BETTER ALTERNATIVES TO SAMPLING PLANS

January 14, 2015 by John N. Zorich Jr.
johnzorich.com johnzorich@yahoo.com

ABSTRACT:

In the medical device manufacturing industry, AQL sampling plans are typically used to ensure quality in incoming product (by incentivizing the supplier to produce high quality product) and in outgoing product (by identifying low quality product). However, AQL sampling plans as commonly used do not consistently achieve either of those goals. For such purposes, there are much better alternatives to sampling plans: instead of AQL criteria for passing/failing a given lot, use %reliability specifications paired with confidence/reliability calculations that are easy to do with an electronic spreadsheet.

NOTE: This article is a text version of the seminar given by the author at ASQ's annual World Conference on Quality and Improvement, in Dallas, Texas, on May 5, 2014.

THE PROBLEM: SAMPLING PLANS ARE NOT OUR FRIENDS

Manufacturers of medical devices in the USA typically perform incoming quality control (IQC) inspection upon receiving purchased product (parts, components, or finished goods) because Food and Drug Administration (FDA) regulations require that "Incoming product shall be inspected, tested, or otherwise verified as conforming to specified requirements" (21CFR820.80). To help implement that requirement, statistically-based sampling plans are almost universally used. The FDA's requirements and guidance about such plans state that "Sampling plans, when used, shall be written and based on a valid statistical rationale" (21CFR820.250). Furthermore, "A manufacturer should be aware of the risks the chosen plan presents....A manufacturer shall be prepared to demonstrate the statistical rationale for any sampling plan used....Sampling plans have a built-in risk of accepting a bad lot....The 'operating characteristic curve' ...can be used to determine the risk a sampling plan presents" (MDQSM).

Type A and Type B sampling plans

There are two major "types" of sampling plans: "Type A" and "Type B. "The purpose of "Type A sampling [is] to accept or reject the immediate *lot* of product at hand", whereas the purpose of "Type B sampling [is] to determine if the *process* which produced the product at hand was within acceptable limits" (Juran and Gryna, p. 25.4; italics and bracketed text were added).

Type A sampling plan probabilities relate to "accepting a lot as a function of *lot quality*;" and therefore "for various values of *lot* fraction defective," the corresponding probabilities are "computed from the hypergeometric distribution." Type B sampling plan probabilities relate to "accepting a lot as a function of *product quality*;" and therefore "for various values of *product* fraction defective," the corresponding probabilities are "computed from the binomial." Type A sampling applies when production is sporadic or when the main concern is to control the "Consumer's Risk...of accepting a lot having...the lot tolerance fraction defective." Type B sampling applies when production is relatively constant or when the main concern is to control the "Producer's Risk" of rejecting "product having a fraction defective...equal to the process average fraction defective." (Dodge and Romig, pp. 56-57)

This article deals focuses primarily on Type A sampling, because consumer protection on a lot-to-lot basis is what is most important in a medical device manufacturing company.

While performing over 500 quality-system audits...

While performing over 500 quality-system audits at more than 200 medical device companies in the United States between 1999 and 2014, the author observed that almost all of medical device companies choose to use either ANSI Z1.4 or Squeglia's C=0 sampling plans for their IQC inspections; but virtually no company was able to explain or demonstrate the FDA-mandated "statistically valid rationale" for the sampling plan used, other than to say "it came from a published sampling-plan booklet."

Those companies were asked by the author for the reason they used such plans; their reply was always some version or part of this answer: "Our sampling plan provides assurance that suppliers provide us with consistently high-quality product, that QC is rejecting low quality product, and that our parts storeroom has a known quality level."

Only a few of those companies could explain how their sampling plan supported any of those three claims. The fact is that neither ANSI Z1.4 nor Squeglia's C=0 supports such claims. Z1.4 states that "the purpose of this standard is, through the economic and psychological pressure of lot non-acceptance, to induce a supplier to maintain a process average at least as good as the specified AQL while at the same time providing an upper limit on the consideration of the risk of accepting occasional poor lots. The standard is not intended as a procedure for estimating lot quality or for segregating lots" [underlining added]. And C=0 states that compared to Z1.4, its plans "provide essentially equal or greater LQ protection at the 0.10 consumer's risk level."

Do sampling plans provide assurance that suppliers provide consistently high-quality product?

Both Z1.4 and C=0 plans are Acceptable Quality Level (AQL) sampling plans in the sense that the user chooses a sampling plan from a table with headings of %AQL. Altho there is no mathematical definition of %AQL, it is generally considered to be the lot percent defective that would on average result in a pass by the sampling plan approximately 95% of the time, which equals a rejection rate of 5%.

With regards to trying to "induce a supplier to maintain a process average," a supplier is financially motivated to maintain his or her process average at no worse than that %AQL level in order to keep lot rejection rates by the customer at less than around 5%. However, the percent defective at which a 5% rejection level occurs in those 2 plans can differ by 10-fold. The OC curves in figure 1 show a 95% pass rate occurs at a defect rate of approximately 5.0% for Z1.4 and at around 0.5% for C=0. Despite this difference, the author has witnessed many companies switch from one of these plans to the other without any consideration of the affect it would have on "inducing a supplier to maintain a process average."

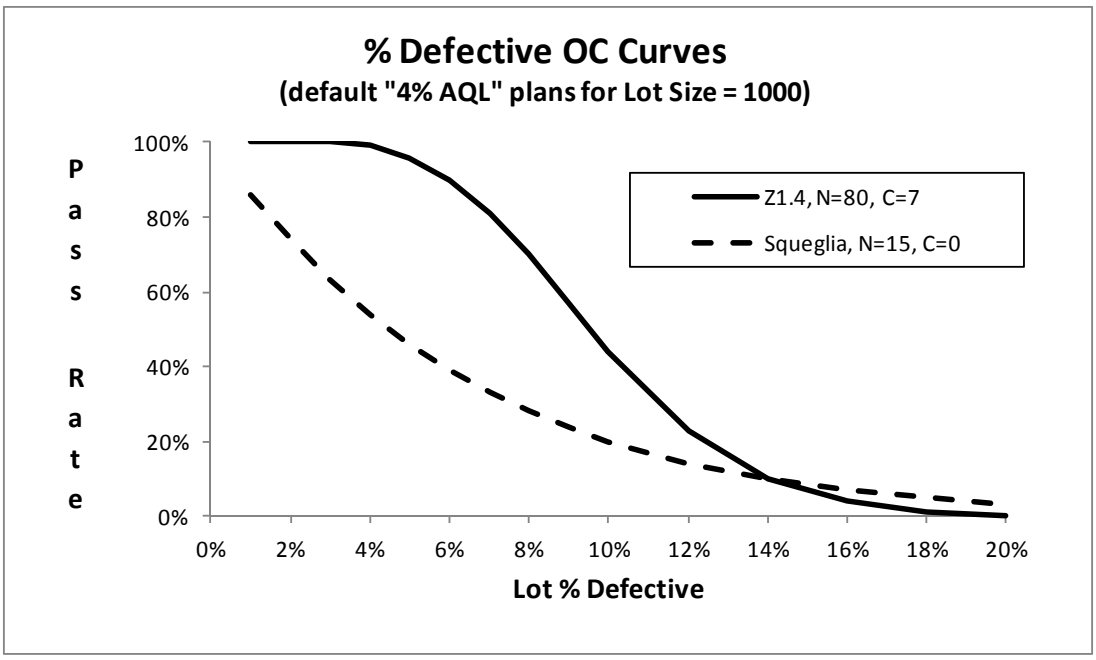

Figure 1

Do sampling plans provide assurance that QC is rejecting low quality product?

The lower end on the right-hand-side of an OC curve indicates what its Lower Quality Level (LQL) or LQ level is. Altho there is no mathematical definition of %LQL, it is generally considered to be the percent of defective product that would on average result in acceptance by the sampling plan about 10% of the time, which equals a rejection rate of 90%. This is the area of the OC curve where consumer's protection is found, since component lots with low-quality are rejected and thus kept out of the production process.

The rejection rate for a given level of low-quality product can differ by 10-fold for a given AQL level, depending on lot size. Table 1 shows a range from 6% to 60%. Similar differences can be found throughout Z1.4 and C=0, as shown in figures 2 and 3. Thus the use of these plans, without controlling the sizes of lots received, can result in a wide variety of protection levels, lot to lot.

Z1.4, general, level II, single, normal, 4% AQL

Lot Size	Sample Size	" C "	*If lot is 15% Defective* Pass Rate is...
20	3	0	60 %
100	20	2	38 %
1,000	80	7	6 %

Table 1: Example comparison of lot-size vs. lot pass rate

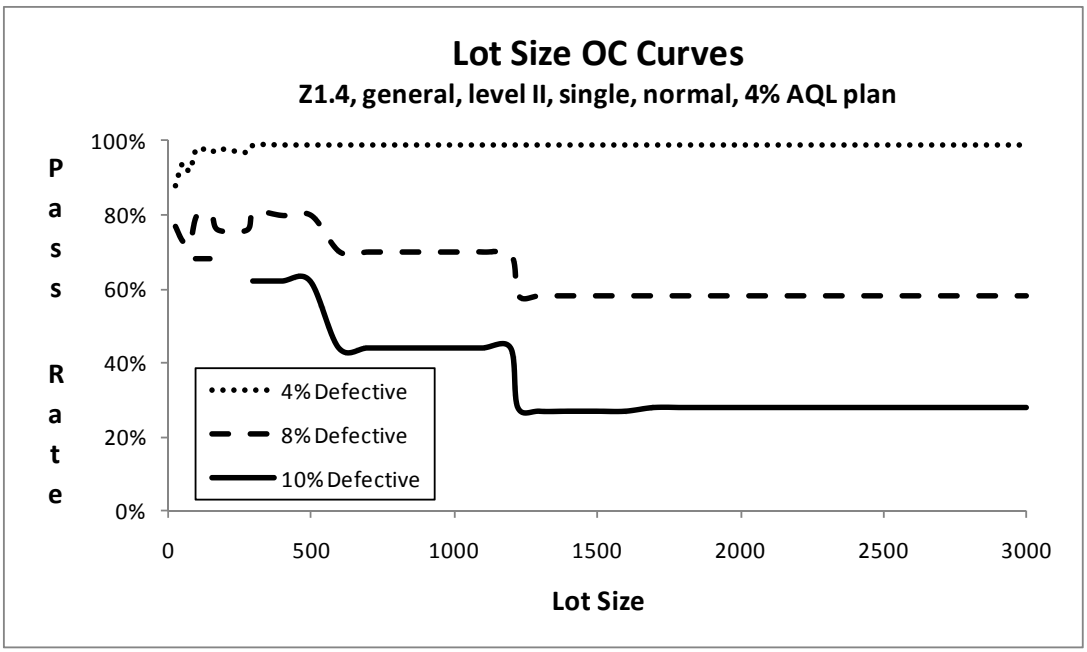

Figure 2: Lot-size OC curve example, based on Z1.4

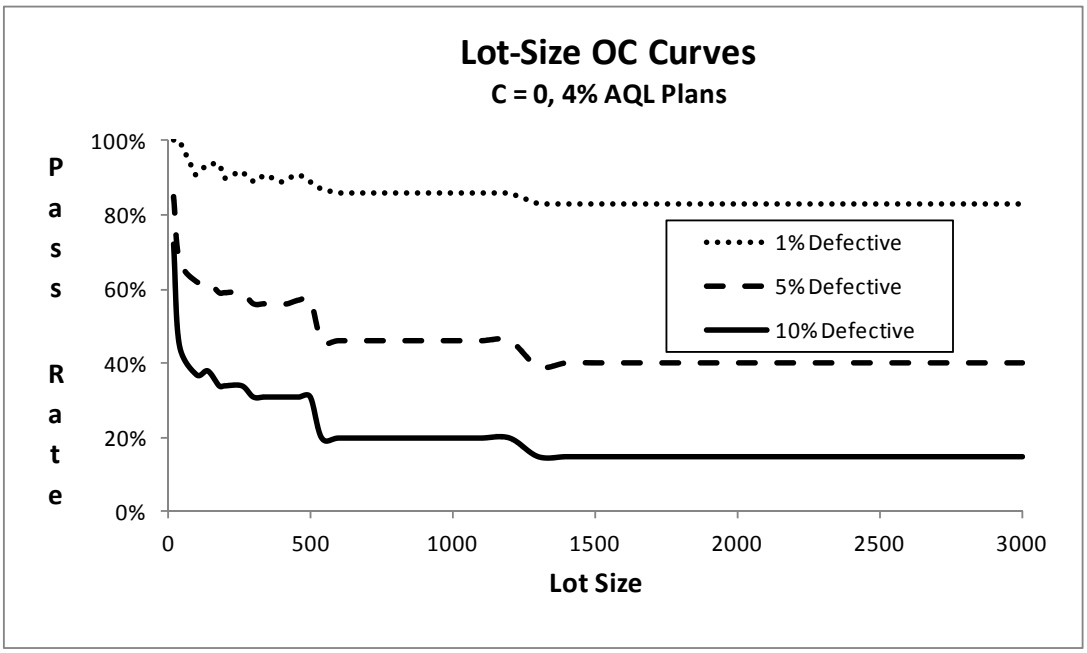

Figure 3: Lot-size OC curve example based on Squeglia (2008).

Do sampling plans provide assurance that our parts storeroom has a known quality level?

The quality of product in the Storeroom (or wherever a company maintains its approved inventory of parts, components, and finished goods) is typically described in terms of AOQ (average outgoing quality). AOQ and AOQL are typically expressed in units of percent defective. AOQ is described as "the expected average quality level of outgoing product for a given value of incoming product quality" where the term “incoming product” means parts, components, or finished goods that have been received for inspection by IQC either from another

department or from an external supplier, and where the term “outgoing product” refers to what is released by IQC for use in production (Quality Glossary, p. 41). AOQL is described as the “maximum limit to the percent of defectives present in the outgoing product” (Juran, p. 78). When a wide range of incoming lot or product percent defective is plotted vs. their corresponding AOQ’s, as shown in figure 4, the peak of the curve identifies the AOQ that equals the AOQL.

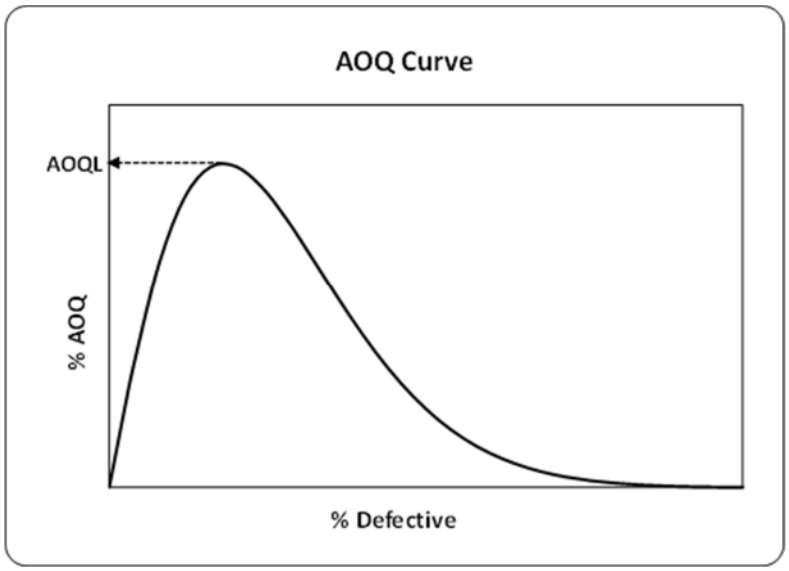

Figure 4: Typical AOQ curve.

The problem is that AOQ and thereby AOQL varies significantly by lot size, as shown in table 2. Thus the use of these plans, without controlling the sizes of lots received, can result in a wide variety of potential storeroom quality.

Approximate Lot Size	Approximate AOQL
12	0.38 %
20	1.4 %
38	2.1 %
70	2.4 %
120	2.6 %
215	1.7 %
390	1.2 %
850	1.1 %
2,200	0.86 %
6,600	0.73 %
22,500	0.61 %
92,500	0.50 %
325,000	0.41 %

Table 2: AOQL for 1%AQL C=0 sampling plan (adapted from Squeglia, 1994, p. 11).

What has been learned by using an AQL sampling plan?

It is unclear how much new information has been acquired after QC completes its inspection of a lot. A pass result at a given AQL does not itself tell us whether the lot has a high or low level of percent

defective, and the quality level in the warehouse in which it will be stored is still unknown. These can be serious problems when dealing with incoming product whose quality level is critical to end-product safety and/or performance.

THE SOLUTION: CONFIDENCE/RELIABILITY SPECIFICATIONS AND CALCULATIONS

We'd have something useful to say about a just-accepted lot if we were to calculate how much of it is in-specification, i.e., if we were to calculate %confidence and %reliability statements for each inspected lot. The term %reliability is used here synonymously with % in-specification, and the term %confidence is used for how much trust we have in our calculation of %reliability. In effect, such a calculation is the lower one sided confidence limit on the %reliability observed in the sample. Using such a calculation has the additional benefit that the sample size itself is always statistically valid because such a limit is automatically mathematically adjusted based upon the chosen sample size and confidence level.

A calculation of %Confidence and %Reliability is relatively simple. For attribute data (i.e., inspection results that provide only pass or fail conclusions), the following formula outputs the lower one sided exact binomial confidence limit of the quality level observed in the sample. This formula outputs the lot's %Reliability at the chosen %Confidence:

=betainv (1 – C , N – F , F + 1)

where...
C = Confidence desired (expressed as a decimal fraction)
N = sample size
F = number of defective items in the sample
betainv = the MS Excel function for the beta-distribution
(per Dovich, Krishnamoorthy, and Tobias and Trindade, the beta distribution is an extremely accurate model for the binomial)

When data are variables and normally distributed, %Reliability can be determined at a chosen %Confidence by using what are called Normal K-Factors or Normal Tolerance Limits tables, or they can be calculated with formulas. Such tables and formulas have been in use for many decades, and can be found in various sources such as Juran and Gryna, Bowker and Lieberman, Dixon and Massey, and NIST. To use such a table, calculate the difference between the sample average and the nearest side of the QC specification. Divide the result by the standard deviation of the sample. The absolute value of that ratio is then sought in the appropriate row on the appropriate table (there are different tables depending upon whether the QC specification is 1-sided or 2-sided). The applicable column's header indicates the %Reliability, such as in table 3.

For example, the following N=10 data set is normally distributed; that has been confirmed by its forming a nearly straight-line on a "normal probability plot" and its having passed three different "significance tests of normality", each with $p > 0.90$:

3.78, 4.18, 4.60, 4.70, 5.11, 5.12, 5.60, 5.75, 5.97, 6.54

The average of those 10 numbers is 5.135, and their standard deviation is 0.849. If their 1-sided QC specification is "≥2.500," then the value to look-up in a 1-sided K-table is

(5.135 – 2.500) / 0.849 = 3.104

The K-table value of 3.104 for N = 10 falls midway between 95% and 99% reliability. The reliability is 96.17% when determined using Excel and a curvilinear interpolation of a 4th-order polynomial.

Table of 1-Sided Normal K-Factors (a.k.a., Normal Tolerance Limits)				
Sample Size	**Reliability Level, at 95% Confidence**			
	90%	95%	99%	99.9%
3	6.158	7.655	10.552	13.857
4	4.163	5.145	7.042	9.215
5	3.407	4.202	5.741	7.501
6	3.006	3.707	5.062	6.612
7	2.755	3.399	4.641	6.061
8	2.582	3.188	4.353	5.686
9	2.454	3.031	4.143	5.414
10	2.355	2.911	3.981	5.203
11	2.275	2.815	3.852	5.036
12	2.210	2.736	3.747	4.900
etc.	etc.	etc.	etc.	etc.

Table 3: K-factor Table (adapted from multiple sources)

For variable data composed of measurements that are not normally distributed but that can be transformed into normality, %Reliability is determined by applying the K-factor-table method to the transformed values. For example, the distribution of the following N=12 data set is considered non-normal based upon its normal probability plot being obviously curved:

8.80, 9.30, 10.30, 10.80, 11.00, 12.40, 13.23, 13.34, 14.98, 17.20, 18.06, 20.80

After the values are transformed by taking their inverse, their normal probability plot looks very straight. The average 1/X value is 0.080 and the standard deviation of those transformed values is 0.021. If the 1-sided QC specification is 5.50, its value after a 1/X transformation is 0.180. The value to look-up in a 1-sided K-table is

$$(0.180 – 0.080) / 0.021 = 4.878$$

The value of 4.878 falls midway between 99% and 99.9% reliability on the K-table row for N=12. The reliability is 99.89% when determined using Excel and a curvilinear interpolation of a 4th-order polynomial.

For variable data comprising measurements that cannot be transformed into normality, reliability can be determined using reliability plotting (Tobias and Trindade). Reliability plots are cumulative distribution probability plots that have had their Y and/or X-axis values transformed in such a way as to linearize the plotted points. Examples of commonly used transformations for the X-axis include square-root, natural log, and inverse. Examples of commonly used transformations for the Y-axis include Ln(1/(1–F)), Ln(Ln(1/(1–F))), Ln(F/ (1–F), and Ln(1/(Ln(1/F))), where "F" represents the median

rank of the %Cumulative expressed as a decimal fraction. The following formula for F is recommended (Tobias and Trindade, p. 164):

F = BETAINV (0.5 , Rank , Sample Size – Rank + 1)

where BETAINV is the MS Excel function; and a Rank of 1 is given to the smallest data value in the sample, a Rank of 2 is given to the next smallest, a Rank of 3 to the next smallest, and so on.

After such linearization, linear regression is employed to determine what transformed %Cumulative on the Y-axis corresponds to the extrapolated transformed X-axis value for the QC specification. The upper 1-sided confidence limit is then calculated for that transformed Y-axis value. Then the confidence limit is reverse-transformed yielding a value in units of untransformed %Cumulative. Finally, that %Cumulative is subtracted from 100% to give the %Reliability at the confidence level that was used to calculate the confidence limit.

For example, the distribution of the following N=10 data set is non-normal based upon its normal probability plot being extremely curved:

3, 6, 9, 16, 27, 39, 80, 150, 300, 2000

Even tho this data produces a line with a high correlation coefficient (0.985) on a log-normal normal probability plot, it has a definite curve to it and therefore cannot be accurately extrapolated by linear regression. However, if the %Cumulative is transformed using Ln(1/(Ln(1/F))), a probability plot vs. Ln(X) is very straight and therefore can be extrapolated accurately to the log of the QC specification of 2.00, as depicted in figure 5.

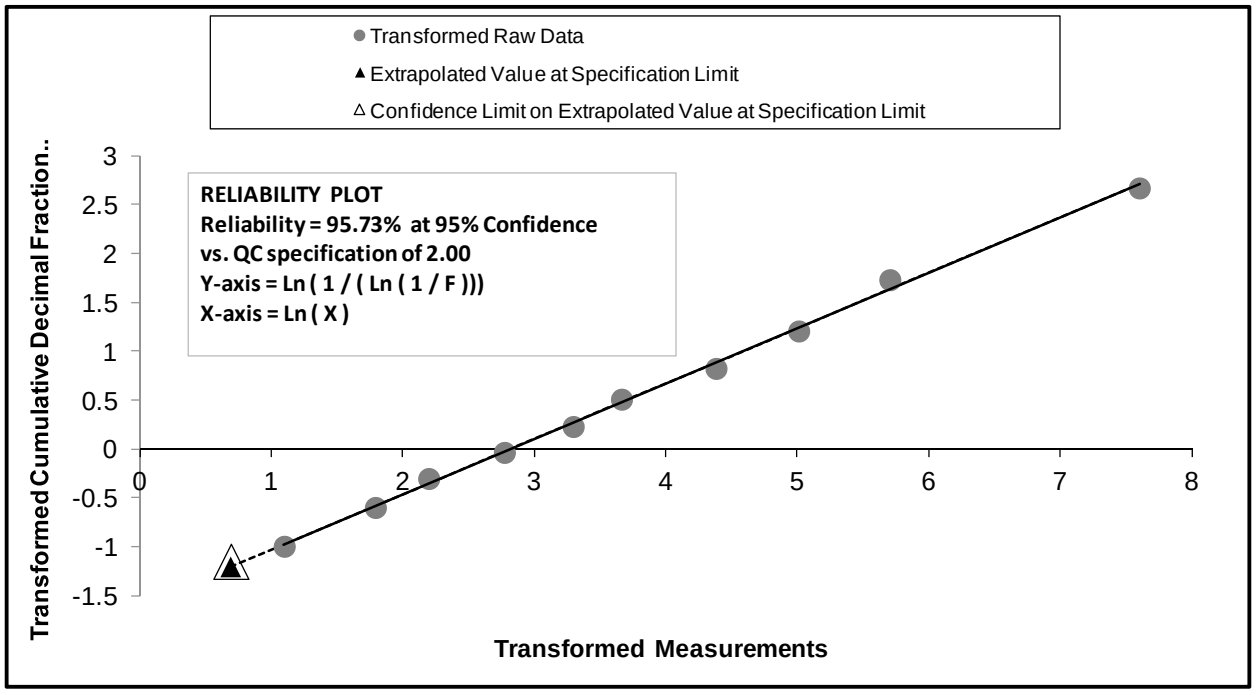

Figure 5: Reliability plotting example.

CONCLUSION AND RECOMMENDATIONS:

Commonly-used AQL sampling plans do not provide much information in regards to the quality level of a given Lot. On the other hand, confidence/reliability calculations do provide explicit information about a given lot's quality and such calculations are as easy to use as an AQL table when an electronic spreadsheet is used. Therefore confidence/reliability specifications are more useful than AQL ones for incoming QC activities in a regulated environment such as a medical device manufacturing company. Instead of a QC specification such as AQL 1.00%, consider using a QC specification such as 97% Reliable at 95% Confidence.

REFERENCES

- 21CFR820. (1996). *Medical Devices: Current Good Manufacturing Practice, Quality System Regulation*
- ANSI/ASQ Z1.4-2008, *Sampling Procedures and Tables for Inspection by Attributes*, Milwaukee, WI: ASQ Quality Press
- Bowker, A. H., and Lieberman, G. J. (1972). *Engineering Statistics*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall
- Dixon, W. J., and Massey, F. J. Jr. (1969). *Introduction to Statistical Analysis,* 3rd ed., New York, NY: McGraw-Hill
- Dodge, H. F. and Romig, H. G. (1959). *Sampling Inspection Tables – Single and Double Sampling*, 2nd ed. New York, NY: John Wiley & Sons.
- Dovich, R. A. (1990). *Reliability Statistics*, Milwaukee, WI: ASQC Quality Press
- Juran, J. M., and Gryna, F. M. (1988). *Juran's Quality Control Handbook*, 4th ed. Columbus, OH: McGraw-Hill Inc.
- Juran, J. M. (1997). Early SQC: A Historical Supplement. *Quality Progress*. 30(9): 73-81.
- Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Boca Raton, FL: Chapman & Hall/CRC press
- *Medical Device Quality Systems Manual.* (MDQSM). (1996). HHS Publication FDA 97-4179
- NIST/SEMATECH. (10/30/2013). *e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook
- Pyzdek, T. (1992). *Quality Engineering Handbook*, Milwaukee, WI: ASQC Quality Press
- Quality Glossary. (2007). *Quality Progress*, 40(6): 39-59
- Squeglia, N. L.(1994). *Zero Acceptance Number Sampling Plans*, 4th ed., Milwaukee, WI: ASQ Quality Press
- Squeglia, N. L. (2008). *Zero Acceptance Number Sampling Plans*, 5th ed., Milwaukee, WI: ASQ Quality Press
- Tobias, P. A., and Trindade, D. C. (2012). *Applied Reliability*, 3rd edition, Boca Raton, FL: Chapman & Hall/CRC press