

A Special Issue Article featuring a paper presented at 11th Annual ENBIS Conference

Using DOE with Tolerance Intervals to Verify Specifications

By Patrick J. Whitcomb and Mark J. Anderson (mark@statease.com)

Given the push for Quality by Design (QbD) by the US FDA and equivalent agencies worldwide, statistical methods are becoming increasingly vital for pharmaceutical manufacturers. Design of experiments (DOE) is a primary tool because “it provides a structured, organized method for determining the relationship between factors affecting a process and the response of that process.”¹

QbD background

QbD revolves around the concept of the “design space”, which is defined by the FDA as the “multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have demonstrated to provide assurance of quality. Working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory post approval change process. Design space is proposed by the applicant and is subject to regulatory assessment and approval.”²

The FDA recommends that the design space be determined via a series of three methods:

- The first-principles approach, combining experimental data and mechanistic knowledge, is used to model and predict process performance.
- DOE is used to determine the impact of multiple factors and their interaction.
- Finally scale-up correlation is used to translate operating conditions between different scales or types of equipment.

Compared to the traditional one-factor-at-a-time (OFAT) method, DOE drastically reduces the number of runs required to determine the desirable settings. There are many different types of designs, each of which offers advantages for certain applications. We will first focus on mixture experiments because of their usefulness for pharmaceutical formulation. At this stage the use of confidence intervals (CI) will be introduced as a refinement for framing QbD design space for functional development – finding a recipe that on average should meet all the needs. Then we will transition to a response surface method (RSM) design for process optimization. This is the time to impose more conservative tolerance intervals (TI) for verifying that the design space will be robust for meeting the manufacturing specifications on every individual unit, not just on average.

Framing a functional design space for a pharmaceutical formulation

Let us consider, for example, the effect of changing the proportions of polymer, drug and three excipients on four responses in a sustained-release tablet based on a hydrophilic

polymer.³ Table 1 shows the ranges of these 5 components undergoing testing for a functional design-space determination.

In this case we assume that a quadratic polynomial, which includes second order terms for curvature, will adequately model the responses listed in Table 2. Aided by software developed of this purpose,⁴ we constructed a 25-run experiment comprised of the following points:

- 15 for the 15-parameter quadratic mixture model (selected via IV-optimality using a point exchange algorithm)
- 5 for testing lack of fit (chosen on the basis of a distance-based space filling algorithm)
- 5 replicates selected via the optimality criterion.

Table 3 shows the resulting formulations (in a standard order) and their efficacy in terms of the experimental results.

By scaling each response from zero (no good) to one (ideal) a utility function can then be computed to provide the overall desirability for any given formulation based on the model predictions. For example, as shown in Table 4, the desirability of T(50%) – the half-life of release – is 1 at a response value of 8 hours – the target. This response’s desirability ramps down to 0 (no good) as it moves away from the targeted value for drug release down to 6 hours (too fast) or up to 10 (too slow); respectively. Figure 1 flags a desirable point for a regular strength formulation with the drug dosage (component E) set to 1%. The expected performance on average falls well within the operational boundaries. However, to account for the uncertainty in the predicted responses derived from this limited experiment, we superimposed one-sided confidence intervals to frame the functional design space shown in Figure 2.

Establishing a QbD design space that falls within tolerance intervals

Ultimately the FDA and other regulators want assurance that all, or nearly all, individual units (not just on average!) will meet specification. Then the objective is to define a processing window that provides a high level of confidence, say 95%, that 99% (a generally acceptable threshold) of the population meet (or exceed) specifications. Here’s the approach taken for the statistical niceties:

- Via empirical DOE model the responses as functions of the process factors and then
- Apply a tolerance interval to “back off” (provide a buffer) from the specifications.

What most engineers and their managers will not appreciate is that a relatively large experiment will be required to create any design space meeting such a rigorous statistical requirement. Thus, it is vital to estimate the number of runs before embarking on the experiment. The following case study illustrates how response surface methods (RSM) and

tolerance intervals can be used to set an operating window where specifications are consistently met. In the process we show how to assess whether the planned experiment will produce the hoped-for result – a QbD design space that satisfies the regulatory authorities.

In this case an optimal RSM design (quadratic) is run with a tableting process on two vital factors:

- A. Granulation time: 3-7 minutes
- B. Lubrification time: 2-8 minutes

The total of these two times must not exceed the range of 7 to 12 minutes – a multifactor constraint.

Three critical responses must all meet their final-product specifications:

1. Dissolution $\geq 75\%$
2. Friability $\leq 0.5\%$
3. Hardness ≥ 10 kP

Table 5 displays the 23-run design and the experimental results. The fitted polynomials produce the predicted boundaries overlaid in Figure 3 (a graphical optimization plot). It exhibits a broad operating window. However, if the tableting process is operated on a boundary, then 50% (half!) of tablets produced fall outside the specification. To develop greater confidence that individual units fall within specifications, manufacturers must back off from the boundary by the tolerance interval (TI) – specifically the half-width. Formulas and factors (“ k_2 ”) for computing a TI can be found in standard texts,⁵ but in general it widens directly with confidence level and proportion of in-specification product. Thus, the more demanding one gets for confidence and/or the nearer to 100 percent the requirements reach, the smaller the QbD design space becomes – all else equal. In fact, we often see the space covered up entirely by the TI due to the experiment being insufficient (too small) for this purpose.

For example, consider the case at hand – the RSM on tableting process – relying on a design created in the same fashion as the previous case (formulating a tablet), that is:

- 6 points for the two-factor quadratic model
- 5 lack of fit points
- 5 replicates.

As detailed by the presentation upon which this article is based,⁶ this 16-run design suffices for functional design-space determination based on a fraction of design space (FDS) analysis.⁷ However, to maintain this FDS for a design space framed by TI, the model points must be over-specified by more than double to 13, which is why the actual experiment required 23 runs (=13+5+5).

It turns out that not only are tolerance intervals much wider than confidence intervals in general (for single samples), but they expand a bit further in the context of multifactor experiments.⁸ This TI “multiplier” (TIM) is quantified for the two-factor RSM design along the top curve in Figure 4, which also displays the single-sample TI (labeled “k2”) and, far below that, the baseline for the confidence interval (CI) – designated by the statistical factor “ $t^*(1/\sqrt{n})$ ”. (See the Appendix for formulas.) Notice that the gap does not close as sample size (degrees of freedom) increase, because, whereas confidence intervals shrink to zero, tolerance intervals asymptote to the Z score for the chosen risk (alpha) level. All of this makes perfect sense to a statistician but to put it plainly: Larger sample sizes are needed for TIs than CIs. Table 6 reinforces this message by quantifying the run increase in designs of more than 2 factors.

Returning to our case study, see the TI-framed, QbD design space in Figure 5. Within this region, the two vital process factors can vary with little risk of producing off-grade product (based on specifying TIs at 95% confidence for 99% proportion).

Summary

In conclusion, a Quality by Design design-space determination maps out a region where in-specification product can be reliably manufactured. This requires allowance for uncertainty about the boundaries by buffering them with tolerance intervals. As illustrated by Figure 6, the FDA allows engineers to create a control space that floats inside the design space – allowing for inevitable process drift from changing raw materials and other sources of natural variability.

Engineers are advised to size their experiment appropriately for the purpose:

- For functional design account for uncertainty in the form of confidence intervals (CI), which reduce the risk of mean results falling outside the allowable operating boundaries.
- To verify that product falls within final manufacturing specifications impose tolerance intervals (TI).

Keep in mind that TIs range far wider than CIs. Therefore many more runs, often about half-again, will be required to accommodate TIs for the manufacturing design space.

References

1. *Guidance for Industry, Q8(R2) Pharmaceutical Development*, U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), November 2009, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Revision 2.
2. Ibid.
3. Lewis GA, Mathieu D, Pahn-Tan-Luu R. *Pharmaceutical Experiment Design*, Informa Healthcare, USA, Inc., 2008: Chapter 10, Section II, "Second example: a hydrophilic matrix tablet of 4 variable components".
4. Design-Expert® software, version 8, Stat-Ease, Inc., Minneapolis, MN.
5. Hahn GJ, Meeker WQ, *Statistical Intervals; A Guide for Practitioners*, John Wiley and Sons: New York, 1991.
6. Whitcomb PJ, Anderson MJ, *Using DOE with Tolerance Intervals to Verify Specifications*, 11th annual meeting of the European Network for Business and Industrial Statistics (ENBIS), University of Coimbra (Portugal) September 4-8, 2011.
7. Whitcomb PJ, "Sizing Mixture (RSM) Designs," American Society of Quality (ASQ) Chemical and Process Industries Division Website <http://asq.org/cpi/>.
8. Gryze SDE, Langhans I, Vandebroek M, "Using the Intervals for Prediction: A Tutorial on Tolerance Intervals for Ordinary Least-Squares Regression", *Chemometrics and Intelligent Laboratory Systems*, 2007, 87:147 – 154.

Tables and Figures

Components	Low	High
A. Lactose	5	42
B. Phosphate	5	47
C. Cellulose	5	52
D. Polymer	17	25
E. Drug	1	2
Total = 100 wt %		

Table 1: Components and ranges (in terms of weight percent) for tableting experiment

Response	Units
1. T(50%)	Hours (H)
2. Release rate	Kinetic order
3. Hardness	kiloPonds (kP)
4. Friability	weight loss in %

Table 2: Response measures for tableting

Std #	Point Type	A: Lactose wt %	B: Phosphate wt %	C: Cellulose wt %	D: Polymer wt %	E: Drug wt %	t(50%) Hr	Release Rate Order	Hardness kP	Friability %
1	Edge	42.00	20.00	20.00	17.00	1.00	6.27	1.32	8.24	0.00
2	Edge	22.33	5.00	52.00	19.67	1.00	6.14	1.17	6.48	1.47
3	Edge	5.00	47.00	26.00	21.00	1.00	11.61	1.26	9.44	1.09
4	Edge	42.00	29.67	5.00	22.33	1.00	8.42	1.48	6.43	3.78
5	Edge	42.00	29.67	5.00	22.33	1.00	7.89	1.41	8.30	4.24
6	Edge	5.00	32.00	37.00	25.00	1.00	14.12	1.05	3.97	0.00
7	Interior	29.67	33.42	16.67	19.00	1.25	8.18	1.49	7.35	1.77
8	Interior	20.17	15.42	40.17	23.00	1.25	8.99	1.24	4.13	0.89
9	Edge	29.67	47.00	5.00	17.00	1.33	6.33	1.66	6.09	4.34
10	Plane	42.00	5.00	29.33	22.33	1.33	5.69	1.27	8.69	1.62
11	Center	23.33	25.83	28.33	21.00	1.50	8.23	1.39	7.56	0.42
12	Edge	5.00	24.33	52.00	17.00	1.67	8.27	1.55	6.48	0.00
13	Edge	5.00	24.33	52.00	17.00	1.67	7.56	1.52	6.61	0.78
14	Interior	14.17	36.42	24.67	23.00	1.75	9.16	1.34	8.39	0.00
15	Edge	17.00	47.00	17.00	17.00	2.00	4.66	1.69	6.10	0.52
16	Edge	17.00	47.00	17.00	17.00	2.00	6.02	1.68	7.22	1.39
17	Edge	30.00	5.00	46.00	17.00	2.00	5.61	1.41	9.17	0.00
18	Edge	30.00	5.00	46.00	17.00	2.00	4.53	1.58	8.04	0.00
19	Edge	42.00	30.00	5.00	21.00	2.00	3.94	1.17	6.20	4.28
20	Edge	42.00	30.00	5.00	21.00	2.00	4.56	1.29	7.79	4.55
21	Plane	22.67	40.00	10.33	25.00	2.00	7.98	1.54	8.86	1.90
22	Vertex	5.00	47.00	21.00	25.00	2.00	10.45	1.53	10.54	0.00
23	Plane	33.33	12.00	27.67	25.00	2.00	9.21	1.40	9.47	0.00
24	Edge	29.00	5.00	39.00	25.00	2.00	7.64	1.38	9.83	0.00
25	Edge	8.67	12.33	52.00	25.00	2.00	9.84	1.30	5.09	0.36

Table 3: Mixture design of experiment with 25 runs (formulations)

Response	Goal	Low	High
1. T(50%)	Target = 8	6	10
2. Shape	Minimize	1	1.5
3. Hardness	Target=6	4	9
4. Friability	minimize	1	2

Table 4: Optimization requirements

Std	Point Type	A: Granulation minutes	B: Lubrication minutes	Disso-lution %	Fria-bility %	Hard-ness kP
1	Vertex	5.00	2.00	69.30	0.30	10.50
2	Vertex	5.00	2.00	66.90	0.30	10.30
3	Vertex	5.00	2.00	67.50	0.36	10.80
4	Vertex	7.00	2.00	61.20	0.31	9.67
5	Vertex	7.00	2.00	64.40	0.30	9.55
6	Vertex	7.00	2.00	64.40	0.28	9.28
7	Interior	4.12	3.27	75.50	0.29	12.20
8	Interior	6.52	3.50	80.20	0.23	10.20
9	Interior	5.40	3.50	82.30	0.19	10.90
10	Vertex	3.00	4.00	65.80	0.49	14.10
11	Vertex	3.00	4.00	68.70	0.57	13.80
12	Vertex	3.00	4.00	68.40	0.49	13.90
13	Interior	4.80	4.87	85.90	0.30	11.20
14	Interior	4.80	4.87	87.30	0.30	11.60
15	Vertex	7.00	5.00	91.80	0.29	10.40
16	Vertex	7.00	5.00	90.40	0.35	10.50
17	Vertex	7.00	5.00	93.50	0.34	10.80
18	Interior	3.50	6.00	77.70	0.53	13.20
19	Vertex	5.27	6.73	92.70	0.45	10.90
20	Vertex	3.00	8.00	56.70	0.96	12.90
21	Vertex	3.00	8.00	60.70	0.96	12.90
22	Vertex	4.00	8.00	72.90	0.80	11.80
23	Vertex	4.00	8.00	74.40	0.77	11.80

Table 5: RSM experiment design and results

Factors	p + 5 + 5 points	Extra runs for TI	Increase in size
2	16	7	44%
4	25	10	40%
6	38	13	34%
8	55	15	27%
10	76	16	21%

Table 6: Extra runs needed for tolerance interval sizing to FDS > 90% (p = # model parameters)

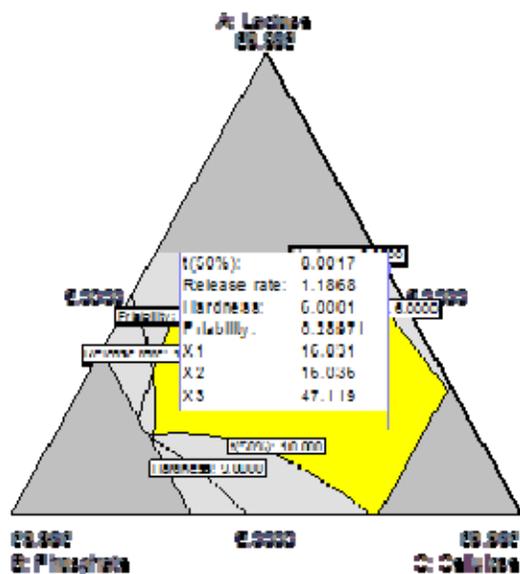


Figure 1: Graphical optimization plot showing most desirable point (1% drug level)

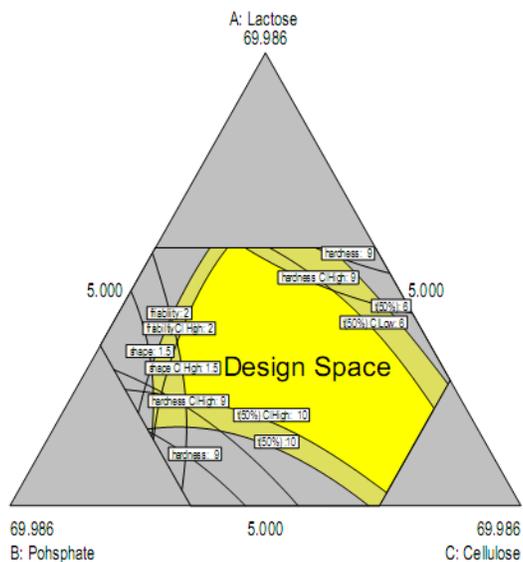


Figure 2: Functional design space framed with confidence intervals

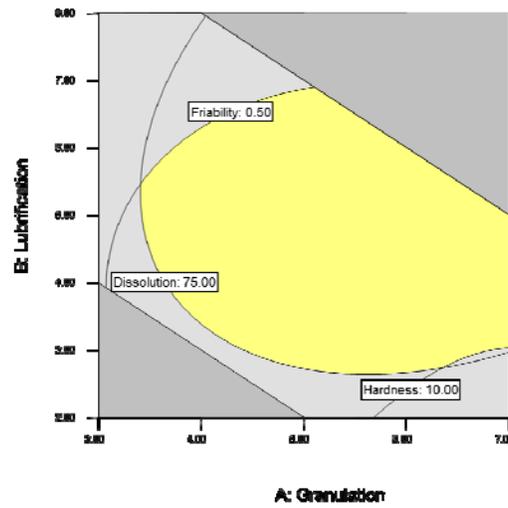


Figure 3: Specifications on graphical overlay

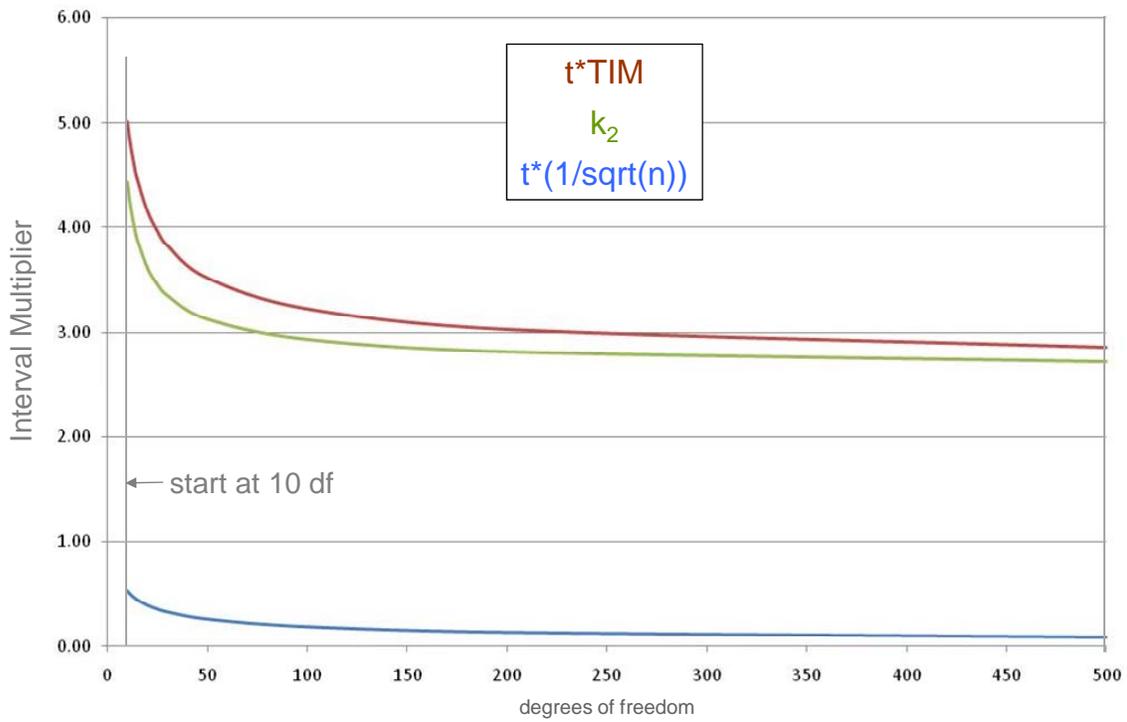


Figure 4: Relative interval widths as a function of degree of freedom

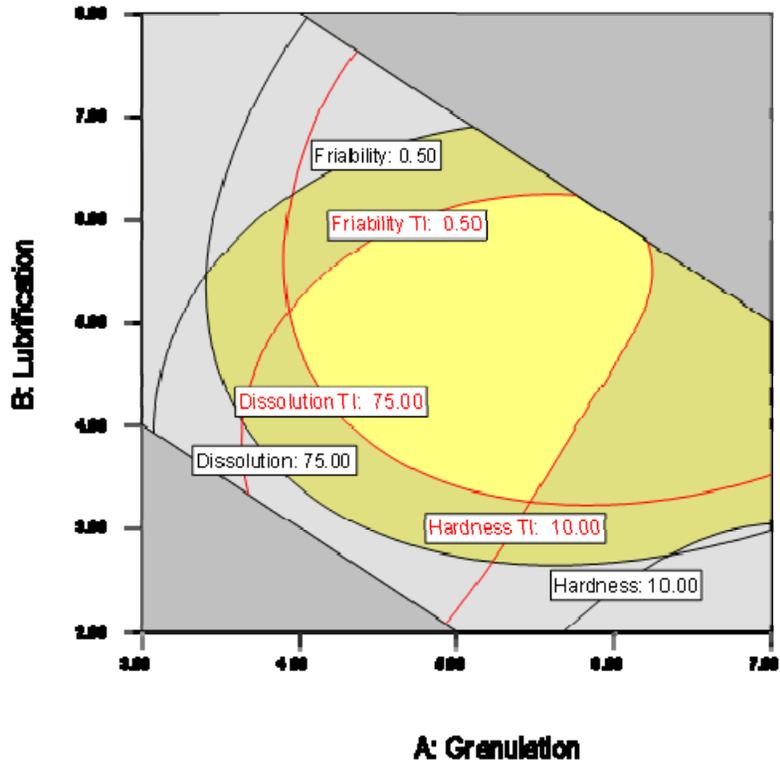


Figure 5: QbD design space for tableting process

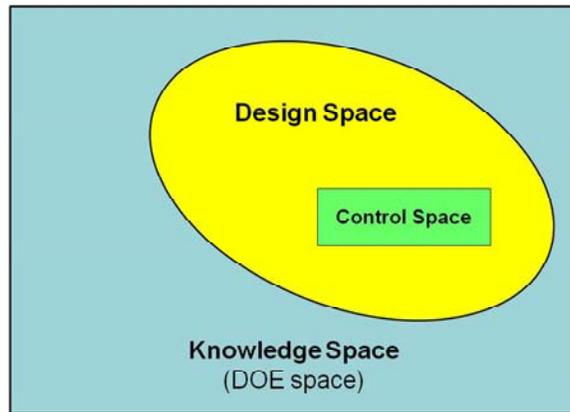


Figure 6: Control space within the design space

Appendix: Interval Formulas

Symbols:

- $\bar{}$ = mean
- $\hat{}$ = predicted
- 0 = subscript designating a point in space (a particular combination of factor levels)
- d = half-width of interval
- k_2 = factor for tolerance (depends on specified confidence and proportion)
- n = sample size
- p = parameters
- P = Proportion
- s = standard deviation
- t = t-statistic (from Student's distribution)
- t' = statistic from non-central t-distribution
- T = transpose
- x = factor
- X = x matrix
- y = response
- α = risk
- Δ = width of one-sided interval
- $\Phi^{-1}(P)$ = Inverse cumulative normal distribution
- χ^2 = Chi-square statistic

$$CI = \bar{y} \pm t \left(\frac{s}{\sqrt{n}} \right)$$

$$CI = \bar{y} \pm d$$

Equation 1: Confidence interval (CI)

$$\bar{y} - k_2(s) \leq P\% \text{ of population} \leq \bar{y} + k_2(s)$$

$$TI = \bar{y} \pm k_2(s)$$

$$TI = \bar{y} \pm d$$

Equation 2: Tolerance interval (TI)

$$\Delta_{CI} = s(t_{n-p,\alpha}) \sqrt{x_0^T (X^T X)^{-1} x_0}$$

Equation 3: One-sided CI width for multifactor experiment (for two-sided use $\alpha/2$)

$$TI = \hat{y}_0 \pm s \left[t_{\alpha, df} \sqrt{x_0^T (X^T X)^{-1} x_0} + \Phi^{-1}(P) \sqrt{\frac{n-p}{\chi^2(n-p, 1-\alpha)}} \right]$$

$$TI = \hat{y}_0 \pm (s)(t_{\alpha, df})(TI \text{ multiplier})$$

$$TI = \hat{y}_0 \pm d$$

Equation 4: Two-sided TI for multifactor experiment

$$\Delta_{TI} = s \left(-\sqrt{x_0^T (X^T X)^{-1} x_0} \right) t'_{n-p, \alpha, \frac{\Phi^{-1}(P)}{\sqrt{x_0^T (X^T X)^{-1} x_0}}}$$

Equation 5: One-sided TI width for multifactor experiment