Inference for Quality-by-Design: Application in the Life-Sciences Industry

Paul van Eikeren, Ph.D.

January 2008

inference for Quality-by-Design

Quality-by-Design

Role of Quality-by-Design in Life-Sciences Industry

Quality-by-Design (QbD) is a proactive, scientific, risk-based paradigm for reliably producing high-quality products. QbD applies quality best practices to development and manufacturing processes and shifts the focus from quality-by-testing to quality-by-design. The U.S. FDA has launched a campaign to promoted QbD across the life sciences industry. Application of QbD speeds up the FDA product review and approval process and has been shown to facilitate swifter approval of original product improvements and process definitions. This helps life-science companies bring new products to market faster and at sharply reduced cost.

The central objective of QbD is to develop product and process understanding. It starts by defining the desired product up front, which involves identifying and defining critical to quality attributes. Quality-by-Design then uses a systematic and scientific approach to achieve process understanding that includes identifying how variability in raw materials and process execution affects performance and quality.

QbD introduces the concept of a Design Space—that is, the multidimensional combination and interaction of input variables (e.g., material) and process parameters that have been demonstrated to provide assurance of quality. Identifying and verifying the Design Space is a principal task of QbD. Three approaches, used alone or in combination, are commonly used. One, a first-principles approach that combines experimental data and mechanistic knowledge, physics and engineering to model and predict performance. Two, a statistically designed experiment approach that employ efficient methods to determine the impact of multiple parameters and their interactions. And three, a scale-up correlations approach that employ a semi-empirical approach to translate operating conditions between different scales or pieces of equipment.

Benefits of Implementing Quality-by-Design

Implementing Quality-by-Design provides significant business benefits including the following:

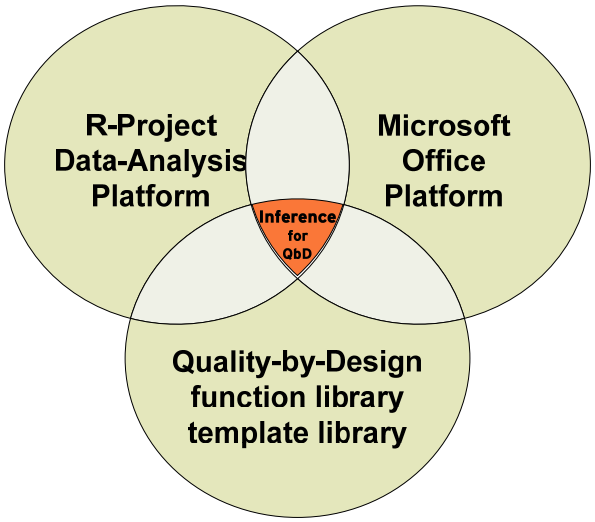
- products designed to meet intended use and corresponding processes designed to consistently meet product critical-to-quality attributes;
- product specifications based on rational, scientific principles;
- increased manufacturing efficiency resulting from reduced cycle times, reduced work in progress, and increased capacity;

- reduced manufacturing cost by using optimized and robust processes that prevent rejects, scrap and reprocessing;
- regulatory flexibility, provided that you are working within the design space, as working within the design space is not considered a change; and
- consistent and improving quality over time using the capability to continually monitor, refine and update the process.

Inference for Quality-by-Design

Overview of Software Solution

Inference for Quality-by-Design (Inference for QbD) is and integrated collection of software and services to enable scientists, engineers and plant people involved in research, development and manufacturing to implement of quality by design practices. It is based on the Inference solutions platform (www.inference.us) developed by Blue Reference, Inc (www.BlueReference.com).



Inference for QbD is comprised of a collection of Desktop Applications, Inter-Viewers, Enterprise Services and Supporting Tools, which bridge Quality-by-Design from development to manufacture. Inference for QbD is built on top of two familiar software platforms: one, Microsoft Office as the user interface, which extends the QbD approach to non-expert users; and two, the R-project data analysis platform as the computational engine, which leverages the "best-of-breed" data analysis platform. Inference for QbD inherits required QbD intelligence from custom libraries of QbD data-analysis functions and QbD application templates.

Business Value of Inference for QbD

Deployment of Inference for QbD provides customers with immediate business value in three principal ways:

1. **Extract product and process knowledge from data.** Product and process knowledge are the basis for development decisions in life-science companies. Inference for QbD

transforms large volumes of low value multi-dimensional data to focused high-value product and process knowledge.

2. **Reduce the time and cost for acquiring product and process knowledge.** Inference for QbD employs modern machine learning techniques to achieve highly efficient and effective extraction of product and process knowledge from data by enabling the following:

   - ***Leveraging existing data***. Inference from QbD can extract valuable process knowledge (data mining) from the rich stores of prior data from both successful and failed development candidates.
   - ***Reducing number of experiments***. Inference for QbD employs adaptive experimental design, which uses existing data in conjunction with limited data from new experiments to achieve the requisite process knowledge.
   - ***Automating QbD workflow***. Inference for QbD improves the productivity of the scientist and engineer by automating the data preparation, documentation, analysis and reporting involved in QbD workflow.
   - ***Streamlining decision making***. Inference for QbD enables the scientist and engineer to be in control of the knowledge extraction process while minimizing the need for IT staff involvement.

3. **Provide an effective vehicle for deploying QbD practices in accordance to FDA guidance.** Inference for QbD enables the application of a structured and consistent approach to QbD across the entire development enterprise by supporting collaboration, searching, version control and security. Additionally, Inference for QbD provides the supporting elements needed to achieve compliance with FDA guidelines for CPG 7132c.08, CMC filing, ICH Q8 and 21 CFR Part 11.

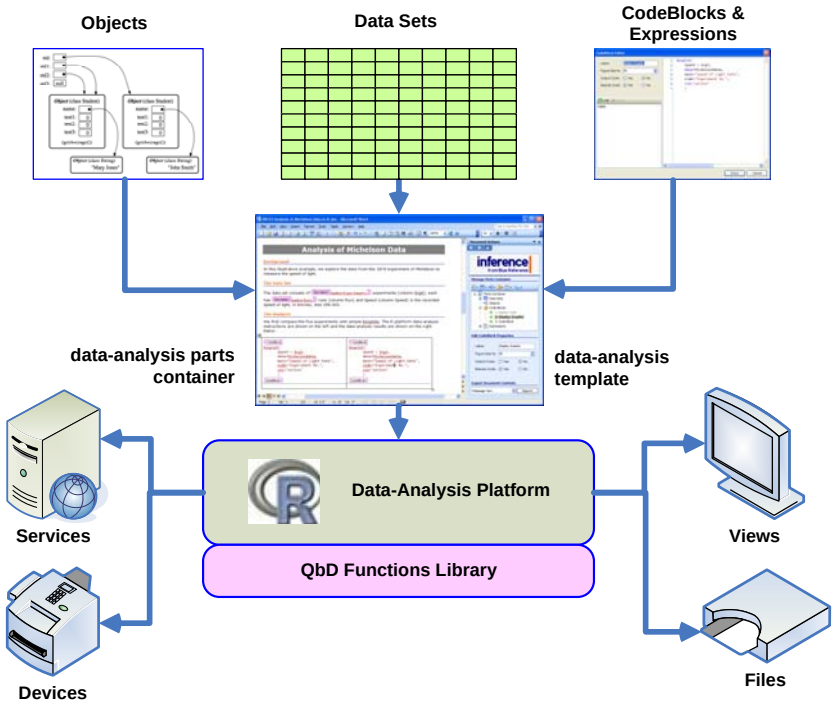
Scope of Inference for QbD

Inference for QbD encompasses all functional elements involved in life-sciences product development from candidate nomination to product launch. Functional areas served by Inference for QbD include the following:

- **Pharmaceutical Development** including Analytical R&D, BioProcess R&D, Chemical Process R&D, Pharmaceutical R&D, Regulatory Affairs and Supporting Function;
- **Pre-Clinical Pharmacology** including Pharmacodynamics and Pharmacokinetics; and
- **Pharmaceutical Manufacturing** including manufacturing process improvments, manufacturing production control, raw material variation control, and real-time process monitoring.

Inference for QbD Software Solution

Built on the Inference Solutions Platform



Inference for QbD is built upon the patent-pending* Inference Solutions Platform (www.inference.us), which is comprised of the following five elements:

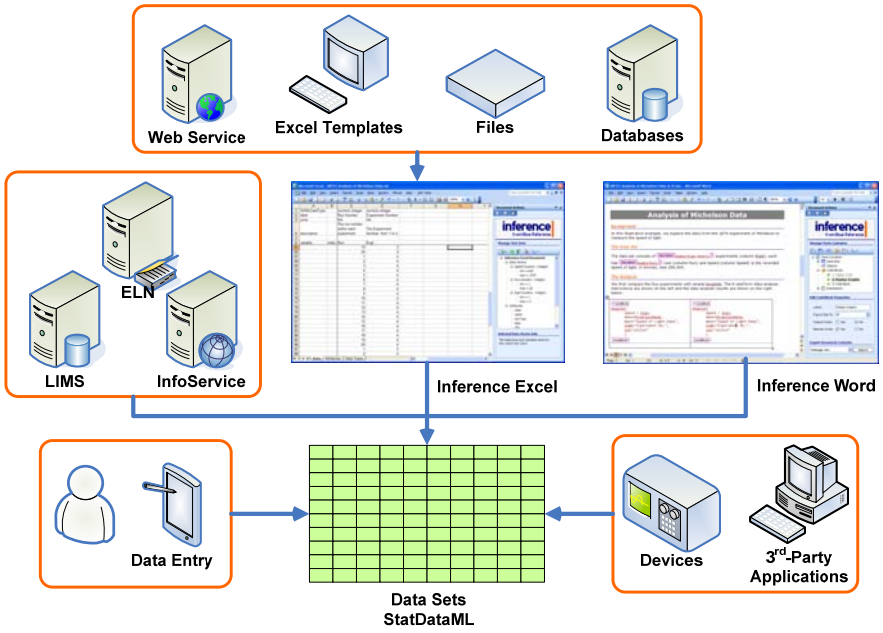
- **Inference Word**. Word 2003/2007 is the principal interface for the user. Word has been extended to support a data-analysis parts container for holding objects, data sets, codeblocks and expressions. The combination of Word text, formatting and the data-analysis parts container comprises a data-analysis template. Such an Inference Word template can be executed in performance of QbD workflow.
- **Objects, CodeBlocks and Expressions**. A collection of computer-readable, scripting instructions that can be executed by the R Data-Analysis Platform.
- **Data Sets**. A collection of data, meta-data, and parameter values processed by the R Data-Analysis Platform in performance of QbD workflow.
- **R Data-Analysis Platform**. The computational engine that executes the collection of scripting instructions.
- **QbD Functions Library**. Pre-built data analysis primitives specifically designed for QbD needs, which can be assembled into CodeBlocks and Expressions for use in QbD workflow.

* U.S. Patent Applications including the following: 11/619,316; 11/619,326; 11/619,336; 11/619,343.

Execution of QbD workflow entails processing the data sets in Inference Word according to the scripting instructions contained in the CodeBlocks and Expressions using the R Data-Analysis Platform in conjunction with the QbD Functions Library to output a results document. Such a results document can take alternate forms depending on the selected export service, which includes views of the results (e.g., web page, Word print preview), files for storage (e.g., PDF), services (e.g., results record to ELN), and devices (e.g., work list to a HPLC controller).

Data Sets Are Pluggable Entities

To be effective in a life-sciences product-development environment, Inference for QbD has to exhibit extreme flexibility with regards to data sources and integration with existing systems and software applications. Towards that end, Inference for QbD uses the open-standard, Statistical Data Markup Language (StatDataML), as its core data format for data sets. StatDataML is a table-oriented data structure similar in concept to an Excel worksheet. StatDataML is an XML-based data format that fully supports the capabilities of the leading data-analysis platforms (e.g., R, S-Plus, Matlab, JMP, SAS, SPSS), scripting platforms (CPython, Ruby, IronPython, IronRuby, dynamic VB, JScript) and computer languages (e.g., C#, VB.NET, F#, and Java).



Inference for QbD provides a range of tools for accessing data sources to assemble and prepare data sets. A principal tool is Inference Excel, which is comprised of Excel that has been extended with an embedded container for holding StatDataML. Inference Excel can be used to assemble and prepare data sets using Excel's built-in data import capabilities from

Excel templates and worksheets, files, databases, web pages and web services. Additionally, Inference Word can export any computed data objects as StatDataML data sets, or alternatively, as Inference Excel documents. Blue Reference also provides a StatDataML software developer's kit that Visual Studio developers can use for converting foreign data streams to StatDataML. Such foreign data streams could originate from data repositories (e.g., ELN and LIMS), devices (e.g., HPLC, instrument output), 3rd party application (e.g., JMP DOE software), or proprietary data entry applications (e.g., Compose desktop application from Symyx ELN).

Employs FDA-Accepted R Data-Analysis Platform

Inference for QbD employs the R data-analysis platform for statistical and graphical computation, which is the premier platform for research and development in statistics. The R data-analysis platform is based on the S-language developed at ATT labs. The R data-analysis platform is comprised of over 1,000 core statistical functions and has available over 1,200 user-developed packages (libraries) of custom functions. The R data-analysis platform is available as an open source implementation (R-project), as supported open source implementations (R+, RPro, RStat), and as a commercial implementation (S-Plus). The US FDA has granted approval for use of the R Data-Analysis Platform on FDA's locked desktop computers. The US FDA has accepted internal use of the R Data Analysis Platform as reflected in the following: (1) it provides an internal course on R for FDA reviewers; (2) it provides an internal tutorial on Statistical Graphics with R for use with clinical trials; and (3) it is engaged in an initiative to use R as the centerpiece of an "Open Toolbox" for clinical statistics. Efforts at the R-project are under way toward providing a validated R. Efforts include providing a documented Software Development Lifecycle and system documentation.

Provides Extensive QbD Functions Library

Inference for QbD includes a comprehensive library of functions comprised of parameterized computational components for use in QbD workflow. Function libraries are extensible and pluggable (packages). And, the user can build their own functions from the "ground up" or assemble from existing primitives and functions.

Export Services Are Pluggable Entities

Execution of an Inference Word document outputs a results document and serves as an integration point with other systems. Results document take alternate forms depending on the requirements of the export service selected. Export services are pluggable entities—that

is, we will continue to provide new export services and new export services can be created using the Inference SDK. Representative export services include the following:

- **Views** (previews in Inference): in browser; in Word; in Acrobat
- **Files** (saved to hard disk): Word (*.doc, *.docx); Acrobat (*.pdf); CodeBlock/Expression Scripts (*.R); Inference Excel (*.xls, *.xlsx); Inference Studio (*.infcontainer)
- **Services** (sent to a remote system): Inference Repository; Nugenesis EDMS (in development); Symyx ELN (in development); CambridgeSoft ELN (in development)
- **Devices** (send to a remote device): Agilent ChemStation (in development)

Includes Pre-built QbD Template Applications

Inference for QbD "out-of-the-box" includes a comprehensive library of pre-built QbD template applications that comprise commonly used QbD workflows. Illustrative classes of template applications include the following:

- **Automated Data Analysis.** Useful when significant input is not required and the same computational protocol needs to be applied many times to experimental designs and data structures that are well-defined. May include built-in explanations of the data-analysis results. Examples include measurement capability analysis, data profile analysis, and analysis for identification of critical process parameters.
- **Scheduled Reports.** Provides static reports that need to be updated on a regular basis. Report integrates data source, free text, computation and formatting. Examples include reports on accelerated API and drug product stability studies.
- **Custom QbD Desktop Applications.** Building inference templates with embedded objects provides a means to execute parameterized computations. For example, a model object that enables forecasting of the performance of a production process as a function of material and process attributes can be added to a data-analysis parts container and accessed via CodeBlocks. Such an Inference Template can be used to deploy simulation models to end users for learning about and probing the behavior of the production process.
- **Streamlined QbD Workflow Applications.** Many QbD workflows involve collaboration among multiple workers, functions and settings. Templates for such applications require long-running and complex workflows involving a centralized information repository with workflow management. Examples include analytical method validation and process validation workflows.

Offers Content and Collaboration Management

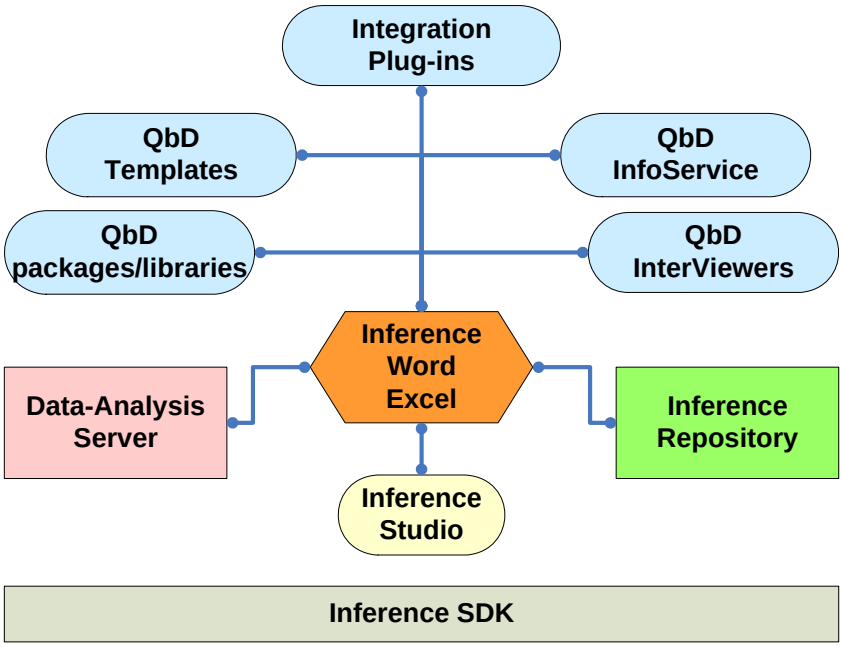
Inference Repository is an integrated suite of server capabilities that can improve QbD effectiveness by providing content management and enterprise search for QbD templates, documents and records, accelerating shared QbD business processes, and facilitating QbD information sharing across the development enterprise. Inference Repository is built upon Microsoft SharePoint Services 3.0 in Microsoft Server 2007, which enjoys over 100 million users, and provides capabilities and corresponding benefits including the following:

- **Manage QbD content and processes**
  - control Inference for QbD templates, documents and records through detailed, extensible policy management
  - centrally store, manage, and access QbD templates, documents and records across the enterprise
  - extend QbD business processes across the enterprise in a consistent fashion
  - streamline everyday QbD workflow
- **Improve QbD implementation**
  - present QbD information in one central location
  - quickly connect people with QbD information
  - share QbD data broadly while helping protect sensitive information
  - take advantage of unstructured QbD personnel relationships to drive better decisions
  - provide infrastructure needed to achieve FDA 21 CFR Part 11 compliance
- **Simplify internal and external collaboration**
  - enhance customer and partner QbD relationships
  - provide ubiquitous enterprise search for QbD information
  - enable QbD workers to work when and where they want
  - connect and share QbD knowledge
  - personalize QbD operations
- **Empower Development IT to make a strategic impact**
  - enable Development IT to focus on more strategic tasks
  - simplify deployment, management and system administration of Inference for QbD
  - enable Inference for QbD system monitoring and usage tracking

Deploying Inference for QbD

Whole Product Solution

Inference for QbD is not just a piece of software; it represents a "whole product solution"—that is, an integrated and comprehensive framework of software and services that fulfills the life-science customers expectations for implementing Quality-by-Design practices. The whole products is comprised of the core desktop applications (Inference Word/Excel/Studio), Inter-Viewers (dfExplorer, modelExplorer), Enterprise Services (Inference Repository, Data-Analysis Services, and QbD InfoServices), and Supporting Tools (Inference Utilities, Integration Plug-Ins, SDK, QbD Function Library, and QbD Template Library).



Flexible Licensing Structure

An Inference for QbD deployment involves four distinct, stakeholder groups, which are briefly described in the text and table below:

Specialists supply QbD strategies to QbD Power Users. Deployment of Inference for QbD enables specialists to focus on their highest-value contribution—the creation and verification of new design and analysis strategies using the R data-analysis platform and the Inference SDK.

Power Users are implementers and provide the link between the specialists and the End Users. Deployment of Inference for QbD enables Power Users to focus on their highest-value contribution—the creation and maintenance of QbD application templates for end users.

End Users, who represent the largest group, run QbD templates and analyze data looking for patterns. Their focus is transparent interpretation of the data, trustworthy open access to results, and transfer of knowledge to team members and stakeholders. Using Inference for QbD enables them to leverage and scale-up the knowledge and skills of Power Users.

Managers view QbD reports automatically generated by Inference for QbD templates. Their interest in Inference for QbD is related to impact, reduced cost, transparent communication, speed of data and information to process understanding, and increased productivity of their staff.

Inference for QbD Audience Groups				QbD Task Areas
Depth of Knowledge:	**Office**	**DAP**	**QbD**	
Specialists (3%)	**M**	**H**	**L**	
• statistician • chemometrician • process modeler • chemical engineer • software developer				• construct new design and analysis strategies • pursue integrations and extensions • create demonstration templates • create verification templates • maintain function template library
Power Users (12%)	**H**	**M**	**H**	
• domain experts • product experts • process experts • analytical dev experts • QbD experts				• maintain QbD template library • create QbD application templates • create QbD reporting templates • create QbD training/tutorial templates • create QbD validation templates
End Users (80%)	**M**	**L**	**M**	
• R&D scientist & engineers • pilot plant workers • manufacturing workers • QA/QC workers				• apply QbD application templates • modify QbD application templates • modify QbD reporting templates
Managers (5%)	**L**	**L**	**L**	
• development managers • quality managers				• run QbD reporting templates

On the basis of the above, the four stakeholder groups are distinguished on the basis of the functions they perform, the benefits they derive from using Inference for QbD, and the elements of the Inference for QbD whole product they need to accomplish their goals. Accordingly, Inference for QbD is licensed under a four tier model, consistent with the differential function, benefit and need of the four stakeholder groups.

Vehicle for Implementation of QbD Practices

Inference for QbD provides a powerful vehicle for enterprise implementation of QbD practices by enabling appropriate QbD workflow in accordance to FDA guidelines.

QbD Actions: Inference Features	Contribution to Process Understanding
Analyze and Optimize Measurement Capability	
• Gage repeatability and reproducibility analysis • Repeated measurements study (ISOPLOT) • Sample size calculations • Measurement optimization (e.g., Simplex)	• Make reliable measurements sufficient to support needed decisions. • Estimate measurement precision (repeatability, reproducibility), accuracy, linearity (in accuracy and precision), and stability. • Discover and characterize sources of measurement variance.
Analyze Raw Material Variation	
• Principal component classification • Random Forest classification • Bayesian ensemble classification	• Identify variation in raw materials due to vendor differences and lot-to-lot differences. • Relate raw material variation to product quality.
Perform Surrogate Critical to Quality Attribute Analysis	
• MVDA/chemometrics modeling and validation • Multivariate calibration • Linear and nonlinear modeling • Random Forest modeling • Bayesian ensemble modeling • Multiway PLS modeling	• Use nondestructive and rapid analysis methods (e.g., FT-NIR) in conjunction with calibration as surrogates for multiple slow analyses of critical quality attributes (e.g., potency, dissolution, disintegration). • Identify quality surrogates for clinical performance. • Obtain time/cost savings for routine release testing and formulation development support.
Visualize and Analyze Historical Process Data	
• Data mining using multivariate approaches • Visual data analysis (e.g., Trellis, Parallel coordinate; box plots) • Data profiling and summary statistics • Partial Least Squares regression analysis • Random Forest analysis • Principal component analysis • Recursive partitioning analysis • Bayesian analysis • Survival analysis (e.g., stability) • Time-series analysis	Employ powerful, easy-to-use methods for • using historical information to establish prior knowledge and guide direction of future experimentation, • using literature, internal R&D, and manufacturing information • identifying Critical to Process Parameters (CPPs), • identifying functional relationships between parameters, • identifying sources of variability in Critical Quality Attributes (CQAs), and • identifying variability control strategies.

Construct Experimental Designs	
• classic DOE designs • algorithm-based designs • adaptive designs	• Employ the most effective method to design experiments which address process understanding questions. • Employ a strategic platform for scientific decision making based on cause-and-effect relationships.
Analyze Experimental Design Execution	
• linear, interaction and quadratic regression models • Partial Least Squares regression models • Bayesian models	• Analyze and visualize results of executing experimental designs. • Determine impact of parameters and their interactions. • Build quantitative models for quality performance of materials and processes.
Define, Test and Visualize Design Space	
• Modeling of critical product quality attributes • Trajectory, contour and 3D plots • Prediction of quality parameters • Monte Carlo simulations	Implement multivariate analytic and mechanistic approach to establish design space and specify control space to ensure specific product quality and consistency using • first-principle approach to model and predict performance, • statistically designed experiments that determine impact of multiple parameters and interactions; and • semi-empirical material/process correlations.
Carry Out Technology Transfer and Scale-Up	
• Product and process design and optimization • Nonlinear systems analysis and modeling • Reaction kinetics analysis and modeling • Batch and continuous process simulation	Use an integrated set of tools for • batch and continuous process simulation, • first-principle and empirical determination of scale-up correlations, and • optimization of process conditions.
Monitor Design Space and Process Robustness	
• Control strategies • Run chart/trend chart • Control charts • Process capability (Cp; Cpk) • PCA, Hotelling's T2 and Q (squared prediction error)	Scope/leverage design space and process robustness by • analyzing process variability and its sources, • defining and testing process robustness by simulation, and • identifying conditions that minimize effect of material and process variables.
Perform Manufacturing Process Improvements In Design Space	
• Pareto charting • Design of Experiments • Regression/correlation/ANOVA • R/T-tests/F-tests • Multidimensional correlations • Batch statistical process control • Multivariate statistical process control • Survival analysis	Employ tools for continuous improvement that capitalize on the Quality by Design approach for • implementing and leveraging regulatory flexibility afforded by the use of design space, and • control of material and process variables to obtain desired quality output.

For More Information

Visit the Inference for QbD website at http://www.inferenceForQbD.com. This Web site features the latest news and information about the Inference for QbD solution, including product information, screencasts, case studies, white papers, information about related technologies and more.

To obtain information on the Inference solution platform, which is used to build the Inference for QbD solution, visit the Inference website at http://www.inference.us.

For additional information contact via e-mail:

Paul van Eikeren, Ph.D.

paul.van.eikeren@bluereference.com