Quality Assessment for the Evaluation of Sample Size Reduction for Continuous In-Process Data

Carola E. Alomar Bermúdez
Manufacturing Engineering
Edgar Torres, Ph.D.
Industrial Engineering Department
Polytechnic University of Puerto Rico

Abstract *— Statistical tests look for evidence that you can reject the null hypothesis and conclude that your program had an effect. With any statistical test, however, there is always the possibility that you will find a difference between groups when one does not actually exist. Power refers to the probability that your test will find a statistically significant difference when such a difference actually exists. In other words, power is the probability that you will reject the null hypothesis when you should. Performing power analysis and sample size estimation is an important aspect of experimental design, because without these calculations, sample size may be too high or too low. If sample size is too low, the experiment will lack the precision to provide reliable answers to the questions that are being investigated. If sample size is too large, time and resources will be wasted, with minimal gain. The purpose of Power Analysis and Sample Size Estimation is to provide you with the statistical methods to answer your questions quickly, easily, and accurately*

Key Terms *— Continuous In-Process Data, Process Sampling, Sample Power Analysis, Sample Size*

RESEARCH OBJECTIVES

The main objective of this design project is to make a quality assessment evaluation of the sample size reduction for continuous in-process data of a container population.

SCOPE

The statistical evaluation for the population sample size (n = 80) adequacy during compression in-process test out of limits testing is presented in this article for a pharmaceutical drug product.

RESEARCH BACKGROUND

One of the main concerns was if the sample size of the intermediate container (n = 80 units) during the compression in-process out of the limits of Stage II was representative of the container population.

As part of the compression manufacturing procedure, in the event of compression in-process results out of limits for weight, thickness or hardness, Stage II sampling and testing of a representative sample, n = 80 units, from the intermediate container is performed to accept or reject (accept with “0”; reject with”1”), the drug product contained in the intermediate container when evaluated against the in-process limits as established in the manufacturing procedure [1]. The intermediate container is emptied after every in-process sampling and testing with acceptable results.

Determining the sample size is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make implications about a population from a sample. In practice, the sample size used in a study is determined based on the expense of data collection and the need to have sufficient statistical power.

The number of caplets to include in the research study, the sample size of the study, is an important consideration in the design project. Sample size is closely tied to statistical power, which is the ability of a study to enable detection of

a statistically significant difference when there truly is one. Power analysis can be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size. It can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size. When comparing two groups of individuals, the power of a statistical test must be sufficient to allow the detection of a statistically significant difference between the two groups if indeed there is a difference. [2]

Sample size and representative samples are two related, but different issues. The sheer size of a sample does not guarantee its ability to accurately represent a target population. Large unrepresentative samples can perform as badly as small unrepresentative samples.

Representative sampling is a type of statistical sampling in which a researcher attempts to select individuals which are representative of a larger population. In statistical sampling, people gather data from a small group and try to extrapolate the results to make generalizations about a larger group. When some parts of the target population are not included in the sampled population, we are faced with selection bias, which prevent us from claiming the sample is representative of the target population. [3]

In this project we are going to be sampling the population in the intermediate container which refers to the process through which a group of representative individuals (caplets or tablets) is selected from a population for the purpose of statistical analysis. Performing this kind of sampling correctly is extremely important, as errors can lead to invalid or misleading data.

In defining a population for study, such a population must be specific enough to provide us a clear understanding of the applicability of the study to a particular situation and understanding of that same population. Therefore, it is important to select the proper method of sampling, the process by which representative individuals are randomly selected to provide insights into the entire population under study. The four methods are:

- Simple random sampling: determines the percentage of the population to be studied, assigning each individual a number and using it arbitrarily.
- Stratified sampling: identifies subgroups in the general population to equal number of individuals within different subgroups for the purpose of comparing their responses to those of other subgroups
- Systematic: the researcher picks every *nth* individual from the population that he or she is studying to gather information.
- Cluster sampling: randomly selects groups rather than individuals to be included in a study. It is a technique in which a larger population is broken up into smaller clusters. [4]

Statistical Approach & Sample Size Evaluation

This design project presents a statistical evaluation to support the intermediate container population sample size of n = 80 units during compression in-process out of limits Stage II.

Currently there are two sampling approaches to evaluate in-process physical test (i.e., weight, thickness, hardness) during the compression process. One of the approaches uses process sampling for routine Stage I testing during the compression run. The other approach uses population sampling for Stage II testing from intermediate containers in the event of out of limits results.

- Process sampling (Stage I in-process testing), is performed to monitor and control the compression process.
- Population sampling (Stage II for the intermediate container under evaluation), is performed for quality decision purposes to accept or reject the population under evaluation.

The focus of this evaluation is to determine the sampling size adequacy of n = 80 units for the population sampling approach, obtained from

intermediate containers during compression Stage II in-process testing.

There are a few factors to be considered for this project:

- Type of data (discrete vs. continuous):
  - Discrete data: can only posses a specific value and can only represent a few values. (It is what it is and its measures are limited). [5]
  - Continuous data: it is measured and represented by an infinite number of values and can possess any value, and has no natural category, meaning we cannot precisely measure its category. (For example weight, thickness and hardness). The type of data under evaluation is continuous.
- Determine what to do:
  - Describe a characteristic for a whole group (e.g., for continuous data, evaluate mean data)
  - Be within a certain precision (± ___ units)
- What confidence level is required (usually 95%)

Precision is how narrow the acceptable range needs to be for an estimate of a characteristic. Similar to the evaluation of confidence intervals (CI), precision is used to indicate the reliability of an estimate. The precision range, act as a good estimate of the unknown population response. Precision is inversely proportional to the square root of the sample size as seen in Equation 2.

In statistics, a confidence interval (CI) is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate. It is an observed interval, which frequently includes the parameter of interest if the experiment is repeated. How frequently the observed interval contains the parameter is determined by the confidence level. More specifically, the meaning of this term is that, if confidence intervals are constructed across many separate data analyses of repeated experiments, the proportion of such intervals that contain the true value of the parameter will match the confidence level [4]. In other words a confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter.

The confidence level is determined by the researcher (not by the data). If a corresponding hypothesis test is performed, the confidence level corresponds with the level of significance (i.e. a 95% confidence level reflects a significance level of 0.05) and the precision range contains the response values that, when tested, should provide an adequate standard level of significance to reflect that the evaluated sample is representative of the unknown population from where the sample was collected.

SAMPLE SIZE EVALUATION

For the specified Stage II compression in-process population sampling description using continuous data, the following sample size formulas were used as part of the evaluation estimates, Equation (1).

$$n = \left(\frac{2s}{d}\right)^2 \quad (1)$$

Where:

d = precision (± ___ units)
n = sample size
s = standard deviation

An evaluation was performed to understand the precision (± ____ units) given by a population sample size of n = 80 units during compression Stage II in-process testing. The sample size (n = 80) was evaluated to establish the precision for each in-process attribute (i.e. weight, thickness, and hardness) and to compare the resulting precision to the acceptable/decision limits for each in process physical test characteristic (decision limits are presented in Table 1 for product A, product B, and product C).

The success criterion for the evaluation was established if the resulting precision was within the decision range for the acceptable in-process attribute.

Equation (2) can be derived from Equation (1):

$$d = \frac{2s}{\sqrt{n}} \quad (2)$$

Where:

n = 80, Compression Stage II in-process testing sample size.

s estimate = (1/6) of the acceptable specification range, this assumption was to consider worst case scenario with respect to "higher/wider" resulting precision (d) for comparison to the decision limit ranges. For each in-process test s = (USL - LSL)/6

Using the previously presented equation, Equation (2), precision (d) given by the sampling size (n = 80) was estimated and compared to the acceptable specification ranges for weight, thickness and hardness in-process characteristics of the evaluated drug products. Refer to table 1 for in-process decision limits and maximum population sizes in intermediate containers for product A, product B and product C respectively.

Table 1
Compression In-Process Decision Limits and Maximum Population Size in the Intermediate Container

Product		Weight	Thickness	Hardness
Product A	Product Cget	0.604 g	5.90 mm	11.5 kp
	Decision Limits (individual)	0.574 – 0.634 g	5.60 – 6.20 mm	8.0 – 15.0 kp
	S = Range/6	10	0.1	1.17
	Maximum Population Size in Intermediate Container	Parameter 74 set for **product A** check master recipe: each 120,000 units (120,000 caplets) x (0.000604 kg/caplet) = 72.48kg 72.48 kg/2 machines sides = 36.24kg in each intermediate container (36.24 kg/intermediate container)/(0.000604 kg/caplets) = **60,000 caplets/intermediate container**		
Product B	Product Cget	0.393 g	4.10 mm	8.0 kp
	Decision Limits (individual)	0.373 – 0.413 g	3.80 – 4.40 mm	5.4 – 10.6 kp
	S = Range/6	6.67	0.1	0.87
	Maximum Population Size in Intermediate Container	Parameter 74 set for **product B** check master recipe: each 180,000 units (180,000tablets) x (0.000393 kg/tablet) = 70.74kg 70.74 kg/2 machines sides = 35.37kg in each intermediate container (35.37 kg/intermediate container)/(0.000393 kg/caplets) = **90,000 tablets/intermediate container**		
Product C	Product Cget	15.54 g (n = 20)	6.0 mm (n = 5)	12.0 kp
	Decision Limits (individual)	14.76 – 16.32 g (n = 20)	5.7 -6.3 mm (n = 5)	9.0 – 14.0 kp
	S = Range/6	0.26	0.1	0.83
	Maximum Population Size in Intermediate Container	Parameter 74 set for **product C** check master recipe: each 46,000 units (46,000 units) x (0.000777 kg/unit) = 35.74 kg 35.74 kg in each intermediate container (35.74 kg/intermediate container)/(0.000777 kg/unit) = **46,000 tablets/intermediate container**		

RESEARCH METHODOLOGY

The methodology which will be followed is the Sample Power Analysis.

Here we discuss the techniques of statistical power analysis, sample size estimation, and advanced techniques for confidence interval estimation. The main goal of first the two techniques is to allow us decide, while in the process of designing an experiment: how likely your statistical test will be to detect effects of a given size in a particular situation, and how large a sample is needed to enable statistical judgments that are accurate and reliable. The third technique is useful in implementing the previous objectives and in evaluating the size of experimental effects in practice.

Performing power analysis and sample size estimation is an important aspect of experimental design, because without these calculations, sample size may be too high or too low. If sample size is too large, time and resources will be wasted, with minimal gain. If sample size is too low, the experiment will lack the precision to provide reliable answers to what is being investigated.

In most situations in statistical analysis, we do not have access to an entire statistical population of interest, either because the population is too large, or the measurement process is too expensive to allow more than a small segment of the population to be observed. As a result, important decisions about a statistical population are made on the basis of a relatively small amount of sample data.

Typically, we take a sample and compute a quantity called a statistic in order to estimate some characteristic of a population called a parameter. In general, the larger the sample size n, the smaller sampling error tends to be. If n is too small, there is not much point in gathering the data, because the results will tend to be too imprecise to be of much use.

On the other hand, there is also a point of diminishing returns beyond which increasing n provides little benefit. Once n is large enough to produce a reasonable level of accuracy, making it larger simply wastes time and money.

To understand power, it is helpful to review what inferential statistics test. When you conduct an inferential statistical test, you are often comparing two hypotheses:

- The null hypothesis: This hypothesis predicts that your program will not have an effect on your variable of interest.
- The alternative hypothesis: This hypothesis predicts that you will find a difference between groups.

Statistical tests look for evidence that you can reject the null hypothesis and conclude that your program had an effect. With any statistical test, however, there is always the possibility that you will find a difference between groups when one does not actually exist. This is called a Type I error. Likewise, it is possible that when a difference does exist, the test will not be able to identify it. This type of mistake is called a Type II error.

Power refers to the probability that your test will find a statistically significant difference when such a difference actually exists. In other words, power is the probability that you will reject the null hypothesis when you should (and thus avoid a Type II error).

The purpose of Power Analysis and Sample Size Estimation is to provide us with the statistical methods to answer our questions, about desirable level of precision for example, quickly, easily, and accurately.

RESEARCH RESULTS

Using the presented analysis, precision (d) given by the sampling size of (n = 80) was estimated and compared to the acceptable decision limit ranges for weight, thickness and hardness in-process characteristics of the evaluated drug products. Refer to Table 2. It can be noticed that all precision (d) results met the success criteria (i.e., the resulting precision was within the decision limit range for the applicable in-process attribute.)

Table 2
Specification Range vs. Precision for n = 80 Sampling Size

Drug Product	In-Process Test	Precision d= ± units	Decision Limit Range/2 (± units)
Product A	Weight	2 mg	30 mg
	Thickness	0.02 mm	0.30 mm
	Hardness	0.3 kp	3.5 kp
Product B	Weight	2 mg	20 mg
	Thickness	0.02 mm	0.30 mm
	Hardness	0.2kp	2.6 kp
Product C	Weight	0.06 g	0.78 g
	Thickness	0.0 mm	0.3 mm
	Hardness	0.2 kp	2.5 kp

The results obtained from this study, were further analyzed modifying the standard deviation and comparing how this affected the sample size. The obtained results (for products A, B and C) were that by increasing the standard deviation, estimated sample size was unchanged, unlike the precision that increased (refer to Table 3). The results were not modified because we want the precision to be low. This is because increasing the precision does not improve accuracy, (accuracy indicates proximity of measurement results to the true value, precision to the repeatability or reproducibility of the measurement.(See Figure 1.)[6]

Table 3
Change in Precision for Product A (60,000 Caplets/Intermediate Container)

Product A	Standard Deviation	Precision (d)	Sample Size
Weight	10	2.3	76
	13	2.99	76
	16	3.68	76
Thickness	0.1	0.23	76
	0.3	0.69	76
	0.6	0.138	76
Hardness	1.17	0.2691	76
	1.20	0.276	76
	1.23	0.2829	76

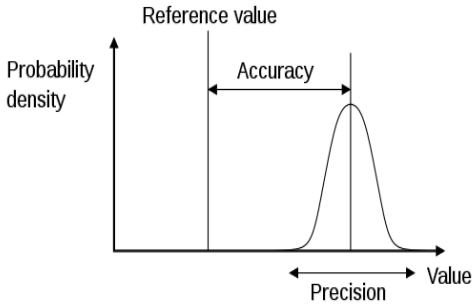

Figure 1
Accuracy vs. Precision

The following tables (Table 4, 5 and 6) and figures (Figure 2, 3 and 4) show the estimated sample size for continuous measures at a 95% confidence level (CI) for product A. The plots represent the precision vs. sample size for a maximum intermediate container population of 60,000 and a standard deviation of 10 for weight, 0.1 for thickness and 1.17 for hardness.

The same procedure was made for products B and C.

Table 4
Estimated Sample Size for Continuous Measurements for Product A (Weight)

Estimated Sample Sizes for Continuous Measures (95% C.I.)			
Enter Population Size Here*	60,000	Precision (d)	Estimated Sample Size
		0.1	24001
Enter Estimated Standard Deviation Here	10	0.2	8572
(If unknown, use 1/6 of the plausible range of the data)		0.3	4139
		0.4	2401
		0.5	1559
Note: This worksheet is used to estimate sample size for continuous data, e.g., cycle time, pressures, weights and ther continuous measures. Sampling error is the expected precision associated with the listed sample size.		0.6	1092
		0.7	807
		0.8	619
		0.9	490
		1	398
		1.1	330
		1.2	277
		1.3	237
* *For process sampling use the total number of units produced in the time period you wish to characterize*		1.4	205
		1.5	178
		1.6	157
		1.7	139
		1.8	124
		1.9	111
		2	100
		2.1	91
		2.2	83
		2.3	76

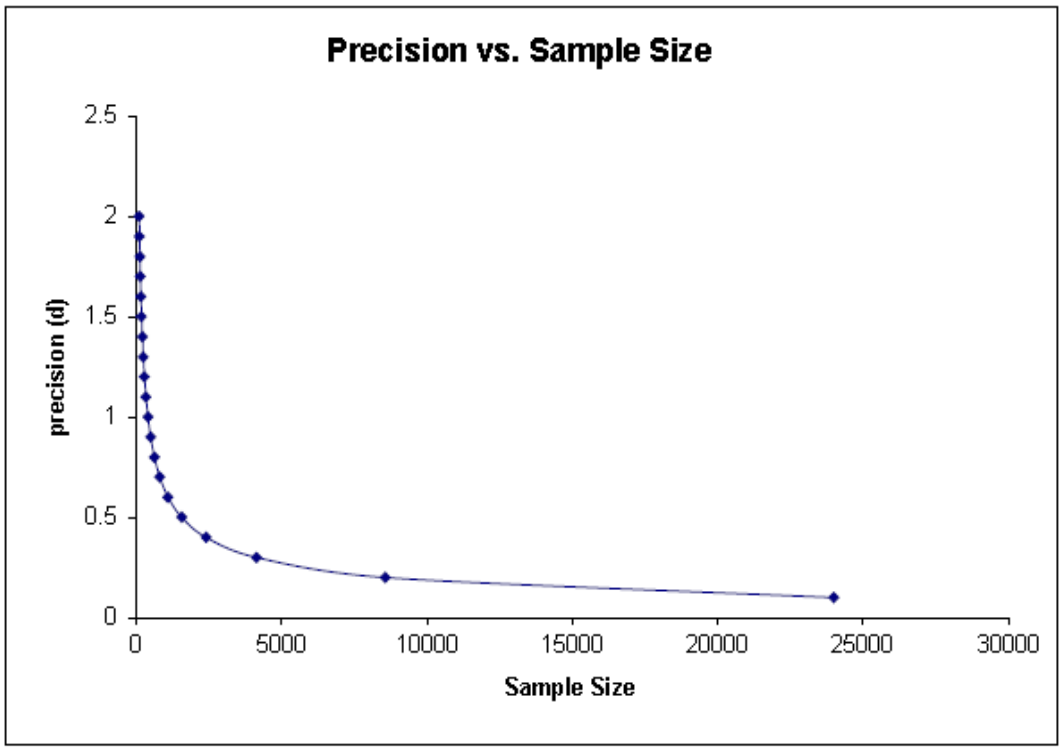

Figure 2
Precision vs. Sample Size for Product A (Weight)

Table 5
Estimated Sample Size for Continuous Measurements for Product A (Thickness)

Estimated Sample Sizes for Continuous Measures (95% C.I.)			
Enter Population Size Here*	60,000	Precision (d)	Estimated Sample Size
		0.001	24001
Enter Estimated Standard Deviation Here	0.1	0.002	8572
(If unknown, use 1/6 of the plausible range of the data)		0.003	4139
		0.004	2401
		0.005	1559
Note: This worksheet is used to estimate sample size for continuous data, e.g., cycle time, pressures, weights and ther continuous measures. Sampling error is the expected precision associated with the listed sample size.		0.006	1092
		0.007	807
		0.008	619
		0.009	490
		0.01	398
		0.011	330
		0.012	277
		0.013	237
		0.014	205
** For process sampling use the total number of units produced in the time period you wish to characterize*		0.015	178
		0.016	157
		0.017	139
		0.018	124
		0.019	111
		0.02	100
		0.021	91
		0.022	83
		0.023	76

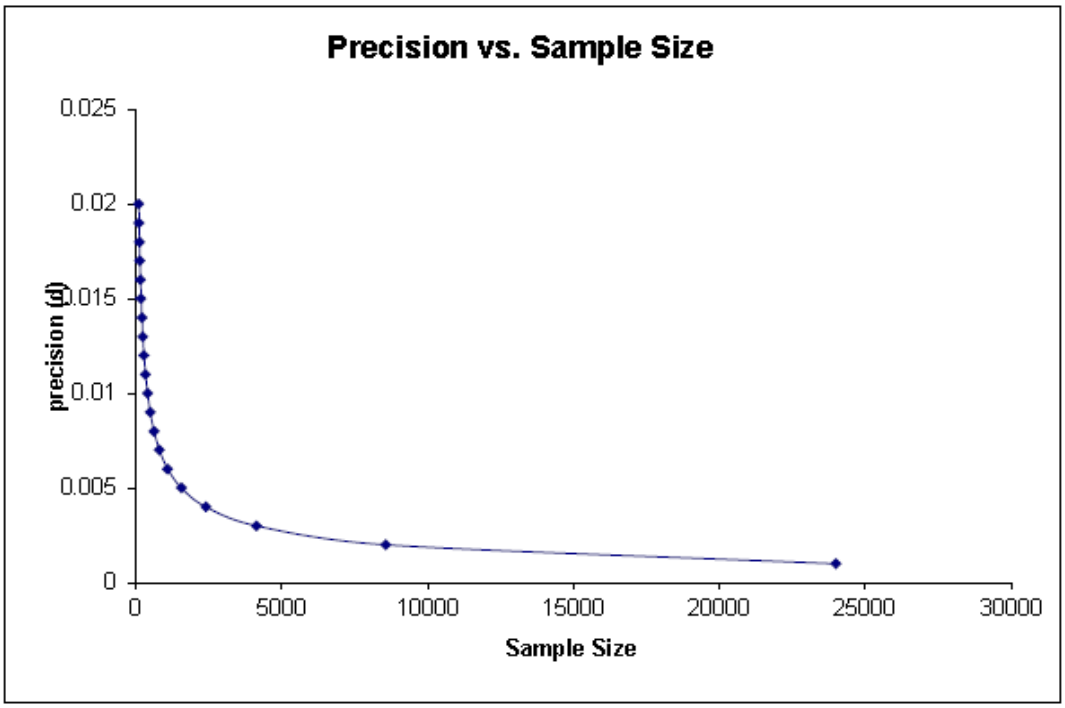

Figure 3
Precision vs. Sample Size for Product A (Thickness)

Table 6
Estimated Sample Size for Continuous Measurements for Product A (Hardness)

Estimated Sample Sizes for Continuous Measures (95% C.I.)			
Enter Population Size Here*	60,000	Precision (d)	Estimated Sample Size
		0.0117	24001
Enter Estimated Standard Deviation Here	1.17	0.0234	8572
(If unknown, use 1/6 of the plausible range of the data)		0.0351	4139
		0.0468	2401
		0.0585	1559
Note: This worksheet is used to estimate		0.0702	1092
sample size for continuous data, e.g., cycle time,		0.0819	807
pressures, weights and ther continuous measures.		0.0936	619
Sampling error is the expected precision associated		0.1053	490
with the listed sample size.		0.117	398
		0.1287	330
		0.1404	277
		0.1521	237
** For process sampling use the total number of*		0.1638	205
units produced in the time period you wish		0.1755	178
to characterize		0.1872	157
		0.1989	139
		0.2106	124
		0.2223	111
		0.234	100
		0.2457	91
		0.2574	83
		0.2691	76

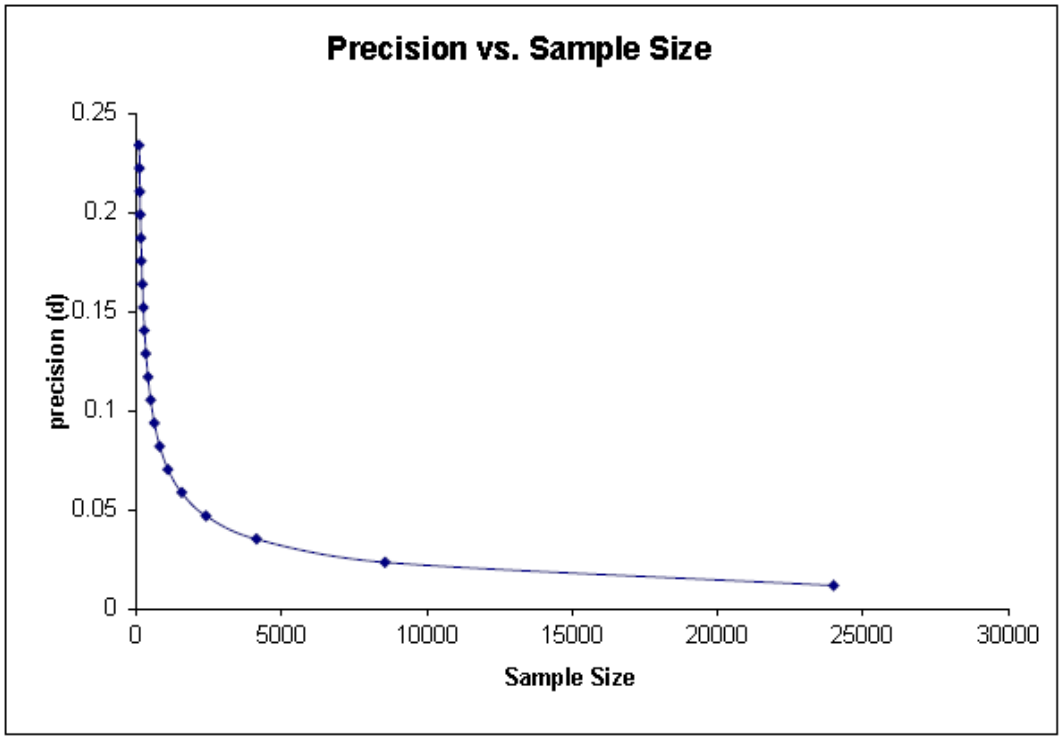

Figure 4
Precision vs. Sample Size for Product A (Hardness)

CONCLUSIONS

The objective of this project was achieved by presenting a statistical evaluation and rationale to support the adequacy of intermediate container population sample size of n = 80 units during compression in-process out of limits Stage II testing for product A, product B, and product C.

All precision results met the success criteria (i.e., the resulting precision was within the decision limit range for the applicable in-process attribute). The continuous data sample size analysis yielded favorable and tight precision range results of weight, thickness and hardness compression in-process tests at a 95% confidence level for product A, product B, and product C drug products. The precision interval range results at a significance level of 0.05, suggests that the evaluated population sample size n = 80 units is considered representative of the rest of the intermediate container population form where the sample was collected.

REFERENCES

[1] General Instructions and Guidelines for the Management of Events During the Compression Run of a lot in the Tablet Press: SOP; 35.0 Version

[2] John, E.; "Sample Size Estimation: How Many Individuals Should Be Studied?" Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University. Retrieved from: http://radiology.rsna.org/content/227/2/309.full

[3] Mora, M.; "Representative Samples, Does Sample Size Really Matter?" Retrieved on 01/010/2013, https://www.surveygizmo.com/survey-/representative-sample

[4] Dale, D.;" Population Sampling Methods for Research Studies: Definitions and Uses" Share Your Voice 04/26/2006. Retrieved from: http://voices.yahoo.com/population-sampling-methods-research-studies-definitions-32408.html?cat=4.

[5] Triola M.F. (2007). Elementary Statistics, USA: Addison Wesley

[6] Stamatis D.H. (2004). Six Sigma Fundamentals, USA: Taylor & Francis Group.