dx.doi.org/10.14227/DT320425P166

# Parametric Tolerance Interval Test for Dissolution Testing of Immediate-Release Solid Oral Dosage Forms

**Steven J. Novick[1*], David LeBlond[2], and Stan Altan[3]**
[1]*Takeda Pharmaceuticals, Cambridge, MA, USA.*
[2]*Robert Singer Consulting, Sonora, CA, IL, USA.*
[3]*Johnson & Johnson Innovative Medicine, Titusville, NJ, USA.*

*email: steven.novick@takeda.com*

## ABSTRACT

**Introduction:** The compendial method specified in the *United States Pharmacopeia* (*USP*) General Chapter Dissolution <711> serves as a standard for batch quality. Although it has been commonly used by industry as a batch release test, it lacks any statistical underpinning. This study proposes the parametric tolerance interval test (PTIT) as a robust risk-based procedure for batch release decisions. The PTIT approach can be calibrated to match the operating characteristics of USP <711> under specific test settings to allow for flexible decision criteria, multiple stages, varying sample sizes, and alpha-spending adjustments if needed. **Methods:** PTIT compares a one-sided, beta-content, gamma-based confidence tolerance limit against a testing limit. Monte Carlo simulations were used to calculate the operating characteristics of USP <711> and PTIT across different testing parameters. The robustness of PTIT was evaluated for deviations from normality, and a Bayesian PTIT variant is introduced, with inference through posterior probabilities. **Results:** Implementing PTIT is recommended by comparing a 95% confidence/85% content tolerance limit to the $Q - 5$ testing limit. This approach allows for other confidence and content levels, as considered appropriate. The operating characteristics align well with USP <711> when the SD of the mean is 3%. PTIT remains robust to slight departures from normality. The Bayesian approach is equally viable while also providing the ability for prior information inclusion as well as consideration of non-normal data distributions. **Conclusion:** The PTIT offers a practical solution for customizing dissolution release testing to specific product and process needs. This underscores the importance of sophisticated statistical approaches to enhance decision-making, transparency, and maintain drug product quality.

**KEYWORDS:** USP <711>, Q testing, parametric tolerance interval test (PTIT), dissolution, in vitro release

## INTRODUCTION

Dissolution or in vitro release testing of solid dose products (e.g., tablets, capsules) is a regulatory and commercial necessity. Regulations mandate that drug product batches meet compendial dissolution specifications prior to distribution, and post-marketing commercial testing tracks batch quality consistency. In vitro release testing also provides insights into the disintegration and release rate of the active pharmaceutical ingredient, which can indicate bioavailability and therapeutic effects.

The *United States Pharmacopeia* (USP) general chapter <711> specifies equipment, media, protocols, and acceptance limits applicable to immediate, extended, modified, and delayed release dosage forms requiring dissolution testing (*1*). As companion sets of guidelines, USP <1092> advises on assay development, and USP <724> extends the concept of standards to transdermal dosage forms (*2*, *3*). Companies are advised to develop their own tailored batch release procedures, ensuring the USP standards are met with high confidence. The United States FDA has explicitly noted that USP <711> and, similarly, USP <905> (uniformity of dosage units) is not intended to provide statistical assurance of quality for the broader batch release testing of dose units (*4*, *5*). Consequently, manufacturers are advised to implement more stringent and statistically grounded release

*Corresponding author

tests, taking into consideration the Biopharmaceutics Classification System properties of the formulation or compound (*6*).

For immediate-release products, USP <711> follows a three-stage zero-tolerance decision rule in which the summary statistics and limits vary across the stages. Because it lacks any underlying parametric model that permits hypothesis-driven inference, it cannot characterize batch quality. Consequently, there have been efforts to develop statistically grounded release tests. The parametric tolerance interval testing (PTIT) was proposed by Tsong and Shen as a consistent underlying model and hypothesis-based batch population inference approach (*7*). Subsequent works by Hauck et al., Dong et al., and Otava et al., refined the PTIT operating characteristics, with calibration enabling alignment with USP <711> stringency (*8–10*).

The FDA supports statistical approaches like PTIT for batch release testing of dissolution and broader quality assurance, as evidenced in its guidance for inhalation and nasal drug products. The current work builds upon the endeavors of Hauck et al., Dong et al., and Otava et al. to provide definitive recommendations for the use of PTIT for hypothesis-driven batch release testing (*8–10*). This study assesses the statistical power of PTIT under a variety of scenarios and illustrates its robustness to deviations from normality. Finally, this study aims to provide a modernized PTIT via Bayesian method to accommodate the possibility of prior information or non-normal distributions.

## METHODS

### Parametric Tolerance Interval Test (PTIT)

Consider the population of solid oral dosage units (without loss of generality, tablets) in a batch. Let $Y$ denote the percent dissolution of a tablet at a predefined time point and let $Q$ denote the dissolution criterion from USP <711>. Assume that $Y \sim N(\mu, \sigma^2)$ are independent normally distributed random variables, where $\mu$ = mean, and $\sigma$ = standard deviation (SD). Because the percent dissolution must fall above 0% and (roughly) below 100%, some care must be taken in making the normality assumption. It is our experience that, over a wide range of time points chosen to describe the dissolution profile, the normality assumption is reasonable. This is frequently the case in the region of $Q$ = 70–80%. Solutions for non-normal distributions will be discussed later.

For lower testing limit $L$, (i.e., $L \leq Q$) and for proportion $p$, a reasonable null ($H_0$) and alternative ($H_A$) hypothesis for batch release testing is given by H1 and visualized in Figure 1.

$H_0$: Less than 100$p$% of tablets > $L$ (H1)

$H_A$: At least 100$p$% of tablets > $L$

Let $q_p(\mu, \sigma^2) = \mu - \Phi^{-1}(p)\sigma$ denote the lower 100(1 – $p$)% quantile of $Y$, and $\Phi^{-1}(p)$ is the inverse of the standard normal cumulative distribution function. To declare $H_A$ in H1, we must have $q_p(\mu, \sigma^2) > L$. If tolerance limit $T_L$ is a lower 100(1 – $\alpha$)% confidence limit for $q_p(\mu, \sigma^2)$ and $T_L > L$, we can state that at least 100$p$% of tablets in the batch are > $L$ with 100(1 – $\alpha$)% confidence. To test the hypotheses in H1, we declare $H_A$ if $T_L > L$.

A lower 100(1 – $\alpha$)% confidence limit for $q_p(\mu, \sigma^2)$ is also called a 100(1 – $\alpha$)%/100$p$% Beta-content tolerance limit for $Y$. Under the normal distribution assumption, a 100(1 – $\alpha$)%/100$p$% tolerance limit is given in Chapters 2 (frequentist) and 11 (Bayesian) by Matthew and Krishnamoorthy (*11*). The procedure of testing H1 with a tolerance interval is called a one-stage parametric

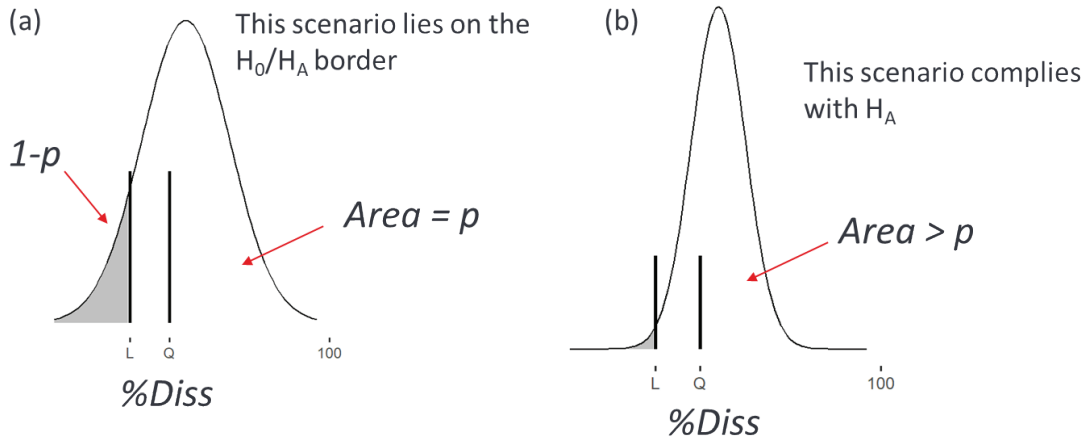


*Figure 1. (**a**) Normal distribution on the $H_0/H_A$ border for testing hypothesis 1 (H1). (**b**) Normal distribution that meets with $H_A$ of $H_1$. Diss: dissolution; $H_0$: null hypothesis; HA: alternative hypothesis; L: lower limit.*

tolerance interval test (PTIT-1). As an alternative test statistic, a Bayesian rule can be applied to accept $H_A$ if the posterior probability that $q_p(\mu, \sigma^2) > L$ is at least 1 - α. The Bayesian paradigm proves especially useful when a non-normal distributional assumption is imposed on the dissolution data. For now, the standard classical statistical approach is explored.

Because USP <711> is a three-stage test, consider a three-stage PTIT (PTIT-3) with sample sizes $n_1 = 6$ for stage 1, $n_2 = 12$ for stage 2, and $n_3 = 24$ for stage 3. Let $Y_j$ denote the percentage of dissolution for the $j^{th}$ tablet, with $Y_j \sim N(\mu, \sigma^2)$; $j = 1, \ldots, 24$. Let $\bar{y}_i$ and $s_i$ denote the sample mean and SD of the full sample at the $i^{th}$ stage, respectively. Testing of multiple stages is adjusted for alpha-spending ($\alpha_1$, $\alpha_2$, $\alpha_3$) to achieve an overall type 1 error, α. The lower 100(1 – α)%/100*p*% tolerance limit for the $i^{th}$ stage is given by Eq. (1).

$$T_L(i) = \bar{y}_i - t^{-1}\left(1 - \alpha_i,\ n_i - 1,\ ncp = \sqrt{n_i}\,\Phi^{-1}(p)\right)\frac{s_i}{\sqrt{n_i}}, \quad (1)$$

where $t^{-1}$ ($\eta$, $\lambda$, $\phi$) is the 100$\eta$% quantile of the non-central T distribution with $\lambda$ degrees of freedom and non-centrality parameter (*ncp*) $\phi$.

Different alpha-spending calculations may be explored and employed, depending on costs, risks, and stage of development considerations. We follow Tsong and Shen, who implemented the alpha-spending approach of O'Brien and Fleming with an overall α = 0.05 so that $\alpha_1 = 0.00009$, $\alpha_2 = 0.00554$, and $\alpha_3 = 0.04824$ (*7, 12*). Another reasonable choice, as performed by Novick et al., is the DeMets and Lan and Pocock alpha-spending function, which yields $\alpha_1 = 0.0179$, $\alpha_2 = 0.0189$, and $\alpha_3 = 0.0279$ and more evenly distributes the risk across testing stages (*13–17*). While O'Brien and Fleming put a larger burden on testing in stages 1 and 2 and may be seen as well-aligned with USP <711>, both alpha-spending methods share an overall type 1 error rate of 0.05. The choice of alpha spending adjustment may also be linked to stage of process validation, as defined by the 2011 process validation FDA guidance (*18*). It makes sense to apply the O'Brien and Fleming adjustment during process validation stages 1, 2, and early 3 (not to be confused with testing stages), when the historical knowledge of the process is still limited (*12*). But later in process validation stage 3, when the historical knowledge of the process has accumulated, relaxing the adjustment to the DeMets and Lan method may be justified (*16*).

At the $i^{th}$ testing stage, $H_A$ in H1 may be accepted if $T_L(i) > L$; otherwise, testing proceeds to the next stage. In this work, if $H_A$ is not accepted at stage 3, the test results in a failure and the batch cannot be released to market.

Without loss of generality, let $Q = 80\%$ for the remainder of this paper. Dong et al. and Otava et al. examined the PTIT under the assumptions laid out in this section, with $L = Q$ across various choices for p (*9, 10*). This PTIT is considered to be overly conservative compared to the operating characteristics specified in USP <711>, given that, for the empirical requirements of stage 3, about 92% (22 out of 24) of dosage units must exceed $Q - 15$. Instead, we consider a PTIT with $L = Q - \delta$ for some $\delta \geq 0$. In the Results section, a Monte Carlo study will explore the values δ and *p* so that, under selected conditions, the PTIT-3 operating characteristic (i.e., probability to declare $H_A$ in Eq. 1) will be similar to USP <711> (i.e., probability to satisfy USP <711> requirements). By careful selection of *Q*, δ, and *p*, the user may ensure that the probability to declare $H_A$ in H1 is not larger than the probability to meet the requirement of USP <711>.

**Confidence Interval Test (CIT)**

For lower testing limit *M* (i.e., $M \leq Q$), a reasonable hypothesis for batch release testing is given by H2.

$$H_0: \mu \leq M \qquad (H2)$$

$$H_A: \mu > M$$

A lower 100(1 – α)% confidence limit for the batch mean is given in Chapter 7 by Ross (*19*).

To test the hypotheses in H2, we declare $H_A$ if confidence limit $C_L > M$. As a comparator to USP <711>, in which one must show that the sample mean $\bar{y} > Q$, because $C_L < \bar{y}$, it follows that it is desirable for $M < Q$.

As with the tolerance limit, the lower 100(1 – α)% confidence limit is modified with alpha-spending for the three stages, as shown in Eq. (2).

$$C_L(i) = \bar{y}_i - t^{-1}(1 - \alpha_i,\ n_i - 1)\frac{s_i}{\sqrt{n_i}}, \qquad (2)$$

where $t^{-1}$ ($\eta$, $\lambda$) is the 100$\eta$% quantile of the central *T* distribution with $\lambda$ degrees of freedom.

In this method, the three-stage procedure of testing H2 with Eq. (2) is called the confidence interval test (CIT). Because USP <711> places requirements on both individual dosage units and the sample mean, it makes sense to require both the PTIT and CIT. That is, one must claim $H_A$ in both H1 and H2 by showing $T_L(i) > L$ and $C_L(i) > M$ at some stage $i = 1$, 2, or 3. Because this is an example

of intersection-union testing, no adjustment to the type 1 error (except for the alpha-spending) is made (*19*). A Bayesian rule can also be applied to jointly accept $H_A$ in H1 and H2 if the posterior probability that $q_p$ ($\mu$, $\sigma^2$) > $L$ and $\mu > M$ is at least 1 - $\alpha$. As with the PTIT, we consider a CIT with $M = Q - \gamma$ for some $\gamma \geq 0$.

### Monte Carlo Simulations for Normally Distributed Data

Monte Carlo simulations were performed to investigate the operating characteristics for meeting the requirements of USP <711> (Table 1), the PTIT alone, and the combined PTIT and CIT (PTIT+CIT). It will be shown in this section that the added value of the CIT is debatable, so the focus of this work will be on the PTIT. Although the sample size in USP <711> is fixed with three stages, the operating characteristics of the PTIT were investigated with larger sample sizes and separately, with only one or two stages.

*Table 1. Operating Characteristics for USP <711> Requirements*

| Stage | Sample size | USP <711> Criteria |
|---|---|---|
| 1 | $n_1 = 6$ | All 6 values > $Q + 5$ |
| 2 | $n_2 = 12$ (6 additional) | Mean of 12 values > $Q$<br>All 12 values > $Q - 15$ |
| 3 | $n_3 = 24$ (12 additional) | Mean of 24 values > $Q$<br>At least 22 of 24 values > $Q - 15$<br>All 24 values > $Q - 25$ |

*Based on information from USP <711> (1).*
*USP: United States Pharmacopeia.*

Unless otherwise noted, $Q = 80\%$ and data were generated as independent $Y_j \sim N$ ($\mu$, $\sigma^2$) ($j$ = 1, …, 24), with 75 < $\mu$ < 90 and $\sigma$ = 0.5, 1, 3, 4.5, 6. For the PTIT with $T_L(i) > Q - \delta$, testing parameters were varied according to $p$ = (0.80, 0.85, 0.90, 0.95), and $\delta$ = (0, 5, 10, 15). For the CIT with $C_L(i) > M - \gamma$, we examined $\gamma$ = (0, 3).

To determine the operating characteristics for the PTIT with $T_L(i) > Q - 5$ as a function of sample size, the sample size was increased at each stage by 1x ($n_1 = 6$, $n_2 = 12$, $n_3 = 24$), 2x ($n_1 = 12$, $n_2 = 24$, $n_3 = 48$) , and 3x ($n_1 = 18$, $n_2 = 36$, $n_3 = 74$). O'Brien and Fleming alpha spending is a function of the relative sample size of stage, so the values of ($\alpha_1$, $\alpha_2$, $\alpha_3$) remain unchanged (*12*). Because the operating characteristics of the PTIT and USP <711> can be matched at $\sigma = 3$, $p = 0.85$, and $\delta = 5$, the main interest is to examine the operating characteristic for 3 < $\sigma$ < 6% to determine if the PTIT can recover its disadvantage for $\sigma > 3\%$.

To study the effect of staged testing on the PTIT, single-stage testing (PTIT-1) was performed with $n_1 = 24$; two-stage testing (PTIT-2) was performed with $n_1 = 12$ and $n_2 = 24$; and PTIT-3 was performed with $n_1 = 6$, $n_2 = 12$, $n_3 = 24$. With an overall $\alpha = 0.05$, there is no alpha-spending adjustment for PTIT-1. For PTIT-2, the O'Brien and Fleming alpha-spending adjustment is $\alpha_1 = 0.000687$ and $\alpha_2 = 0.049771$. For PTIT-3, $\alpha_1$, = 0.00009, $\alpha_2 = 0.00554$, and $\alpha_3 = 0.04824$. The CIT was only examined in three-stage testing with the same sample sizes and alpha-spending adjustments as PTIT-3. Each simulated scenario was run 10,000 times.

### Monte Carlo Simulations for Non-Normal Data

Although the normal distribution may be a reasonable choice for most dissolution data, it is plausible that dissolution distribution for some products may deviate from this assumption. We examine the robustness of the PTIT with $T_L(i) = Q - 5$ ($\delta = 5$) and $p = 0.85$ to such deviations by characterizing the operating characteristics of the PTIT under a skew normal (SN) and a T distribution (see supplemental material for functional forms). In the SN probability density function, $\beta$ controls the skewness, $\epsilon$ is the location parameter, and $\omega$ is the scale parameter (*20*). In the T probability density function, $n$ denotes the degrees of freedom, $\epsilon$ is the location parameter, and $\omega$ is the scale parameter (*21*). Relative to the normal distribution, the SN with a negative skew parameter is skewed to the left, which places more probability in the left tail, and the T distribution puts more probability in both tails.

To illustrate the skew and extra tail probability, a Monte Carlo simulation was performed to explore the robustness of the proposed PTIT-3 to the SN and T distributions relative to the normal distribution. The means and tail probabilities less than $Q = 80\%$ were matched across all three distributions for each scenario. Means ranged from ($Q + 1$) < $\epsilon$ < 90%, and tail probabilities are 0.01, 0.1, 0.2, and 0.3. The skewness settings for the SN distribution were $\beta$ = –4, –3, –2, –1, 0 (where –4 = large skew and 0 = no skew). The degrees of freedom for the $T$ distribution are $\gamma$ = 3, 5, 10, 25, $\infty$ (where 3 = larger tail probabilities, $\infty$ = normal tail probabilities). In all cases, the scale parameter was derived from the other parameters.

### Bayesian Method for PTIT and CIT

Equations (2) and (4) provide a lower 100(1 – $\alpha$)%/100$p$% tolerance limit and a lower 100(1 – $\alpha$)% confidence limit for the mean, respectively, using a frequentist construct specifically for the normal distribution. Although Bayesian analysis may directly calculate the posterior distribution to meet $H_A$ in H1 and H2, for a Bayesian analogue to the frequentist system, one may construct a lower 100(1 – $\alpha$)%/100$p$% Bayesian tolerance limit by calculating the lower 100$\alpha$% quantile of the posterior distribution $T_L = \mu - \sigma\Phi^{-1}(p)|\boldsymbol{Y}$, where $\Phi^{-1}(p)$ is the inverse

cumulative distribution function of the standard normal, and ***Y*** denotes the sampled dissolution data. A lower 100(1 – α)% credible limit ($C_L$) for μ may substitute for the confidence limit. With the Jeffreys' prior, Matthew and Krishnamoorthy show that $T_L$ is equal to H2 and $C_L$ is equal to Eq. (3) (*11*). However, depending on applications and justifiable prior knowledge, Bayesian analysis may leverage different prior distributions, which would then affect the values of $T_L$ and $C_L$.

For PTIT-3, one might use the alpha-spending procedure suggested for frequentist testing and calculate the lower 100(1 – $\alpha_i$)% quantile $T_L(i) = \mu - \sigma\Phi^{-1}(p)|\boldsymbol{Y}_i$ and a $C_L(i)$, a lower 100(1 – $\alpha_i$)% credible limit for μ, where $\boldsymbol{Y}_i$ denotes the cumulative sampled dissolution data at the $i^{th}$ stage. It may be antithetical to use an alpha-spending schema because Bayesian probabilities, unlike *p*-values, are not calculated with conditioning on $H_0$ (Bayes factors are a notable exception).

For consistency with the frequentist approach, an analogous test can be constructed using the Bayesian versions of $T_L(i)$ and, if desired, $C_L(i)$. From these, one may construct Bayesian PTIT and PTIT+CIT procedures. Note that for the PTIT+CIT, the Berger and Hsu intersection-union procedure does not extend to Bayesian hypothesis testing (*22*). Bayesian analysis would instead calculate the joint posterior probability (*Pr*) of $H_A$ directly via Eq. (3).

$$p_i = Pr\,(\mu - \sigma \times \Phi^{-1}(p) > L \text{ and } \mu > M|\boldsymbol{Y}_i) \qquad (3)$$

Then, at the $i^{th}$ stage, if $p_i > 1 - \alpha_i$, $H_A$ is declared; otherwise, move to the next stage.

Bayesian statistics may also extend the PTIT and CIT to other distributions. Let $Y_j \sim F(\boldsymbol{\theta})$, for some distribution *F*(.) with parameter vector **θ**, *j* = 1, 2, …, 24 (or some other sample size) and let $g(\boldsymbol{\theta})$ denote the mean of the distribution. For the PTIT, a lower 100(1 – α)/100*p*% Bayesian tolerance limit is given by the lower 100(1 – α)% quantile of the posterior distribution $T_L = F^{-1}(\boldsymbol{\theta}, p)|\boldsymbol{Y}$ and $C_L$ may be given as the lower 100(1 – α)% posterior quantile of $g(\boldsymbol{\theta})$. Thus, the generalization of Eq. (3) is given by Eq. (4).

$$p_i = Pr\,(F^{-1}(\boldsymbol{\theta}, p) > L \text{ and } g(\boldsymbol{\theta}) > M|\boldsymbol{Y}_i) \qquad (4)$$

As before, if $p_i > 1 - \alpha_i$, $H_A$ is declared at the $i^{th}$ stage; otherwise, move to the next stage.

For normally distributed data and vaguely informative priors, the Bayesian method should perform similarly to the frequentist procedures described in earlier sections. The Bayesian method is demonstrated in the results with SN and T-distributed computer-generated data with (β = -3, ϵ = 88, ω = 4) and (γ = 5, ϵ = 85, ω = 2), respectively.

## RESULTS

### Monte Carlo Simulations

#### *USP <711> and PTIT*

To determine the operating characteristics for satisfying the requirements USP <711> and the PTIT alone using Eq. (1) to test H1, a Monte Carlo simulation was conducted. The operating characteristics for stage 3 (overall probability) are provided in Figure 2.

From Figure 2, calibration of the PTIT with USP <711> can be determined in several places. For example, the operating characteristics match up well when σ = 3, *p* = 0.85 and δ = 5. From our experience, σ = 2–3% stands as a typical range, with 1–2% and 3–5% representing tight and variable dissolution methods, respectively. Producers with an SD that falls outside of the range of these simulations are encouraged to conduct their own set of simulations to examine the operating characteristics for their specific analytical circumstances. In this work, the PTIT with σ =3, *p* = 0.85, δ = 5 stands as a reasonable point of comparison against USP <711>. Given the set of parameters, the PTIT procedure rewards lower SD and penalizes larger SD compared to USP <711>. This is a desirable feature of the PTIT. Another potential PTIT choice is σ = 2.5, *p* = 0.90, δ = 5. Earlier, *p* = 0.92 and δ = 15 was suggested to be a reasonable choice, but, from Figure 2, one can infer that this scenario would be far too liberal to match with USP <711> until σ = 6, which represents a highly variable dissolution method. In practice, for batch release characterization and testing, one should choose a PTIT that is more conservative than the USP <711> criteria (Table 1).

The operating characteristics for each stage of the PTIT with *p* = 0.85 and δ = 5 are shown in Supplementary Figure S1. Across all stages, relative to USP <711>, the PTIT is more liberal with small SD values and more conservative with larger SD values.

#### *USP <711> and PTIT + CIT*

To determine the operating characteristics for satisfying the requirements of USP <711> and the PTIT+CIT, using Eq. (1) to test H1 and Eq. (2) to test H2, a Monte Carlo simulation was conducted. The operating characteristics for stage 3 (overall probability) are provided in Figure S2 with $C_L(i) > Q$ (γ = 0). As expected, one cannot calibrate the PTIT+CIT to match with USP <711> for any value of *p* or δ when σ ≥ 2.5. In Figure 3, the operating characteristic of the PTIT+CIT with $C_L(i) > Q - 3$ (γ = 3) is compared to the PTIT alone, with δ = 5, 10. When δ = 5, there is no

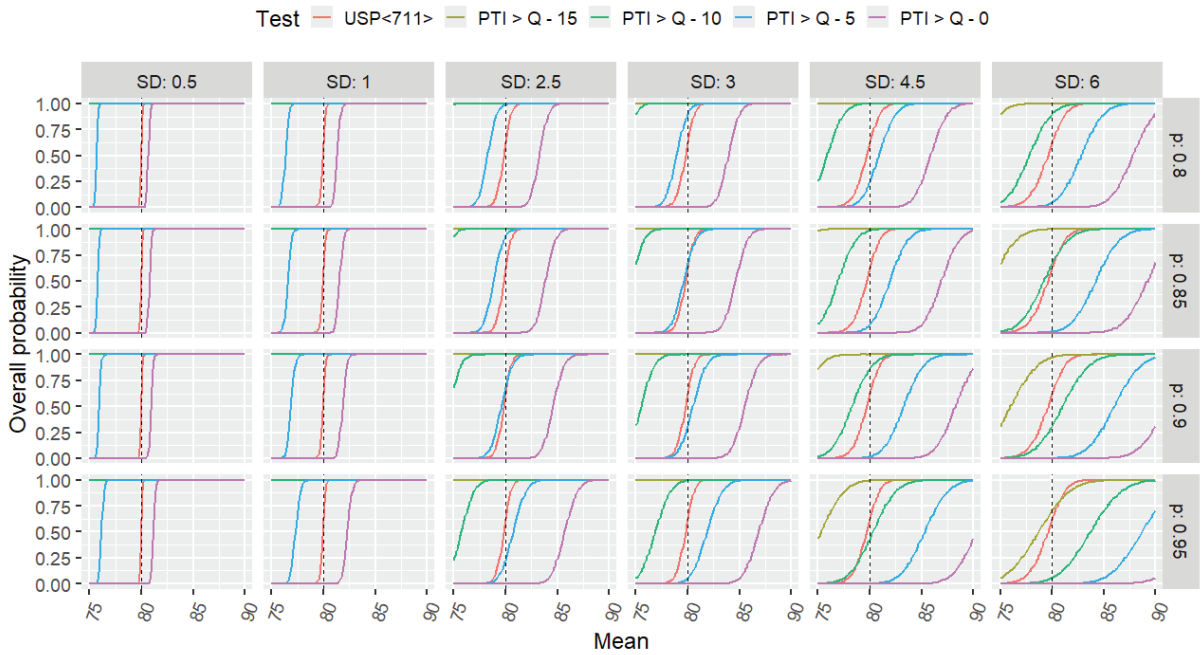


*Figure 2. Overall operating characteristics for satisfying USP <711> and three-stage parametric tolerance interval (PTI) tests for normally distributed data with a population mean and standard deviation.*

significant difference between the PTIT+CIT and PTIT alone except when σ ≥ 1, which represents a rare tight dissolution assay. The differences are made clear when δ = 10, with decreasing preference for the PTIT alone as the SD increases. Considering the recommended PTIT settings of $p$ = 0.85 and δ = 5, the CIT does not contribute to the stringency of the test procedure, so it is an unnecessary test; however, it also appears to do little harm.

### *PTIT as a Function of Sample Size*

The sample size for USP <711> is $n_1$ = 6, $n_2$ = 12, and $n_3$ = 24. The operating characteristics for PTIT as a function of

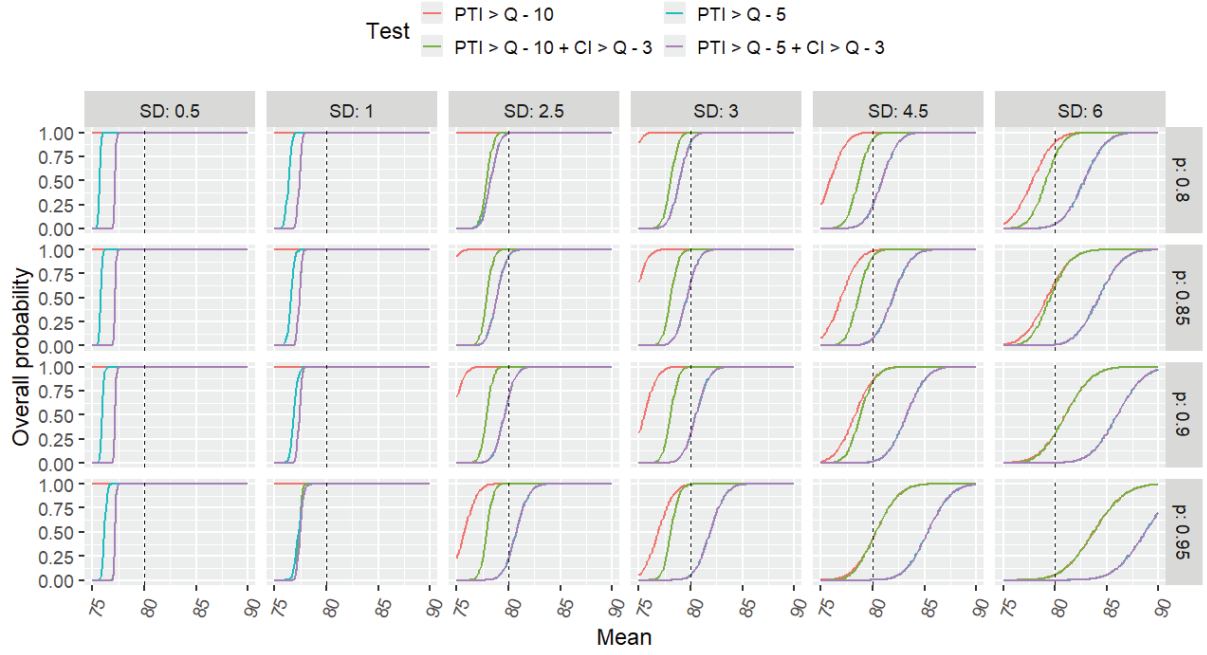


*Figure 3. Overall operating characteristics for satisfying three-stage parametric tolerance interval (PTI) test alone and combined with the confidence interval (CI) test ($C_L$ (i) > Q -3) for normally distributed data with a population mean and standard deviation.*

sample size are shown in Figure S3. Doubling and tripling the sample size improves the operating characteristic of the PTIT but cannot match that of USP <711> for σ ≥ 4.5%.

***PTIT with Multi-Stage Testing***

As shown in Figure S4, the number of stages does not appear to affect the probability to satisfy H1 with the PTIT. Because staging may affect efficiency, the expected number of dosage units was calculated. For single-stage testing, the number of units is always 24. For two-stage testing, the expected number of units is $12 + 12 \times (1 - p_1)$, where $p_1$ is the probability to meet the requirements of the PTIT in stage 1. For three-stage testing, the expected number of units is $6 + 6 \times (1 - p_1) + 12 \times (1 - p_2)$, where $p_i$ is the probability to meet the requirements of the PTIT in stage *i*, (*i* = 1 or 2). The expected number of dosage units are shown in Figure S5, which indicates that multi-stage testing generally requires fewer dosage units, making it the more efficient option. This advantage must be balanced against the requirement of representativeness.

**Robustness of the PTIT to Non-Normality**

Figure S6 shows a normal distribution with parameters μ = 85 and σ = 3.04; an SN distribution with parameters β = –4, ϵ = 88.3, and ω = 4.2; and a T distribution with parameters γ = 3, ϵ = 85, and ω = 2.1, each possessing a tail probability below *Q* = 80% of 0.05. The scenarios β = 0 for SN and γ = ∞ for T represent the normal distribution. The operating characteristics are shown in Figure S7. It appears that skewness and excess tail probability both drive operating characteristic probabilities lower. Thus, for the SN and T distributions, it may be inferred that the PTIT shows robustness to deviations from normality.

**Bayesian Methods for Non-Normality**

To demonstrate the Bayesian method, 24 observations were generated and split into three stages, respectively, from the SN (β = -3, ϵ = 88, ω = 4) and T (γ = 5, ϵ = 85, ω = 2) distributions. The results are provided in Table S1 and Figure S8. The mean for both distributions is 85%, and the 5% and 95% quantiles of the two distributions are similar.

The PTIT, CIT, and probability $p_i$ from Eq. (4) were calculated by correctly assuming the SN and T distributions. For model fitting, vaguely informative prior distributions are given by the following, where HC = half-Cauchy, Γ is the gamma distribution with parameters shape (sh) and scale (sc), and *Q* = 80%.

- SN: $\beta \sim T(\gamma = 3, \epsilon = 0, \omega = 1)$; $\epsilon \sim N(\mu = Q, \sigma = 10)$; $\omega \sim HC(0, 1)$
- T: $\gamma \sim \Gamma(sh = 2, sc = 0.1)$; $\epsilon \sim N(\mu = Q, \sigma = 10)$; $\omega \sim HC(0, 1)$

Parameter estimates (posterior medians) with 95% credible limits for the SN and T distributions are provided in Tables S2 and S3, respectively. Results of testing are given in Table 2, which shows that the SN-generated data fails stages 1 and 2 but passes in stage 3. The T-generated data would fail stage 1 but pass at stage 2. The same conclusion was drawn using PTIT+CIT and Eq. (3) for the assessment.

## DISCUSSION

Dissolution testing for the purpose of assuring drug product quality has a long history as part of pharmaceutical company's overall control strategy. USP <711> sets forth a compendial standard of quality and has often been used for batch release testing (*1*). Although this practice has been criticized by both the scientific community and the FDA, the limited literature on the topic has had little influence in changing industry practices. Consequently, this study provides an updated view of an existing statistically based decision procedure.

The PTIT statistical approach for batch release has been previously proposed for content uniformity and more recently, for dissolution. The current study was built upon this approach and proposes a flexible PTIT statistical procedure that permits varying the decision rule criterion, the number of stages and sample sizes, and proposed a Bayesian counterpart with a decision criterion supported by a posterior probability.

*Table 2. Results of Bayesian PTIT, CIT, and Posterior Probability for Batch Release Testing with Q = 80%*

| Distribution | Stage | $\alpha_i$[a] | $1-\alpha_i$ | $T_L(i)$ Must be > $Q - 5$ | $C_L(i)$ Must be > $Q - 3$ | $p_i$ Must be > $(1 - \alpha_i)$ |
|---|---|---|---|---|---|---|
| SN | 1 | 0.00009 | 0.99991 | 58.9 (fail) | 72.3 (fail) | 0.957 (fail) |
| | 2 | 0.005544 | 0.994456 | 74.5 (fail) | 79.9 | 0.992 (fail) |
| | 3 | 0.048242 | 0.951758 | 78.6 | 83.2 | > 0.999 |
| T | 1 | 0.00009 | 0.99991 | 63.3 (fail) | 76.0 (fail) | 0.988 (fail) |
| | 2 | 0.005544 | 0.994456 | 79.1 | 82.9 | > 0.999 |
| | 3 | 0.048242 | 0.951758 | 81.4 | 84.5 | > 0.999 |

[a]*alpha-spending values from O'Brien and Fleming (11).*

It is encouraging to observe the increasing regulatory acceptance of Bayesian approaches. A Bayesian PTIT approach can offer three advantages:

1. The Bayesian perspective supports patient-centric risk-based release decisions by quantifying batch quality probabilistically.

2. When prior knowledge about underlying model parameters (e.g., mean and SD) can be justified from representative historical studies, the Bayesian paradigm provides distributional tools for expressing that knowledge quantitatively and incorporating it seamlessly into the decision process.

3. For products that require more complex modeling (e.g., non-normal, hierarchical, or nonlinear models), non-Bayesian approaches may require approximations or even be intractable. Bayesian methods are less dependent on analytical derivations and provide exact solutions to any desired degree of Monte Carlo accuracy.

An alpha spending adjustment based on the O'Brien and Fleming method was implemented in the multiple stage testing to accommodate sequential testing (*12*). For convenience and for comparative purposes, this study assessed operating characteristics using the same sample sizes as given in the USP <711> three-stage test with $Q$ = 80% at chosen values of σ and $p$ (proportion above $Q$). Given the set of parameters, the PTIT procedure rewards lower variability and penalizes larger variability compared to USP <711>. For typical parameter values, it is a more stringent test procedure than the USP <711> rules. The addition of a simultaneous test on the batch mean value was found to provide little, if any, advantage in forming a more informative or more stringent test. Robustness of the PTIT procedure was studied through the assessment of mild skewness and wide tails. For both cases, the PTIT procedure showed robustness to departures from normality, especially in those cases where the mean was close to $Q$. Finally, a Bayesian version of the proposed test was detailed, with the possibility of the incorporation of appropriate prior information and non-normal data distributions. Inference is then provided in terms of the posterior probabilities.

Although PTIT procedures have been proposed previously, we are not aware of any approved drug product that employs this approach to assure conformance to the USP <711> standard. It is important to understand that there is always some probability that a given dataset passes the PTIT as we have described it but fails to meet USP <711> criteria (*1*). The operating characteristic curves in this work demonstrate that the probability to declare HA with the PTIT can be no larger than the probability to meet the USP requirements.

This study is not proposing to change or replace the USP <711> compendial standard. The intent is to propose a coherent statistical framework for batch release decisions that, if passed, will provide assurance that the test batch meets the existing compendial standard with similar or smaller probability. This PTIT test is framed as a batch release decision tool, but it seems reasonable that a similar PTIT, with appropriately adjusted parameters, may also be useful for other purposes, such as developmental or investigational decision making.

## CONCLUSION

The need for a statistically based decision procedure for dissolution release testing was the motivation for developing this procedure, especially in view of the widespread but inappropriate application of USP <711> for batch release by companies. The proposed Bayesian PTIT approach promotes patient centric decision-making by allowing customizable criteria, direct risk control, and the ability to integrate historical data. It provides strict evaluation standards, ensuring a rigorous risk control strategy with good performance characteristics relative to the USP <711> criteria. The proposed PTIT method offers a robust statistical framework for reliable drug product quality assurance and is easily adapted to conform to companies' risk tolerance practices specific to the product and the process.

## DISCLOSURES

The authors received no financial support for this work and have no conflicting interests.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article and may be found at https://osf.io/k2s85/files/52yft.

## REFERENCES

1. <711> Dissolution. In *USP–NF*. United States Pharmacopeia, 2023. DOI: 10.31003/USPNF_M99470_03_01.
2. <1092> The dissolution procedure: development and validation. In *USP–NF*. United States Pharmacopeia, 2023. DOI: 10.31003/USPNF_M643_05_01.
3. <724> Drug release. In *USP–NF*. United States Pharmacopeia, 2023. DOI: 10.31003/USPNF_M99490_05_01.
4. <905> Uniformity of Dosage Units. In *USP–NF*. United States Pharmacopeia, 2022. DOI: 10.31003/USPNF_M99694_01_01.
5. Questions and Answers on Current Good Manufacturing

Practice Regulations: Production and Process Controls. U.S. Food and Drug Administration. May 22, 2024. https://www.fda.gov/drugs/guidances-drugs/questions-and-answers-current-good-manufacturing-practice-regulations-production-and-process (accessed Oct 16, 2025).
6. M9 Biopharmaceutics Classification System-Based Biowaivers; Guidance for Industry. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 2021.
7. Tsong, Y.; Shen, M.; Shah, V. P. Three-stage sequential statistical dissolution testing rules. *J. Biopharm. Stat*. **2004**, *14* (3), 757–779. DOI: 10.1081/BIP-200025689.
8. Hauck, W. W.; Foster, T.; Sheinin, E.; Cecil, T.; Brown, W.; Marques, M.; Williams, R. L. Oral dosage form performance tests: new dissolution approaches. *Pharm. Res*. **2005**, *22* (2), 182–187. DOI: 10.1007/s11095-004-1184-4.
9. Dong, X.; Tsong, Y.; Shen, M.; Zhong, J. Using tolerance intervals for assessment of pharmaceutical quality. *J. Biopharm. Stat*. **2015**, *25* (2), 317–327. DOI: 10.1080/10543406.2014.972512.
10. Otava, M., Jacquart, S.; Altan, S. Dissolution testing strategies for large sample sizes and applications in continuous manufacturing. *Dissolut. Technol*. **2024**, *31* (3), 128–134 DOI: 10.14227/DT310324P128.
11. Mathew, T.; Krishnamoorthy, K. *Statistical Tolerance Regions: Theory, Applications, and Computations*; John Wiley & Sons: Hoboken, NJ, 2009.
12. O'Brien, P. C.; Fleming, T. R. A multiple testing procedure for clinical trials. *Biometrics* **1979**, *35* (3), 549–556. DOI: 10.2307/2530245.
13. Novick, S.; Christopher, D.; Dey, M.; Lyapustina, S.; Golden, M.; Leiner, S.; Wyka, B.; Delzeit, H. J.; Novak, C.; Larner, G. A two one-sided parametric tolerance interval test for control of delivered dose uniformity. Part 1--characterization of FDA proposed test. *AAPS PharmSciTech* **2009**, *10* (3), 820–828. DOI: 10.1208/s12249-009-9270-x.
14. Novick, S.; Christopher, D.; Dey, M.; Lyapustina, S.; Golden, M.; Leiner, S.; Wyka, B.; Delzeit, H. J.; Novak, C.; Larner, G. A two one-sided parametric tolerance interval test for control of delivered dose uniformity--part 2--effect of changing parameters. *AAPS PharmSciTech* **2009**b, *10* (3), 841–849. DOI: 10.1208/s12249-009-9269-3.
15. Novick, S.; Christopher, D.; Dey, M.; Lyapustina, S.; Golden, M.; Leiner, S.; Wyka, B.; Delzeit, H. J.; Novak, C.; Larner, G. A two one-sided parametric tolerance interval test for control of delivered dose uniformity--part 3--investigation of robustness to deviations from normality. *AAPS PharmSciTech* **2009**c, *10* (3), 829–840. DOI: 10.1208/s12249-009-9271-9.
16. DeMets, D. L.; Lan, K. K. G. In *Recent Advances in Clinical Trial Design and Analysis*; Thall, P. F., Ed.; Kluwer Academic Publishers: Boston, 1995.
17. Pocock, S. J. Group sequential methods in the design and analysis of clinical trials. *Biometrika* **1977**, *64* (2), 191–199.
18. Process Validation: General Principles and Practices; Guidance for Industry. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Veterinary Medicine, 2011.
19. Ross, S. M. Introduction to probability and statistics for engineers and scientists, 6th ed.; Academic Press: London, 2020.
20. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Statist*. **1985**, *12*, 171–178.
21. Hill, G. W. ACM Algorithm 395: Student's t-Distribution. *Comm. ACM*, **1970**, *13* (10), 617–619.
22. Berger, R. L.; Hsu, J. C. Bioequivalence trials, intersection–union tests and equivalence confidence sets. *Statist. Sci*. **1996**, *11* (4), 283–319. DOI: 10.1214/ss/1032280304.