Process Characterization Essentials Part I: **Process Understanding and Health Authorities Guidance**

Thomas A. Little Ph.D.
2/16/2017

President, Thomas A. Little Consulting, Bioassay Sciences
12401 N Wildflower Lane
Highland, UT 84003
1-925-285-1847
drlittle@dr-tom.com

Health Authorities Guidance on Process Characterization

There are many places in International Council for Harmonization (ICH) and US Pharmacopeial Convention (USP) guidance documents that describe the need for a well-characterized process to support the development and communication of process understanding. Demonstrating process understanding is essential for submissions to the health authorities. Understanding appropriate process parameter sensitivities, set points, ranges, and in process controls are critical to drug development. Process understanding improves confidence that the factors influencing the drug substance and drug product are well understood and maintained at acceptable levels.

ICH Q8 states (1):

> *"Process development studies should provide the basis for process improvement, process validation, continuous process verification (where applicable), and any process control requirements. Where appropriate, such studies should address microbiological as well as physical and chemical attributes. The knowledge gained from process development studies, can be used, as appropriate, to justify the drug product specification.*
>
> *"The manufacturing process development program or process improvement*

program should identify any critical process parameters that should be monitored or controlled (e.g., granulation end-point) to ensure that the product is of the desired quality."

FDA's process validation guidance states (2),

"Designing an efficient process with an effective process control approach is dependent on the process knowledge and understanding obtained. Design of experiment (DOE) studies can help develop process knowledge by revealing relationships, including multivariate interactions, between the variable inputs (e.g., component characteristics or process parameters) and the resulting outputs (e.g., in-process material, intermediates, or the final product). Risk analysis tools can be used to screen potential variables for DOE studies to minimize the total number of experiments conducted while maximizing knowledge gained. The results of DOE studies can provide justification for establishing ranges of incoming component quality, equipment parameters, and in-process material quality attributes."

PROCESS MODEL DEVELOPMENT

Knowledge about how process input (X) factors influence product (Y) responses relative to critical quality attributes (CQAs) is fundamental to defining and defending process understanding. Ultimately, knowledge must be in the form of an equation (either empirical or based on well-established scientific principles) to be useful. Process characterization equations are typically multiple factor, including main effects, interactions, and quadratic terms and may be either linear (in their coefficients) or nonlinear. Process models may also be generated based on individual measurements, statistics (mean and standard deviation), or based on the parameters of a curve (linear, exponential, square root, 4 parameter logistics (PL) curves, 5PL, etc.)

The following are common steps in developing process understanding and generating a mathematical/empirical model of the process. The primary focus of this paper is the steps associated with process model development. Below are the primary steps needed to characterize a process:

- CQAs and high-level risk assessment
- Defined process with risk
- Low-level risk assessment
- DOE or retrospective analysis
- Model building and statistical significance

- Effect size and critical parameter identification
- Model equation.

MODEL USAGE

Model usage will be discussed in Part II of this paper. Once the process model has been developed, it can be used to further optimize the process, select set-points, evaluate the design space, establish limits on the inputs and outputs from the process, and develop the control strategy. The following are some of the most common usages once the model has been developed:

- Set-point selection and process optimization
- Model verification (confirmatory experimental runs at the optimum)
- Design space and control plan
- Edge of failure analysis
- Tolerance design for process limits
- Capability and tolerance design for product quality attributes.

CQAS AND HIGH-LEVEL RISK ASSESSMENT

The first step in process characterization is to determine the business case, all relevant CQAs, and associated limits. Determine why the characterization is needed and what knowledge deficit it will fill. ICH Q9 *Quality Risk Management* (3) recommends a risk-based approach when determining which of the many unit operations require characterization. From a process characterization point of view, the question is: What are the CQAs that will be influenced by each process step? A high-level risk assessment is made from CQAs to multiple unit operations (upstream cell culture or downstream purification). The next risk question is: Which unit operation requires characterization?" If characterization is not performed, then how are prior knowledge, lack of influence, or platform approach justified or demonstrated? The result of the high-level risk assessment is a list of unit operations that require characterization to mitigate the risk (4) (Figure 1).

Drug Substance High Level Risk Assessment

Product, Project or CMC Activity:

Date:

Background, Problem, Business Objectives and Goals: **Team Leader** and **Team Members:**

What is the problem you are trying to solve, risk needing assessment? What is the background, purpose and/or goals? Begin point and end point?

Risk Question(s):

What is/are the specific product, process or assay development risk question(s) that need to be assessed?

CQAs and Unit Operation Correlation **Select High, Medium or Low to evaluate the potential influence the unit operation may have on the CQA**

Release Testing or Characterization	CQA/Assay Name	USL	Target	LSL	Unit Op 1 (replace unit op with operation name)	Unit Op 2	Unit Op 3	Unit Op 4	Unit Op 5	Unit Op 6	Unit Op 7	Unit Op 8
Release	Total Protein				High	Medium	Low					
Release	HMW											
Release	Endotoxin											
Release	HCP											
Release	Concentration											
Characterization	Mass Spec											

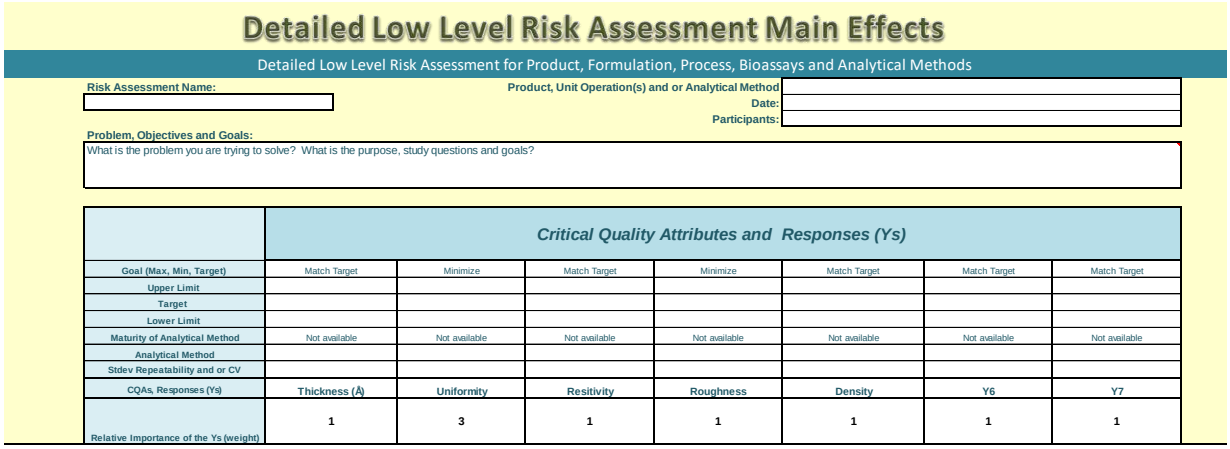
Figure 1: High-level risk assessment.

DEFINED UNIT OPERATION WITH RISK

From the high-level risk assessment, there are a defined set of unit operations with development risk that require characterization. Make sure the process is well defined/understood with all defined equipment, equipment settings of interest, sequence of operations, process holds, and materials used. Make sure critical inputs (upstream process outputs) and materials are well defined. Process definition and details will be used in the low-level risk assessment prior to DOE definition.

LOW-LEVEL RISK ASSESSMENT

Low-level risk assessments are used to rationalize the selection of factors, responses, operational ranges, and model terms to be used in the DOE (see Figure 2). A clear line of sight between CQAs and the potential impact and influence of each material and process parameter aids in parameter selection (5). Factors can be controllable (continuous, categorical or mixture) or uncontrollable (uncontrolled or covariate). From the low-level risk assessment, one should now know how to design the experiment.



Detailed Low Level Risk Assessment Main Effects

Detailed Low Level Risk Assessment for Product, Formulation, Process, Bioassays and Analytical Methods

Risk Assessment Name:

Product, Unit Operation(s) and or Analytical Method

Date:

Participants:

Problem, Objectives and Goals:

What is the problem you are trying to solve? What is the purpose, study questions and goals?

	Critical Quality Attributes and Responses (Ys)						
Goal (Max, Min, Target)	Match Target	Minimize	Match Target	Minimize	Match Target	Match Target	Match Target
Upper Limit							
Target							
Lower Limit							
Maturity of Analytical Method	Not available	Not available	Not available	Not available	Not available	Not available	Not available
Analytical Method							
Stdev Repeatability and or CV							
CQAs, Responses (Ys)	**Thickness (Å)**	**Uniformity**	**Resitivity**	**Roughness**	**Density**	**Y6**	**Y7**
Relative Importance of the Ys (weight)	**1**	**3**	**1**	**1**	**1**	**1**	**1**

Figure 2: Low-level risk assessment.

DOE OR RETROSPECTIVE ANALYSIS

Characterization can be accomplished with a DOE (prospective analysis) or via a retrospective analysis. It is generally not recommended to use retrospective analysis for process characterization as it does not allow for sufficient operational range and complex model building is typically not possible from the factors of interest.

DOE design has three components: DOE design linked to a low-level risk assessment, DOE fraction of the design space evaluation, and sample size and power analysis.

DOE Design

DOE generation needs to be linked to the risk assessments and business objectives. D-Optimal custom designs are most common depending on the problem complexity. D-Optimal designs are preferred as they place most of the runs at the corners of the design space (better signal) and more reliably estimate the coefficients of the model. I-Optimal designs are not recommended for characterization studies as they place too many runs at the center of the design space. Make sure to include factors that may affect the process at scale if the experiments are run at small scale.

Screening studies are recommended only when trying to characterize materials, and pre-DOE single factor studies are recommended when factor ranges are poorly understood prior to designing a multiple factor study. Definitive screening studies are not recommended unless they exactly match the risk assessment (no interactions, main effects and quadratic only). If studying stability, growth rates, reaction rates etc. make sure to add the multiple time points as Ys and not Xs. The time points will later be added to the model and crossed with all other factor terms in the model.

Make sure to add some additional runs to the DOE design to account for the effects of uncontrolled factors that may influence the response, can be measured during the run (*in situ*) and latter may be added to the model.

DOE Fraction of the Design Space

Programs such as SAS/JMP have tools (6) to evaluate the design the computer generated. Generally, two to three runs more than the minimum design (saturated) are sufficient to characterize the process and generate the design space. The additional runs are not center points in the design they are added to complete the design space and make it more orthogonal.

The prediction variance for any factor setting is the product of the error variance (RMSE) and the relative prediction variance computed from the DOE design. Before any DOE is run, the error variance is unknown, so the prediction variance is also unknown; however, the ratio of the prediction variance to the error variance is not a function of the error variance. This ratio, called the relative variance of prediction, depends only on the design and the factor settings and can be calculated before acquiring the data.

Fraction of the design space (Figure 3) is a good method to evaluate the entire experimental design. Good designs will have over 95% of the prediction variance below 1. A good practice is to check this plot prior to running the DOE. Add one or two more runs if more than 5% of the curve is above 1.

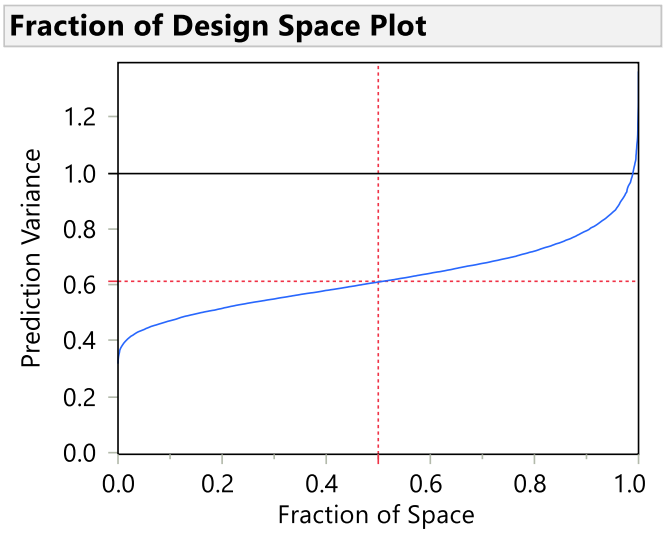


Figure 3: Fraction of the design space plot.

DOE Sample Size and Power

Power is the ability to reliably detect change in the process. The effect of the factors at the time the DOE is designed is unknown; however, it is possible for any characterization DOE to evaluate the level of signal-to-noise (SN ratio, anticipated coefficient, or *t*-test) the DOE can detect a significant signal with the associated power (likelihood of detection). Power analysis and correction are done before the study has been considered acceptable and sufficient.

Figure 4 assumes the smallest SN ratio of for all terms, including main factors, quadratics, and interactions in the model that will have power above 95%. The study design is evaluated to determine if it has sufficient power to correctly detect changes in the design space. Values of 2–3 will reliably detect weak signals from the process, 4–5 medium signals, and 6+ only strong signals. Adding additional runs will lower the SN ratio and improve power. The intercept is not a consideration in the evaluation of power. Ultimately power can be controlled two ways, 1) add more runs (reduce the noise) or 2) increase the operational range of the factor (boost the signal).

Power Analysis		
Significance Level	0.05	
Anticipated RMSE	1	
Term	**Anticipated Coefficient**	**Power**
Intercept	3	0.406
X1	3	0.997
X2	3	0.997
X3	3	0.997
X4	3	0.997
X5	3	0.999
X1*X2	3	0.992
X1*X3	3	0.992
X1*X4	3	0.995
X2*X3	3	0.968
X2*X4	3	0.992
X3*X4	3	0.992
X1*X1	5.7	0.951
X2*X2	5	0.959
X3*X3	5	0.957
X4*X4	5.7	0.952

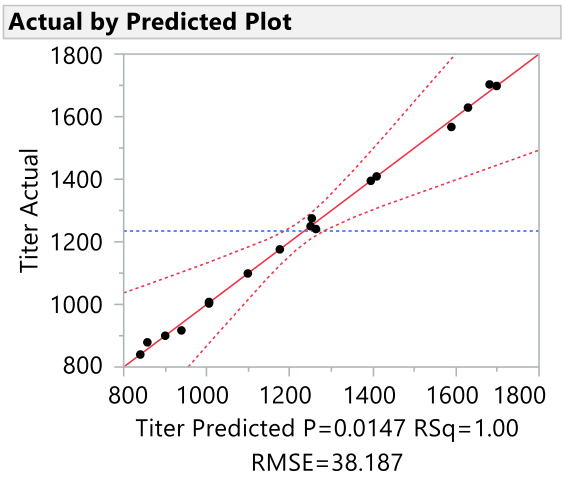
Figure 4: Power analysis.

MODEL BUILDING AND STATISTICAL SIGNIFICANCE

There are two options when building a model. Analyze the factors as an uncoded multiple regression analysis, or analyze the factors as coded, such as:

Uncoded	Coded
$y=\beta_o+\beta_1X_1$	$y=$ Mean + ½ Effect$((X_1$-Midpoint)/1/2 Range of $X_1)$

Both models provide the same estimation; however, uncoded the coefficients are incomparable and coded they are comparable as they are all in units of Y and not in change in Y relative to the change in X (slope or rate). Generally coded is preferred for characterization purposes (see Figure 5).



Summary of Fit

RSquare	0.997887
RSquare Adj	0.983093
Root Mean Square Error	38.18704
Mean of Response	1234.647
Observations (or Sum Wgts)	17

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	14	1377111.4	98365.1	67.4542
Error	2	2916.5	1458.3	**Prob > F**
C. Total	16	1380027.9		0.0147 *

Figure 5. Overall model and analysis of variation (ANOVA).

When building the multifactor model, the Adjusted RSquare measures the amount of variation explained by the change in the factors. Root mean squared error is the amount of residual variation in units. F ratio and Prob > F indicate if the model is significant (not zero).

Effect tests (see Figure 6) are used to evaluate each term in the model for significance. Non-significant factors may be removed prior to finalizing the model. Model simplification is desirable; however, not required. Model simplification improves confidence intervals and the likelihood of significance detection so it is a best practice.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Load (OD)	1	1	898729.72	616.3070	0.0016 *
Temp	1	1	15845.59	10.8662	0.0810
NaCl	1	1	2441.27	1.6741	0.3250
Flow Rate	1	1	25245.56	17.3122	0.0532
Load (OD)*Load (OD)	1	1	21134.94	14.4934	0.0626
Load (OD)*Temp	1	1	8749.10	5.9997	0.1340
Temp*Temp	1	1	479.89	0.3291	0.6241
Load (OD)*NaCl	1	1	40671.44	27.8906	0.0340 *
Temp*NaCl	1	1	97524.36	66.8777	0.0146 *
NaCl*NaCl	1	1	2099.45	1.4397	0.3530
Load (OD)*Flow Rate	1	1	8473.53	5.8108	0.1375
Temp*Flow Rate	1	1	79691.83	54.6490	0.0178 *
NaCl*Flow Rate	1	1	2314.39	1.5871	0.3348
Flow Rate*Flow Rate	1	1	142.18	0.0975	0.7844

Figure 6: Effect tests.

EFFECT SIZE AND CRITICAL PROCESS PARAMETER IDENTIFICATION

Finally, effect size and critical process parameters are identified (see Figure 7). To determine if a parameter is critical, evaluate the full effect (change in Y) and divide it by the tolerance (two sided limits), margin (one sided limit) or mean (no specification). If the resulting ratio is more than 20%, it is critical and may result in out-of-specification (OOS) events if not controlled.

Term	1/2 Effect	Prob>|t|	Multiplier	Full Effect	% of Tolerance	Critical Process Parameter
Load (OD)	278.22	0.0016 *	2	556.44	111.29	CPP
(Load (OD)-150)*(Load (OD)-150)	113.41	0.0626	1	113.41	22.68	CPP
(Temp-27.5)*(NaCl-7.35294)	96.94	0.0146 *	2	193.88	38.78	CPP
Temp	38.27	0.0810	2	76.54	15.31	Non Critical
(Load (OD)-150)*(Flow Rate-23.3824)	30.36	0.1375	2	60.71	12.14	Non Critical
(Temp-27.5)*(Temp-27.5)	12.09	0.6241	1	12.09	2.42	Non Critical
(Flow Rate-23.3824)*(Flow Rate-23.3824)	-9.22	0.7844	1	-9.22	1.84	Non Critical
(NaCl-7.35294)*(Flow Rate-23.3824)	-14.16	0.3348	2	-28.31	5.66	Non Critical
NaCl	-14.71	0.3250	2	-29.41	5.88	Non Critical
(Load (OD)-150)*(Temp-27.5)	-35.34	0.1340	2	-70.68	14.14	Non Critical
(NaCl-7.35294)*(NaCl-7.35294)	-45.19	0.3530	1	-45.19	9.04	Non Critical
Flow Rate	-50.99	0.0532	2	-101.98	20.4	CPP
(Load (OD)-150)*(NaCl-7.35294)	-59.34	0.0340 *	2	-118.69	23.74	CPP
(Temp-27.5)*(Flow Rate-23.3824)	-105.96	0.0178 *	2	-211.93	42.39	CPP

Figure 7: Critical process parameters.

MODEL EQUATION

FDA and the European Medicines Agency have requested that the model equations be added to development reports and submissions to allow the health authorities the ability to do their own modeling and simulation as wanted. Below is an example of a model from a process characterization study:

1176 + 271.15625 * ((:Name("Load (OD)") - 150) / 50) + 56.4375 * ((:Temp - 27.5) / 2.5) + -18.3566176470588 * ((:NaCl - 7.5) / 2.5) + -49.65625 * ((:Flow Rate - 22.5) / 7.5) + ((:Name("Load (OD)") - 150) / 50) * (((:Name("Load (OD)") - 150) / 50) * 113.408088235294) + ((:Name("Load (OD)") - 150) / 50) * (((:Temp - 27.5) / 2.5) * -35.3382352941177) + ((:Temp - 27.5) / 2.5) * (((:Temp - 27.5) / 2.5) * 12.0882352941176) + ((:Name("Load (OD)") - 150) / 50) * (((:NaCl - 7.5) / 2.5) * -59.34375) + ((:Temp - 27.5) / 2.5) * (((:NaCl - 7.5) / 2.5) * 96.9375) + ((:NaCl - 7.5) / 2.5) * (((:NaCl - 7.5) / 2.5) * -45.1911764705883) + ((:Name("Load (OD)") - 150) / 50) * (((:Flow Rate - 22.5) / 7.5) * 30.3566176470588) + ((:Temp - 27.5) / 2.5) * (((:Flow Rate - 22.5) / 7.5) * -105.963235294118) + ((:NaCl - 7.5) / 2.5) * (((:Flow Rate - 22.5) / 7.5) * -14.15625) + ((:Flow Rate - 22.5) / 7.5) * (((:Flow Rate - 22.5) / 7.5) * -9.21691176470584)

SUMMARY

Process characterization and model building are essential skills and required for modern drug development. Linking CQAs, risk assessment, analytical methods, DOE design, and process understanding are skills that must be nurtured and applied within the development team. Generation of a reliable process equation that models the process variables and provides detailed process understanding is the goal of process characterization.

Part II of this paper will carefully explore the use and application of the model developed from a well-characterized process.

References:

1. ICH, Q8(R2) *Pharmaceutical Development* (ICH, 2009).FDA,

2. *Process Validation: General Principles and Practices*, Stage 1 Process Design, Building and Capturing Process Knowledge and Understanding (CDER, January 2011).

3. ICH, Q9 *Quality Risk Management* (ICH, 2006).

4. T. Little, Quality Risk Management Templates, High-Level Risk Assessment, http://qualitybydesignconsulting.com/tools/index.php

5. T. Little Low-Level Risk Assessment, http://qualitybydesignconsulting.com/tools/index.php

6. SAS/JMP 13.0, www.jmp.com