PharmaSUG 2019 - Paper SI-327

Begin with the End of Validation: Adapting Quality-by-Design Approach in Statistical Programming to Ensure Quality and Compliance Excellence

Linghui Zhang, PRA Health Sciences

ABSTRACT

In many ways, quality and compliance are essential in the pharmaceutical industry. Statistical programming conducts the variety of activities of data management, analysis, and report in the entire data flow in drug development, including preclinical and clinical research, regulatory submission, and post marketing surveillance. Quality and compliance are both crucial components of statistical programming. To achieve high quality, validation process is developed to identify data issues, and is usually performed before data and data-related products are released finally. However, the validation process is time- and resource-consuming. It's quite challenge to fix the data issues after validation in complex clinical trial programs, under tough timeline and temporary workforce shortness.

In order to "get it right first time", Quality-by-Design (QbD), a process oriented method was applied to manage risks in quality and compliance, and to advance product and process quality in statistical programming. QbD is a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and control based on sound science and quality risk management. QbD is also a regulatory expectation. FDA and ICH published several guidelines to direct pharmaceutical manufacturing.

The paper talks about adapting QbD approach in statistical programming to identify critical issues relevant to quality and compliance prospectively. The QbD elements and steps will be introduced and followed by the challenges of implementing QbD approach in statistical programming. This paper is designed to the programming leads and managers, but all levels of programmers can benefit from learning the QbD approach.

INTRODUCTION

Quality-by-Design (QbD) is not a new concept and has been wildly used to advance product and process quality in many industries. QbD was first proposed by quality guru Joseph M. Juran in 1992 and its basic principles include that quality could be planned, and that most quality crises and problems relate to the way in which quality was planned in the first place. FDA adopted QbD principals and approaches to ensure pharmaceutical quality throughout in drug development and manufacturing.

Quality is the fundamental of the safety and efficacy for every single drug product. Both regulatory agencies and pharmaceutical sponsors are passionate about quality to assure consumers have access to medicines with desired performance. To meet the needs of pharmaceutical market and ensure the regulatory policies are rooted in the cutting edge of pharmaceutical sciences, in August 2002, FDA announced a significant initiative, "Pharmaceutical Current Good Manufacturing Practices (CGMPs) for the 21st Century: A Risk-Based Approach". In this initiative, FDA proposed to integrate quality systems and risk management approaches into its existing programs to facilitate industry application of modern manufacturing technologies. The quality management system is based on science and risk management is the concept pf QbD. With that goal, QbD was introduced as a systematic method for manufacturers to achieve the desired quality.

ICH Quality Vision came on the heels of FDA initiative in July 2003 to "develop a harmonized pharmaceutical quality system applicable across the lifecycle of the product emphasizing an integrated approach to quality risk management and science" [(Brussels July 2003)]. Soon after that, FDA and ICH made an agreement to develop an internationally harmonized plan for developing a pharmaceutical quality system based on an integrated approach to risk management and science. ICH established three Expert Working Groups to develop a harmonized pharmaceutical quality system based on an integrated approach to risk management and science, then published three guidelines Q8, Q9, and Q10.

ICH Q8 introduced the basic principles in the life-cycle of pharmaceutical product under the guidance of QbD approach. The good practices for pharmaceutical development was described and incorporated with the elements of risk and QbD approach. ICH Q9 introduced the Quality Risk Management (QRM) which was integrated into decisions by industry and regulators regarding quality, including CGMP compliance. ICH Q10 described the Pharmaceutical Quality System (PQS) that facilitate establishment and maintenance of a state of control for process performance and product quality, and continual improvement. Implementation of the ICH Q8, Q9 and Q10 guidelines can improve drug quality and efficiency of pharmaceutical manufacturing throughout product lifecycle.

QBD APPROACH

The definition of QbD and relevant concepts are listed in Table 1 (Annex Glossary in ICH Q8 R2, 2009). To achieve the highest level of quality, the foundation of quality management is moving beyond Quality by Testing (QbT) to QbD. As ICH Q8 stated, "*It is important to recognize that quality cannot be tested into products, i.e., quality should be built in by design.*" Planning with the end in mind is the basic principle of QbD and sets the strategic foundation for drug development and process. In a QbD system, the quality of a drug product is achieved by the desired performance which is predefined to meet patient requirements on safety and efficacy. The quality profile and factors significantly impacting on the product quality are identified and evaluated. After well understanding the product design, the product process can be designed to consistently meet the requirements of quality and control these variations impacting on quality. As the result, the quality is built in throughout the product development, but not just relied on testing the end product alone.

Quality: The suitability of either a drug substance or a drug product for its intended use. This term includes such attributes as the identity, strength, and purity (ICH Q6A).
Quality by Design (QbD): A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management. (ICH Q8)
Quality Target Product Profile (QTPP): A prospective summary of the quality characteristics of a drug product that ideally will be achieved to ensure the desired quality, taking into account safety and efficacy of the drug product. (ICH Q8)
Critical Quality Attribute (CQA): A physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality. (ICH Q8)
Critical Process Parameter (CPP): A process parameter whose variability has an impact on a critical quality attribute and therefore should be monitored or controlled to ensure the process produces the desired quality. (ICH Q8)
Control Strategy: A planned set of controls, derived from current product and process understanding that ensures process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring and control. (ICH Q10)
Lifecycle: All phases in the life of a product from the initial development through marketing until the product's discontinuation.

Table 1. Glossary of QbD

IMPLEMENTING QBD APPROACH

There're four key steps of implementing QbD approach with emphasis on pharmaceutical manufacturing process.

Define the Desired Product Performance

The desired product performance is that the drug product should possess in order to reproducibly deliver the therapeutic benefit promised in the label to the consumer (J. Woodcock). The desired product performance is achieved by the quality characteristics summarized as the Quality Target Product Profile

(QTPP). Examples of QTPP include formulation characteristics, indications, dosage strength(s), delivery systems, adverse reactions, etc. The QTPP is the fundamental of the entire drug discovery and development process, the following product and process design are more objective and efficient with the clearly predefined QTPP.

Design the Product

Based on the QTPP, the Critical Quality Attributes (CQA) can be determined and evaluated in product design. This step focuses on understanding the sources of variation, detecting the presence and degree of variation, evaluating the impact of variation on the process and ultimately on product attributes, and controlling the variations. Take the drug substance as example, the physical, chemical, and biological properties of the drug substance can influence the performance of the drug product and its manufacturability, should be identified and discussed. All these properties of the substance are CQAs which should be considered, assessed scientifically and controlled within an appropriate limit, range, or distribution.

Design and Discover the Product Process

After well understanding the quality specifications and product design, a flexible and robust process can be designed and developed to meet the CQAs and produce the desirable QTPP. Besides the QTPP and CQA, other factors to consider for process design are type of process, equipment, material transfer, and manufacturing variables. These factors are investigated as the Critical Process Parameters (CPPs). The issues in product process can be anticipated once CPPs are identified and assessed.

Develop the Control Strategies

Under the enhanced understanding on the design and development of product and process, the corresponding control strategies can be established to control and monitor the all critical attributes and parameters identified in the upstream steps. The control strategy enables the best set of attributes and parameters are implemented to produce the final product. At the end, a management system focusing on continually improving the process capability and Developing the Control Strategy should be planned and implemented.

The first two steps focus on understanding the product. QTPP and CQA are two elements shaping the product design. The last two steps emphasis on understanding and controlling the process.

ADVANTAGE OF APPLYING QBD IN PHARMACEUTICAL MANUFACTURING

There’re many advantages of properly implementing QbD approach in pharmaceutical manufacturing.

First of all, QbD can assure product quality through design and performance-based specifications. QbD can facilitate continuous improvement in product management and life cycle.

Second, QbD can reduce the cost of drug development and approval. The examples of manufacturing cost reduced by QbD are lost batches, manufacturing deviations, the internal costs of trouble shooting, investigation and remediation. Besides the manufacturing cost, QbD can also reduce the regulatory cost. QbD provides the ability to meet FDA submission guidelines and expectations as well as the opportunities for flexible regulatory approaches. By implementing QbD, less queries and inspections of manufacturing sites from regulatory agency are required.

Third, QbD can significantly reduce the time from drug development to market. QbD supports more efficient use of development time and helps on determining more reliable supply. QbD can enhance the quality of regulatory reviews through standardized review questions, thereby can reduce the review and approval time.

At last, besides improving all the three elements of quality-time-cost (QTC) triangle, QbD also supports on achieving compliance as applying QbD is regulatory expectation and requirement in pharmaceutical development and manufacturing.

ADOPTING QBD IN STATISTICAL PROGRAMMING

WORKING ENVIRONMENT, EXPECTATIONS ON QUALITY AND COMPLIANCE IN STATISTICAL PROGRAMMING

In pharmaceutical industry, the statistical programmers participate in a variety of activities relevant to the entire data flow in clinical development. These activities include but not limit in data management, analysis and report, publication, regulatory submission, and post marketing surveillance. The statistical programmers collaborate with professionals from multiple disciplines including database manager, clinical manager, statistician, pharmacometrician, medical writer, and scientists in clinical, bioanalytical, and translational medicine team. In this matrix working environment, the statistical programmers support to the investigation on the safety, efficacy, and other characteristics of a medical product by providing data-related products and services. Thus, as expected and required in the drug product, quality and compliance are two essentials in the products and process of statistical programming, and could be improved by applying QbD approach. Generally, quality of statistical programming can be defined as products and services that deliver intended performance. Compliance can be defined as meeting established guidelines, policies and procedures, e.g. ICH Good Clinical Practices (GCPs), 21 CFR Part 11, company policies/SOPs, etc.

However, different from drug products which are material objects with quality specifications measured by physical, chemical, biological properties, or other characteristics, the products of statistical programming are dataset, data-related table, figure and listing (TFL), and software application. Apparently, the quality specifications of programming product are different from these of drug product. It's helpful to discuss the specifications on quality and compliance of programming products before thinking about implementing QbD to statistical programming.

As programming product with high level of quality, the dataset and TFL must be accurate, complete, and consistent. The dataset and TFL must support the investigation on the medical product. The software application must carry out the exact functions as predefined, therefore its output is usable and reliable. Moreover, the programming product and its generation process must be compliant with internal and external SOPs and guidance.

Besides the differences on quality specifications due to product properties, the expectations on quality vary from the end users. Depending on the role of stakeholders in clinical trial, the quality and compliance of programming products can be perceived in multiple dimensions. A statistician expects a dataset for analysis is error-free, complete, and supporting the statistical model. The trial director overseeing the progress of a clinical trial perceives the programming service is productive, efficient, well-coordinated, accountable to unexpected and urgent events, and overall programming process follows the industry SOPs and regulatory guidance.

ADAPT QBD TO STATISTICAL PROGRAMMING

Due to the differences on quality specifications and product process between manufacturing and statistical programming, it's not appropriate to just copy and paste QbD approach established and applied in manufacturing to statistical programming. Instead, the programmers need to adapt the strategy, basic elements, and key steps of QbD to statistical programming environment. The strategy of implementing QbD is to achieve quality through leveraging information and knowledge of product and process, reducing risks and uncertainties, controlling and improving the process. This strategy can be implemented to statistical programming through the basic elements and key steps including the following:

Define the desired product performance and identify the QTPP

The QTPP of programming products and service should contains quality expectations for datasets and TFLs, functional specifications for software applications, compliance requirements on all types of programming products and processes, and timely delivery. The data quality expectations for dataset and TFL are accuracy, completeness, consistency, and free of errors and duplicates. The functional specifications define the intended capabilities of the software application. The programming products and their processes must be compliant with internal SOPs/external standards and guidance. The timely delivery is required for programming service. This paper will focus on the datasets and TFLs, the most common products of statistical programming for clinical trial research.

Determine the CQAs impacting on QTPP and desired product performance

Many factors can impact on QTPP of a programming product, such as the source data, the complexity of the programming product, time line, workforce, etc. Let's take a deep look at these factors.

The quality and integrity of source data are of course the key attributors of QTPP for analysis dataset and TFLs which contain the variables and records derived and calculated from source data. Depending on the analysis purpose and study process, the source data for datasets and TFLs can come from Case Report Form (CRF), , CDISC datasets in standard (e.g. SDTM, ADaM) format, data files provided by external vendors in nonstandard formats (e.g. files in excel, csv, text format). In real practice, programmers usually take tremendous effort and time to check data issues in the source data, e.g. edit check on CRF data, reconciliation on non-CRF data, and Pinnacle 21 verification on CDISC data. Despite these checking procedures, it's very common that some data issues can't be caught until final datasets and TFLs are generated. The data issues carried over to final products often cause updating on source data and rework of the entire programming process. Therefore, the statistical programmers are strongly encouraged to generate data checking list besides the checking procedures discussed above. For example, the issues in Table 2 are common in source data, but are often neglected.

Missing values
- Not coded AE terms, MH terms, CM decodes
- AE terms with missing severity
- "Specify if other" is missing when "Other" is selected.
- The exact time information is missing for AEs that occurred on the same day as the first dosing date
- Missing demographic variables: age, race, gender, etc.
- Missing or incomplete date (and time)

Date (and time) variables are not in sequential order.
- Start date (and time) is later than end date (and time).
- Visit number vs actual visit date (day)
- Nominal day (and time) vs actual date (and time)
- Dispense time is later than the dosing time.
- Protocol deviation start date is after the study disposition date.
- Any date about EX, disposition, findings is after the date of completed the study or early discontinuation.

Duplicates
- AE group ID
- Dosing records on the same date (and time) with the same route and amount
- Disposition records or incorrect DSDECOD/DSTERM was selected.
- Records on the same date (and time) with identical findings.

PK data
- PK concentration is greater than zero and not BLQ at the time point prior to first dose.
- PK metabolite is greater than zero and not BLQ at the time point prior to first dose if it is collected,
- Unusual PK concentrations
  - Peak is not observed on PK curve, e.g. close to zero or PK curve is flat.
  - Missing distribution and elimination phase on PK curve.
  - PK concentration is elevated again in elimination phase with no additional dose.

Other issues
- Systolic blood pressure is less then Diastolic blood pressure.

Table 2. Common Issues in Source Data

The complexity of product is considered a CQA even though no causal relationship between complexity and quality of a programming product. The quality requirements of a product should never be reduced or limited due to its high complexity level. Yet, it is more likely to make mistakes or fail to follow compliance requirements when generating complex products. Thus, the programmers should pay more attention on the product with high complexity level to reduce the risk of quality and compliance issues. It is a good exercise to evaluate the complexity level of a programming product, and this exercise also helps on understanding the product specifications, designing the development and validation plan, allocating the time and workforce efficiently.

The complexity of a programming product has multiple dimensions. For example, the overall complexity of a dataset can be determined based on the questions in Table 3. Then the development and validation plan can be designed based on the complexity. For datasets with high complexity level, the programmer

having more experience may be assigned to the task. Correspondingly, the independent programming with “proc compare” to generate identical output can be the validation plan to ensure the quality of the completed dataset. If its complexity level is low, the only the derived variables and records can be verified when the time and workforce are limited. Programming team can evaluate the complexity according to the study design, analysis purpose, data and TFL specifications, and use the evaluated complexity to guide the programming development and validation.

1. How many derived variables are required in the dataset?
   □1-10 □11-20 □21-30 □31-40 □>40
2. How many derived records per subject?
   □1-3 □4-6 □7-9 □>10
3. What are the derivation rules and programming logic for derived variable and record?
   □LOCF □WOCF □Mean □Max □Min □Other (Specify):
4. Is any statistical method employed to generate such variables or records?
   □No □Yes
   If yes, what is the statistical method? Is the statistical method simple or complicate?
5. What is the data structure if it’s not CDISC data?
   What is the complexity level of the data structure? □ Low □ Middle □ High

Determine the overall complexity level of the dataset based on the questions above:
□ Low □ Middle □ High

Table 3. Questions to Evaluate Complexity of Analysis Dataset

Similar to complexity, the time and workforce spent on the product are the risk factors impacting on quality and compliance, the more time and more programmers, the better quality and compliance in the completed product. However, the time line is always tight and the workforce is always insufficient in reality. Although these two factors are usually not negotiable, programming team can focus on how to use them more efficiently. For example, prioritize the time and workforce on the milestone deliverables, deliverables with high complexity level or supporting the primary analysis.

Determine the programming process and critical process parameters (CPPs)

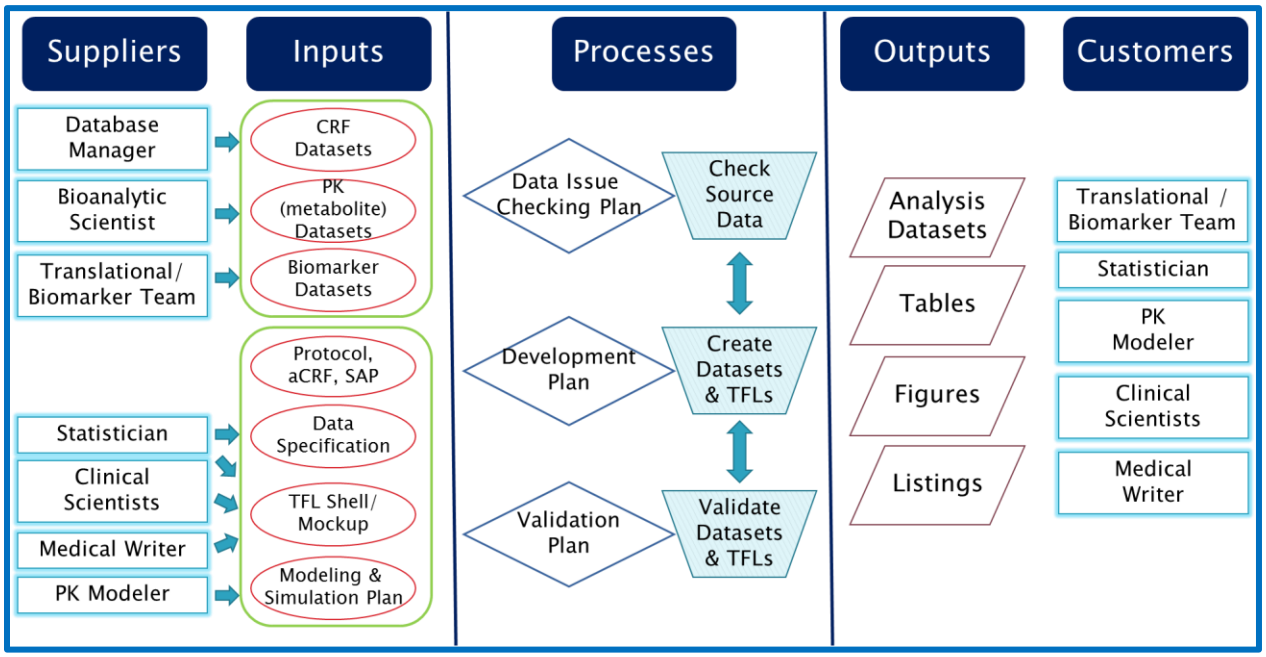

Figure 1. The Work Flow in Statistical Programming in a SIPOC Diagram

A typical work flow in statistical programming can be mapped to a SIPOC (Supplier, Input, Process, Output, and Customer) diagram in Figure 1. In the programming process, programmers will check issues in source data, generate and validate analysis datasets and TFLs upon specifications and shells. Despite

a fairly simply process, there're a few factors impacting on the quality of the products and the overall efficiency of the programming process.

First, this process needs clear and efficient communications between programmers, suppliers and customers. The programmers provide the feedback on issues in source data to the corresponding data suppliers, and confirm their understandings on the data specifications and TFL shells with document suppliers. Besides the communications with suppliers, the programmers also discuss the timeline, priority of the deliverables with customers, collect the feedbacks on the draft outputs, and update the final deliverables for the customers. Apparently, the communication is a CPP in the programming process. It is a good practice to have a communication plan agreed with the entire study team. This communication plan should contain at least three elements: 1) key contacts from programming and other functional teams; 2) the effective and reliable communication path (e.g. instant message, email, phone, meeting, etc.); 3) the structure of meetings (e.g. frequency, duration, attendees, topic, etc.). Setting up a clear and efficient communication plan is the very first step while starting on a new study and project.

Next, the programming team need clear, flexible, and practicable plans to guide the programming activities. At least three plans should be considered: source data checking plan, development plan, and validation plan. Each programmer should know his or her role in every plan and play it well.

In the source data checking plan, the programmers should know timeline of availability on source data and confirm with the suppliers if any checking or reconciliation has been conducted on the source data. If so, ask the suppliers to share their checking result and use the list in Table 2 as additional checking.

The programming team can make development plan to allocate the time, workforce, and priority on every deliverable based on the well understanding on the data specification, TFL shells, the milestone deliverables, and the dependencies between datasets and TFLs. In this plan, team can also decide if existing macros and programs can be used or simply modified to fit the new specifications, or team need to scratch new macros and programs. In addition to document the working procedures, programming team may communicate with suppliers and customers about any updates anticipated. During the programming process, there're many reason for suppliers to update source data, data specifications, and TFL shells. For example, fixing the issues in source data, adding new variables and records in data specifications, changing the TFL shells to present the result in a better way. These back and forth updates often cause more or less rework, cost more time on programming, thereby risk the quality and compliance in deliverables. It's a positive strategy to discuss the potential updates and rework with suppliers and customers in advance, then prepare for potential rework in the development plan to make it more robust and flexible.

With the insights on the requirements for deliverables and entire programming flow, the validation plan can include different validation methods: 1) independent programming with "proc compare" to produce identical output; 2) independent programming to verify key variables and records; 3) validate key variables and records by developer. For deliverables to be submitted to regulatory agency or deliverables containing critical variables for study analysis, method 1 is the best strategy to ensure the quality though it is costly and time consuming. If the programming outputs are used for making critical decisions and merely simple, then it's appropriate to only verify the key variables or records by method 2 or 3. The programming team can determine the validation methods based on the priority, complexity of the deliverable, and the availability of time and workforce.

Conduct and Control the Programming Process, and Manage the Lifecycle of Programming

While crafting the code, all programmers no matter their experience level, should follow Good Programming Practices (GPP) and craft well-structured and well-documented programs. The manager or lead programmer should oversee the programming activities and examine if all the programming plans are applied properly. Moreover, it is important to build a "quality culture" within programming team. For example, set up routine discussions on improving quality and compliance, share the lessons learned after programming products are completed finally.

CONCLUSION

QbD is a systematic approach to achieve the quality through planning with the quality in end product, designing the process and building the quality into the product. QbD approach is based on sound science and quality risk management. The strategy, basic elements, and key steps of QbD can be adapted to statistical programming to ensure the quality and compliance in programming products.

REFERENCES

Juran, J.M. 1992. Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services.

FDA. Aug 2002. Pharmaceutical Current Good Manufacturing Practices (CGMPs) for the 21st Century: A Risk-Based Approach.

ICH. May 2006. Q8 Pharmaceutical Development.

ICH. June 2006. Q9 Quality Risk Management.

ICH. April 2009. Q10 Pharmaceutical Quality System

Pharmaceutical Quality/Manufacturing Standards (CGMP). https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/ucm064971.htm

J. Woodcock. The concept of pharmaceutical quality. Am. Pharm. Rev. November–December, 1–3, 2004.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Linghui Zhang
PRA Health Sciences
zhanglinghui@prahs.com