD (via a lower SD), when compared to % RSD (which would be measured against the sample mean value); leading to an experimental bias.
Page 22. Section 3.3.4: Power Considerations		Add formula(s) or at least a reference of how to derive the sample size in section 3.3.4.	

Page, Line or Section	Original Text	Proposed Change	Justification of Proposed Change
Page 33, line 4 from the bottom	~~*"Intuitively it seems preferable to avoid replication in favor of testing more levels to elicit the true calibration relationship."*~~	Delete this sentence or refer to publications which support the statement.	A balanced proportion between numbers of test levels and replicates is necessary to support the choice of an appropriate model as well as a reasonable precision of the estimated error variation.
Page 36, line 6.	*"Response normalized values are normally distributed. This assumption ensures the accuracy of the confidence level of the TOST and is an underlying assumption of the AICc statistic."*	Change to: "Response normalized values are normally distributed. This assumption ensures the accuracy of the confidence level of the TOST and is an underlying assumption of the AICc statistic, **as it is implemented in section 6.8**."	The basic definition of AICc is 2k*n/(n-k-1)+2ln(L) where L is the maximized value of the likelihood function for the model. It is possible to derive the likelihood for other distribution than the normal. So AICc does not require a normal distribution, but it makes the calculations simpler.

*** *** ***

ABOUT IPAC-RS

International Pharmaceutical Aerosol Consortium on Regulation and Science (IPAC-RS) is an international association that seeks to advance the science, and especially the regulatory science, of orally inhaled and nasal drug products (OINDPs) by collecting and analyzing data, and conducting joint research and development projects. Our members include innovator and generic companies that develop, manufacture or market orally inhaled and nasal drug products for local and systemic treatment of a variety of debilitating diseases such as asthma, chronic obstructive pulmonary disease and diabetes. We aim to build consensus and contribute to effective regulations and standards by sharing the results of our research through conferences, technical journals, and discussions with regulatory bodies.

Current members include 3M, Actavis, AstraZeneca, Boehringer Ingelheim, Catalent, Chiesi, GlaxoSmithKline, Hovione, MannKind Corporation, Merck & Co., Inc., Mylan, Novartis, Sunovion, Teva, and Vectura.