Research article

Risk assessment and integrated process modeling–an improved QbD approach for the development of the bioprocess control strategy

Daniel Borchert[1], Diego A. Suarez-Zuluaga[2], Yvonne E. Thomassen[2] and Christoph Herwig[3,]*

[1] Exputec GmbH, Mariahilferstraße 88a/1/9, 1070 Vienna, Austria
[2] Intravacc, Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, The Netherlands
[3] Vienna University of Technology, Research Area Biochemical Engineering, Gumpendorferstraße 1a, 1060 Vienna, Austria

* **Correspondence:** Email: christoph.herwig@tuwien.ac.at; Tel: +43158801166400.

Abstract: A Process characterization is a regulatory imperative for process validation within the biopharmaceutical industry. Several individual steps must be conducted to achieve the final control strategy. For that purpose, tools from the Quality by Design (QbD) toolbox are often considered. These tools require process knowledge to conduct the associated data analysis. They include cause and effect analysis, multivariate data analysis, risk assessment and design space evaluation. However, this approach is limited to the evaluation of single unit operations. This is risky as the interactions of the operations may render the control strategy invalid. Hence, a holistic process evaluation is required. Here, we present a novel workflow that shows how simple data analysis tools can be used to investigate the process holistically. This results in a significant reduction of the experimental effort and in the development of an integrated process control strategy. This novel QbD workflow is based on a novel combination of risk assessment and integrated process modeling. We demonstrate this workflow in a case study and show that the herein presented approach can be applied to any biopharmaceutical process. We demonstrate a workflow that can reduce the number of factors and increase the amount of responses within a Design of Experiments (DoE). Consequently, this result demonstrates that experimental costs and time can be reduced by investing more time in thoughtful data analysis.

Keywords: Quality by Design; improved QbD; cause and effect analysis; quantitative risk assessment; design space; integrated process model; control strategy

1. Introduction

These are important instructions and explanations. Thank you for your cooperation. There are three process validation stages suggested by the Food and Drugs Administration (FDA) [1]:

• Stage 1: Process characterization—Establish control strategy and identify critical process parameters (CPPs)

• Stage 2: Process performance qualification (PPQ)—Show that control strategy is sufficient

• Stage 3: Continuous process verification—Monitor that the process is under control

Stage 1 aims to identify the relations of the process parameters (PP)s, operated within a particular range, to the critical quality attributes (CQAs) [2] by using process knowledge and experimental data. As suggested by the international conference of harmonization (ICH) Q9 guideline [3], it is current good practice to conduct a risk assessment (RA) before experiments are planned. Since such a procedure is common practice in process development, it should still be used in process validation to ensure the traceability of the conducted experiments. Within the Q9 guideline, the RA target is "the identification of hazards and the analysis and evaluation of risks associated with exposure to those hazards" [3]. Hence, the impact of each PP on each CQA is evaluated within that approach. The information required to perform a RA [3] consists of historical data, process expert's knowledge and literature. Once the PPs with the biggest influence on the CQAs are determined via the RA, their influence is experimentally confirmed. In the pharmaceutical industry, this is often done via design of experiment (DoE) studies [4] of each unit operation. The exact amount of PPs investigated within a DoE, as well as the aim of the design, are the most crucial factors which must be carefully selected [5]. This unit operation isolated view of the process is often reported as contra-productive because they often interactively influence each other [6]. Therefore, it is recommended to change the process characterization from a unit operation oriented perspective to a holistic one. This holistic process evaluation or the reduction of separate unit operation can either be done by a Life Cycle assessment [7] or by continuous processing [8–10], perfusion [7,9] or QbD [11]. All of these techniques have the purpose to reduce the overall experimental costs.

Next to characterizing the process, the "Stage 1" goal is to identify an appropriate process control strategy. This strategy identifies the necessary proven acceptable ranges (PARs) for the critical PPs (CPPs) to achieve the required product quality [12]. This is best done by using a QbD approach [13]. Figure 1 displays the QbD approach as proposed by the CMC Biotech working group [13] used for "Stage 1". Afterward, Stage 2 will be pursued to prove the chosen control strategy, while finally, Stage 3 is used to monitor and continuously improve the process. Within this work, the procedure, methods and workflow suitable for Stage 1 will be discussed in more detail. Stage 2 and 3 are not further considered within this work.
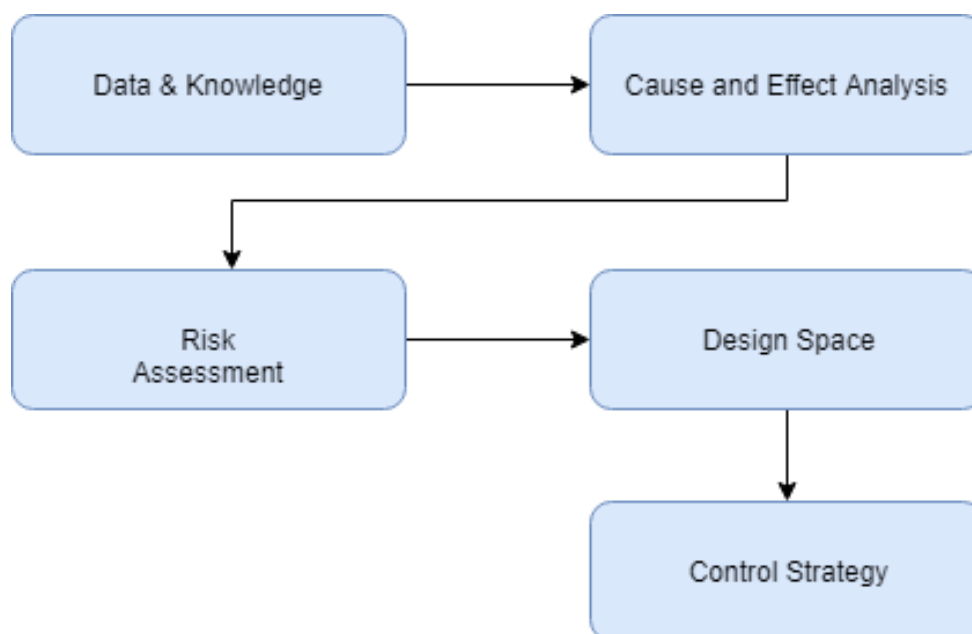
Figure 1. QbD approach following the suggestion from the CMC Biotech working group [13].

Figure 1 represents the five major parts of the QbD approach for Stage 1. Data & Knowledge represent the need for data and process knowledge and act as the start of the approach. Cause and Effect Analysis symbolizes the need for statistical analysis approaches to identify the impact of PPs on CQAs. Risk Assessment is the evaluation tool used to estimate the PP effect on the CQA based on statistical and process knowledge. Design Space represents the identified operation range of the CPPs within the process. Finally, Control Strategy serves the aim of the assessment and indicates the required PP control action (with specified PARs) to produces the required CQA quality. Although this approach has shown good results in the last years [13], it requires much time and costs to fulfill each step. This is mainly because the approach has been focused on each unit operation independently.

Moreover, this per unit operation approach bears a significant risk, which is that an isolated establishment of the control strategy may fail because the interactions between the unit operations may not be taken into account [11].

In order to avoid the above mentioned issues, we identified improvement capabilities using basic data science approaches. Within this work, an improved QbD approach for process characterization, Stage 1 [1], is shown. The suggested approach focuses on improvements in the steps Cause and Effect Analysis and Risk Assessment and how the holistic interactive process impacts these two procedures. These improvements result in reduced experimental effort and in the development of an integrated process control strategy. The herein presented QbD workflow is based on a novel combination of RA and integrated process modeling (IPM). We demonstrate it via a case study and show that the approach can be applied to a biopharmaceutical process anticipated to become a platform process. In this case study, we were able to reduce the number of factors considered within a DoE significantly, from at least eight to three.

2. Materials and method

2.1. Data

The dataset used for the herein conducted case study was provided by the Institute for Translational Vaccinology (Intravacc). The data represent development studies for an inactivated polio vaccine platform process. Sabin poliovirus type 2 was produced with an animal component free media on Vero cells (USP phase). Further, five downstream processing (DSP) unit operations were conducted to purify the virus. The key performance indicator (KPI) and the impurities were determined at the end of the USP and at each DSP unit operation.

2.2. Software

Within the case study, two commercially available software tools were used to perform the cause and effect analysis. SIMCA version 13.0.3.0 (Umetrics AB, Umea, Sweden) was used for raw data analysis (RDA). To perform feature based analysis (FBA), inCyght® Web version 2020.03 (Exputec GmbH, Vienna, Austria) was used. The required uni- and multivariate statistical analysis tools were already implemented in the software. The data were preprocessed with inCyght® and MS Excel 2016 (Microsoft, Redmond, WA) before data analysis with SIMCA. Python 3.7 (Python Software Foundation [14]) was used for the RA simulation and plotting.

3. Results

This section is presented in two parts. First, a comparison of the current QbD (cQbD) with the improved (new) QbD (nQbD) approach is given. Second, the results of the analysis for a case study based on the nQbD approach are described.

3.1. Comparison of cQbD and nQbD approach

The cQbD approach is well established and often used in industry. In parallel, data science applications increased in importance within the past years [15]. They are used within the cQbD approach to increase process knowledge and to save time and costs. The nQbD approach applies a combination of existing and novel data science and analysis methods, resulting in an improved QbD approach. Figure 2 summarizes both approaches side by side.
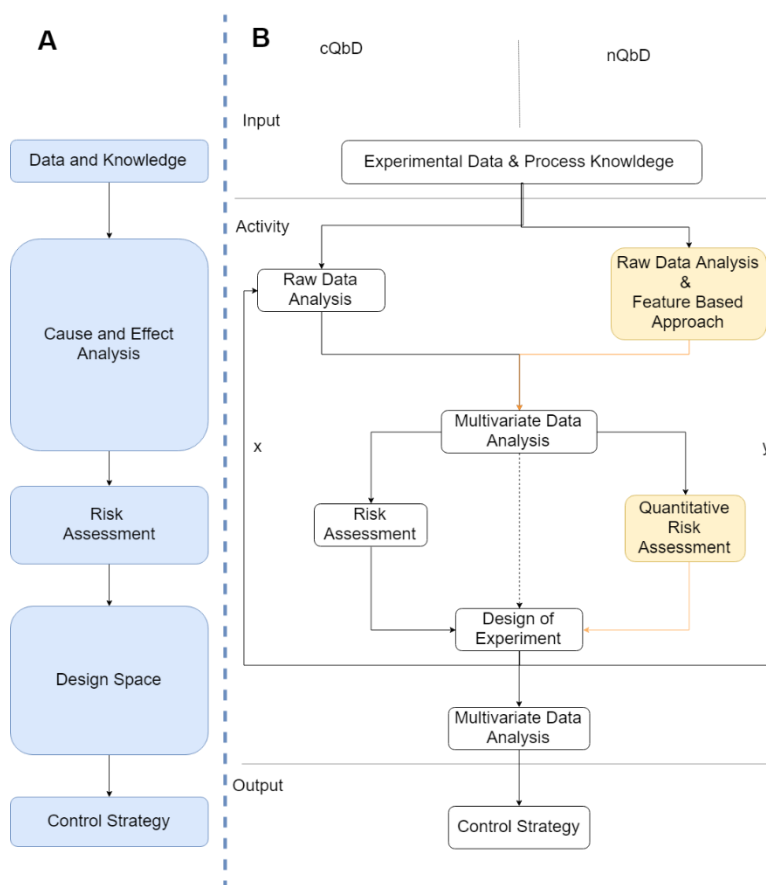
Figure 2. Comparison of cQbD and nQbD approach. Panel A shows the general QbD approach as described in the introduction (Figure 1). Panel B shows the input, output and data analysis activities related to the different QbD approaches. On ehe left side of panel B, the cQbD approach is presented. While, on the right side of panel B, the nQbD approach is displayed. Orange marked cells or pathways represent the improvements discussed in this work. The input and output of the approaches are similar.

The basic concept of the QbD approach, as shown in Figure 1, has not been changed with the suggested approach (Figure 2A). Within Figure 2, if a particular strategy differs, a separate box for each approach is displayed. The first evaluation of Figure 2B indicates that the input and output of both approaches are similar, while in the activity section, differences are observed. Within this section, the entire workflow and in more detail, the differences between the current and the new QbD approaches will be explained below.

3.1.1. Input

3.1.1.1. Data and Knowledge

At the beginning of the analysis process, knowledge and historical data are often available. We assume that the key performance indicators of the process, often called the responses, are already defined according to the ICH Q8 guideline [16]. The more is known about the process, the better and

faster the analysis approach will be. Therefore, comprehensive process knowledge is crucial for the following steps. In former times, it was often reported [17] that just one data format was used for data analysis since it was often not possible to merge all available data types. This reduced data gathering results in a significant loss of information and thus process knowledge. The holistic data storage and analysis of all available data types are often very complex and time-intensive. Many different data storage and analysis tools are currently available; its benefits and appropriate analysis tools were recently reviewed by Steinwandter et al. [15].

3.1.2. Activity

3.1.2.1. Cause and Effect Analysis

At the beginning of the workflow, it is crucial to identify the reason for the response variance. This is done via a cause and effect analysis (often also called root cause analysis [18,19]).

cQbD Approach. Currently, two approaches are commonly followed to perform the cause and effect analysis. Raw data analysis (RDA) [20] is currently the most prominent tool used in industry. While the feature-based approach (FBA) has increased in importance within recent years [21,22]. Nevertheless, independent of the approach, the workflow to identify influencing parameters effecting a CQA can be done within seven steps (Figure 3).
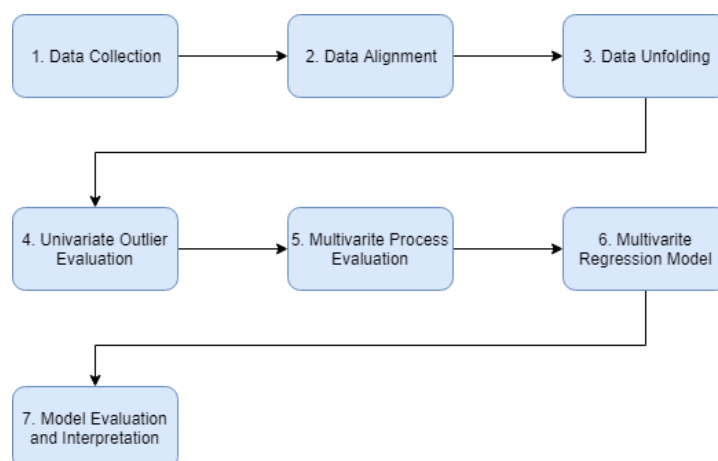

Figure 3. Seven phases of the cause and effect analysis workflow according the methodology presented by Borchert et al. [18].

The RDA and FBA approaches differ primarily in step 3 (Data Unfolding). While the RDA analysis uses all available data holistically, the FBA uses reasonable information mined from the process data [18]. Furthermore, both approaches use just a particular data format [17,20–22], time series data for RDA and features for FBA. The remaining data are often not included in the analysis. Since not all available data are considered in data analysis, a loss of process knowledge by using each approach individually is assumed. Independent of the approach, the multivariate model generation and result interpretation might be hard for bioprocess engineers. Hence, they rely on statisticians to do such an analysis. On the other hand, statisticians are often not trained in the context of biopharmaceutical production, which can result in a multivariate data analysis (MVDA), which is

not useful to apply. This problem regularly results in the bioprocess expert using software tools to apply and understand data analysis algorithms and their results.

nQbD Approach. Even when process scientists possess both process and statistical knowledge, the gap of using all available data for data analysis still exists. In this contribution, a comprehensive data evaluation workflow that enables the use of all data holistically and the mining of reasonable and adequate information [18] is presented. As shown in Borchert et al. [18], a combination of RDA and FBA is the crucial element of the analysis approach. The workflow is based on three elements:

1. Hypotheses generation
2. Process knowledge
3. Hypotheses testing

The RDA covers elements 1 and 2. Generally, the variance in the CQA is detected holistically, but the real causing factor needs to be further investigated. The first hypotheses can be established during data analysis. By subsequently applying the RDA, the hypotheses can be verified, or new ones can be generated. Based on that outcome, the FBA can be used to test the generated hypotheses (Element 3) via reasonable information mining from the previously investigated time series data. The mined information will be evaluated by applying standard uni- and multivariate data analysis approaches. A final multivariate regression model ends up in the identified root cause, which is often displayed using an Ishikawa (Fishbone) diagram.

This unique analysis capability by using all available data was already applied to a vaccine production process using measurements from the upstream unit operation and is shown to be useful for [19]:

- Deeper process understanding
- Holistic process data evaluation
- Determination for additional experiments to increase process knowledge
- Reduction in the experimental effort by investing more time in data analysis

Consequently, applying this approach allows us to learn more from the data. This means that a stable basis for the next step in process characterization is established.

3.1.2.2. Risk Assessment

The next step in the QbD approach is risk evaluation. Within an RA, usually, three factors [3] are often estimated for each PP on each CQA.

- Severity (S)—Measures the feasible value of harm
- Occurrence (O)—Represents the probability of occurring the harm
- Detectability (D)—The capability to identify the harm

Usually, a team of process experts rates each factor, within the predefined PP ranges, based on their process knowledge and experience.

cQbD Approach. The standard approach evaluates a RA using the Risk Priority Number (RPN). This method multiplies the S, O and D values to obtain the RPN. Since it is often the case that the three factors are ranked from one to ten, the RPN can, therefore, range from 1–1000. Consequently, each PP has an RPN and can be sorted based on it. The higher the RPN, the higher the PP is considered to (negatively) impact the CQA when operating outside the PP target range. Usually an arbitrary RPN number is defined as the threshold to determine if the PP is criticality affecting the CQA.

This approach is prevalent in the industry. However, from a mathematical point, it is not appropriate. The reasons for this are [23]:

- Multiplication of ordinal scaled values
- Gaps within the scale
- Different meaning although RPN is equal
- Significant influence by small changes in RPN

In the past years, several approaches were developed to tune the RPN approach [24]. These tools are mainly developed in order to avoid the current pitfalls. Those multi-attribute decision approaches are based on sophisticated algorithms like grey relation [25] or the order of preference by the similarity of the ideal solution [26]. Such tools are handy for their benefit but are often hard to interpret for biopharmaceutical process experts

nQbD Approach. This approach translates the S estimation of the RA to model effect size and to use the O estimation to predict the operating parameter within a defined judged range. A simple linearization is used for the model size definition and standard truncate and selection methods are used to estimate the parameter distribution within the PP judged range [27]. In addition to the conventional RA, a new estimation needs to be entered into the RA, the critical delta of CQA [%] [27]. Like for S and O, this needs to be done for each unit operation and CQA. This value can be gathered from process knowledge, experimental data, or literature.

Beyond this novel S and O interpretation, this implements "integrated process modeling" (IPM) [28] into it. This enables a holistic process assessment of the CQA for all unit operations. Figure 4 shows a three unit operations example.
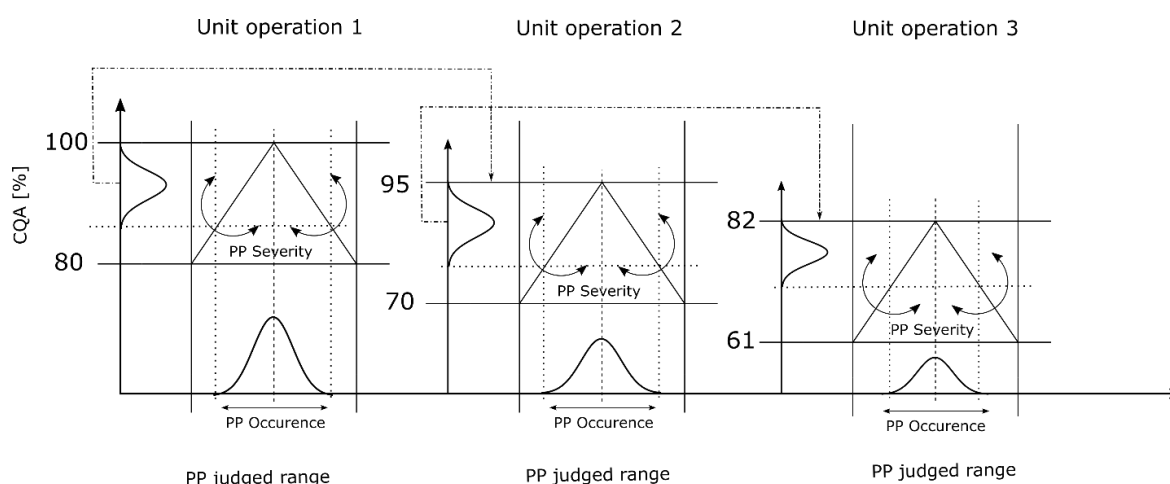

Figure 4. Novel risk assessment evaluation approach. The plot shows an example process with three unit operations. It can be seen that the severity estimation of the RA results in the slope of the effect in CQA reduction [%] and that the occurrence estimations results in an estimated value within the displayed sample distribution within the judged range. The end of the previous unit operation is estimated as starting material of the next one, and 100% of the relative CQA is used at the beginning. The delta CQA [%] will be calculated by including the required variable critical loss per CQA [%] and can vary between CQA and unit operation. Image adapted from Borchert et al. [27].

Figure 4 shows the tool based on three-unit operations. The holistic simulation, starting from the first until the last unit operation, is further referred to as a cycle. In order to simulate the process reproducibility as close as possible to the real one, a Monte Carlo simulation [29] was applied. The method simulated 1000 such cycles [28], including all PP for each CQA independently. In order to evaluate the PP critically onto the process, this simulation has to be repeated by excluding one PP from the simulation, until each PP has been removed once. After that, the resulting CQA distribution at the last unit operation results in the CQA reduction per PP.

This result allows a quantitative statement and sorting of PPs accordingly. PPs criticality can be identified by evaluating their simulated impact on the quality attribute. And, it can now be estimated based on the impact not in a single unit operation but on the whole process. Furthermore, this approach can avoid all incorrectness of the RPN approach:

• No multiplication of ordinal scales values, since S and O are translated to model effect size and parameter distribution

• No gaps in the critically PP evaluation scale since no multiplication of S, O and D. Although gaps are present in that scale, this is expected according to the applied S and O transcription

• The different impact from S and O assessment is demonstrated. We show the different impact with the conducted in-silico study within Borchert et al. [27]

• Small criticality changes by small changes in S and O. We show that small changes in the S and O assessment do not influence the assessment that much as before

The third factor "D" can be included after the simulation. Since factor D is often described as the ability to recognize a failure before it causes damage [3], a final re-sorting step can be applied. Therefore, the entire PP evaluation is split into equal blocks. These blocks are then further sorted by its D evaluation. This re-sorting can cause that PPs with a lower impact on the process appear as more critical than before although the S and O linearization shows less quantitative impact [27].

This method implements a novel rational decision-making tool, which avoids all of the known disadvantages of the RPN approach. Moreover, it enables a quantitative evaluation of the whole process.

3.1.2.3. Design Space

The last step in the activity section is the design space evaluation. The design space is used to describe the relationship between the input parameters of the process and the CQA [16]. According to the ICH Q8 (R2) guideline, the design space is defined as the multivariate combination and interaction of input variables and process parameters that have been demonstrated to provide assurance of quality. The design space can be a sophisticated mathematical operation or a combination of factors of the multivariate model [16]. Beyond its interpretation, the most important fact is that if the process is operated in that space, the CQA will meet the defined product quality.

The prior conducted MVDA is used to increase process knowledge and the RA results in the first hypotheses of PP effects to the CQA. The evaluation of such hypotheses is commonly done using a DoE approach, since targeted experiments will be conducted. The appropriate design is often very sophisticated since many factors have to be considered. Nevertheless, a DoE will be often conducted for a particular purpose; the three primary reasons for a DoE are:

1. Screening—Identify the most influential factors and their appropriate ranges
2. Optimization—Identify the process optimum

3. Robustness testing—Proof that the process is under control

Independent of the aim of the DoE, experiments must be conducted with a particular objective. After defining it, the next and most crucial step for DoE design is the factor selection. Depending on the number of factors that should be investigated in the experimental design, the number of required experiments can increase drastically and should be selected carefully. The number of factors, which will be considered in a DoE, are often taken from prior RA evaluation.

After DoE experiments, MVDA is conducted to identify the relations between the changes in the investigated factors on the CQA. In that case, regression analysis is often conducted [4]. The analysis procedure in this stage of the process is similar for cQbD and nQbD and therefore not further discussed in this work

cQbD Approach. The Fishbone diagram, from a prior root cause analysis, is one strategy used to identify the number of factors that have to be considered within a DoE [30–32]. Another strategy, and the most prominent, is the use of a RA [31].

Both approaches, Fishbone and RA, often result in many factors that must be included within DoE study. Such DoEs are costly and time consuming. Since these are critical factors within process assessment, the current design space assessment approach needs to be optimized.

The knowledge gathered from DoE can be used for control strategy definition or can be reassessed in the cause and effect analysis (Figure 2). This recycling step can occur multiple times (see Figure 2B, arrow x) so enough process knowledge is gathered and sufficient information for defining the control strategy is available. Hence, many experiments might be required to fulfill the aim.

nQbD Approach. Using the quantitative RA, it is feasible to:

- Evaluate the PP criticality based on a quantitative PP assessment
- Reduce the number of factors required for the DoE
- Evaluate the process holistically

Since the number of factors that are considered in the DoE can be optimized, the power of the model can increase drastically with fewer experiments. This reduction in experimental effort results in specific process knowledge, which can fill existing knowledge gaps and thus reduces the amount of knowledge recycling (Figure 2B, arrow y).

Furthermore, the herein developed RA evaluation approach can be used to simulate the entire process. This new evaluation tool enables a holistic process assessment leading to one holistic control strategy based on fast and reasonable design space evaluation.

3.1.3. Output

3.1.3.1. Controll Strategy

The QbD approach aims to identify the control strategy of the process in scope. A control strategy is according to the ICH Q11 [33] guideline defined as a planned set of controls, derived from current produce and process understanding, that assures process performance and product quality (ICH Q10). Furthermore, it is mentioned within that guideline that, a control strategy should ensure that each drug substance CQA is within the appropriate range, limit, or distribution to assure drug substance quality. The control strategy assures that the CQA meets drug substance specifications over the entire process.

The essential element of the QbD approach, the process knowledge, will be used within all stages of the process lifecycle approach. Thus, it makes it crucial not only for process characterization but also for the entire product lifecycle.

cQbD Approach. According to the ICH Q11 recommendation, a CQA can have several control strategies over the process. This recommendation often results in many independent control strategies at a particular unit operation. Such independent control strategies are used to assess the CQA quality over the process. Furthermore, they are used to take counter measurements in case of any abnormalities. Nevertheless, such independent strategies are hard to maintain and time-consuming for the process operators and should be avoided in the future.

nQbD Approach. Since our approach lays the basis for one holistic process control strategy, it could be possible to save experimental effort and time in the future. This significantly reduces the number of control strategies from many different to one holistic. Nevertheless, this needs to be assessed in further studies, which is out of the scope of this contribution.

3.2. Case Study

A case study was conducted to prove the applicability and to demonstrate the potential of the nQbD approach. Each step within the activity part, as shown in Figure 2B, was separately considered and core results are present here.

In order to follow the herein suggested nQbD approach, we consider two variables from the second DSP unit operation in more detail and track their progress through the entire case study. These variables are the transmembrane pressure (TMP) and the concentration factor (CF) of that unit operation, further referred to as UO2_TMP and UO2_CF, respectively.

3.2.1. Cause and Effect Analysis

All five DSP unit operations were investigated independently based on the workflow from Borchert et al. [18]. Finally, it was possible to get a significant multivariate model for each unit operation and the effects on the KPI and impurities could be identified. Figure 5 shows the resulting Ishikawa diagram for the second DSP unit operation.
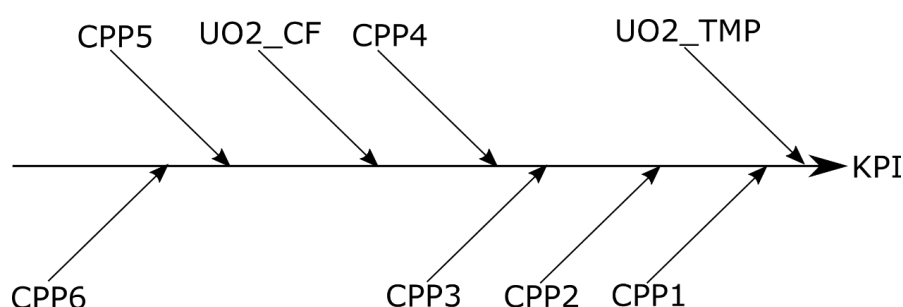

Figure 5. Ishikawa diagram from the second DSP unit operation on the process KPI. All the significant PPs are shown and considered as CPP. The higher the effect of a CPP, the closer it is to the KPI. The PP effect on the KPI is indicated by the arrow direction. Positive effecting variables are shown above the horizontal line, while negative effecting PP is shown below.

Considering Figure 5, eight PP of unit operation 2 have a significant impact on the KPI and are considered as CPP. It can be seen that the parameter UO2_TMP shows the most significant impact on the KPI. The other parameters, as well as UO2_CF, show significant impact, but less than UO2_TMP and have to be considered as essential parameters in the next steps of analysis. This result visualization provides a first detailed understanding of each unit operation. Such information allows it to increase process knowledge and to evaluate the next step more accurately.

3.2.2.   Risk Assessment

Based on the existing data and the previously gathered process knowledge, it was possible to evaluate the critical delta for each unit operation and CQA which would be loss (for KPI) or increase (for Impurity). This maximum change in CQA, which is still acceptable for the process, can be estimated based on process knowledge or calculated based on real data. In case of considering real data, the mean of the existing CQA measurement is taken and then three standard deviations from the mean can be considered as acceptable. This acceptable limit depends mainly on the investigated process. In our case, to calculate the acceptable change of the CQA at DS (DSCQA), the mean value was taken and compared with the lowest acceptable boundary of the CQA, according to the process regulation. Table 1 shows the result of this evaluation conducted for the KPI, Impurity 1 and Impurity 2. Detail measurements used for the calculations are not shown due to confidential reason and will be published separately.

Table 1. Acceptable (critical) delta CQA [%] assessment for the considered CQAs in this case study. The star sign values are estimation since no impurity measurement are included in the available data.

Unit operation	Acceptable ΔCQA [%] KPI	Acceptable ΔCQA [%] Impurity 1	Acceptable ΔCQA [%] Impurity 2
USP	10	5	7
DSP UO1	10	20	20
DSP UO2	11	5*	5*
DSP UO3	7	22	20
DSP UO4	8	7	5
DSP UO5	4	2	2
Drug substance	2.5	10	5

Based on the performed RA and the evaluation shown in Table 1, the IPM simulation could be conducted as suggested by Borchert et al. [27]. Based on that result, PPs can be stated as critical if a particular PP exceeds the DSCQA acceptable limit. It is often the case that the investigated KPI should remain as high as possible at the end of the process and that the impurity concentration should be minimized. Both aspects can be considered within the conduced simulation since relative values of the CQAs are considered in this approach.

3.2.3.   Design Space

The new suggested RA evaluation tool provides a more reasonable result interpretation as the standard used RPN approach [27]. Based on it, an overview of the quantitative contribution

assessment is shown in Figure 6. The plot shows the estimated CQA reduction for each PP assessed within the investigated RA.

	KPI	Impurity 1	Impurity 2
UO5_Var5	0.79 %	1.56 %	0.02 %
UO5_Var4	2.07 %	0.022 %	0.02 %
UO5_Var3	1.07 %	2.04 %	0.01 %
UO5_Var2	1.7 %	1.82 %	0.01 %
UO5_Var1	0.31 %	1.99 %	0.33 %
UO4_Var5	0.5 %	3.5 %	1.43 %
UO4_Var4	2.13 %	4.38 %	1.33 %
UO4_Var3	0.35 %	2.84 %	1.24 %
UO4_Var2	2.29 %	2.46 %	0.09 %
UO4_Var1	2.08 %	2.73 %	0.43 %
UO3_Var5	1.47 %	10.03 %	5.64 %
UO3_Var4	1.48 %	9.77 %	1.97 %
UO3_Var3	0.53 %	7.77 %	1.94 %
UO3_Var2	0.11 %	9.79 %	1.84 %
UO3_Var1	0.41 %	6.47 %	1.46 %
UO2_Var5	0.97 %	1.3 %	0.75 %
UO2_CF	0.75 %	2.72 %	0.3 %
UO2_Var3	2.7 %	2.71 %	0.29 %
UO2_Var2	0.4 %	2.63 %	1.33 %
UO2_TMP	2.88 %	2.28 %	0.03 %
UO1_Var4	0.95 %	9.77 %	4.01 %
UO1_Var3	0.79 %	6.14 %	3.98 %
UO1_Var2	1.18 %	9.12 %	3.94 %
UO1_Var1	0.46 %	9.2 %	3.88 %

Figure 6. CQA reduction overview for each CQA and PP assessed within the RA. For each PP and each unit operation within the RA, the impact on the CQA reduction is estimated and shown. Black background color indicates a low impact of a particular PP, while a white background color shows a significant CQA reduction. The red marked cells on the left side show the unit operation PP, which exceeds the criticality boundary for the investigated CQA (Table 1). In total, three different PPs from two different unit operations are presented as being critical for the process.

Figure 6 demonstrates that three parameters from two different unit operations can be seen as potentially critical for the entire process, since these variables are above the acceptable delta CQA [%], compare Table 1. Two out of three factors are, UO2_TMP and UO2_Var 3, since they have a higher delta CQA as accepted in the KPI assessment, and one, final UO3_Var5, should be included since delta CQA is higher as the acceptable limit for Impurity 1 and 2 assessment. Hence, those would be the PP selected for a follow-up DoE evaluation.

With the conduced case study, it was possible to identify the critical PPs from the entire process by using a rational PP evaluation. Furthermore, it was shown, based on the two tracked variables UO2_TMP and UO2_CF, such a result would not to be feasible when unit operations were assessed just individually. It has to be considered that PPs affect the entire process and not just the performance of the particular unit operations, as shown beginning of the analysis (Figure 5). Furthermore, it was shown that a critical CQA assessment, as shown in Table 1, is necessary to conduct a holistic process assessment. This awareness is crucial when the number of experiments should be significantly reduced.

Prior QbD studies resulted in eight crucial factors for DoE for the KPI (Figure 5) and no crucial factors for the selected impurities. The reduction of factors for a subsequent DoE study can reduce the experimental effort and can increase the power of the DoE. Details result cannot be shown for confidential reason at the moment and will be published separately.

4. Discussion

This work aimed to demonstrate possible improvements within the QbD approach, which finally affect the process control strategy. Standard data science tools are combined to accelerate time to market by reducing experimental effort.

We established an improved QbD approach by using simple statistical tools and by combing existing approaches used for root cause analysis (RCA) [18]. Beyond the demonstration of new process knowledge gathering techniques [18,19], a novel RA evaluation approach was developed [27] and demonstrated. The new approach allows that the current RA approach, the RPN approach, is no longer necessary since a more realistic PP criticality estimation is now present.

4.1. Benefits of the improved QbD Approach

The primary benefit of this work is that a roadmap, beyond the appropriate use of data science tools, was established to:

- See the big picture of the process—Consider the process holistically and not just each unit operation separately.
- Think beyond the current process boundaries—Evaluate the meaning of particular PP weighing values.

The herein developed improvements of the QbD approach for control strategy can, therefore, be summarized as:

- The identification of PPs affecting a CQA is crucial for process characterization and is an essential element within the working routine for process scientists within the biopharmaceutical industry.
- A simple combination of two existing root cause analysis approaches can significantly increase the process knowledge. Further, all available data from USP or DSP unit operations can be included in the data analysis, which was not feasible before. Previous studies are just able to focus the analysis on one format, as done by Sagmeister et al. [21] and Golabgir et al. [22] on feature data and Kirdar et al. [20] on time series data.
- The improved process gathering capabilities enable fast and meaningful hypotheses generation and testing in one comprehensive analysis workflow. Furthermore, this is just possible since all available data can be used holistically within the analysis approach.
- Following the RPN approach is not further required for RA evaluation. The quantitative risk assessment described here allows a more realistic factor interpretation and identification of the effects the PPs on the CQAs. Furthermore, it was shown that the mathematical incorrectness of the RPN would be avoided with the developed approach. Besides that, the estimated factors have a different impact on the PP criticality.
- A more reliable CPP assessment enables a reduction in experimental effort for design space evaluation since meaningful CPPs can be identified and further investigated, compared to section Design Space. The holistic process evaluation improves the experimental design planning and can reduce costs and time since less PP needs to be investigated in a DoE study. The design space evaluation is out of the scope of this contribution and, therefore, not discussed in this section.

Within this contribution, a case study with DSP data was conducted to prove the suggested workflow improvements. It was shown that several improvements could be shown:

- Identification of significant multivariate model on DSP data, section Cause and Effect Analysis
- Reduced amount of PP which can be stated as pCPP, section Risk Assessment
- Reduced amount of experiments for a DoE study, section Design Space

These presented achievements were feasible using the improved QbD approach. Based on that result, reduced time to market is assumed since meaningful impacts on process CQA can be identified and fewer experiments are needed with the DoE study. This reduction has an impact on the patients since drugs can be developed more efficiently concerning time to market.

4.2. Drawbacks of the improved QbD Approach

For a comprehensive workflow elaboration, it should be pointed out that the limitation of the presented approach can be summarized as:

- The gathering of data with different formats and subsequent data set preparation is often very time and resource-intensive.
- Advance statistical know-how needed to interpret MVDA results.
- Use of commercially available software tools: The RCA workflow suggestion within section Cause and Effect Analysis uses commercially available software tools, which require licenses. These resources are often difficult to purchase for a single user. Beyond the cost factor, the time factor for software training should not be neglected and need to be considered before a useful application of that workflow.
- Programming skills for RA evaluation: The quantitative PP assessment tool presented section Risk Assessment is currently based on a python script level and is not available in a commercially available software tool. To apply that approach, Python, or any other programming language, need to be understood on an advanced skill level to establish the required script.
- Just main effects are evaluated with the presented RA evaluation approach: Since the presented workflow based on the linear assumption of the model effect size based on the severity estimation, it was just feasible to resolve main effects. Interaction and quadratic impacts of PPs to the CQA are currently not considered in the presented work by Borchert et al. [27] but could be added if prior knowledge suggests such effect of the PP.

Although there are some limitations, we assume that modern Manufacturing Science and Technology (MSAT) Departments, as well as CMC Groups, can apply the herein presented approach. This is because process knowledge, skillset and rational thinking are usually available within the teams evaluating the process characterization.

5. Conclusion

We prove with the conducted case study, that:

- It was possible to increase the process knowledge by considering historical data
- Meaningful time series assessment and subsequent rational information mining result in more accurate process knowledge as before
- Risk assessment linearization returns quantitative impact of PP on particular CQAs
- Holistic process assessment allows the identification of critical unit operation and PP for design space evaluation

• The knowledge gathered from unit operation based assessment has to be collected and assessed holistically for a comprehensive process evaluation

As a whole, this work suggests a nQbD approach for process characterization purposes. The cQbD approach is commonly used and well established in the biopharmaceutical industry and has, therefore, a significant potential in its use. The correct interpretation of MVDA is crucial for further process characterization cycles. In order to accelerate and to facilitate the interpretation routine, reasonable data mining and a good process understanding are essential. It was shown that these steps need to be conducted already at the beginning of the data analysis and are crucial for the remaining analysis workflow. One solution for data mining procedures and process knowledge gathering is presented in this work. It shows that already simple steps with basic statistical methods can be applied to fulfill this aim, as long as this knowledge can be transferred to risk assessment. This transfer further facilitates the holistic process assessment by conducting the suggested linearization of Severity and Occurrence estimations to comprehensively assess the process design space leading to a new control strategy evaluation.

The scope if the work was to shown simple improvements on existing data science methods within QbD. Although machine learning or artificial intelligence algorithms may result in other results when included in the QbD, it is often difficult for process experts to follow such black box model approaches if they have not sufficient statistical or mathematical background. Moreover, interpretation of the relation between the variables might be difficult when using these algorithms.

Within a next step in the research field, we would suggest to include mechanistic model for data generation. Such models can be used to simulate unit operation where process data are missing or how additional unit operation within the process can affect the process holistically.

Concluding, it was shown that it is possible to improve the current and commonly used QbD approach with simple adjustments to decrease the overall process characterization efforts and thereby potentially accelerate time to market.

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

Author contributions

DBO and CHE had the idea for the improved QbD approach. DBO wrote the manuscript and did the case study. DSU and YTH assisted in writing a reviewed the manuscript. CHE assisted in writing the manuscript.

References

1. FDA, Guidance for Industry, 2011. Available from: https://www.fda.gov/downloads/drugs/guidances/ucm070336.pdf.
2. Katz P, Campbell C (2012) FDA 2011 process validation guidance: process validation revisited. *J GXP Compliance* 16: 18–29.
3. Guideline IHT (2005) Quality risk management Q9.
4. Politis NS, Colombo P, Colombo G, et al. (2017) Design of experiments (DoE) in pharmaceutical development. *Drug Dev Ind Pharm* 43: 889–901.
5. Burdick RK, LeBlond DJ, Pfahler LB, et al. (2017) Process Design: Stage 1 of the FDA process validation guidance, In: Burdick RK, LeBlond DJ, Pfahler LB, et al., *Statistical Applications for Chemistry, Manufacturing and Controls (CMC) in the Pharmaceutical Industry*, Cham: Springer International Publishing, 115–154.
6. Doran PM (2013) *Bioprocess Engineering Principles*, Boston: Elsevier.
7. Bunnak P, Allmendinger R, Ramasamy SV, et al. (2016) Life-cycle and cost of goods assessment of fed-batch and perfusion-based manufacturing processes for mAbs. *Biotechnol Prog* 32: 1324–1335.
8. Diab S, Gerogiorgis DI (2018) Process modelling, simulation and technoeconomic evaluation of crystallisation antisolvents for the continuous pharmaceutical manufacturing of rufinamide. *Comput Chem Eng* 111: 102–114.
9. Langer ES, Rader RA (2014) Continuous bioprocessing and perfusion: wider adoption coming as bioprocessing matures. *BioProcessing J* 13: 43–49.
10. Fisher AC, Kamga MH, Agarabi C, et al. (2019) The current scientific and regulatory landscape in advancing integrated continuous biopharmaceutical manufacturing. *Trends Biotechnol* 37: 253–267.
11. Herwig C, Glassey J, Kockmann N, et al. (2017) Better by Design: Qualtiy by design must be viewed as an opportunity not as a regulatory burden. *The Chemical Engineer* 915: 41–43.
12. Herwig C, Garcia-Aponte OF, Golabgir A, et al. (2015) Knowledge management in the QbD paradigm: manufacturing of biotech therapeutics. *Trends Biotechnol* 33: 381–387.
13. CMC Biotech Working Group (2009) A-Mab: a Case Study in Bioprocess Development.
14. Welcome to Python.org. Available from: https://www.python.org/.
15. Steinwandter V, Borchert D, Herwig C (2019) Data science tools and applications on the way to Pharma 4.0. *Drug discov today* 24: 1795–1805.
16. Guideline IHT (2009) Pharmaceutical Development Q8 (R2).
17. Thomassen YE, Van Sprang ENM, Van der Pol LA, et al. (2010) Multivariate data analysis on historical IPV production data for better process understanding and future improvements. *Biotechnol Bioeng* 107: 96–104.
18. Borchert D, Suarez-Zuluaga DA, Sagmeister P, et al. (2019) Comparison of data science workflows for root cause analysis of bioprocesses. *Bioproc Biosyst Eng* 42: 245–256.
19. Suarez-Zuluaga DA, Borchert D, Driessen NN, et al. (2019) Accelerating bioprocess development by analysis of all available data: A USP case study. *Vaccine* 37: 7081–7089.
20. Kirdar AO, Green KD, Rathore AS (2008) Application of multivariate data analysis for identification and successful resolution of a root cause for a bioprocessing application. *Biotechnol Prog* 24: 720–726.

21. Sagmeister P, Wechselberger P, Herwig C (2012) Information processing: rate-based investigation of cell physiological changes along design space development. *PDA J Pharm Sci Technol* 66: 526–541.

22. Golabgir A, Gutierrez JM, Hefzi H, et al. (2016) Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. *Biotechnol Adv* 34: 621–633.

23. Bowles JB (2003) An assessment of RPN prioritization in a failure modes effects and criticality analysis. *The 2003 Proceedings Annual Reliability and Maintainability Symposium*, IEEE, 380–386.

24. Sharma KD, Srivastava S (2018) Failure mode and effect analysis (FMEA) implementation: a literature review. *J Adv Res Aeronaut Space Sci* 5: 1–17.

25. Sharma RK, Kumar D, Kumar P (2007) Modeling system behavior for risk and reliability analysis using KBARM. *Qual Reliab Eng Int* 23: 973–998.

26. Braglia M, Frosolini M, Montanari R (2003) Fuzzy TOPSIS approach for failure mode, effects and criticality analysis. *Qual Reliab Eng Int* 19: 425–443.

27. Borchert D, Zahel T, Thomassen YE, et al. (2019) Quantitative CPP evaluation from risk assessment using integrated process modeling. *Bioengineering* 6: 114.

28. Zahel T, Hauer S, Mueller EM, et al. (2017) Integrated process modeling—a process validation life cycle companion. *Bioengineering* 4: 86.

29. Bonate PL (2001) A brief introduction to Monte Carlo simulation. *Clin Pharmacokinet* 40: 15–22.

30. Wechselberger P, Sagmeister P, Engelking H, et al. (2012) Efficient feeding profile optimization for recombinant protein production using physiological information. *Bioprocess Biosyst Eng* 35: 1637–1649.

31. Abu-Absi SF, Yang LY, Thompson P, et al. (2010) Defining process design space for monoclonal antibody cell culture. *Biotechnol Bioeng* 106: 894–905.

32. Liu H, Ricart B, Stanton C, et al. (2019) Design space determination and process optimization in at-scale continuous twin screw wet granulation. *Comput Chem Eng* 125: 271–286.

33. Guideline IHT (2011) Development and manufacture of drug substances (chemical entities and biotechnological/biological entities) Q11.