Determining Criticality—Process Parameters and Quality Attributes Part II; Design of Experiments and Data-Driven Criticality

By Mark Mitchell

A practical roadmap in three parts that applies scientific knowledge, risk analysis, experimental data, and process monitoring throughout the three phases of the process validation lifecycle.

The most recent FDA (1) and International Conference on Harmonization (ICH) (2-4) guidance documents advocate a new paradigm of process validation based on process understanding and control of parameters and less on product testing. Consequently, the means of determining criticality has come under greater scrutiny. The FDA guidance points to a lifecycle approach to process validation (see Figure 1).

![Figure 1: Process validation lifecycle.](image-url)
In Part I of this series, the author introduced the concept of continuum of criticality and applied it to the concepts of critical quality attributes (CQAs) and critical process parameters (CPPs). In the initial phase, the CQAs had their criticality risk level assigned according to the severity of risk to the patient. Applying a cause-and-effect matrix approach, the potential impact of each unit operation on the final product CQAs was assessed and each unit operation was thoroughly analyzed for its directly controllable inputs and outputs. Finally, a qualitative risk analysis or a formal failure mode effects and criticality analysis (FMECA) was conducted for each of the identified process parameters. The purpose of this assessment is to provide a focus for the downstream process characterization work required to complete process validation Stage 1 (process design).

This initial risk assessment is performed prior to the baseline characterization work and can be used as the primary means of determining the criticality of process parameters under the following conditions:

- When a platform process that possesses similar properties and process to another commercial product (e.g., new strength or new dosage form)
- When there is a significant body of published data on the process
- When experimental studies and commercial data are available, such as when the process validation lifecycle is applied to a legacy product to substantiate the initial assessment.

In these cases, this initial assessment can be further bolstered through the addition of an uncertainty component to the traditional risk score. For example, a high-risk critical parameter with low uncertainty (due to substantial supporting data) may not require further study, but a medium-risk parameter with high uncertainty may require further experimentation to quantify the risk to product performance.

The challenge facing most organizations is how to effectively evaluate the impact of potentially hundreds of process parameters on product performance to determine what is truly critical. Few companies have the time or resources to design experimental studies around all potentially critical process parameters. The initial risk assessment provides a screening tool to sort out the parameters that have low or no risk.

**Design space and design of experiments**

The goal is to increase process knowledge by providing a mechanistic understanding of the relationship between process parameters, raw material attributes, and CQAs. This is defined as both the demonstration of impact and the quantification of the contribution of each parameter to the product’s performance. Through this exercise, it will be possible to identify the process design space. The ICH guidance defines three elements—knowledge space, design space, and control space—to establish a process understanding (see Figure 2) (2).
ICH Q8 defines design space as, “The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality.”

The design space is part of an enhanced process development approach referred to as quality by design (QbD). Prior to QbD, pharmaceutical development did not require the establishment of functional relationships between CPPs and CQAs. Consequently, process characterization experiments were primarily univariate (one factor at a time [OFAT]), showing that, for a given range of a process parameter (referred to as proven acceptable range or PAR), the CQAs meet acceptance criteria. While univariate experiments can provide some limited knowledge, a compilation of OFAT studies cannot typically define a design space because it cannot substantiate the importance or contribution of the parameter to the product CQA being evaluated. To do this, multivariate studies must be performed to account for the complexities of interactions when several CPPs vary across their control ranges.

Design spaces can be developed for each unit operation or across several or all unit operations. Although it may be simpler to develop for each unit operation, downstream unit operations may need to be included to sample and test the appropriate CQAs. For example, to perform a multivariate study on a fermentation unit operation, additional processing through cell lysis and purification unit operations is needed so that CQAs may be sampled and tested. The challenge faced by most development programs is how to efficiently and cost-effectively derive maximum process understanding in the fewest number of studies. To do this, a staged approach using multiple studies is most efficient.

**A staged design of experiment approach**

The following is an example of a simple staged design of experiment (DOE) approach. More complex DOE designs and strategies may be required, but these designs are typical:

- **Screening (fractional factorial, Plackett-Burman).** To identify or screen out process parameters that have no significant impact on a CQA. Screening designs can test main (individual impact and contribution) effects of each parameter being evaluated.
- **Refining (full factorial).** Having dropped out parameters, which do not impact the product CQAs, the refining step tests both main effects and interactions between the remaining parameters and generates first-order (linear) relationships between process parameters and CQAs. The criticality level of a CPP is
determined from the quantitative impact on the CQA shown in the modeled relationship.

- **Optimization (central composite, Box-Behnken).** To generate response surfaces and illustrate second-order (quadratic) relationships between process parameters and CQA. This analysis allows optimal set points for the design space or control space to be identified to target desired CQA (or performance attributes) values.

The DOE design assists in determining which parameters are studied and what set point value is used for each experimental run. The initial risk assessment, using prior knowledge and scientific principles, provides an expected relationship as to which CQAs and their related in-process controls will be affected by the given process parameters. Although the focus is on quality impact, process performance attributes (no quality impact) should also be sampled and measured as appropriate. This step is especially important during the optimization stage because a trade-off may be required in terms of optimizing quality and performance attributes.

For process validation Stage 1 process characterization studies, analytical methods for measuring CQAs may not yet be fully validated, but still, must be scientifically sound. The level of accuracy and precision of the analytical method or measurement system must be well understood because they directly impact the quantitative decision process when interpreting study results early in the process-design stage. Techniques of measurement system analysis such as Gage repeatability and reproducibility (Gage R&R) studies are recommended because they provide information on the variability of the measurement system. The Gage R&R study provides a quantitative measurement of the measurement tools contribution to variation for any measurement made. Typically, a percent contribution from R&R variability must be < 20% and demonstrate at least five distinct categories for the method result to be meaningful. The distinct categories are the number of discernable groups of measurements that can reliably span the range of the CQA.

Including replicate runs in addition to the study experimental design provides crucial data for estimating the underlying variability of the study. This is because, during each run, small unmeasured and uncontrolled variations always occur and may influence the result. Two otherwise identically configured runs may produce slightly different responses due to changes in environment, equipment, measurement, sampling, and operators, among others. Even deliberately fixed parameters (those not under study) may not be exactly identical from run to run. Together, these are called “noise” factors and are important in discerning true responses (i.e., “signal”) caused by the changing parameters from the inherent variability. Differences between sets of replicate runs allow for the quantification of this variability. Large changes in the responses between replicates may indicate either an unstable experimental platform (such as poor run-to-run control) or that a low-risk CPP or non-CPP, may have a higher impact on the CQAs than originally assessed.

Where a raw material has a critical material attribute (CMA) of medium-risk to high-risk impact to CQAs, it should be included as a parameter of the study where possible. Multiple lots or lots with extreme variation in CMAs may not always be available during early development or characterization studies. This limitation frequently is one of the primary drivers of establishing the continuous process verification (CPV) program in Stage 3 to monitor the future impact from this raw material variation. For large studies, multiple lots of raw materials may be required. Consideration should be given to either proportional mixing of the raw material lots for each run or use of a statistical technique called blocking, which incorporates change of material lots into the experimental design.

In each design, the choice must not only be made on the number of parameters to be studied, but at how many levels (i.e., set points within the range) and how many times a particular set of conditions is repeated (replication).

The number of levels is related to the mathematical relationship between the parameters and the CQA measured (e.g., two levels for linear or three for quadratic). For screening designs, it is typical to use only two levels (minimum and maximum of the range); for these designs, any known non-linear relationship may have
to be mathematically transformed. For refining design, center points (midpoint of ranges for all parameters) are added to estimate variability and to detect potential curvature.

Because of the cost involved or availability of API, it is not possible to perform all experimental studies at commercial scale (such as with fermentors of 5000 to 25,000 L), hence, most biotech process development programs rely heavily on modeling the process at smaller or intermediate scale. Some process parameters may be independent of the scale or may have simple models to account for scale changes. Scale itself may be considered a parameter. Establishing similar run conditions at multiple scales is an important consideration when trying to qualify the comparability between full-scale and small-scale experiments. Substantial prior experience with scaling particular unit operations may provide key information such as dimensionless parameter groups and scaling equations.

Areas considered for experimental scale include, but are not limited to:
• Aspect ratios of bioreactors and mixing tanks
• Impeller number, size, and location
• Aeration method and effectiveness of oxygen transfer
• Location of addition ports and effect on mass transport and uniformity
• Temperature control and heat-transfer surface area
• Location of instrument sensors and control-loop tuning parameters.

**Screen, refine, and optimize**
The advantage of a screening design is that it can handle a fairly large number of parameters in the fewest number of runs. The disadvantage is that the interaction effect of each CPP on a CQA cannot be directly determined because the experimental parameters are confounded. Confounding refers to a scientific state where there is insufficient resolution in the experiment to resolve the interaction effects from the main contributions of each parameter studied. However, at the screening stage, the objective is to eliminate as many parameters from the potential list of CPPs as possible so that the true process design space can be determined in the refining studies.

At this stage, the criticality of parameters has not yet been verified and parameter control ranges (proven acceptable range) have not yet been determined. Although it is usually the goal to meet the CQAs’ acceptance criteria to ensure product quality, the purpose here is to show how the process responds to the parameters even if the CQAs may not meet their criteria.
Figure 3: Screening design of experiment (DOE) Pareto of parameters for critical quality attribute (CQA) (aggregates). Temp is temperature; Osm is osmolality; Med conc is medium concentration; Inoc conc is inoculum concentration; DO is dissolved oxygen.

Figure 3 is an example output chart from a screening study. This Pareto chart shows the standardized effect or relative impact, for each of eight process parameters on a CQA. A reference line at 2.45 is the threshold below which the parameter's effect is not statistically significant for this study (p-value > alpha of 0.05). In this example, six of the parameters may be screened out of further studies, provided they do not produce significant effects for other CQAs. A similar approach can be used for process performance attributes (non-CQAs) to evaluate parameters that impact process performance, but not quality; these non-critical parameters are frequently called key parameters. If these investigations had been conducted as OFAT studies, it would be impossible to quantitatively determine which parameter had an impact on the product's CQA and to what extent. Through the use of a DOE, it is possible to measure both and define the level of variation explained by the parameters evaluated based upon the data observed.

Once the screening DOE has been completed, parameters that have not shown strong responses to any of the CQAs are now kept constant or well controlled to reduce the number of parameters for refining studies. By employing a full factorial design, all main effects and interactions are separated with regard to the CQA responses; there is no confounding in a full-factorial design. Center-point conditions (runs at the midpoints of all parameter ranges) are recommended because they can be used to detect if significant curvature (non-linear relationship) exists in the response to the parameter and they provide replication to determine the inherent variability in the study.

Figure 4: Refining design of experiment (DOE) Pareto of parameters and interactions for critical quality attribute (CQA) (impurities).

Figure 4 is an example output chart from a four-parameter, full-factorial study. The Pareto chart shows a threshold line. Two parameters and one two-factor interaction are statistically significant for this CQA. All parameters and interactions below this threshold are not statistically significant and their effects have no more impact than the inherent run-to-run variation.
Because these parameters and interactions are not significant, they may be treated as random noise and the model for this attribute is reduced as shown in Figure 5. A mathematical model was generated using the significant factors (pH and temperature) from Figure 3 and the significant interaction (pH and dissolved oxygen [DO]):

\[
\text{Impurities} = \text{Constant} + \alpha(\text{pH}) + \beta(\text{temperature}) + \gamma(\text{DO}) + \delta(\text{pH})(\text{DO})
\]

where: \(\text{Constant}\) is the intercept generated by the DOE analysis, \(\alpha, \beta, \gamma, \text{ and } \delta\) are the coefficients generated by the DOE analysis for each parameter or interaction.

Positively signed coefficients indicate the CQA increases with an increase of the parameter; negatively signed coefficients indicate the CQA decreases with an increase of the parameter. The model equation is a regression, or best fit, from the data for the experiment, and therefore, is valid for the specific scale conditions of the experiment including the ranges of the parameters tested. Models are tested for their “goodness of fit” or how well the model represents the data. The simplest of these tests is the coefficient of determination, or R-squared. Low R-squared values (such as below 50%) indicate models with low predictive capability, that is, the parameters evaluated across the defined range do not explain the variation seen in the data.

This model only represents what would be expected on average for this CQA from the unit operation(s) tested in the study. Even so, the model is a fit to the most likely mean. Recognizing that any model has uncertainty, the model can also be represented with a confidence interval (e.g., 95%) around that mean. Individual runs will also show day-to-day variation around that mean. A single-run value for the attribute cannot be predicted, but a range in which that value will likely fall can be predicted. This range for the single-run value is called the prediction interval (e.g., 95%) for the model. Empirical models such as these are only as good as the data and conditions from which they are generated and are mere approximations of the real world.

Despite the limitation, these empirical models relate not only what parameters have a statistical impact on a CQA, but also the relative amount of that impact. The range through which the parameter is tested in the study has an important relationship to the model generated. For example, perhaps the parameter temperature was initially assigned as high risk. If temperature is only tested through a tight range, the parameter may have little to no impact to CQAs in the study; its effect may be no greater than the inherent variability. If temperature is not statistically significant for the range studied (i.e., its PAR), it is designated as a non-CPP, but only for that PAR. If the temperature should ever move outside the studied PAR, there is a potential risk that it could have a quality impact become critical.
Some organization quality groups rely on the original risk assessment of the process parameter. If this parameter’s severity was initially rated high, this parameter can remain designated as critical but should be designated as a low-risk CPP as long as the parameter is in its PAR. Parameters outside the PAR would be considered outside the allowable limits for that process step because there has been no study of the parameter outside of this range.

If curvature is detected during earlier DOE stages, or if the optimization of any CQA or process performance attribute is needed, then response-surface experimental designs are used. These designs allow for more complex model equations for a CQA (or performance attributes). Two of the simpler response-surface designs are the central composite and Box-Behnken. Both designs can supplement existing full-factorial data. The central-composite design also extends the range of parameter beyond the original limits of the factorial design. The Box-Behnken design is used when extending the limits is not feasible. The empirical models are refined from these studies by adding higher-order terms (e.g., quadratic, polynomial). Even if these higher-order terms are not significant, adding more levels within the parameter ranges will improve the linear model.

Because most empirical models are developed with small-scale experiments, the models must be verified on larger scale and potentially adjusted. Applying the knowledge of scale-dependent and scale-independent parameters while developing earlier DOE designs reduces risk when scaling-up to larger pilot-scale and finally full-scale processes. The models from small-scale studies predict which parameters present the highest impact (risk) to CQAs. Priority should be given in the study design to those high-risk parameters, especially if they are scale-dependent. Since the empirical models only predict the most likely average response for a CQA, several runs at different parameter settings (e.g., minimum, maximum, center point) are required to see if the small-scale model can still apply to the large-scale process.

**Significance and criticality**

Statistical significance is an important designation in assessing the impact of changes in parameters on CQA. It provides a mathematical threshold of where the effects vanish into the noise of process variability. Parameters that are not significant are screened out from further study and excluded from empirical models.

A CQA may be affected by critical parameters in several different unit operations (see the Cause-and-Effect Matrix in Part 1 of this article [5]). Characterization study plans may not be able to integrate different unit operations into the same DOE study. Consequently, several model equations may exist for a single CQA; each model is composed of parameters from different unit operation. The relative effect of each parameter on the CQA can be calculated from these models using the span of the PAR for each parameter. The relative impact of each parameter on the CQA is based on the range of its acceptance criteria. Sorting each parameter from highest to lowest, the criticality of each parameter can be assigned from high to low. Table I is an example of one method for assigning the continuum of criticality.

The steps in determining the continuum of criticality for process parameters are summarized as follows:

- Show statistically significance by DOE
- Relate significant parameters to CQAs with empirical model
- Calculate impact of all parameter from model(s) for CQA
- Compare the parameter’s impact on the CQA to the CQA’s measurement capability
- Assign parameter risk level based on impact to CQA
- Update initial risk assessment for parameters.

<table>
<thead>
<tr>
<th>CPP risk level</th>
<th>% Change in CQA as CPP spans PAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>High risk</td>
<td>&gt; 25%</td>
</tr>
<tr>
<td>Medium risk</td>
<td>10% to 25%</td>
</tr>
<tr>
<td>Low Risk</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td></td>
<td>Below measurement capability</td>
</tr>
</tbody>
</table>
Low risk in risk assessment (not in DOE)
Non-CPP
   Not significant in DOE
   No risk in risk assessment (not in DOE)

Table I: Example of criticality risk assignment for process parameters. CPP is critical process parameter; CQA is critical quality attribute; PAR is proven acceptable range; DOE is design of experiment.

As process validation Stage 2 (process qualification) begins, criticality is applied to develop acceptance criteria for equipment qualification and process performance qualification. Finally, in process validation Stage 3 (continued process verification), criticality determines what parameters and attributes are monitored and trended.

In the third and final part of this article, the author applies the continuum of criticality for parameters and attributes to develop the process control strategy and study its influence on the process qualification and continued process verification stages of process validation.

References
2. ICH, Q8(R2) Harmonized Tripartite Guideline, Pharmaceutical Development, Step 4 version (August 2009).
3. ICH, Q9 Harmonized Tripartite Guideline, Quality Risk Management (June 2006).

Mark Mitchell is principal engineer at Pharmatech Associates.

References