Tolerance Intervals – an Adaptive Approach for Specification Setting

Brad Evans
Associate Director, Pharm Sci & PGS Statistics
Pfizer R&D
Fall Technical Conference 2018
Specifications…

– Ideally defined ahead of time
  • Clinically relevant
  • Build the process to meet
  • Indices to measure performance, stability, etc. (previous talk)
– Ensure safety and efficacy
– But we DON’T always know these ahead of time, so we need

– “data driven specifications”
Ideally, specifications cover the vast majority of results, so true 3 sigma limits might be acceptable.

Very wide = unacceptable?

First 10 batches much narrower than true +/- 3 sigma.

The sample limits are narrower than the process limits, so future values are out of specification.
Why isn’t +/- 3SD enough?

In the (very long) run, mean +/- 3 SD will cover 99.73% of the population if we have a “Normal” or bell-shaped data distribution.

A data-driven specification is set based on 5-15 batches (typically)

Mean +/- 3 SD almost always does the cover the data *we have in hand.*

But it needs to cover **future data, from the same manufacturing process.**

The Tolerance Interval is built to account for sampling variation.

In practice, the following variability will also occur:
- New batches of raw materials
- Changes to the assay / method transfers / site transfers.
- Process improvements
Contrast to SPC, which uses +/- 3SD

Control chart (typically)
• \( n \geq 25 \)
• Continuous process, with subsamples \((xbar, r), (xbar, s)\)
• A precise measurement system
• **Spec is pre-specified, so \( Cpk \gg 1 \) means “highly capable process”**
• Out of *Control* \(-\) go investigate

Pharmaceutical Data-driven Specifications (typically):
• \( n \sim 5-15 \)
• Batch process, \( n=1 \) measurement/batch
• Measurement device (assay) may be a significant source of variability
• **Spec is data driven, so \( Cpk \gg 1 \) amounts to “specs too wide”**
• Out of *Specification* \(-\) (potentially) dispose of the batch
Sample SD relative to true Sigma
(sample variance is scaled Chi-sq)

10% ~ ½

25% ~ 7/10ths

10% ~ 2/3

25% ~ 8/10ths

10% ~ 3/4ths

25% ~ 17/20ths

10% ~ 4/5th

25% ~ 7/8ths
In the long run....

Each connected line is the “running” standard deviation, (first five, first six, first seven..., first 50)
But there are extremes...
Why a Tolerance Interval is Proposed

- Specifications are a commitment that *future* batches will land in the specified window
- “Data based Specification Setting” – the situation where a specification is NOT known ahead of time
- A Tolerance Interval (TI) is one way to calculate a range intended to include a fixed % of the population (coverage) with some specified confidence, and depends on:
  - What data is selected
  - Confidence and coverage used to determine the multiplier
Tolerance Interval

Goal: cover 99.73% of the distribution (+/- 3 sigma)

Point estimates for coverage

Confidence limits around our estimates

For greater coverage the estimates (blue commas) would be farther apart.

Greater confidence $\rightarrow$ wider ( )

Sample mean and sample SD both will bounce around. Formula = mean +/- k*sd, $k = k(n, cov, conf)$
Interplay of (n, coverage, confidence)

Each column is a Coverage / Confidence combination, with color coding on the multiplier.

High confidence, high coverage *but small* $n$ we get multipliers of 5 or 6 or 12+ (upper right)

If we back off coverage and confidence *but large* $n$ we get multipliers < 3 (lower left)

<table>
<thead>
<tr>
<th>Cov</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conf</td>
<td>0.9</td>
<td>0.95</td>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
<td>0.9973</td>
<td>0.9973</td>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>3.49</td>
<td>4.27</td>
<td>4.16</td>
<td>5.09</td>
<td>6.61</td>
<td>5.47</td>
<td>7.88</td>
<td>6.69</td>
<td>6.37</td>
<td>7.80</td>
<td>10.35</td>
<td>12.06</td>
<td>6.77</td>
<td>8.36</td>
<td>9.73</td>
<td>6.77</td>
<td>7.22</td>
<td>8.41</td>
<td>7.52</td>
</tr>
<tr>
<td>7</td>
<td>2.90</td>
<td>3.37</td>
<td>3.46</td>
<td>4.01</td>
<td>4.61</td>
<td>4.54</td>
<td>5.50</td>
<td>5.27</td>
<td>5.29</td>
<td>6.14</td>
<td>7.22</td>
<td>8.41</td>
<td>5.98</td>
<td>6.53</td>
<td>6.37</td>
<td>5.98</td>
<td>6.53</td>
<td>6.37</td>
<td>5.98</td>
</tr>
<tr>
<td>8</td>
<td>2.74</td>
<td>3.14</td>
<td>3.27</td>
<td>3.74</td>
<td>4.15</td>
<td>4.29</td>
<td>4.94</td>
<td>4.91</td>
<td>5.00</td>
<td>5.72</td>
<td>6.49</td>
<td>7.56</td>
<td>5.41</td>
<td>5.98</td>
<td>6.97</td>
<td>5.41</td>
<td>5.98</td>
<td>6.97</td>
<td>5.41</td>
</tr>
<tr>
<td>9</td>
<td>2.63</td>
<td>2.97</td>
<td>3.13</td>
<td>3.53</td>
<td>3.82</td>
<td>4.11</td>
<td>4.55</td>
<td>4.65</td>
<td>4.79</td>
<td>5.41</td>
<td>5.98</td>
<td>6.97</td>
<td>5.18</td>
<td>5.61</td>
<td>6.53</td>
<td>5.18</td>
<td>5.61</td>
<td>6.53</td>
<td>5.18</td>
</tr>
<tr>
<td>10</td>
<td>2.54</td>
<td>2.84</td>
<td>3.02</td>
<td>3.38</td>
<td>3.58</td>
<td>3.97</td>
<td>4.27</td>
<td>4.44</td>
<td>4.62</td>
<td>5.18</td>
<td>5.61</td>
<td>6.53</td>
<td>4.99</td>
<td>5.32</td>
<td>6.20</td>
<td>4.99</td>
<td>5.32</td>
<td>6.20</td>
<td>4.99</td>
</tr>
<tr>
<td>11</td>
<td>2.46</td>
<td>2.74</td>
<td>2.93</td>
<td>3.26</td>
<td>3.40</td>
<td>3.86</td>
<td>4.05</td>
<td>4.29</td>
<td>4.49</td>
<td>4.99</td>
<td>5.32</td>
<td>6.20</td>
<td>4.84</td>
<td>5.09</td>
<td>5.93</td>
<td>4.84</td>
<td>5.09</td>
<td>5.93</td>
<td>4.84</td>
</tr>
<tr>
<td>12</td>
<td>2.40</td>
<td>2.65</td>
<td>2.86</td>
<td>3.16</td>
<td>3.25</td>
<td>3.77</td>
<td>3.87</td>
<td>4.16</td>
<td>4.38</td>
<td>4.84</td>
<td>5.09</td>
<td>5.93</td>
<td>4.72</td>
<td>4.90</td>
<td>5.71</td>
<td>4.72</td>
<td>4.90</td>
<td>5.71</td>
<td>4.72</td>
</tr>
<tr>
<td>13</td>
<td>2.36</td>
<td>2.59</td>
<td>2.81</td>
<td>3.08</td>
<td>3.13</td>
<td>3.69</td>
<td>3.73</td>
<td>4.05</td>
<td>4.30</td>
<td>4.72</td>
<td>4.90</td>
<td>5.71</td>
<td>4.61</td>
<td>4.74</td>
<td>5.52</td>
<td>4.61</td>
<td>4.74</td>
<td>5.52</td>
<td>4.61</td>
</tr>
<tr>
<td>14</td>
<td>2.31</td>
<td>2.53</td>
<td>2.76</td>
<td>3.01</td>
<td>3.03</td>
<td>3.62</td>
<td>3.61</td>
<td>3.96</td>
<td>4.22</td>
<td>4.61</td>
<td>4.74</td>
<td>5.52</td>
<td>4.52</td>
<td>4.61</td>
<td>5.37</td>
<td>4.52</td>
<td>4.61</td>
<td>5.37</td>
<td>4.52</td>
</tr>
<tr>
<td>15</td>
<td>2.28</td>
<td>2.48</td>
<td>2.71</td>
<td>2.95</td>
<td>2.94</td>
<td>3.57</td>
<td>3.51</td>
<td>3.88</td>
<td>4.15</td>
<td>4.52</td>
<td>4.61</td>
<td>5.37</td>
<td>4.44</td>
<td>4.50</td>
<td>5.24</td>
<td>4.44</td>
<td>4.50</td>
<td>5.24</td>
<td>4.44</td>
</tr>
<tr>
<td>16</td>
<td>2.25</td>
<td>2.44</td>
<td>2.68</td>
<td>2.90</td>
<td>2.87</td>
<td>3.52</td>
<td>3.42</td>
<td>3.82</td>
<td>4.10</td>
<td>4.44</td>
<td>4.50</td>
<td>5.24</td>
<td>4.38</td>
<td>4.40</td>
<td>5.12</td>
<td>4.38</td>
<td>4.40</td>
<td>5.12</td>
<td>4.38</td>
</tr>
<tr>
<td>17</td>
<td>2.22</td>
<td>2.40</td>
<td>2.64</td>
<td>2.86</td>
<td>2.81</td>
<td>3.47</td>
<td>3.35</td>
<td>3.76</td>
<td>4.05</td>
<td>4.38</td>
<td>4.40</td>
<td>5.12</td>
<td>4.32</td>
<td>4.31</td>
<td>5.02</td>
<td>4.32</td>
<td>4.31</td>
<td>5.02</td>
<td>4.32</td>
</tr>
<tr>
<td>18</td>
<td>2.19</td>
<td>2.37</td>
<td>2.61</td>
<td>2.82</td>
<td>2.75</td>
<td>3.44</td>
<td>3.28</td>
<td>3.71</td>
<td>4.00</td>
<td>4.32</td>
<td>4.31</td>
<td>5.02</td>
<td>4.26</td>
<td>4.23</td>
<td>4.93</td>
<td>4.26</td>
<td>4.23</td>
<td>4.93</td>
<td>4.26</td>
</tr>
<tr>
<td>20</td>
<td>2.15</td>
<td>2.31</td>
<td>2.56</td>
<td>2.75</td>
<td>2.66</td>
<td>3.37</td>
<td>3.17</td>
<td>3.62</td>
<td>3.93</td>
<td>4.21</td>
<td>4.16</td>
<td>4.85</td>
<td>4.11</td>
<td>4.05</td>
<td>4.75</td>
<td>4.11</td>
<td>4.05</td>
<td>4.75</td>
<td>4.11</td>
</tr>
</tbody>
</table>

Propose: Confidence increases with N Multiplier decreases with N
Two sided multipliers vs. Sample Size

The multipliers are strictly decreasing

$K_2 = \# \text{ of Standard Deviations}$

Sample size →
The confidence is strictly increasing

These are the confidence levels for 99.73% coverage

For 99% coverage they would be higher

For 95% coverage, they would be much higher

However, the limits produced would be the same because the multiplier would be the same
TI gives us a *distribution* of expected risk...

**OOS – how high / how likely**

- *Blue*: 1% chance
- *Green*: 5% chance
- *Black*: 10% chance

Business risk can now be considered

Larger *n* has less chance of high OOS
N=5, multiplier = 4.74, Confidence = 75%

Lower left tail of Coverage: risk of undercovering
# of Batches:=5

10% chance to have ~2.5% OOS

5% chance to have ~7% OOS

1% chance to have ~25% OOS
$N=10$, multiplier = $3.94$, Confidence = $76.3\%$

**Lower left tail of Coverage: risk of undercovering**

# of Batches: $= 10$
N=15, multiplier = 3.71, confidence = 77.5%

Lower left tail of Coverage: risk of undercovering

# of Batches:=15
N=20, multiplier = 3.6, confidence = 78.8

Lower left tail of Coverage: risk of undercovering
# of Batches: 20
N = 25, multiplier = 3.53, confidence = 79.7%
All combined…

For any $N$, higher confidence moves curve to the left
(\textit{but adds white space beyond current data})

For any confidence, larger $N$ makes the curve flatter / steeper
(\textit{but timelines and logistics may not allow})
Back to the example: TI Proposed Limits

Limits at +/- 3 Sample Standard Deviation
3 Sigma limits are 70-130

Conclusion: setting +/- 3 SD limits will small n very likely leads to high Out of Spec rates

This proposed approach dampens the risk

<table>
<thead>
<tr>
<th>Specification Set at N=10</th>
<th>Limits</th>
<th>Cpk at 25 Lots</th>
<th>Cpk at 50 Lots</th>
<th>Cpk at 75 Lots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean +/- 3SD</td>
<td>(80.2, 23.8)</td>
<td>0.77</td>
<td>0.70</td>
<td>0.67</td>
</tr>
<tr>
<td>Proposed TI</td>
<td>(73.8, 130.2)</td>
<td>0.91</td>
<td>0.91</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Cpk – a standard metric for how well a process can meet a Specification. Higher is better
But what about Stability data…..

• All Batches get tested upon manufacture
• Many batches are put “on stability”
  • Samples pulled, analyzed at pre-defined intervals
  • Some attributes change over time
  • Many do not change
If no trend exists, we can model the stability data as:
Attribute = Batch effect + Assay variation

By doing a variance components breakdown, we can get
Var(Total) = Var(Manufacturing) + Var(Assay)

Then we create a second Tolerance Interval using SD(Total)

It can (and does) happen that the stability data indicates a larger Var(Total) than the release data alone
Stability data – no trend

95% Assay / 5% Manufacturing
True limits are 70-130
40 data points from 8 batches
More variation seen in the post release data
Stability data (no trend)

24 Month value from blue batch – predictable

Initial value from the next batch – more variable

Despite 40 data points, there are 8 batches
When a trend exists, the proposal is:

\[ TI(\text{Release}) + \]

Estimated total change over time +

Uncertainty allowance that goes with total change

\[ TI(\text{Release}) + T_{\text{months}} \times \hat{\beta} + k \times se(T_{\text{months}} \times \hat{\beta}) \]
General workflow

**Release data only**

- Tolerance Interval = mean +/- k*SD

**Release + Stability with no discernable trends**

- Variance components analysis to quantify both process variability and assay variability. SD is calculated as the square root of the sum of these two variance components.

- T.I. = mean +/- k*SD

- Wider limits are chosen based on: TI using just Release, TI using Stability

**Release + stability with discernable trends**

- Regression + T.I., adjusting the idea in Allen Dukes Gerger (1991)
Other considerations

How to chose the slope?

One slope for all batches?

Separate slope for each batch?

Random effects models?

Bayesian perspective:
  Similar compounds
  Assay variation prior knowledge
Considerations: over-ride the default model?

Real Trend?

Or artifact of the *ALWAYS* unbalanced design?

Green batches started higher (for known reasons).

Blue started lower (known reason)
Over-ride the default model? (noisy slopes)
Summary

• Product Specific Specifications are set when possible, but there are times when Data Driven Specifications are needed

• A solution is needed, balancing:
  • Want a high chance to cover future data,
  • Limiting the “white space” beyond current data

• A tolerance interval approach, with increasing confidence and decreasing multipliers as sample size increases

• All data is considered, with or without trend

• Suggestions / comments?
Acknowledgements

Al Annamalia
Mohamed Badreddine
Anthony Carella
Aili Cheng
Brent Harrington
Charlie Kish
Greg Larner
Richard Montes
Vicki Morris
Tony Okinczyc
Vaibhav Patil
Mel Perry
Andrew Rugaiganisa
Les Van Alstine
Kim Vukovinsky
Ke Wang
Jenna Zhang


ICH Topic Q 6 A Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances
