Automated Decision-Making Tool That Determines the Most Economic Sampling Plan for Process Validation of Pharmaceutical and Biomedical Products

Andres I. Gola

ABSTRACT
This paper proposes an automated digital system for the selection of the most cost-friendly process validation sampling plan. The system is based on the US Food and Drug Administration’s mandates and other standardized procedures. The algorithm is designed to deliver one output as a function of multiple input parameters without further interaction with the users. The system will reduce any possible subjective elections and will make the output a completely objective and quantitative result.

INTRODUCTION
Process validation is defined in FDA’s process validation guidance as “the collection and evaluation of data, from the process design stage through commercial production, which establishes scientific evidence that a process is capable of consistently delivering quality product. Process validation involves a series of activities taking place over the lifecycle of the product and process.” Process validation activities are defined by FDA in the following three stages, although some activities might occur in multiple stages (1):

- “Stage 1–process design: the commercial manufacturing process is defined during this stage based on knowledge gained through development and scale-up activities.
- “Stage 2–process qualification: during this stage, the process design is evaluated to determine if the process is capable of reproducible commercial manufacturing.
- “Stage 3–continued process verification: on-going assurance is gained during routine production that the process remains in a state of control” (1).

The guidance states, “Before any batch from a process is commercially distributed for use by consumers, a manufacturer should have gained a high degree of assurance in the performance of the manufacturing process such that it will consistently produce active pharmaceutical ingredients (APIs) and products meeting those attributes relating to identity, strength, quality, purity, and potency. The assurance should be obtained from objective information and data from laboratory-, pilot-, and/or commercial-scale studies. Information and data

ABOUT THE AUTHOR
Andres I. Gola works as a quality engineer at GCI, Inc. and is currently studying MS biomedical engineering at the University of Miami. He may be reached by e-mail at a.gola@umiami.edu.
should demonstrate that the commercial manufacturing process is capable of consistently producing acceptable quality products within commercial manufacturing conditions” (1).

FDA states that information and knowledge from product and process development is necessary for “establishing an approach to control of the manufacturing process that results in products with the desired quality attributes” (1).

However, it has been shown through time that biomedical and pharmaceutical firms have failed to adequately validate their processes. These failures are not due to poor product design or production processes. Rather, they point to the use of inadequate sampling plans that not only fail to demonstrate the proper conformance level with a degree of confidence, but also fail to take into account the operational costs of inspection.

The presented system is intended to solve this problem by creating a standardized, quantitative approach that weighs inspection cost vs. units cost in order to deliver an optimal sampling plan.

**SYSTEM DESIGN**
The algorithm was designed to deliver the most objective economic option possible. In accomplishing this, the interaction with the user is limited. The system only requires the input of lot tolerance percent defective (LTPD), the process acceptable quality level (AQL), the type of inspection (attribute or variable), the units, and inspection costs. Figure 1 shows an overview of the algorithm.

The pre-processing consists in the calculation of parameters and the introduction of data. The processing consists of selection and filtering of all those sampling plans that match the specifications. Finally, the post-processing performs the selection of the final sampling plan comparing items cost with inspection cost.

Failure Mode Effects Analysis and Lot Tolerance Percent Defective
Failure mode effects analysis (FMEA) is a bottom-up procedure in product development intended to ease the process of identifying failure modes with relatively high probability of occurring and severity of consequences. It allows corrective actions to be focused where they will produce the greatest impact.

Based on FMEA, in order to calculate the adequate LTPD for the sampling plan, a series of questions will prompt the users to where they are going to assign the following values for the algorithm to calculate the risk factor number (RPN). The algorithm takes into consideration severity, occurrence, and detectability.

The parameters are shown in Figure 2. Severity classification is assigned for each failure mode of each unique item and entered on the FMEA matrix, based upon system level consequences (2). A small set of classifications, usually having four severity levels, is used. In our algorithm, the following are used:

- **Negligible**—no potential for injury or discomfort
- **Minor**—may cause minor injury or discomfort to patient
- **Major**—may require medical or surgical intervention to prevent serious injury
- **Critical**—may cause death or serious injury to a patient.

Frequency estimation may be qualitative or quantitative. For qualitative assessment, a mishap probability code or number is assigned and entered on the matrix, as follows:
• Highly unlikely to occur—the probability of the event occurring is so low that it can be assumed that the event will not reoccur
• Unlikely to occur—no record of previous occurrence or expected to reoccur, but occurrence is theoretically possible
• Likely to occur—has occurred in the past and can be expected to reoccur if no action is taken to correct or prevent the problem
• Highly likely to occur—expected to occur or has occurred frequently (multiple times) in the past.

For each component and failure mode, the ability of the system to detect and report the failure in question is analyzed. One of the following will be entered on each row of the FMEA matrix:
• Readily detectable
• May be detectable
• Not detectable.

Risk priority number (RPN) calculation is an alternate method to criticality analysis.

RPN = detectability * severity * occurrence

Example:
The severity, frequency and detectability for a valve assembly process are as follows:
Severity → Critical = 4
Frequency → Unlikely to occur = 2
Detectability → May be detectable = 2
This gives an RPN of 16

However, a future improvement will be to represent detectability and occurrence using a probability measure and severity using a cost measure. Thus, the system presented herein leverages the known laws of probability, and RPN can then be interpreted as expected future cost of unit failure.

With these values, the algorithm will filter the appropriate sampling plan options as shown in the Table, correlating the RPN and LTPD using a database.

After the LTPD is calculated, the algorithm discards all sampling plans that don’t meet this requirement.

Process Acceptable Quality Level
The acceptable quality level (AQL) is the highest average failure rate, expressed as a percentage of process output that is still considered acceptable (3).

In a quality control procedure, a process is said to be at an acceptable quality level if the appropriate statistic used to construct a control chart does not fall outside the bounds of the acceptable quality limits.

Users of the algorithm can enter the AQL as a percentage (if using attribute inspection) or introduce the upper specification limit, lower specification limit, mean value, and standard deviation of the process (if they are performing variables inspections).

In the second case, a variable that characterizes product performance is typically used to determine whether a product’s performance or characteristic is satisfactory. The parameter must fall within the interval bounded by the lower specification limit (LSL) and the upper specification limit (USL) as shown in Figure 3.

Ppk is a measure of how close the process is to the nearest spec relative to the variation, as follows in Equation 1.

\[
Ppk = \frac{\text{Distance from mean to nearest limit}}{3 \times \text{Standard Deviation}}
\]  
[Equation 1]

Pp compares width of process (6 s) to width of spec (USL - LSL), as follows in Equation 2.

\[
Pp = \frac{\text{Upper Spec Limit} – \text{Lower Spec Limit}}{6 \times \text{Standard Deviation}}
\]  
[Equation 2]

Table: RPN and validation actions relations.

<table>
<thead>
<tr>
<th>RPN</th>
<th>Validation Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 5</td>
<td>Verification*</td>
</tr>
<tr>
<td>6 - 10</td>
<td>Moderate LTPD</td>
</tr>
<tr>
<td>11 - 15</td>
<td>Minimal LTPD</td>
</tr>
<tr>
<td>≥ 16</td>
<td>Needs Improvement</td>
</tr>
</tbody>
</table>

*Whenever the severity is critical, even when we believe the occurrence is low and detection high, this should have a corresponding low LTPD. Example: An RPN of 16 for attribute data requires an LTPD of 1% and a confidence interval of 95%.

Figure 3: Sigma Value Gaussian Distribution.
With the Ppk and Pp calculated, the algorithm can continue its selection of sampling plans.

The graphical user interface is shown in Figure 4.

Taking this into account, the algorithm continues its filtering. In this case, the algorithm will select all those sampling plans where the AQL is close to the estimated process defective level (above and below). This will guarantee that few lots will be rejected or fewest numbers of samples will be required to provide a certain level of confidence.

Items and Inspection Costs
This module of the system calculates the cost of each piece that may be rejected in the sampling plan. Also, the cost of inspection per unit will be calculated based on the salary of the inspectors and the time required to inspect each unit. This information will be requested as shown in Figure 5.

With this information, the algorithm will correlate and balance both costs in order to choose the best sampling plan in the “Evaluation and Decision” module.

Attribute and Variable Sampling Plans
Variable sampling plans require fewer samples than attribute sampling plans while providing a similar level of protection. This translates into lower material cost, less work, and less time for variable sampling plans.

On the other hand, the attribute plan decision rule will reject if too many points are “out”. Typically, the maximum number of defectives allowed in the sample is calculated with the binomial distribution. Because this sampling plan is performed using binary information (i.e., go/no-go gage), there is less information, and as a consequence, larger sample sizes will be required. However, the use of attribute metrics decreases the inspection time, which means less money is expended on operators.

Based upon the chosen parameters, the algorithm will filter only those sample plans that correspond to it (4).

Evaluations and Decision
The evaluation and decision algorithm is the most complex module in the system. The main steps in this algorithm are shown in Figure 6.

As an example, Figure 7 shows the final array of information that the algorithm will contain. After doing all costs calculation, the average is obtained. At this point the system chooses the least expensive option that corresponds to sampling plan number three.

Optimal Sampling Plan
Finally, the chosen sampling plan is presented in a graphical user interface (Figure 8) to the users that shows the
plan and the associated costs and provides the AQL, which states the process defective level required to have a 95% chance of passing the sampling plan.

In this example, the software delivered a attributes sampling plan but no variable sampling plan, because the user chose “attribute” on the window showed in Figure 4.

Furthermore, the program always presents both a single and double attribute sampling plan for the user to take into consideration. The single plan is cheaper if the lots are accepted. However, the double plan is cheaper if the first lot is rejected.

**DISCUSSION AND CONCLUSIONS**

Choosing an appropriate sampling plan for process validation is a challenging task for engineers. These plans not only have to provide a high percentage of acceptances, but also minimize the number of units needed to assure the appropriate AQL and LTPD. Unfortunately, selecting one sampling plan among a vast collection of tables and graphs can hinder the process and blur the view of the entire scenario.

In other words, selecting a sampling plan is confusing due to the large amount of data that engineers have to handle. Depending on the type of inspection procedures, the number of units built, and the cost of each piece, good options can become too expensive or just unviable.

Thus, the system presented in this paper provides an easier and more reliable way for performing this task by automatizing the selection process. This tool asks the users for input parameters in an ordered manner and reducing as much as possible any subjective evaluations.

However, the presented solution is just the tip of the iceberg and a more complete application should be developed in order to deliver a complete sampling plan from just raw information. Keep in mind that RPN is subjective. Human expert judgement and opinion may have important contributions in the sampling plan selection.

**ACKNOWLEDGMENTS**

The author thanks Roberta Goode for her guidance, collaboration, and support during the process of learning, writing, and editing this article.

**REFERENCES**


**ARTICLE ACRONYM LISTING**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQL</td>
<td>Acceptable Quality Level</td>
</tr>
<tr>
<td>FMEA</td>
<td>Failure Mode Effects Analysis</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>LSL</td>
<td>Lower Specification Limit</td>
</tr>
<tr>
<td>LTPD</td>
<td>Lot Tolerance Percent Defective</td>
</tr>
<tr>
<td>RPN</td>
<td>Risk Priority Number</td>
</tr>
<tr>
<td>USL</td>
<td>Upper Specification Limit</td>
</tr>
</tbody>
</table>

**Figure 7:** Optimal sampling plan GUI.

**Figure 8:** Optimal sampling plan GUI developed using Mathworks MATLAB.