A comprehensive methodology for performing Continued Process Verification

Rodrigo Carrillo
INFOTEC
Aguascalientes, Mexico
rodrigo.s.carrillo@gmail.com

Carlos Minutti
INFOTEC
Aguascalientes, Mexico
carlos.minutti@infotec.mx

Paulina Lagunes
INFOTEC
Mexico City, Mexico
paulina.lagunes@infotec.mx

***Abstract*— In pharmaceutical manufacturing, an error has a direct impact on public health, which is why governments require the validation process to ensure product quality, safety and effectiveness. However, when this process is not well-designed, it can result in either lack of control or multiple alarms that require investigation, leading to increased costs or potentially overlooking cases that require attention. Therefore, a framework that balances cost and process knowledge is needed. This work proposes an integrated methodology to analyze manufacturing processes, resulting in process knowledge while reducing the need for unnecessary investigations, thereby enhancing effectiveness of the method. A case study is presented to demonstrate how the classic approach increases costs and yields unreliable results, showing that the groups size defined can lead to higher variability requiring investigation when it is not needed. Furthermore, it produces investigation costs in situations where comprehensive data analysis would have been sufficient. The recommended methodology aims to address these limitations and provide a structured approach for pharmaceutical manufacturers.**

A. Introduction

The Pharmaceutical industry is one of the most regulated industries around the world, due to the potential impact of its products to the overall population. Validation is a process required on the manufacture of pharmaceutical products due to its contribution to the quality, safety and efficacy of the final product. Continued Process Verification (CPV) is the third phase of the Validation Process and has the objective of ensuring continuously, that the production processes maintain a state of control (validated state) during commercial manufacturing, as stated by the FDA (Food and Drug Administration) on its Guidance for Industry on Process Validation [1], by COFEPRIS (Comisión Federal para la Protección Contra Riesgos Sanitarios) at the NOM-059-SSA1-2015 [2] and by the EMA (European Medicines Agency) on its Guideline on process validation for finished products [3].

The current reference for performing data analysis to comply with this regulatory requirement is the 2011 FDA (Food and Drug Administration) Guidance for Industry on Process Validation [1], the EMA Guideline on process validation for finished products [3] and publications such as PDA's Roadmap for the Implementation of Continued Process Verification [5].

Also, Gnoth et al., 2007 [4] offer the usage of Process Automated Technology (PAT), stating advantages like understanding the influence from various process variables using real time data, as a consequence, being able to control the process and assure long-term consistency; this approach is documented by the FDA on its Guidance for Industry to use PAT[1]. While PAT usage provides the required data to comply with the regulatory requirements, not all the equipment used at the pharmaceutical processes was designed to generate real time data (no dataloggers), therefore, the available data is limited to external tests, and statistical inference is needed to control the results from the unit operations.

The Parental Drug Association (PDA), 2012 [5], offers the usage of time series plots along with X-bar and R Control charts as an effective tool to identify special cause variation. By adding the usage of the Nelson Rules (Nelson, 1984 [6]) across runs of data plotted over similar periods of time, it enables the simple and quick identification of patterns over time that might create short-term inconsistencies in the final result. The PDA approach is practical but dependent on statistical knowledge and the criteria of the data analyst; personal interpretation and wrong usage of statistical tools, such as data transformation or reporting continuous probability distributions while having discrete data, like microbiology analysis of bacterial colonies over critical systems (such as Air, Steam, Water, among others), are mistakes committed due to personal criteria.

De-Felipe, 2017 [7], mentions that constantly monitoring compliance within product specifications using capability indexes can lead to a robust process control over time. This proposal relies on external quality tests over the product to recognize deviations or unacceptable variation, which may result in delays in performing process adjustments.

The pharmaceutical industry requires a proven method to execute the Continued Verification Process; assuring compliance of regulatory international requirements, applicability with any manufacturing technology, and being a matter of public health, it should not depend upon companies' internal criteria or expertise, to be performed.

Process control was designed by Dr Shewhart, as mentioned by Wheeler, 1992 [8], where he recognized that every process has variation as part of its nature. He stated that the goal of any Process Manager is to understand the sources of variation and minimize its impact on the product. He categorized two types of errors that management can commit during the effort to understand variation:

Mistake type 1 (over-react): assuming that a process result is due to special cause variation when it's a consequence of common cause variation (natural variation).

Mistake type 2 (under-react): assuming that a process result is due to common cause variation when it is a consequence of special cause variation (assignable causes).

Consistently choosing one approach or the other to manage a process may increase risks or additional costs of investigation. Therefore, the focus should be on achieving a balance between the risk of not recognizing the process variables and product variation (risk of the unknown) and the cost of investigating and understanding the sources of variation (cost of knowledge).

B. Proposal

A method to execute CPV is proposed, considering U.S.A., European and Mexican regulatory requirements and being independent from the technological capabilities, while also balancing the cost of knowledge and the risk of unknown; for the pharmaceutical massive medicine manufacture. Fig. 1 shows the reference model to balance risk and cost of knowledge.

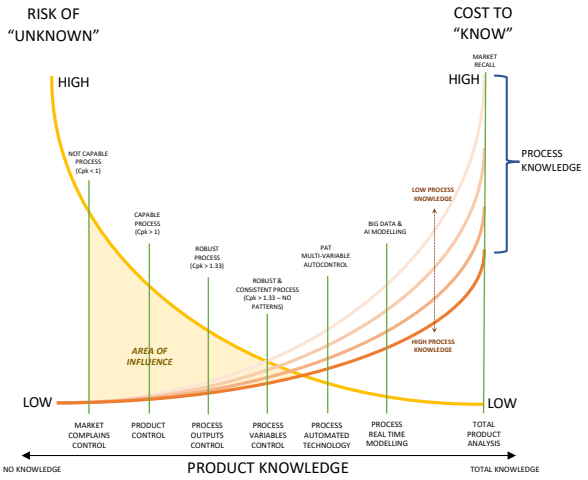

Fig. 1. Risk Vs Cost of Knowledge balance model

We recommend a model that consists of seven steps, each of which have specific tools, activities, and deliverables. It builds a step by step data analysis guidance that allows pharmaceutical manufacturing companies to develop product and process knowledge. Fig.2 shows the steps related to the mentioned methodology.

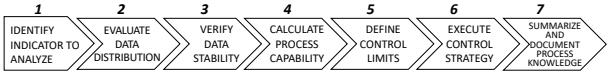

Fig. 2. Continued Verification Process method

Step 1: Identify indicator to analyze. During this activity, the individual responsible of data analysis gathers information about the specific process or product indicator to analyze and validate the data sources as well as the sampling method, creating a general process and product context.

Step 2: Evaluate data distribution. The goal of this phase is to understand the process/product behavior by relating the data to a probability distribution. This allows for statistical inferences to be made with controlled risk.

Step 3: Verify data stability. The objective is to understand the process/product behavior over time in order to identify patterns that may lead to special cause (non-inherent to the process) variation.

Step 4: Calculate process capability. Once the process is related to a probability distribution, and stability is assessed, the data set is compared against the specification limits: This allows for the assess of the probability of compliance with product/process expectations.

Step 5: Define control limits. The principle of control is to identify any deviations from the process as soon as possible to prevent out-of-specification products. A reference for this identification is assigned to the process data, thereby avoiding the use of personal criteria.

Step 6: Execute control strategy: Once the process/product has been assessed and control limits defined, the next activity is to apply the analysis to real process data, this involves applying the control reference established on Step 5 and determining whether a root cause investigation is required (special cause variation) or the process is performing as expected (common cause variation).

Step 7: Summarize and document process knowledge: The final goal of the Continued Process Verification is to generate knowledge that can be applied to the process improvement cycle. Therefore, the last step involves summarizing the analysis results and consolidating the knowledge obtained.

By consistently applying the specified method, pharmaceutical manufacturing companies are able to scientifically control their processes, detect special causes of variation and avoid relying on personal criteria when making decisions about process performance. This method considers the available manufacturing technology and establishes a standardized approach for analysis, reporting, and ultimately, manufacturing.

C. Case Study

The method is tested on a data set from a compression process of a selective antagonist of the leukotriene D_4 (LTD_4) receptor, used to decrease the early and late airway response to allergens (dust mite extract) related to a placebo [9]. The Table 1 shows the product specifications.

TABLE I. Product Summary

Product summary on Tablet compression process for 10 mg tablets.	
Description	***Result***
Production batch size	1,000,000 pieces
Visual characteristics	white, rounded, biconvex, scored tablet
Weight specification	200.00 mg +/- 5.0 % (190.00 – 210.00 mg.).
Compression speed	less than 150,000 tablets per hour
Compression force	7.4 – 9.0 KN.
Tablet weight	200.00 mg.
Tablet width	3.80 mm.
Tablet hardness	8.5 Kp.

The variable under analysis is: Tablet weight; defined as a critical output from the compression process and a significant contributor to the critical product variable: active ingredient content (a product effectiveness driver).

Step 1: Identify the indicator to analyze.

To understand what the data represents from the process under analysis, two activities are defined:

1.1 Understand data nature. The goal is to obtain information about the context of the analysis to be performed. A Context table (Fig. 3.) is created to document the process and product information, creating alignment between business requirements, product and process specifications; and establishing the basis for analysts' criteria. Fig. 3 presents the example of Context table applied for the case study.

Context table	
Indicator:	Tablet weight
Process or product indicator:	Process
Quality attribute:	Amount of products' weight after compression process
Representation of the data:	Weight of a 10 mg compressed tablet, at the end of compression process
Measurement unit:	Milligrams of product per tablet
Sampling method:	Non random sample method using the internal procedure XXX. 1312 samples were recorded along with 24 samples from the beginning of the process.
Sample extraction point:	32 pieces are extracted at the end of the compression equipment assuming one tablet per awl.
Sampling frequency:	During the 9 hour process, it was required to extract one round of 32 tablets every 15 minutes recording the order per awl
Sample extraction responsible:	John Doe, Operator Night Shift
Transforming from sample to data:	Not needed, the weight is measured directly on a scale
Measurement responsible:	Jane Doe, Quality Analyst Night Shift
Measurement system validation:	Measurement was performed at the analytic scale number XXXX model TF52L8TB. Scale sensitivity 0.00001 grm. Last calibration expiration date: 4 months before data analysis.

Fig. 3. Context table

1.2 Define data type: the analyst must identify if the analysis will be based on continuous or discrete data to select the proper statistical tools. For the studied case it was identified that the data has 7.70% of different results (101 different results over 1312 data points) showing "enough" variation (as tested previously using Zagar 2022 [10] data set) to consider the data set as continuous.

Step 2: Evaluate data distribution.

To identify if the data set can be modelled with a probability distribution, the following activities are recommended:

2.1 Identify outliers. The analyst should locate and segregate data points that are statistically distant from the overall data set. A box plot is generated as the first assessment tool for outliers, in case that outliers are observed, a Grubbs Test is suggested to statistically confirm (or reject) the presence of outliers.

As a result, for the studied case, based on the box plot and confirmed by the Grubbs test, no outliers are present on the data set. Fig. 4 shows the box plot applied to the case study where no outliers can be observed, and the Grubbs test used where a P-value of 1.000 is obtained demonstrating that No Outliers are present on the sample.

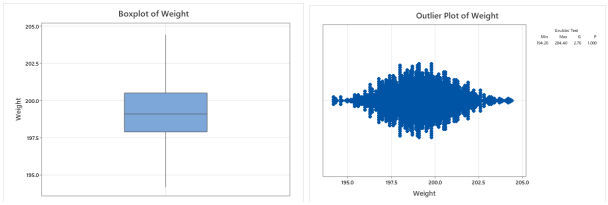

Fig. 4. Box plot and Grubbs test

2.2 Identify data probability distribution: in order to make statistical inferences about the population based on a sample, the probability distribution for the data set must be defined. A Goodness of fit test is performed to evaluate the probability of the data fitting a known distribution.

The result for the case study is that a Normal Distribution fits the data set (P-value > 0.05), so the Normal Distribution is used to make inferences about the population. The sample size is 1312, Mean = 199.19 mg, Standard Deviation = 1.89 mg, Range 194.20 to 204.40 mg. Fig. 5 presents a summary report as well as the goodness of fit test where the result of a Normal Distribution test is observed with a P-value of 0.191.

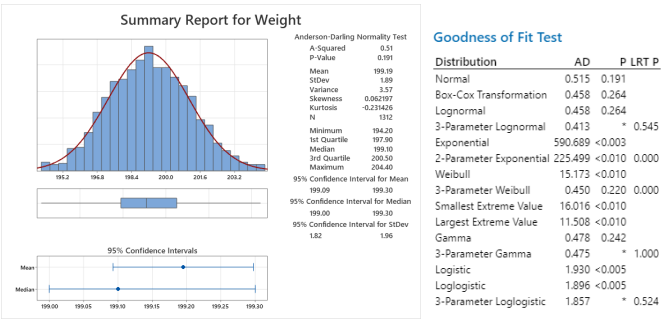

Fig. 5. Distribution and Goodness of Fit Test

Step 3: Verify data stability.

To evaluate the manufacturing process consistency over time, an analysis of patterns is done, by following the next activities:

3.1 Arrange data in chronological order and verify periodicity. To conduct a runs analysis, the data to analyze must be sorted in chronological order, and the periodicity through the sampling should be consistent.

3.2 Perform stability analysis. A Runs Analysis as used by Swed, 1943[11], may assist in diagnosing the nature of a cause.

In the case study, a Run chart analysis is conducted, with each sampling comprising 32 tablets per group, corresponding to each compression awl (41 groups of tablets were collected and extracted over a difference of 15 minutes each). As a result of the analysis, a cluster is identified (approx. P-value 0.004) running from group 11 to group 20. An investigation over the process logbook is conducted, identifying that during this period of time, there were no machine stoppages reported (continuous processing). Fig. 6 shows the Run chart analysis over the case study.

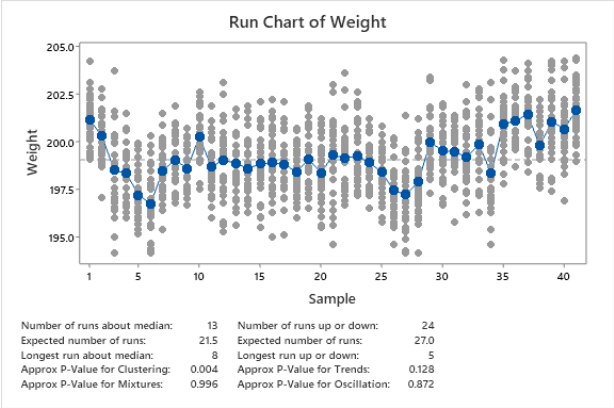

Fig. 6. Run chart of data grouped by awls sampling

Step 4: Calculate process capability.

To assess compliance of the process against product specifications, a capability analysis is performed.

4.1 Calculate process capability: Using the specification limits documented in the context table and the distribution identified from Step 2, a capability analysis is performed. This analysis calculates a ratio between the spread of the sample and the tolerance (difference between the specification limits), known as Cp - Cpk and Pp - Ppk indices (Normal Distribution). The result of this analysis is an estimation of probable results (from the population) out of specification.

Performing the capability analysis over the case study, considering the product specifications and the sampling groups (from Steps 1 Context table) and Normal Distribution (as identified in Step 2); it is determined that the Capability Ppk (as actual performance, according to Mahapatra, 2020[12]) is 1.62. This indicates that the process (population inference) and its variation are capable of complying with the specifications. Fig 7 shows the capability analysis for the case study.

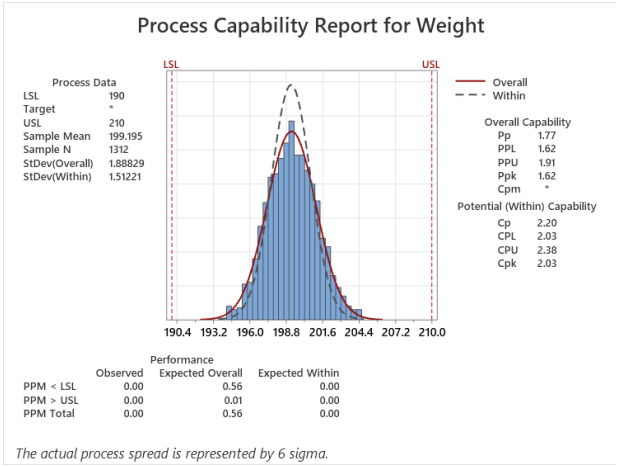

Fig. 7. Capability analysis of data set

Step 5: Define control limits

The goal of controlling a process is the early identification of potential sources of variation that are not inherent to the process (special cause variation). To establish a reference for control, the following activities are defined:

5.1 Calculate control limits: Wheeler, 1992 [13], states that a process behavior chart is sensitive enough to be used in monitoring a process. The chart based on the case study information is X-bar Chart, which utilizes the average for each group of data to assess the variation of the center, and the range of each subgroup to assess the variation of the spread.

In defining the control limits for the data set of the case study, while applying the sampling groups for the process (32

tablets one per awl), it was found that 43.90 % of the data groups (18 out of 41) are outside the proposed X-bar control limits. This would necessitate investigation into all of these groups. Fig. 8 reflects the X-bar chart using the 32 tablets per awl as groups.

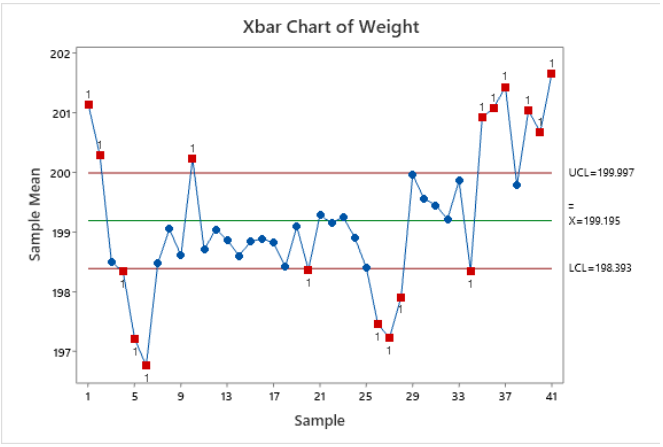

Fig. 8. X-bar chart

By using the identified control limits, we encounter an example of Mistake 1 (over-reacting) mentioned in the introduction. The control reference leads to investigate all the groups beyond the reference, but in practical terms, all of them comply with the specification and have “enough” distance (as analyzed on step 4).

In order to balance the “Risk of No Knowledge” and the “Cost of Knowledge” (Figure 1), the control limits must serve as a reference for special cause variation while also make business sense about the amount of detailed knowledge required. To understand the influence that the subgroup size has on the width of control limits, an analysis varying the subgroups size within the same data set was performed (going from 1 to 32), recording the percentage of data groups shown as out of control (beyond limits).

A quadratic regression dependency analysis (Figure 9) was performed (P-value = 0.000 and R^2 = 91.5 %) showing that there is a relationship between the subgroup size and the percentage of points out of the control limits; the widest limits are obtained when all the data is used as a single subgroup, resulting at 2.36 % of data points out of control limits (31 out of 1312). Fig. 9 shows the dependency analysis done by a quadratic regression relating the percentage of groups obtained being out of control limits against the subgroups size used to perform the calculation for the control limits.

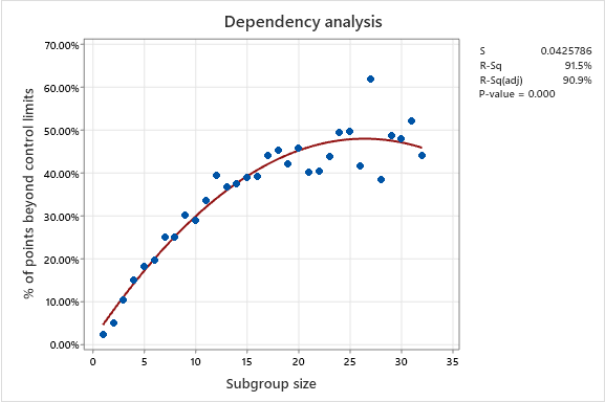

Fig. 9. Dependency analysis of variables

Considering the previous steps results - data set Normally Distributed, Not Stable (explained cluster pattern) and Capable process Ppk = 1.62- it is recommended to use the control limits calculated using all the data set as a single group of information. This yields a Lower Control Limit (LCL) of 193.52 and an Upper Control Limit (UCL) of 204.86. These limits fall within the specification limits and are still able to detect data points to be investigated.

Step 6: Execute control strategy.

As previously defined, the control limits to use are LCL = 193.52 and UCL = 204.86. Applying an X-bar chart shows that no data groups fall outside the control limits. Additionally applying the Nelson Rules, 3 groups are marked as a cluster (investigation performed in Step 3). Fig. 10 presents the X-bar chart with the control limits using a subgroup size of 1.

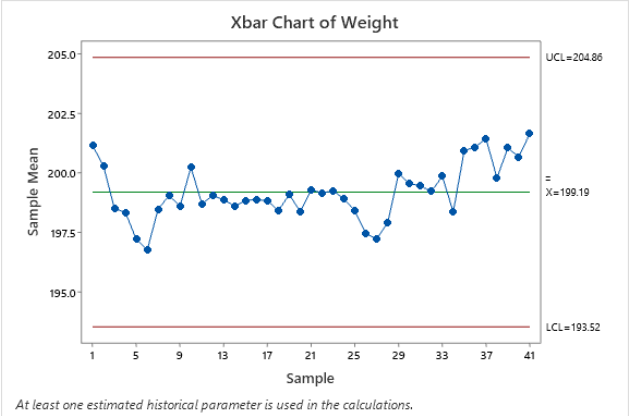

Fig. 10. X-bar chart

Step 7: Summarize and Document process knowledge

Consolidating the results for all the previous steps, it was found out that:

The process probability distribution is a Normal Distribution (Anderson Darling Test P-value 0.191).

The Process is not stable due to clusters (Approx. P-value for Clustering 0.004) from group 11 to 20.

The Process is capable with a Ppk result of 1.62.

Control limits were set at LCL = 193.52 and UCL = 204.86.

The process is under control as no groups fall outside of the control limits.

The knowledge obtained from this analysis is that the tableting equipment is able to decrease its variation when working without stoppages. Comparing the results from groups with stoppages (1 to 10 and 21 to 41) against the groups without stoppages (11 to 20), an Equal Variances test and One-Way ANOVA test were performed, showing that: There are statistically significant differences between the variances of both groups (Multiple comparisons test P-value = 0.000), and there are statistically significant differences between the means of both groups (Welch's test P-value = 0.000). It is demonstrated that the results for the compression equipment are statistically significant different when the equipment is not experiencing stoppages. To improve the process, a maintenance project team is created to decrease equipment stoppages. Fig. 11 shows the Test for Equal variances with the standard deviation chart comparison.

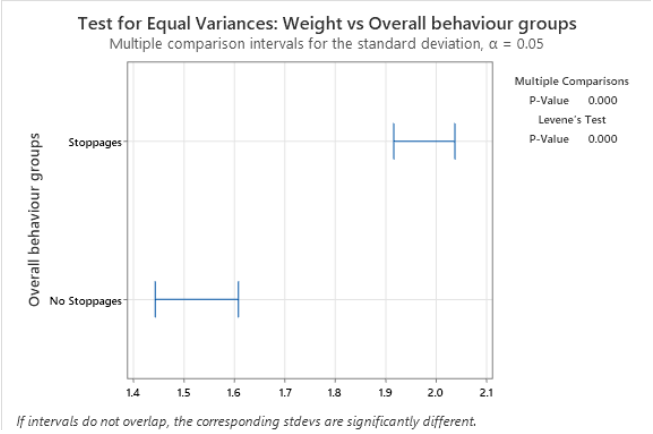

Fig. 11. Test for Equal Variances

Table II shows the Welch's test for comparing means.

TABLE II. WELCH'S TEST

Welch's Test				
Source	***DF Num***	***DF Den***	***F-Value***	***P-Value***
Overall behavior groups	1	695.084	28.66	0.000

Table III shows the ANOVA summary table.

TABLE III. ANOVA TABLE

ANOVA table				
Overall behavior groups	***N***	***Mean***	***St. Dev.***	***95 % CI***
Stoppages	992	199.333	1.974	(199.210, 199.45)
No Stoppages	320	198.768	1.519	(198.601, 198.935)

D. Results

By using the proposed method, the following results were obtained:

Step 1: Provided process and product context to those responsible for performing the analysis.

Step 2: Identified a probability distribution to make inferences about the product population based on a sample.

Step 3: Assessed the behavior of a process over time to identify potential patterns.

Step 4: Evaluated the compliance of the process and product results against the specifications.

Step 5: Established the reference for process and product control.

Step 6: Executed the analysis of process control and provided input information to improve the process.

Step 7: Documented the findings, creating process knowledge for future reference and decision-making.

The Risk Vs Cost of Knowledge balance model serves as a reference for process managers and business leaders to understand the importance of generating process knowledge as part of their responsibilities, based on a scientific approach. The use of a structured method decreases the risk of bias due to personal criteria and allows consistency on the analysis to execute Continued Process Verification.

E. Conclusion

The method presented for performing Continued Verification Process enables the pharmaceutical manufacturing companies to have a step-by-step reference for analyzing their processes and products to detect variation; using a scientific approach and creating new process knowledge. The balance between risk and cost of knowledge (investigation) was established and proved to be useful. As demonstrated in the case study, the definition of Control Limits depends upon the number of subgroups defined for the analysis, this definition should not be a personal criterion and a standard must be defined. A statistical dependency between the subgroup size and the percentage of subgroups beyond control limits was demonstrated. This method can be used by the pharmaceutical manufacturing industry to consistently comply with the Validation process required by the USA, European and Mexican regulatory agencies. The current method is limited to the gathering of continuous data as an output from the unit operations or quality analysis; further study should be done for microbiological analysis and discrete data variables that arises as part of the CPV process.

References

[1] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), & Center for Veterinary Medicine (CVM). (2011). Process-validation--General-Principles-and-practices.pdf. Retrieved from https://www.fda.gov/files/drugs/published/Process-Validation--General-Principles-and-Practices.pdf .

[2] Secretaría De Salud, Consejo De Salubridad General, Instituto Mexicano Del Seguro Social, & et al. (2016). Norma Oficial mexicana nom-059-SSA1-2015, Buenas Prácticas de Fabricación de Medicamentos. Retrieved from https://dof.gob.mx/nota_detalle.php?codigo=5424575&fecha=05%2F02%2F2016#gsc.tab=0

[3] Committee for Medicinal Products for Human Use (CHMP), & Committee for Medicinal Products for Veterinary Use (CVMP). (2016). Guideline on process validation for finished products. Retrieved from https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-process-validation-finished-products-information-and-data-be-provided-regulatory-submissions-revision-1_en.pdf

[4] Gnoth, S., Jenzsch, M., Simutis, R., & Lübbert, A. (2007). Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. Journal of Biotechnology, 132(2), 180-186. doi:10.1016/j.jbiotec.2007.03.020.

[5] Parental Drug Association, 2012. Technical Report No. 59 Utilization of Statistical Methods for Production Monitoring. Bethesda, MD: Parental Drug Association.

[6] Lloyd S. Nelson (1984) The Shewhart Control Chart—Tests for Special Causes, Journal of Quality Technology, 16:4, 237-239, DOI: 10.1080/00224065.1984.11978921

[7] De-Felipe, D., & Benedito, E. (2017). Monitoring high complex production processes using process capability indices. The International Journal of Advanced Manufacturing Technology, 93(1-4), 1257-1267. doi:10.1007/s00170-017-0591-8

[8] Wheeler, D. and Chambers, D., 1992. Understanding Statistical Process Control. 2nd ed. Knoxville, Tennesee: SPC Press, p.xi.

[9] Markham, A., & Faulds, D. (1998). Montelukast. Drugs, 56(2), 251–256. doi:10.2165/00003495-199856020-00010

[10] Žagar, J., & Mihelič, J. (2022). Big Data Collection in pharmaceutical manufacturing and its use for product quality predictions. Scientific Data, 9(1). doi:10.1038/s41597-022-01203-x.

[11] Swed, F. S., & Eisenhart, C. (1943). Tables for Testing Randomness of Grouping in a Sequence of Alternatives. The Annals of Mathematical Statistics, 14(1), 66–87. doi:http://www.jstor.org/stable/2236004.

[12] Mahapatra, A. P., Song, J., Shao, Z., Dong, T., Gong, Z., Paul, B., & Padhy, I. (2020). Concept of process capability indices as a tool for process performance measures and its pharmaceutical application. Journal of Drug Delivery and Therapeutics, 10(5), 333–344. doi:10.22270/jddt.v10i5.4288.

[13] Wheeler, D. and Chambers, D., 1992. Understanding Statistical Process Control. 2nd ed. Knoxville, Tennesee: SPC Press, p.157.

[14] Boyer, M., Gampfer, J., Zamamiri, A., & Payne, R. (2016). A roadmap for the implementation of continued process verification. *PDA Journal of Pharmaceutical Science and Technology*, *70*(3), 282–292. doi:10.5731/pdajpst.2015.006395