2017

Process Capability Calculations with Nonnormal Data in the Medical Device Manufacturing Industry

James Walter Kwiecien
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

James Kwiecien

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Robert Levasseur, Committee Chairperson,
Applied Management and Decision Sciences Faculty

Dr. Thomas Spencer, Committee Member,
Applied Management and Decision Sciences Faculty

Dr. David Gould, University Reviewer
Applied Management and Decision Sciences Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2017

Abstract

Process Capability Calculations with Nonnormal Data in the Medical Device Manufacturing Industry

by

James W. Kwiecien

MBA, University of St. Thomas, 1979

BS, Loyola University, 1969

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

May 2017

Abstract

U.S. Food and Drug Administration (FDA) recalls of medical devices are at historically high levels despite efforts by manufacturers to meet stringent agency requirements to ensure quality and patient safety. A factor in the release of potentially dangerous devices might be the interpretations of nonnormal test data by statistically unsophisticated engineers. The purpose of this study was to test the hypothesis that testing by lot provides a better indicator of true process behavior than process capability indices (PCIs) calculated from the mixed lots that often occur in a typical production situation. The foundations of this research were in the prior work of Bertalanffy, Kane, Shewhart, and Taylor. The research questions examined whether lot traceability allows the decomposition of the combination distribution to allow more accurate calculations of PCIs used to monitor medical device production. The study was semiexperimental, using simulated data. While the simulated data were random, the study was a quasiexperimental design because of the control of the simulated data through parameter selection. The results of this study indicate that decomposition does not increase the accuracy of the PCI. The conclusion is that a systems approach using the PCI, additional statistical tools, and expert knowledge could yield more accurate results than could decomposition alone. More accurate results could ensure the production of safer medical devices by correctly identifying noncapable processes (i.e., processes that may not produce required results), while also preventing needless waste of resources and delays in potentially life-savings technology, reaching patients in cases where processes evaluate as noncapable when they are actually capable.

Process Capability Calculations with Nonnormal Data in the Medical Device Manufacturing Industry

by

James W. Kwiecien

MBA, University of St. Thomas, 1979

BS, Loyola University, 1969

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

May 2017

Dedication

I dedicate the successful outcome of my doctoral efforts to my wife, Susan M. Kwiecien, Ph.D. Without her patience, support, and counsel none of this would have been possible.

Acknowledgments

The completion of this dissertation would not have been possible without the help, support, and encouragement of many people throughout the process. First, I want to thank Dr. Robert Levasseur for accepting the chair role during the approval process for my proposal. I will always be grateful for his willingness to accept this role so late in the process, for quick the quick turnaround times on my submissions, and for the advice and help without which my success would not have been possible. Second I am grateful to Dr. Thomas Spencer for serving on my committee and particularly for providing valuable advice on sampling that helped me produce more meaningful research.

I also want to thank Dr. David A. Gould, my University Research Reviewer, whose insight and knowledge also contributed to the quality of this document. In addition, I want to thank Dr. Kimberly L. Ross, my first mentor, whose growing work and family obligations led to her leaving Walden, and Dr. Christos N. Makrigeorgis, who provided guidance and encouragement as my mentor for several years, but who transitioned to a different role within Walden. I am also grateful for the efforts of other Walden faculty members who contributed to my success.

Finally, I want to thank two highly influential teachers from my distant past, Dr. Albert C. Claus, and the late Father Gerald T. Hughes, S.J. They taught me how to think clearly “in this, the most perfect of all possible universes.”

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Manufacturing firms in the highly regulated medical device industry rely on process capability indices (PCIs) to ensure that the processes used to manufacture products are capable and under control. However, analysis of nonnormal data with a PCI can result in false indicators. From a management viewpoint, misreading the degree of control of processes can result in large fines and possibly even felony convictions.

From a social change perspective, this study was important for three specific reasons. First, correcting a capable process when PCIs give a false reading wastes resources. This waste of resources needlessly increases the cost of the devices. Second, along with wasting resources, time spent improving an already capable process could delay the introduction of potentially life-saving devices to the health community. Third, if PCIs overestimate the capability of a process, patient wellbeing is at risk. These conditions represent factors that contribute to the health care system inefficiencies.

This chapter contains the following: (a) the history of PCIs, the Food and Drug Administration (FDA) regulatory environment, and the important role PCIs fill in the medical device manufacturing environment; (b) a formal problem statement along with a description of the method suggested to answer the research questions generated by the problem statement; and (c) an examination of the significance of this study with regard to its potential contribution to theory, practice, and positive social change.

Background of the Study

This section contains (a) a very short description of medical device manufacturing and testing, (b) an account of the function of PCIs including the history of their

development, (c) a summary of the issues with applying them to everyday problems involving the analysis of nonnormal data, and (d) an explanation of how the lack of training in advanced statistical techniques exacerbates the problems with using PCIs. The section concludes with a description of the development of a tool that could remedy this situation under some conditions, providing a potentially valuable addition to engineering methods.

Medical Device Manufacturing

Medical devices, like many other products, are composed of various components, which themselves may be made up of other components. Figure 1 is an illustration of a simplified block diagram of a generic device.

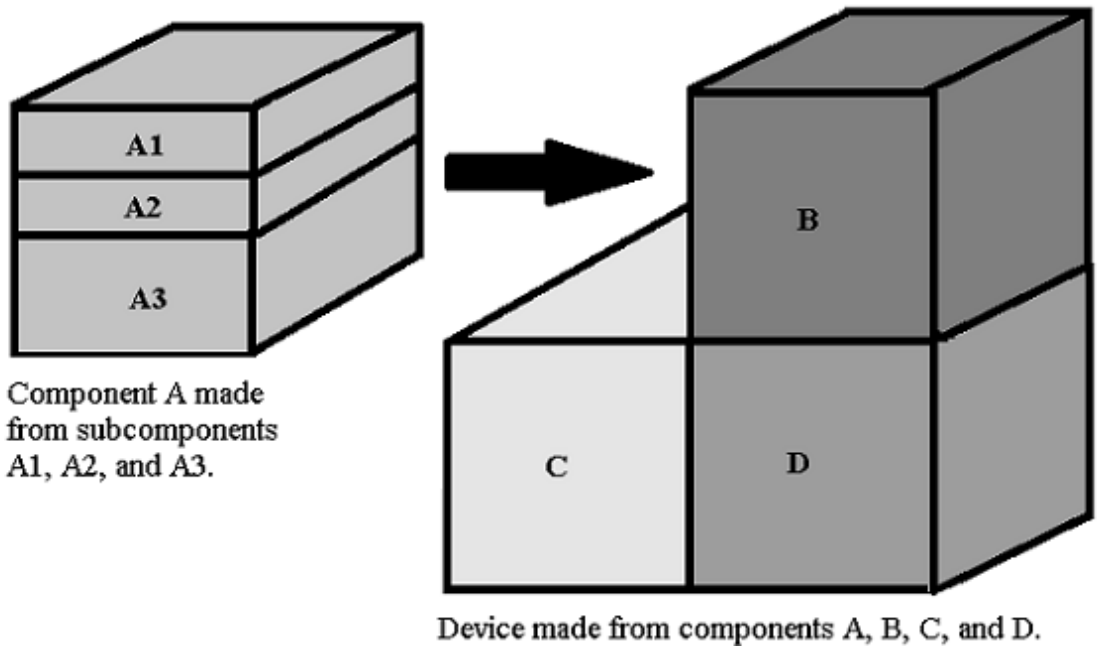

Figure 1. Simplified block diagram of a manufactured product.

Performance characteristics of the components used in the manufacture of a medical device, as well as the finished device itself, must be capable of meeting their

design requirements. To ensure that this is the case, testing that yields quantifiable results is the preferred method. There are two approaches to demonstrate that components or devices meet design requirements (International Organization for Standardization [ISO], 2016).

The first, and preferred method, is to validate the manufacturing process. This is a procedure that ensures that the process is capable of consistently producing output that meets specifications. In cases where process validation is not possible, regulatory bodies require verification of the output, usually through an inspection of the output, which might include sampling for more intense inspection that can include destructive testing (ISO, 2016).

The goal is to ensure process output that is capable of consistently meeting the specifications. A common tool used to monitor process output is the PCI. The primary motivation for this dissertation was to examine the use of PCIs in the measurement of key quantifiable test results.

Process Capability Indices

Description. PCIs are statistical tools used to determine if a process can produce output that consistently meets specifications (Kotz & Johnson, 1993). Intrinsically, the benefit of keeping a process under control is a simple idea to grasp, but there may be different reasons beyond the obvious for requiring this condition. First, if an output fails to meet its desired specifications, it could require rework to become usable, or, in other cases, scrapped. Both of these situations can have a severe effect on both the cost and capacity of an operation. The goal of process design is to build a process that can produce

output correctly, not just the first time, but almost every time. The use of PCIs can be a valuable tool in determining if this goal is achievable.

A second reason for the use of a PCI is the need to meet customer requirements. The third reason is, in some cases, manufacturers must demonstrate that processes are capable of producing their desired output by regulatory agencies (Peña-Rodríguez, 2013). The FDA has the authority to shut down a noncomplying operation and impose substantial fines upon, or even to bring criminal charges against, the management of such activities (Maximum Penalty Amounts, 2014). Another option available to the FDA is to impose a product recall in the case of products that present a danger to public safety (FDA, 2017). Such recalls may not only be expensive in terms of the product removed from user's inventories, but may also have an effect on a company's reputation and stock price. In the worst case, a nonconforming product might result in the injury or death of patients.

Development. The earliest reference to PCIs in Kotz and Johnson's (2002) exhaustive bibliography of references on the topic is the third edition of Juran's *Quality Handbook* published in 1974. This reference would date the use or discussion of this tool to the period prior to the great wave of interest in Japanese management and statistical methods that characterize the period beginning in the early nineteen eighties. Juran's reference (as cited in Kane, 1986) is to the most basic of the PCIs, C_p, now defined as:

$$C_p = \frac{USL - LSL}{6\sigma}, \tag{1}$$

where *USL* and *LSL* refer to the Upper and Lower Specification Limits respectively, and σ is the standard deviation of the process. Other authors use *UTL* and *LTL*, where the *T* refers to tolerance rather than specification (Shewhart, 1931/2015). A normally distributed process output with a known standard deviation forms the underlying assumption (Kane, 1986).

Upon the publication of the fourth edition of Juran's Handbook in 1988, Gryna (1988) expanded the section on PCIs to include the work of Kane (1986), who explicitly added C_{pk}, defined as:

$$C_{pk} = Min\left\{\frac{\mu - LSL}{3\sigma}, \frac{USL - \mu}{3\sigma}\right\}. \tag{2}$$

In *Equation 2*, μ refers to the mean of the process. The equation requires the calculation of both terms within the braces, and then the selection of the smaller of the two values as the value of C_{pk}. For clarification, since Kane's work, the μ and σ used in the expressions for C_{pk} are short-term values.

Equation 2, in terms of *CPU* and *CPL*, where C_{pk} is the minimum of the two, becomes,

$$CPL = \frac{\mu - LSL}{3\sigma}, \tag{3}$$

$$CPU = \frac{USL - \mu}{3\sigma}, \tag{4}$$

$$C_{pk} = Min\{CPL, CPU\}. \tag{5}$$

The Problems with PCI Use

Many objections to the concept of PCIs began to surface shortly after the publication of Kane's work. The literature review section of this dissertation contains a detailed discussion of these objections, but one of the major objections is the underlying assumption of normality in the output of a process (Gunter, 1989a). If the output were not normal, would the users of the PCI be astute enough to recognize that the index was unusable? Would they recognize the need for further steps before the use of the capability result?

Hogg (1985) documented that in training programs for SQC, American engineers lacked the statistical knowledge required to implement these techniques, and recommended methods to remedy this situation. Ten years later, Kettenring (1995) noted that this problem still existed. Eighteen years after Kettenring published some statistical training recommendations, Romeu (2013) verified that the problem defied solution, noting that engineering curricula are already very full, demanding, and growing.

There is little room for the addition of statistical coursework to curricula in spite of the need for engineers to learn statistics to take advantage of the latest Six Sigma and lean methods. Romeu recommended that engineers independently learn more advanced statistical methods. Sleeper (2007) pointed out that even in statistics-intense training, such as Six Sigma, engineers do not learn ways of addressing more complex issues. These complex problems include the analysis of nonnormal data.

Available statistical tools could contribute to this problem. For example, Minitab is the most commonly used tool in Six Sigma statistical analysis (Brook, 2006). Until the

advent of Minitab 16 (2010), and its *Assistant* feature, Minitab output would show capabilities without indications of possible problems. This calculation occurred regardless of the underlying distribution, or lack of distribution. The user interpreted the results without help. With Minitab 16's *Assistant*, the software offers automated feedback on the validity of the results.

If an engineer is fortunate enough to take a statistics course, it may only cover basic statistics, with the distributions taught commonly being limited to the normal, *t*, *F*, binomial, and chi-squared. When faced with nonnormal data, an engineer, exposed only to these basic distributions, may not be aware of the techniques for fitting data to more appropriate distributions. Undergraduate courses generally do not cover other methods for addressing nonnormality, for example, transformations (Field, 2012).

Given the demands on an engineer's available time for learning, it may not be possible to teach the techniques needed to address the calculation and interpretation of PCIs for all but the simplest, normally distributed data (Romeu, 2013). But, it may be possible to develop tools to highlight and quantify areas of high risk. These tools could supplement the Process Failure Mode Effects Analysis (PFMEA) commonly created during the development of medical products. The goal for this study was to provide one such tool, addressing a need revealed by a review of the literature.

Problem Statement

Release of unsafe medical devices into the market threatens patient lives and wellbeing (FDA, 2013; Nagreha & Parmar, 2011). A statistical quality approach offers a partial solution to this problem (Food and Drug Administration [FDA], 2011; Global

Harmonization Task Force [GHTF], 2004; ISO, 2016). The most frequently used tool to measure the output quality of a medical device manufacturing process is the PCI (Peña-Rodríguez, 2013). PCIs, as originally proposed by Kane (1986) require normal data. However, measured data from many processes do not exhibit normality (Sleeper, 2007). The FDA can impose penalties of over $10 million for violations of the Food and Drug Act (FDA, 2011a).

No articles written about PCIs specifically focus on their use in the field of medical manufacturing. A thorough search of the available literature resulted in no articles that specifically addressed the use of PCIs in medical device manufacturing. Medical manufacturing, because of the FDA requirement for lot traceability, offers a unique structure to overcome problems of nonnormality.

Purpose of the Study

The purpose of this empirical quantitative study was to develop a framework that evaluates the ability of a PCI to accurately measure medical device test data under a scenario where output data combines the effects of mixed production lots of components. The study was comparative in nature, involving an examination of the performance of the most commonly used PCI, C_{pk}, using simulated process data by calculating precise capabilities and then comparing these values with the results generated from nonnormal data adjusted indices. Data in this study consisted of combinations of data from different distributions to represent the situation where a lot of raw material used in a process consists of material from several supplier lots. The simulated data represent test values from some production test and are the independent variables. Using simulated data

negates the influence of, or need to control for, any independent variables, because only the final value of a test resulted from the simulation. The value of the calculated PCI was the dependent variable of interest in this study. Successful completion of this study gives evidence of the applicability of PCIs to results that may come from the combination of distributions. The findings could lead to further research that will provide tools that are more accurate for practice in the future.

Research Questions and Hypotheses

The quantitative research question of this research was how accurately does the calculated value of C_{pk}, under the assumption of normality, reflect the actual probabilities of nonconformance from simulated distributions representing the mixture of components from different upstream production batches in a subsequent process? This question reflects the real-world situation in which a production line uses components from several different lots. A PCI, as used in industry, is primarily a point value (Porter & Oakland, 1991). A quality manual or protocol may state that the C_{pk} value must be greater than 1.33, 1.50, or some other value. This question leads to three formal research questions.

The first series of tests involve the comparison of calculated values of the PCIs based on samples taken from each distribution set to a required value.

Research Question 1: Do the PCIs calculated from samples of the combined distributions meet the industry standard?

$H_0 1$: $PCI_C \geq 1.33$,

$H_a 1$: $PCI_C < 1.33$.

Peña-Rodríguez (2013) suggested the value 1.33. PCI_C will be the C_{pk} calculated using sample sizes at levels of 10, 30, 59, and a value determined from each of the 12 distributions using the method described by Mathews (2010). The data in all research questions represents the test results from a test conducted after the assembly of the tested component from parts from different lots. The test addressed both raw and normally transformed data.

The second series of tests involve a comparison of the calculated positions on the *x*-axis of 4 standard deviations, equivalent to a C_{pk} of 1.33 to the value calculated from the parameters of the 12 combined distributions and their transformed values.

Research Question 2: Do the values calculated from samples taken from a combined distribution exceed the actual values required to meet the industry standard?

$H_0 2$: $x_{pci} \geq x_{pdf}$,

$H_a 2$: $x_{pci} < x_{pdf}$.

x_{pci} is the x-axis value calculated from the required value of the PCI, and x_{pdf} is the value calculated directly from the combined probability density functions. The results of this test should mirror those of Research Question 1, but shows the percentage difference between the actual and calculated values. The tests addressed both raw and transformed data. Previous tests have used the combined distributions. The third series of tests involved testing the components of the combined distributions individually and comparing the PCI calculated from these distributions with the standard C_{pk} of 1.33.

Research Question 3: Do the data values from the lateral distributions, isolated from the underlying normal distributions, meet the industry standard?

$H_0 3$: $PC_{gamma} \geq 1.33$,

$H_a 3$: $PC_{gamma} < 1.33$,

$H_0 3$: $PC_{lognormal} \geq 1.33$,

$H_a 3$: $PC_{lognormal} < 1.33$,

$H_0 3$: $PC_{Weibull} \geq 1.33$,

$H_a 3$: $PC_{Weibull} < 1.33$.

In these hypotheses, PC_{gamma} is the C_{pk} calculated from the gamma distributions, $PC_{lognormal}$ is the C_{pk} calculated from the lognormal distributions, and $PC_{Weibull}$ is the C_{pk} calculated from the Weibull distributions. These tests evaluate the suggestion that lots constructed using different lots of components require individual tests. This individual testing might compensate for the effect of a fattened tail on the value of the standard deviation used to calculate the capability index. Testing addressed both raw and transformed data.

The operationalization of the above RQ into a null and alternative hypothesis required the calculation of the PCIs from sample sizes of 10, 30, 59, and a sample size needed to achieve an a priori specified power 1 - beta. The value of 10 represents a low convenience value. A sample size of 30 represents the situation where statistics students correctly learn that this is the number at which a *t* distribution approximates a normal distribution and use it as a default sample size. The sample size of 59 achieves a 95% level of confidence and a 95% reliability level (Lipson & Sheth, 1973). The calculated sample size based on power was indeterminate until the generation of simulated results as described in the next chapter.

Tests on the variance/standard deviation and the chi-square distributions determined the outcome of the hypothesis testing. This testing examined if a Type II error was occurring in the calculation of the PCI by using the definitional equation for C_{pk} to calculate the values for the sample PCIs. The probabilities as defined by the generating cumulative distribution functions yielded the actual values of the indices. From the definitional equation for C_{pk}, it was readily apparent that the value of the index is directly proportional to the distance between the mean of the samples and the specification limit, and inversely proportional to the standard deviation of the sample. This establishes cause and effect of the data on the value of the PCI.

Theoretical Foundation

The theoretical mathematical underpinnings of this dissertation begin with the work of the famous astronomer Newcomb (1886), who hypothesized that data from combinations of several distributions can compose the observations of natural phenomena. In actuality, the data was from just one distribution. Newcomb was examining the effect of errors that occurred during the observation of astronomical phenomena. Thus, the data from an astronomical observation of a celestial body from three different observatories, or using three different instruments, might appear to come from three different normal distributions. In reality, they came from just one. Of particular concern was the identification and elimination of outliers because of their contribution to this effect.

A review of the literature suggested that this concept remained dormant until Tukey (1960) reintroduced it. Rather than examining normal distributions with different

means and standard deviations, Tukey looked at a combination of two normal distributions with the same mean and different standard deviations. The sample data drawn from the two distributions has a probability of ε that it comes from one distribution, and of $(1-\varepsilon)$ that it comes from the other.

Tukey's (1960) concept of the contaminated or mixed normal distribution often serves as the introductory idea in the literature addressing the topic of modern or robust statistics. For example, Wilcox (2012, p.2) expressed it as,

$$H(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi\left(\frac{x}{K}\right), \tag{6}$$

where K is a constant greater than zero. As Wilcox indicated, it is possible to determine the parameters of the distribution in *Equation 6*.

Using the foundation of a contaminated or mixed distribution, the next step examined the effect of mixing three distributions that are not necessarily normal and then consider the impact that such a combination may have on the calculation of C_{pk}. The expectation was that a mixture of data from different distributions may give misleading results, even if data taken individually from each distribution provided satisfactory index values. Overstatement of index values may produce a risk to patient safety; if understated, a waste of resources could result from fixing a process that is performing adequately. Such a waste of resources needlessly contributes to the increase in the cost of medical devices. A delay wastes time correcting an already capable process. This could result in a postponement in the introduction of possibly life-saving devices to medical practitioners.

The management framework for this study combined principles of scientific management, statistics applied to quality control, and systems theory; it rested on the work of three scholars. The first, Taylor (1911/1998), applied the principles of scientific management to work by measuring output and setting standards. The second, Shewhart (1931/2015), applied statistical techniques to the problem of production quality. The last, von Bertalanffy (1969), first formulated the systems approach.

To protect the public from the potentially hazardous effects of poorly manufactured medical devices, the FDA (2009) combined the measurement approach pioneered by Taylor with the statistical analysis quality techniques first advanced by Shewhart to develop a system to evaluate the output of a production activity for the manufacture of medical devices. The FDA system relies on an extension to Shewhart's earlier work by Kane (1986) and requires the calculation of a PCI, generally C_{pk}, to evaluate the output of a process. In cases of nonnormality of data, the accuracy of this index can deteriorate leading to either an overestimation or an underestimation of the quality of the process output (Gunter, 1989a, 1989b, 1989c, 1989d). Neither situation is desirable.

Nature of the Study

This study involved the use of simulated data generated from specific distributions using the R (2016) statistics application. This entailed combining the data from different distributions, calculating C_{pk} for each combination of distributions, and comparing these to the real results. It was possible to calculate accurate values of the

probabilities of nonconformance because of the known distributions used to generate the data.

The use of real-world data was impractical for this study. Such real-world data could take two forms: (a) data specifically generated for this study, and (b) existing data available from my work as an engineer in the medical device field. In either case, because of the focus on nonnormal data, the actual character of the PCIs that would describe the process is indeterminate because the nature of the underlying distribution(s) is unknown. In the case of extant data, I inquired and was told permission to use it would be impossible to obtain because of the possibility of any future litigation with the data as evidence. Using simulated data was a strength, not a shortcoming, because it allowed more precise control over the data generated for analysis.

The literature reveals a mix of both real-world and simulated data used in the study of PCI behavior. Pearn and Chen (1997) used real data from the manufacture of electrolytic capacitors. Pearn, Wu, and Wang (2005) did the same with an application toward audio speaker production. Niavarani, Noorossana, and Abbasi (2012), Ye, Ma, and Wang (2011), and several others used simulated data. In a practitioner-oriented book emphasizing curve fitting, Bothe (2001) focused on the use of real-world data. Simulation appears to be the preferred method in articles that study the performance of different indices under varying conditions. Strongly mathematically oriented research often focuses on results derived solely from the mathematics rather than testing the derived expressions on any real-world data (University of Bristol, School of Mathematics, Institute of Pure Mathematics, 2017).

Another advantage of using simulated data is the control of independent variables. In this study, the result, C_{pk}, was the dependent variable. A simulation generated data, representing test results, to calculate the value of this PCI, and the dispersion of these values reflects this effect of the independent variables. For example, consider a wire assembly that has a solder joint on one end, and a crimped connection on the other. After making a batch of cables, a sample is pull tested to ensure that it meets a minimum specification. Independent variables, in this case, might include measurement error, differing operator techniques, various settings on crimp machines, and a range of raw materials, among others. The net result of the effects of all of these independent variables is the difference in the values of the pull testing that exists between the different samples. Simulation of this cable testing would give one result, the ultimate value of the pull test.

Definitions

Terms used in this dissertation are:

Acceptable quality level (AQL): "A specified quality level for each lot such that the sampling plan will accept a stated percentage (say 95 percent) of submitted lots having this quality level" (Juran & Gryna, 1974, p. 25.5).

Industry standard: "An established standard, norm, or requirement in a particular area of business" (Industry standard, n.d.). For C_{pk} the de facto standard is 1.33 (Peña-Rodríguez, 2013).

Installation qualification (IQ): "Establishing documented evidence that process equipment and ancillary systems are capable of consistently operating within established limits and tolerances" (FDA, 1996, p. 4-2).

Operational qualification (OQ): The validation step in which "process parameters should be challenged to assure that they will result in a product that meets all defined requirements under all anticipated conditions of manufacturing, i.e., worst case testing" (GHTF, 2004, p. 10).

Process capability index (PCI): A statistical index developed to "establish the relationships between the actual process performance and the manufacturing specifications" (Pearn & Kotz, 2006, p. 7).

Process performance qualification: "Establishing documented evidence that the process is effective and reproducible" (FDA, 1996, p. 4-2). *PQ*, for Performance Qualification, is the common term for this qualification.

Product performance qualification (PPQ): "Establishing documented evidence through appropriate testing that the finished product produced by a specified process(es) meets all release requirements for functionality and safety" (FDA, 1996, p. 4-2).

Quality: "Fitness for use" (Montgomery, 2013, p. 6).

Tolerance (or specification) limits: "Set by engineering design function to define the minimum and maximum values allowable for the product to work properly" (Pearn & Kotz, 2006, p. 4). Lower limits are often abbreviated *LSL* or *LTL*, and upper limits as *USL* or *UTL*.

Validation: "Confirmation by examination and provision of objective evidence of consistent fulfillment of a particular requirement for a specific intended use can be consistently fulfilled" (FDA, 1996, p. 4-2).

Target (T): The value of the perfect output from a process as set by engineering.

Assumptions

An assumption in this study was that process output is often nonnormal in nature. My experience and the voluminous literature addressing analytical techniques for nonnormal data were the basis for this assumption. Articles regarding the calculation of PCIs, when the data may be nonnormal, are just a small subset of the literature surrounding nonnormality.

While the data from a process may be nonnormal, one assumes that the practitioner has examined the process to remove all controllable sources of variation within it. An engineer should not look at the output from a process using a PCI, find that it is not normal and move on. Nonnormal, or uncontrollable output, is a serious source of concern. Examining a process using a PCI is an iterative process, and only after executing all reasonable efforts to bring a process to normality should practitioners accept nonnormal output (Sleeper, 2007).

The assumption that normality is the desired output from a process was a basis for this study. In some cases, for example, reliability studies where the Weibull distribution is the expected output, this may not be true. In these cases, other methods for measuring process capability are appropriate (Dodson, 2006; Tobias & Trindade, 2012).

Another assumption was that C_{pk} is often used to measure process capability output in spite of nonnormality. C_{pk} is the one most frequently used PCIs because it has become the de facto standard to measure process capability (Peña-Rodríguez, 2013).

A last assumption was that the specification limits set by the engineer are appropriate. For example, engineering drawings often contain a block titled *Tolerances*

Unless Otherwise Specified (French & Vierck, 1972). If designers are relying on these *block tolerances*, they may not reflect the dimensions required to be consistent with the design requirements. Similarly, other process outputs may not have correct tolerances. In the literature reviewed for this study, Mathews (2010) was alone in pointing out the need for accurate specification limits as described above.

Scope and Delimitations

The scope of this study was the evaluation of the PCI C_{pk} when applied to data that is nonnormal because it was composed of data from two nonnormal distributions combined with data generated from a central normal distribution. The data composing the distribution should be identifiable by the source distribution representing the ability to do this data decomposition by lot identifier in a practical application of the method. Distributions used for simulated data generation are the normal, Weibull, lognormal, and gamma. The normal component is always the central distribution and the nonnormal distribution data provides the lateral distributions intended to fatten the tails of the combined distribution. The scope reflects the use of these distributions in simulation studies found in the literature. Transformation methods used are the Box-Cox, square root, inverse, inverse square root, and asinh (Rivera, Hubele, & Lawrence, 1995).

The scope of this study was only testing that generates quantitative data of the measurement of some component characteristic, for example, length or breaking force. It does not apply to a process that relies on qualitative judgments of product quality. The proposed method would not be applicable to measuring the aesthetics of surface finish

unless there is also quantitative definition. “Scratches after finishing can be no more than 0.001 inches deep,” is such a definition.

Additionally, the study scope only includes processes where components are identifiable and separated by lot or job number. Components mixed without identification before a downstream process would make the tool developed and evaluated in this study inapplicable.

The scope of this study does not include PCIs based on any other techniques than comparing the number of standard deviations from the closest specification limit. For example, it does not include methods based on yield (Grau, 2011, 2012). Because this study is mathematical in nature, it should be generalizable to any process that produces measurable output similar to that of the output of the simulation.

Limitations

This study was a quantitative statistical exercise and did not involve the use of a questionnaire, a survey instrument of any kind, or existing data. No coding of results was necessary; thus, there was no dependence on the training or judgment of coders. The numeric values of the results, rather than the opinion of the experimenter, classified them as conforming or nonconforming. The classification as conforming or nonconforming depends on the output of the definitional equation for C_{pk}. Definitional equations are inherently externally and internally valid. Definitional equations will give incorrect results when applied improperly, but the results will be consistent with the mathematical method used in the formulation of the equation.

As shown by their formulas, there are an infinite number of frequency distributions. For example, the standard normal distribution has a mean of zero and a standard deviation of one. An experimenter could not list all of the normal distributions with a standard deviation greater than zero and less than or equal to one. It would be impossible to do so because there are an infinite number of values between these two limits; mathematically, it is an uncountable set. Selecting particular parametric distributions, with specific parameter values, limited the output from this study to combinations of distributions that are *close enough* for the results to be applicable.

A possible limitation of the study arose from the decision to distribute the lateral distributions symmetrically around the mean of the underlying normal distribution. This may have been responsible for the low number of failures with the application of the Anderson Darling test to the combination distributions. This finding is a possible topic for further research and discussed further in Chapter 5.

The R random number generator in the R (2016) statistics program generated the values used in this study. From the total data generated, I drew samples using the R *sample* function. This tool eliminated any bias in sample selection.

Significance of the Study

This section contains a discussion of the significance of this study with respect to theory, practice, and social change. From a theory viewpoint, the results of this study may extend the application of PCIs specifically to the medical manufacturing field. For practitioners, this study provides another tool that engineers in regulated industries can apply in addressing nonnormal data. Finally, from a social change prospective, the

application of the findings of this study could possibly reduce the cost of medical devices, reduce time to market for some devices, and help prevent dangerous devices from reaching the market.

Significance to Theory

Looking at the early development of PCIs, it is impossible to find any significant theoretical breakthroughs from their development and use. Rather, they are a repackaging of existing statistical tools that give a more easily understood result. They are tools used to judge if a process can be depended on to produce consistent output that conforms to the desired specifications. These specifications themselves are engineering constructs of some characteristic of the part or assembly indicating the bounds within which it is suitable for use (Kane, 1986).

In the case of a finished device, for example, a pacemaker, an engineer possibly can use PCIs to evaluate if that device will furnish signals at the correct time and with the proper amplitude and frequency. In the case of a locating pin, they may indicate if the parts produced by a process will consistently fit into a mating hole. An evaluation of the risk that arises from the requirement that the output of a process must meet a particular value of C_{pk}. was a primary goal for this study. The focus was on the medical device field. Peña-Rodríguez's (2013) value guidelines are the standard of comparison.

After an initial flurry of activity over the first few decades following their introduction, the study of PCIs appears to have slowed. There is a dearth of recent literature. If the results of this study indicate excessive risk, it is possible that it could

spur the development or adaptation of more suitable measures of the reliability of process output.

Significance to Practice

When manufacturing medical products, it is essential that the completed product will do no harm to a patient. To accomplish this goal, a production line for a new product undergoes several validation steps. Validation is also required for product line relocations (FDA, 1996).

The first of these is the Installation Qualification (IQ), which establishes the ability of the equipment used in production to safely function within the operating parameters of the process. The second step is the Operational Qualification (OQ), which establishes the ability of the process to produce acceptable product under all anticipated operational conditions. The Performance Qualification (PQ) is the third step. This validation step verifies the ability of the process to produce acceptable product under normal operating conditions. It is during this validation activity that the engineer uses a PCI to evaluate production output (FDA, 1996).

A more accurate method of evaluating the output of a process will allow manufacturing, process, and quality engineers to avoid two different mistakes. From the mathematics, if a PCI gives a value that is lower than the real value, fixing a process that is already producing acceptable product may waste time. If a PCI gives a value higher than the real value, the result may be a compromise of patient safety.

Significance to Social Change

There are many reasons why the cost of medical care is increasing. One may be the incorporation of high technology into devices that can routinely perform procedures that would have been unimaginable only a few decades ago. As these devices become increasingly complex, more potential failure modes may be possible. Production lines that may exceed the devices themselves in their level of complexity manufacture these products.

The output from the production lines must meet the requirements of the next user in the supply chain (Montgomery, 2013). This condition is true whether the output is just a component for the next process step or is the finished device ready for patient use. PCIs are the tool the medical industry relies on to ensure this readiness for the next step (Peña-Rodríguez, 2013).

The application of PCIs to data unsuitable for analysis by this family of tools can have several outcomes depending on the application framework. It is possible that an accurate quantification of the process capability will result in spite of the unsuitable data. However, two other outcomes are more likely, either the PCI will indicate that the method is more capable or less capable than it is. The examination of these assertions was a goal of this study.

In the first case, the result might be a line shutdown when unusable parts move to the next process step. An alternative would be that inadvertently using nonconforming parts might cause field failures, possibly in an operating room. The results from this scenario could range from an inconvenienced medical team to a dead patient. In the

second case, the PCI would indicate that the parts from the process are not meeting specification. In this case, there may be a waste of engineering and manufacturing resources fixing a problem that does not exist. In addition, there could be a delay in potentially life-saving devices reaching practitioners while an already capable process is refined.

In either case, there is a cost to society. This cost is either in the quality of, or in the addition of more costs to, already expensive goods and services. This study has the potential to contribute to a remedy for both of these situations, and that could be a contribution to positive social change.

Summary

This chapter contains an explanation of the early development of PCIs, and their rise in importance as American industry implemented and expanded SQC methodologies. Firms implemented SQC to compete with foreign goods thought by the public as having better quality than their domestic equivalents. The chapter also includes an explanation of the widespread requirement for the use of these indices in medical device manufacturing by both domestic and international regulatory agencies.

Also described are some of the shortcomings of PCIs, for example, their reliance on normality and in-control processes, as well as the lack of sufficient statistical training on the part of most engineers to overcome these shortcomings. The chapter concludes with a description of the significance of the research question from theoretical, practical, and social change perspectives, and proposed a method to answer it.

The focus of this study came from working as an engineer in the medical device field. The specific idea arose from the difficulties encountered using capability indices to validate processes to meet both customer and regulatory agency requirements. The next chapter contains a discussion of the literature reviewed in an attempt to find a solution to this problem.

Chapter 2: Literature Review

In spite of the best efforts of the FDA, ISO, and GHTF to guide the production of medical devices, large numbers of FDA Class I recalls still occur. Class I recalls occur when a shortcoming represents a serious threat to the wellbeing or even the life of a patient. One hundred and nineteen Class I recalls in the period ranging from 2008 to mid-2011 were detailed by Nagreha and Parmar (2011). The FDA (2013) reported 307 Class I recalls during the years 2003 to 2013. These numbers indicate there is some failure occurring that exposes many patients to such serious risk.

The chapter begins with a review of literature relevant to the development and use of PCIs, their limitations, required data adjustments needed before their application, and examples of their use in both real-world and simulated data. Next, is a discussion of the theoretical foundations of PCIs, as they evolved from the application of statistical methodology to monitor the state of a production process in the next section. A person examining the origins of PCIs may link them to SQC and the design of experiments.

The following section includes a review of the literature to determine the ability of the most significant indices to provide a consistent level of process control. The focus was on those occasions when the input to the process may vary. The intent of the regulation of medical devices is to ensure patient treatment with a device that works consistently and safely. The question was: can PCIs as currently used contribute to this goal?

Literature Search Strategy

The use of PCIs is relatively recent. A definition of C_p appeared in the third edition of *Juran's Quality Control Handbook* in 1974. It was the publication of Kane's (1986) article that aroused the interest of researchers. Not many books were found for this review that specifically deal with process capability as opposed to books that discussed process capability as part of a broader quality perspective or as chapters in quality handbooks.

Of the books found, three—Bothe (2001) and Wheeler (2000a, 2000b)—were practitioner-oriented. The focus of these books was on practical applications of indices to real-world situations. Their subject was existing indices, rather than an extension of basic applications of statistical methods to process capability studies through the development of new indices. Wheeler's books contain no reference section, in contrast to Bothe's book, which does.

Kotz and Johnson (1993), Kotz and Lovelace (1998), and Pearn and Kotz (2006) wrote books that might be of more interest to researchers. The authors of these books traced the mathematical evolution of many PCIs, from the most general to very specialized indices, and detailed their strengths and weaknesses. These books contained extensive bibliographies.

A book that falls between these two categories was Sleeper's (2007) book. The purpose of this book was to acquaint the advanced practitioner with some less well-known frequency distributions. These distributions might arise in the application of Six

Sigma methodology to real-world problems. This application of Six Sigma methodology included the use of PCIs in non-normal data situations.

Kotz et al. (2002), Spiring, Leung, Cheng, and Yeung (2003), and Yum and Kim (2011) wrote three exhaustive bibliographies of the literature of PCIs. The authors list of Kotz et al. (2002) contains the names of many of the most prominent researchers in the field. The bibliographies formed the starting point for this literature review providing for the identification of articles relevant to the research topic, input into Google Scholar's *Cited by* feature to locate more recent articles published since the bibliographies. Also, many publishers' journal repositories also offer a *Cited by* feature.

Process capability has not been an active area of doctoral level research. Kotz et al. (2002) identified only two doctoral dissertations. Spiring et al. (2003) found six. My search efforts included ProQuest Dissertations & Theses Global to find more dissertations. Other databases searched for articles were ABI/Inform Complete, Academic Search Complete, Business Source Complete, Current Index to Statistics, Emerald Management, ERIC, MEDLINE, SAGE, and Science Direct.

Because this is a relatively new area of research, the search input did not include a date range for the search. Search terms included *ab(process capability)*, *(ab(process capability) AND ab((index OR indices)) AND ab((nonnormal OR non-normal)))*, *(ab(process capability) AND ab((index OR indices)) AND ab((nonnormal OR non-normal))) AND adv(Kotz, Samuel)*, *(ab(process capability) AND ab((index OR indices)) AND ab((nonnormal OR non-normal))) AND adv("Johnson, Norman L")*, *ab(process capability) AND ab((index OR indices)) AND ab((contaminated OR nonnormal)) AND*

ab((contaminated OR non-normal)), and *ab(process capability) AND ab((index OR indices)) AND ab((nonnormal OR non-normal))*

Of these, *ab(process capability) AND ab((index OR indices)) AND ab((nonnormal OR non-normal))*, gave the best results for dissertations, returning six. The search yielded 13 dissertations from all sources for review.

The structure of the publication of the articles reviewed appears to follow a cluster rather than a tree structure. In other research areas with tree structures, Article *A* appears, and Article *B* will come next, building on Article *A*. *B* will generate *C*; *C* will inspire *D*, and so on. In the process capability literature, two articles, in particular, Kane (1986) and Pearn, Kotz, and Johnson (1992) appeared to be at the center of the cluster. These two articles generated articles developing a particular application extension, but this extension will not be very long. Observation of this pattern during this review showed that new articles will go back to the center of the cluster and start anew. Depending on the area of research, authors may write new articles adding to the group. Because of the limited evolutionary nature of this structure, it appears that early articles are much more relevant to current thinking than might be the case in other areas of research.

Theoretical Foundation

PCIs are evaluation tools based on the application of statistical methodology forming the principles of SQC. SQC is the application of specialized statistical methods to evaluate product quality. Results of the application of SQC guide the adjustment of the process to achieve quality standards. Continual monitoring can control that process to

ensure that production of quality products will continue in the future. Shewhart's (1931/2015, 1939/1986) research provided the foundation and Deming (1950/1966), Juran (1988), and Ott (2000) built on this. The Six Sigma approach, as developed by General Electric, includes several of the tools defined by these authors and many new ones (Pyzdek, 2003).

In applications of SQC, two questions immediately arise. First, is the process capable (Grant & Leavenworth, 1980)? Answering this question requires two very different measures. The designer of the entity defines the first, the specification limit. For example, if a locating pin is to slide into a hole in a part, then the tolerances, as expressed by the specification limits, on both diameters must allow the parts to mate. The more precise the fit required, the tighter the tolerances on both parts (French & Vierck, 1972).

Consider the production of a locating pin. There will be a *USL* specifying the maximum diameter, and an *LSL* defining the minimum diameter. This difference between the two is the tolerance spread, expressed mathematically as *USL* - *LSL*. A process produces the pin. This manufacturing process will produce parts with some variation around the target diameter. The range of the variation around the desired dimension is the natural tolerance of the process. In a capable process, the natural tolerance, expressed as the distance between three σ above and three σ below the mean value of the parts produced by the process, will fit within the tolerance spread. The mean of the process merely serves as a point defining the center of a six-standard deviation interval (Grant & Leavenworth, 1980).

For example, the target diameter of the pin may be 0.250 inches. The process may produce pins with a mean of 0.250 inches and a standard deviation of 0.007 inches. Then the natural tolerance of the process would stretch from {[0.250 + 3(0.007)] - [0.250 - 3(0.007)]} inches = 0.042 inches, or 6σ.

Capability does not imply control. Capable refers to the process spread while control refers to process location. If the target diameter of the pin was 0.3125 ± 0.025 inches [(0.042 < 0.050 = 2(0.025)], the process would still be capable. The process is not in control because it is producing pins with a diameter of 0.250. The specification is 0.3125. The natural spread of the process is tight enough to fit within the specifications. For a process to be in control, the natural spread around the mean must fit within the tolerance spread centered approximately on the target dimension. Good process design allows the production of acceptable parts within the specification limits. Engineering specifies these limits on a process whose outputs are going to fall in a natural spread. This spread happens independently of the engineer's desired outcome (Montgomery, 2013). This discussion does assume a normal distribution around the process mean.

The difference between capability and control may be subtle but is important. Figure 2 is an illustration of the various cases possible and is typical of those found in SQC textbooks to illustrate these concepts, for example, Grant and Leavenworth (1980). PCIs are the tools used by practitioners to determine how consistently a process can produce quality parts because they can look at both capability and control.

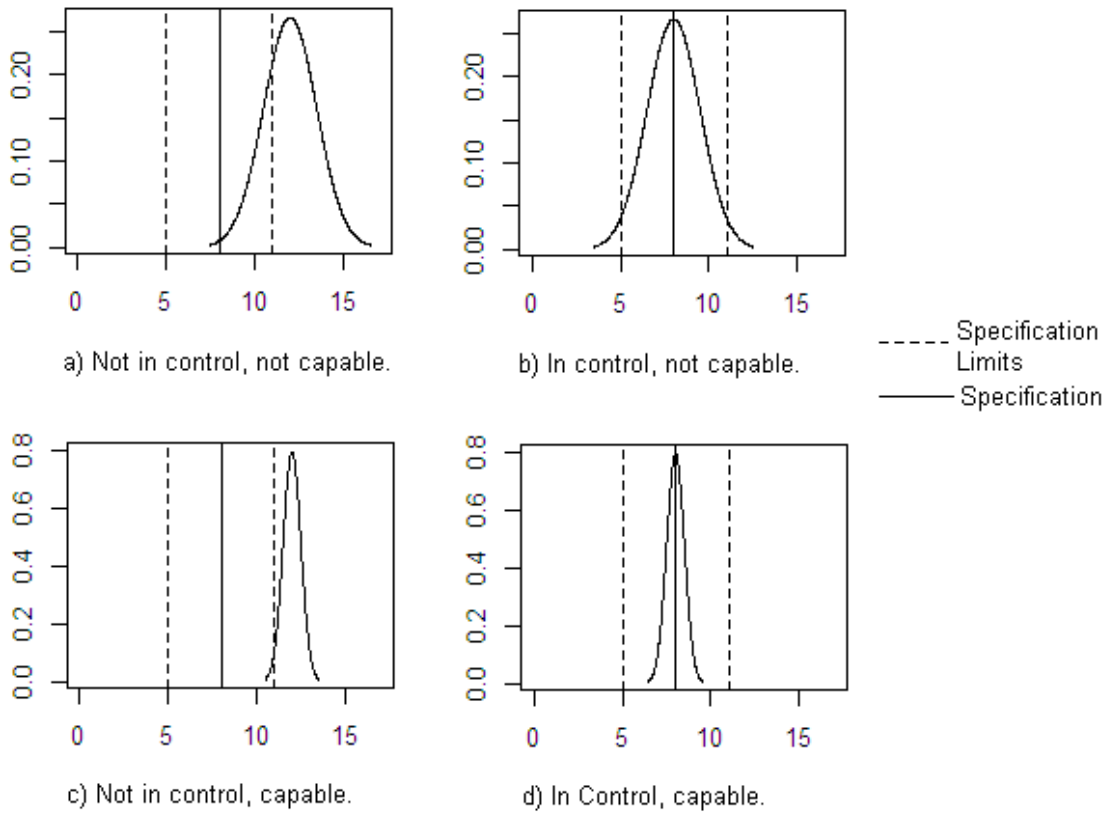

Figure 2. Examples of control and capability for different processes.

Panel *a* is an illustration of a process that is not capable and not in control. The natural spread would not fit within the tolerance limits. The centering of the process moved the spread of the process further outside these limits.

In panel *b*, the mean of the normal process shown is eight, and the standard deviation is 1.5. The *USL* is 11, and the *LSL* is five. Consequently, four standard deviations lie between the specification limits. This range means that 9.121% of the parts will lie to the right of the USL, and another 9.121% will lie to the left of the *LSL*.

If the data in panel *b* represents process output, it would have a C_p of 0.667. Because the process mean is at the midpoint between the *USL* and *LSL*, *CPU* and *CPL* are both equal to 0.667, which is the same as the value of C_p. Because *CPU* and *CPL* are

equal, C_{pk} is also equal to 0.667. If the mean of this process was 8.5, and everything else remained the same, C_p would remain at 0.667 because it is not sensitive to location. However, *CPU* would decrease to 0.556 while *CPL* would increase to 0.778. Because C_{pk} is the minimum of *CPU* and *CPL*, it would drop from 0.667 to 0.556. The proportion of parts below the LSL would drop to 5.99% while the percentage of parts above the USL would increase to 13.33%.

Panel *c* is data from a process that is capable. The natural spread of the process is less than the tolerance spread. The location of the process mean moves the output of the process outside of the tolerance spread.

In Panel *d*, eight is still the process mean, but the standard deviation is 0.500. In this case, the C_p and C_{pk} would be two, and a tiny percentage of parts, less than 0.003%, would lie outside the specification limits. With a shift in the mean to 8.5, C_p would remain at two, *CPU* would decrease to 1.667, and *CPL* would increase to 2.333. This increase would result in a C_{pk} of 1.667. Parts to the left of the *LSL* would decrease to 0.0002%, and parts to the right of the *USL* would increase to 0.043%. This example and the previous one, show that PCIs use the same statistical theory as SQC.

Another theoretical foundation of PCIs is the pioneering work of Fisher in applying an experimental approach to the study of crop yields (Fisher, 1935/1971). Fisher's daughter (Box, 1980, 1987) wrote descriptions of Fisher and Gosset's early work in experimental design. The first of these articles is mathematical, the second is anecdotal.

Fisher's (1935/1971) work evolved into a methodology known as the design of experiments (DOE). Box and others (Box & Draper, 1969; Box, Hunter, & Hunter, 1978) contributed to the method. Taguchi (1987) further expanded the methodology by factoring the cost of quality into the DOE methodology. New PCIs incorporated Taguchi's concepts of the cost of poor quality (Boyles, 1991).

DOE has evolved into a complex area of study incorporating a wide variety of methods to optimize performance (Lawson, 2014; Montgomery, 2001). In simple terms, DOE is a method for structuring experiments, varying parameters, and linking the changes in the output of the experiment to the changes in the parameters. The FDA has issued a nonbinding recommendation for DOE use in the determination of process parameters (FDA, 2011b; Pluta, 2014).

Literature Review

Regulation of Medical Device Manufacturing

Essential to best practices in the manufacture of medical devices is adherence to multiple regulations from different agencies. These regulations are cited here as literature because they provide the basis of the need for the research conducted in this study. Violation of these regulations can result in fines and/or imprisonment of those responsible for the nonconformance (FDA, 2014).

Advances in medical technology have conquered once devastating diseases and increased life expectancy in much of the world. However, even in cases where a complex device functions properly, the failure of a minor subsystem can have potentially disastrous consequences. Even the failure of a system unrelated to the functioning of the

device, the sealing of the lid to the tray containing the assembly, can expose patients to dangerous contamination (Mays, 2008).

Because of the importance of patient safety, medical device manufacturing is a very highly regulated industry in the United States and abroad. In the United States, the primary regulatory agency is the U.S. Department of Health and Human Services, Food and Drug Administration (FDA). Other countries have similar agencies, and the International Organization for Standardization (ISO, 2016) has issued a standard, ISO 13485, to set requirements for device manufacturing. The Global Harmonization Task Force (GHTF) had as its mission creating uniform device regulation worldwide. The International Medical Device Regulators Forum (IMDRF) replaced the GHTF, and issued guidance documents to accomplish the original task. The activities of these agencies also include the procedures for the approval of new medical devices. Relevant to this study are the manufacturing validation procedures that must occur, and the statistical tools used during these activities. The intent of these analytical tools is to ensure proper and reliable manufacture of devices.

The FDA provides no specific guidelines on the procedures to follow to ensure a process is under control. The FDA offers general guidance in Quality System (QS) Regulation/Medical Device Good Manufacturing Practices, 21CFR820.75 (2013). This guidance indicates that a process must be under control under specific circumstances,

Subpart G--Production and Process Controls

Sec. 820.75 Process validation.

(a) Where the results of a process cannot be fully verified by subsequent inspection and test, the process shall be validated with a high degree of assurance and approved according to established procedures. The validation activities and results, including the date and signature of the individual(s) approving the validation, and, where appropriate, the major equipment validated, shall be documented.

(b) Each manufacturer shall establish and maintain procedures for monitoring and control of process parameters for validated processes to ensure that the specified requirements continue to be met. (p. 151-2)

ISO 13485, Medical devices -- Quality management systems -- Requirements for regulatory purposes (2016), offers similar guidance,

7.5.6 Validation of processes for production and service provision

The organization shall validate any processes for production and service provision where the resulting output cannot be or is not verified by subsequent monitoring or measurement and, as a consequence, deficiencies become apparent only after the product is in use or the service has been delivered.

Validation shall demonstrate the ability of these processes to achieve planned results consistently. (p. 19)

In its final report, the recommendations of the GHTF (2003) are similar,

Each process should have a specification describing both the process parameters and the output desired. The manufacturer should consider whether the output could be verified by subsequent monitoring or measurement (**A**). If the answer is

> positive, then the consideration should be made as to whether or not verification alone is sufficient to eliminate unacceptable risk and is a cost-effective solution (**B**). If yes, the output should be verified, and the process should be appropriately controlled (**C**).
>
> If the output of the process is not verifiable then the decision should be to validate the process (**D**); alternatively, it may become apparent that the product or process should be redesigned to reduce variation and improve the product or process (**E**). In addition, a change in a manufacturing process may result in the need for process validation even though the process formerly only required verification and control. (p.7)

PCIs have become the standard to verify process control. Specifically writing about FDA regulations, Peña-Rodríguez (2013) indicated that a C_{pk} of 1.33 evidences an in-control process. An examination of available medical device manufacturer quality manuals verified this finding. However, several manufacturers suggest that a value of 1.33 is only a starting point and that a mature process should have an even higher value.

Process Capability Indices

The first indices, C_p and C_{pk}. As described in the Introduction, the first reference to capability analysis was in the third edition of Juran's Quality Handbook published in 1974 (Kotz et al., 2002). In the fourth edition, Gryna (1988) discussed the capability index that defines this index as the tolerance spread divided by 6σ as described earlier. More importantly, Juran and Gryna introduced into their handbook the capability

concepts developed by Kane (1986). Kane's article is seminal because it laid the foundation for all further developments and extensions of PCIs.

Kane (1986) pointed out that one of the major problems with C_p is that while it measures capability, it gives no indication at all of if a process is in control. Sullivan (1984) had reached the same conclusion. A C_p of 1.33 would indicate that the tolerance spread is $1.33(6\sigma) = 8\sigma$ wide. For simplicity, set the desired output at 25 units of some measure, hundredths of an inch, millimeters, ounces, etc. Assuming a σ of one, an *LSL* of 21 and a *USL* of 29 would result in a C_p of 1.33 because the difference between the limits is eight. However, a process with a mean of 50 units and a standard deviation of one would also have the same C_p.

In his development of C_{pk}, *Equation 2*, Kane (1986) used the mean, μ, to locate the output of the process. He also divided the tolerance spread into two intervals, μ - *LSL* and *USL* - μ. The two intervals reduced the denominator of the expression by a half. It now equaled 3σ instead of the 6σ of the natural tolerance. This division then defined two new indices, $CPL = (\mu - LSL)/3\sigma$, and $CPU = (USL - \mu)/3\sigma$. C_{pk} is the minimum of these two indices. For C_{pk} to take on a positive value, the value of σ must be between the specification limits.

Determine the midpoint of the specification range by adding the *LSL* and the *USL* and dividing by two (Kane, 1986, pp. 45-46),

$$M = \frac{(USL + LSL)}{2}. \tag{7}$$

Kane used an *m* to indicate the value in *Equation 3*. Authors writing since then have followed the convention of using *M* for this value. This use of *M* is a convention followed throughout this dissertation.

Using these results, Kane defined a factor, *k*, used to relate C_p to C_{pk} (p. 46),

$$k = \frac{|M - \mu|}{\frac{USL - LSL}{2}}, \tag{8}$$

$$C_{pk} = C_p(1 - k). \tag{9}$$

Kane stated this without proof. Bothe (2001) does provide a proof of the relationship.

Kane's (1986) other contribution is the identification and use of a target value, *T*, in evaluating a process. This *T* value is the desired value of the outcome of the process. If the specification calls for a value of 0.250 inches, *T* = 0.250 inches. This use is in contrast to the previous use of μ, the mean of the process output regardless of how close it is to the desired value. An implicit assumption Kane made in his derivation is the centering of *T* between the *LSL* and the *USL*, although he does allow for cases where this is not true. Other authors, discussed later, examined in more detail situations where this assumption did not hold. Using the target value, *T*, Kane offered an alternative formulation of *k* (p. 46),

$$k = \frac{\mu - T}{Min\{T - LSL, USL - T\}} \tag{10}$$

Through substitution in *Equation 5*, Kane (p. 48) showed,

$$C_{pk} = Min\{CPL, CPU\}, \tag{11}$$

where,

$$CPL = \frac{T - LSL}{3\sigma}\left\{1 - \frac{|T - \mu|}{T - LSL}\right\} \quad (12)$$

$$CPU = \frac{USL - T}{3\sigma}\left\{1 - \frac{|T - \mu|}{USL - T}\right\} \quad (13)$$

Kane (1986) offered several cautions in the use of the indices he documented. First, there may be pressure to apply the indices too soon before the process is under control. His equations are based on a mature, in-control process as indicated by the use of μ and σ rather than $\overline{X}$ and s. Second, he acknowledged that sampling often yields the values of these parameters and that the size of the samples could affect their values and, consequently, the values of the indices. Third, he was concerned about the difficulty of calculating indices; the later availability of cheap computing power remedied this concern. Fourth, he expressed reservations about the normality of the processes, and the effects nonnormality would have on the accuracy of the indices. Fifth, he recognized that manufacturing methods could change, even over the short term. The specific example he used addressed how tool wear might affect a process and cause changes to the values of the indices over time.

Kane (1986) clearly recognized that PCIs had some serious shortcomings. Other researchers identified this as well. Gunter (1989a, 1989b, 1989c, 1989d) wrote a four-part series criticizing C_{pk} in *Quality Progress*. His concerns about the use of the index reflected some of those expressed by Kane (1986).

Kane recognized that the index was only suitable for processes that fit a normal distribution. Gunter reinforced this concept and elaborated by noting that relying on the

central limit theorem, a frequently applied tool in SQC, does not present an acceptable solution. Instead, he suggested that data transformations, if they apply to the data under analysis, might offer a solution, as might robust methods. Kane expressed concern about the effect of different sample sizes on the calculated values of the indices. Gunter also cited sampling as a possible source of inaccuracy, but his reservation focuses on sampling error rather than on sample size. Gunter, also, like Kane, recognized the use of this index in calculating capabilities for processes that were not yet under control to be a major source of error.

Interestingly, by the time of Gunter's article, others were beginning to suggest alternatives to C_{pk} to overcome these difficulties. For his criticism of the use of the central limit theorem, Gunter referred his readers to another article in which Chan, Cheng, and Spiring (1988a) mathematically showed the inapplicability of this theory to the calculation of C_{pk}. However, Gunter did not discuss the alternative the latter article suggested, C_{pm}.

The incorporation of a target value, C_{pm}. Kane (1986) did introduce a PCI expression incorporating a target value, *T*. Taguchi (as cited in Boyles, 1991) was the first to use the measured deviation from *T* as a penalty factor in the calculation of an index. According to Boyles, Taguchi focused on the cost of poor quality, either as a financial impact on the firm or to society in general. Thus, Taguchi's formulation incorporated a cost factor. Adding the cost factor differed from the other authors reviewed here who used PCIs as methods to monitor the performance of a process without examining any cost considerations. Taguchi introduced his work at the American

Statistical Association Annual Meeting in 1985. The first journal article incorporating *T* reported the research of Chan et al. (1988a). Their research was independent of Taguchi's earlier research (Boyles, 1991).

Chan et al. (1988a) developed a new process indicator that reflected the effect of the process mean missing the target value. It also incorporated the natural variation around the process mean. The focus was on expanding the ability of C_p to include a penalty of the value of the indicator based on the distance of the process mean from the target. They did this by defining a new index, C_{pm}, defined as (p. 164),

$$\hat{C}_{pm} = \frac{USL - LSL}{6\sigma'}, \tag{14}$$

where the definition of σ' is (p.164),

$$\sigma' = \sqrt{E(X-T)^2}\ . \tag{15}$$

σ' is estimated with *Equation 16*,

$$\hat{\sigma}' = \sqrt{\sum_{i=1}^{n} \frac{(x_i - T)^2}{n-1}}\ . \tag{16}$$

Referring to *Equation (*1), we would estimate C_p as (Kane, 1986, p. 42),

$$\hat{C}_p = \frac{USL - LSL}{6s}, \tag{17}$$

where *s* is the sample standard deviation,

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}\ . \tag{18}$$

Comparing *Equations 14* and *16*, it is apparent that the difference between the two is that the denominator of $\hat{C}_p$ increases with the sample standard deviation resulting in a lower value. Conversely, a smaller standard deviation will result in a higher value of $\hat{C}_p$. The value of $\hat{C}_p$ depends only on the specification limits, set by engineering, and the value of the standard deviation, a process characteristic. It is thus independent of the desired output value from the process, that is, *T*.

$\hat{C}_{pm}$, on the other hand, is calculated by using the average distance from *T*. The closer the distribution of the output is to *T*, the smaller the denominator becomes, and the larger $\hat{C}_{pm}$ becomes. The greater the spread of the output is with respect to *T* becomes, the smaller the resultant value of $\hat{C}_{pm}$. If the mean of the process is equal to *T*, or at least very close when compared to the tolerance interval, the C_{pm} is equal to C_p. Chan et al. (1988a) showed that (p. 164),

$$C_{pm} = \frac{[USL - LSL]}{6\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{C_p}{\sqrt{1 + \frac{(\mu - T)^2}{\sigma^2}}}. \tag{19}$$

They also showed that the bias of $\hat{C}_{pm}$ as an estimator of C_{pm} asymptotically approaches zero as the sample size increases and does so more rapidly than $\hat{C}_p$ does for C_p. They also pointed out that while C_p does not change as the mean of the process changes, C_{pm} does and does so in a manner similar to C_{pk}. Boyles (1991) indicated that this is true only by meeting their assumption that $\mu = T$.

Boyles (1991) explored several aspects of PCIs involving a targeted process value, T. Consider C_p and the case where $M = \mu$ with a normal distribution assumed. In this instance, a C_p of one would place the center of the distribution midway between the specification limits. Calculate a Z-score for a specification limit by taking the distance between μ and the specification limit and dividing the result by the σ. $\Phi(Z)$ represents the cumulative distribution for this value. Applying this to the calculation of C_{pk}, Boyles expressed the percentage conforming as (p.18),

$$\%Yield = 100\left[\Phi\left(\frac{USL-\mu}{\sigma}\right)-\Phi\left(\frac{LSL-\mu}{\sigma}\right)\right]. \tag{20}$$

He also showed that this establishes upper and lower bounds on the yield of a process with C_{pk} values calculated for both the upper and lower specification limits. He concluded that C_p provides an estimate of the yield the process could achieve, while C_{pk} indicates the actual yield limits of the process for particular values of this PCI.

Boyles (1991) noted interesting behavior for both C_{pk} and C_{pm} using a plot of the value of these indices as of function of μ and σ. For any constant σ, C_{pk} will reach its maximum when $\mu = M$, *Equation* **Error! Reference source not found.**. At this point, $C_{pk} = C_p$ while at or beyond both specification limits, $C_{pk} = 0$. At any constant value of μ,

$$\lim_{\sigma \to 0} C_{pk} = \infty \tag{21}$$

Because C_{pk} can increase without bound in this situation, Boyles suggested that it does not serve well as an indicator of the centering of the process or the distance between the mean of the process and T.

Boyles (1991) considered C_{pm} to be a better indicator of process centering. He analyzed C_{pm} similarly to the analysis of C_{pk}. Consider C_{pm} as a function of μ and σ. For a constant value of σ, C_{pm} will reach its maximum value when $\mu = T = M$. C_{pk} equals zero at or beyond the specification limits. C_{pm} will approach zero as the distance between μ and T increases,

$$\lim_{|\mu - T| \to \infty} C_{pm} = 0 . \tag{22}$$

Boyles (1991) indicated that there is an upper bound to C_{pm} as σ approaches zero (p. 20),

$$C_{pm} < \frac{USL - LSL}{6|\mu - T|} . \tag{23}$$

However, this is only true in cases where μ is not equal to T. C_p would be a line parallel to the $\mu - C_{pk}$ or $\mu - C_{pm}$ planes. This line intersects the vertex of the angle of the plot of C_{pk} or tangent to the circles formed by the plot of C_{pm}. Cutting planes construction parallel to the $\mu - \sigma$ plane, results in those found in Boyles (1991).

Refining the target value, C_{pmk}. Pearn et al. (1992) found inconsistencies in both the prior work of both Kane (1986) and Chan et al. (1988a). These discrepancies concerned the percentage of non-conformance versus the value of the PCI. They considered cases where T lies within the specification limits but is not equal to M. Half the distance between the specification limits is (p.217),

$$d = \frac{USL - LSL}{2} \tag{24}$$

If $d = 6\sigma$, and $T = 3[(USL) + LSL)]/4$ (p. 218), then if $\mu = T - d/2 = M$ (p. 218) and $\mu = T - d/2 = USL$ (p. 218), it can be shown that the values of C_{pm} are equal to $2/\sqrt{13} = 0.555$ for both values of μ. This is due to the $(\mu - T)^2$ term in the denominator of the expression for C_{pm}. While the values of C_{pm} are equal, the percentages non-conforming are 0.27% and 0.50% respectively as shown in Figure 3.

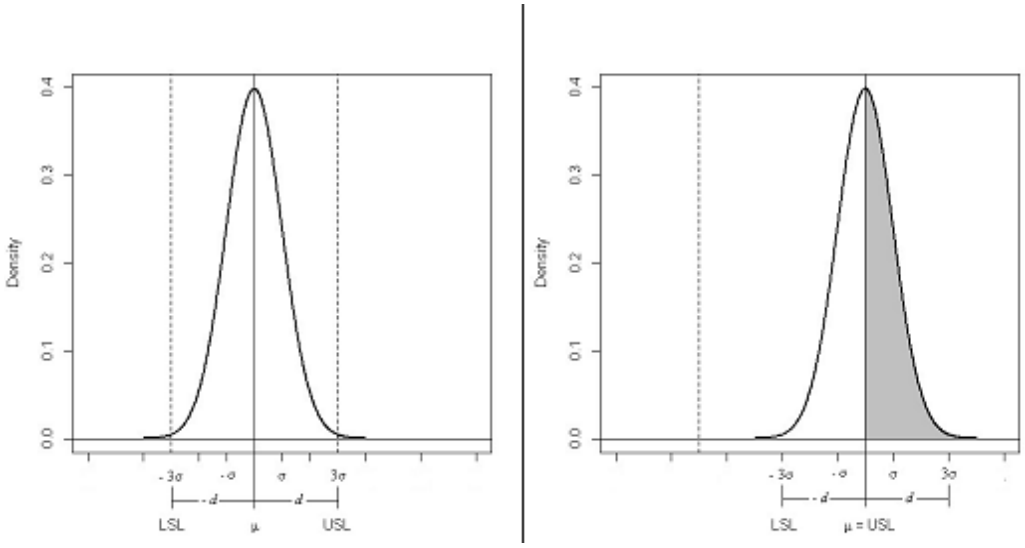

Figure 3. Process with identical C_{pm} values, but differing percentages conforming.

To overcome these problems, while restricting their proposal to the case where $T = M$, *Equation 7*, they developed a new index, C_{pmk}. Defining σ' as, $\sigma' = \sigma^2 + (\mu - T)^2$ (p. 217),

$$C_{pmk} = \frac{\min(USL - \mu, \mu - LSL)}{3\sigma'} = \frac{C_{pk}}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}, \tag{25}$$

with an estimator of (p. 221),

$$\hat{C}_{pmk} = \frac{d - \left|\overline{X} - M\right|}{3\sqrt{\frac{1}{n}\left\{\sum_{i=1}^{n}\left(X_i - \overline{X}\right) + n\left(\overline{X} - T\right)^2\right\}}}, \tag{26}$$

where (p. 217),

$$d = \frac{USL - LSL}{2} \tag{27}$$

As is the case with C_{pm}, deviations from T induce a penalty in the form of a larger value of the index. The authors indicated that C_{pmk} is the most sensitive of the four indices to values that deviate from the target value, T, followed by C_{pm}, C_{pk}, and C_p. They further identified C_p as the first-generation index, C_{pk} and C_{pm} as the second generation, and C_{pmk} as belonging to a third generation.

A unifying index, $C_p(u, v)$. Writing after Boyles (1991), and Pearn et al. (1992), Vännman (1995) developed a unified approach to PCIs. To avoid the problems highlighted by Pearn et al. (1992), he assumed that $T = M$ and normality (p. 807),

$$C_p(u,v) = \frac{d - u\left|\mu - M\right|}{3\sqrt{\sigma^2 + v(\mu - T)^2}} \tag{28}$$

This equation can generate all of the capability indices considered so far,

$$C_p(0,0) = C_p; \qquad C_p(1,0) = C_{pk}; \quad C_p(0,1) = C_{pm}; \qquad C_p(1,1) = C_{pmk}, \tag{29}$$

by using different values of u and v.

An advantage of Vännman's (1995) approach is that the possible adjustment of the values of u and v to increase or decrease the sensitivity of the index. This adjustment is to the distance between M and T. Vännman pointed out that this is especially important

when dealing with a small value of σ. Use of Vännman's expression of capability indices in terms of u and v has become very common in the literature.

A fourth generation index, C_{psk}. A year before Vännman's (1995) article appeared, Benson (1994) completed a doctoral dissertation that proposed a fourth generation index. What makes this dissertation noteworthy is that Samuel Kotz, who, along with Pearn and Johnson, made major contributions to the development and exploration of indices, was Benson's co-chair. In fact, in the index Benson proposed, C_{psk}, sk stands for Samuel Kotz.

Benson's (1994) index was similar to that of Vännman's with the addition of another parameter, w, (p. 44),

$$C_p(u,v,w) = \frac{d - w|\mu - T| - u|\mu - M|}{3\sqrt{\sigma^2 + v(\mu - T)^2}}. \tag{30}$$

Benson did refer to previous work by Vännman discussing the index. Benson showed that the inclusion of the w parameter allows for the case $M \neq T$. Under assumed normality, the addition of the w parameter extracted an additional penalty. This penalty applied to the differences between the mean and the target, T.

Unlike other indices, Benson's did not assume that the target lies at the midpoint of the specification range. Like Vännman's (1995) index, $C_p(u, v)$, appropriate values of u, v, and w will yield the other indices.

Basic indices conclusions. This section has examined the basic capability indices. While their evolution has made them more sensitive to process irregularities, one thing remained constant: An assumption of normality is the basis for these indices. With a

failure of that assumption, the results of a capability analysis may be misleading at best and dangerous at worst. In the medical field, relying on incorrect results can endanger patient lives and expose a manufacturer to serious legal and financial liability.

After the development of the basic indices, Kotz and Lovelace (1998) found "The avalanche" (p. 95) of indices began. The reference is to indices developed to cope with the shortcomings of the basic versions. New indices continue to be developed (Lupo, 2015). The next section will address methods to compensate for the lack of normality in the use or development of PCIs including addressing nonnormal data in general.

Overcoming Nonnormality

Fortunately, problems with nonnormal data are not restricted to the examination of the output from manufacturing processes. Instead, they are attracting considerable attention from statisticians and researchers in other fields who have developed methods for addressing this problem. The methods developed in other fields can also apply to PCI calculations. While the study of process capability is relatively recent, the study of nonnormality is not. Pearson (1894, 1901, 1916) opened this subject to review. Prior to Pearson's research, scholars held that all probability distributions were normal but with differing amounts of skew (Department of Statistics, University of Minnesota, Morris, n.d.). Pearson developed descriptions for several different types of frequency distributions, including the normal distribution, according to their skew and kurtosis. Statistician, including those studying PCIs, widely use Pearson's method and its variants.

Transformations, including percentile methods. One of the most widely used methods for addressing nonnormality is that of transforming the data. If the application of

a transform achieves normality, then a practitioner can apply a capability index to the transformed data and specification limits to assess the capability. Two methods, the Johnson, and the Box-Cox transformations, are very widely used. Many statistical packages, for example, Minitab (2010), incorporate both of these methods.

The first method discussed, developed by Johnson (1949), is similar to the approach taken by Pearson (1894, 1901, 1916) in that he also used the moments to develop his frequency curves. He provided a method to translate curves so that they will coincide with Pearson curves, ideally that representing the normal distribution. Although the transformation is simple in appearance, (Johnson, 1949, p. 152), it is rather complex,

$$z = \gamma + f\left(\frac{x-\xi}{\lambda}\right). \tag{31}$$

Like Pearson, the skew and kurtosis are calculated. Johnson (1949) then provided a lookup table for these values, depending on the type of curve under analysis, allowing the determination of the other parameters, and the data transformation. In total, he developed three transformations S_B, S_L, and S_U. Slifker and Shapiro (1980) provided a more detailed explanation of the procedure and its application. A frequently cited article using this method is Pyzdek's (1992).

The other common transformation is the Box-Cox (Box & Cox, 1964, p. 214).

$$y^{(\lambda)} = \begin{cases} \dfrac{(y+\lambda_2)-1}{\lambda_1} & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases} \tag{32}$$

With different values of λ, the transform takes on different characteristics. For example, a value of 0.50 will result in a square root transform, and a value of -0.50 is the reciprocal square root transform. Wu, Lin, Yang, and Pearn (2014) reported a recent application of the Box-Cox transformation to C_{pk} calculations.

The need to address general problems of nonnormality in statistical analysis led to the development of these methods. Along with these, other types of simple transforms exist. For example, a trigonometric function can be applied to the data and the specification limits. It is important to note that making any adjustments, for example to the specification limits, require back transformation before use. Kabacoff (2015) expressed caution about justifying transformations before applying them. Other methods, discussed later, address the problem of nonnormality as it applies to quantifying process capability.

Applications based on Pearson probability distributions. Clements (1989) developed a method based on Pearson's (1894, 1901, 1916) system using the calculations for this method done by Gruska, Mirkhani, and Lamberson (1979). Clements's method first required the calculation of the mean, $\overline{X}$, the sample standard deviation, s, the skewness, S_k, and the kurtosis, K_u, for the data. He used these values to create standardized values. Users apply the values by looking them up in the appropriate tables published in his article, adapted from the tables found in Gruska et al (1979).

The values from the tables correspond to the 0.00135 and .99865 percentiles that are the values for 3σ in either direction from the mean of a normal distribution. Clements designated these as L_p' and U_p' respectively. Taken individually, they corresponded to the

values needed to calculate *CPU* and *CPL*; when combined, they gave the 6σ value required to calculate C_p.

Next, one looked up the value of the median in another provided table. M' designates this value after adjusting the sign for either positive or negative values of skewness. The values of L_p, U_p and M are calculated from these values (Clements, 1989, p. 97).

$$L_p = \overline{X} - sL_p', \tag{33}$$

$$U_p = \overline{X} + sU_p', \tag{34}$$

$$M = \overline{X} + sM'. \tag{35}$$

These represented the values of the percentiles and the estimated median of the distribution. Given upper and lower tolerance (*UTL* or *LTL*) or specification limits (*USL* or *LSL*), represented as U_t and L_t respectively, the PCIs were calculated as (p.97),

$$C_p = \frac{(U_t - L_t)}{(U_p - L_p)}, \tag{36}$$

$$CPL = \frac{(M - L_t)}{(M - L_p)}, \tag{37}$$

$$CPU = \frac{(U_t - M)}{(U_p - M)}, \tag{38}$$

$$C_{pk} = Min(CPL, CPU). \tag{39}$$

These equations correspond to *Equations 1*, *3*, *4*, and *5*.

Publication of Clements' (1989) article followed that of Chan et al. (1988a) by a year. It, understandably, did not include any of the calculations found in that article for

C_{pm} equivalents. Because it preceded Kotz et al. (1993), it did not contain any references to C_{pmk}. It is an important article that laid the foundation for other research that expanded Clements' methodology.

Pearn and Kotz (1994) filled this C_{pm} gap by modifying the indices C_{pm} and C_{pmk} to incorporate Clements' method. They used the percentile points generated through the application of Clements' method and the median, M, in place of the mean, μ. The results were (Pearn & Kotz, 1994, p. 142),

$$\hat{C}_{pm} = \frac{(USL - LSL)}{6\left[\left[\frac{(U_p - L_p)}{6}\right]^2 + (M - T)^2\right]^{\frac{1}{2}}}, \tag{40}$$

$$\hat{C}^*_{pm} = \frac{\min[USL - T, T - LSL]}{3\left[\left[\frac{(U_p - L_p)}{6}\right]^2 + (M - T)^2\right]^{\frac{1}{2}}}, \tag{41}$$

$$\hat{C}_{pmk} = \min\left[\frac{USL - M}{3\left[\left[\frac{(U_p - M)}{3}\right]^2 + (M - T)^2\right]^{\frac{1}{2}}}, \frac{M - LSL}{3\left[\left[\frac{(M - L_p)}{3}\right]^2 + (M - T)^2\right]^{\frac{1}{2}}}\right] \tag{42}$$

.

Pearn and Chen (1995) refined the method. Instead of treating using $U_p - M$ and $M - L_p$ as the value of 3σ, they replaced the two 3σ intervals with $(Up - L_p)/2$. The resulting equations, now including expressions for $\hat{C}_p$ and $\hat{C}_{pk}$, were (p. 387),

$$\hat{C}_p = \frac{USL - LSL}{U_p - L_p} \tag{43}$$

$$\hat{C}_{pk} = \frac{\min(USL - M, M - LSL)}{\frac{(U_p - L_p)}{2}} \tag{44}$$

$$\hat{C}_{pmk} = \frac{\min(USL - M, M - LSL)}{3\left[\left[\frac{\left(U_p - M\right)}{6}\right]^2 + \left(M - T\right)^2\right]^{\frac{1}{2}}}. \tag{45}$$

Expressed in Vännman's (1995) notation, these are (p. 387),

$$\hat{C}_p(u,v) = \frac{d - u|M - m|}{6\sqrt{\left[(U_p - L_p)/6\right]^2 + v(M - T)^2}}. \tag{46}$$

In this same work, Pearn and Chen (1995) proposed a method for asymmetric tolerance intervals. Independently, Vännman (1997) also addressed this topic by building on his original (1995) work by using different values of *u* and *v*. Pearn, Chen, and Lin (1999) refined the research they had done in the asymmetric case by incorporating some of the ideas of Vännman (1997). This incorporation of Vännman's work resulted in a set of indices that outperformed all earlier efforts for the asymmetric case.

Applications based on Burr cumulative distributions. Most of the applications of percentile methods have used Pearson probability distribution curves. The evaluation of process capability can also use the distribution curves of Burr (Burr, 1942, 1973; Burr & Cislak, 1968; Zimmer & Burr, 1963).

Burr and Cislak (1968) proposed the equation (p. 629) to explain its use,

$$F(x) = 1 - (1 + x^{c})^{-k} \quad x \geq 0 \quad c, k > 0$$
$$= 0 \qquad\qquad\quad x < 0 \qquad (47)$$

Given a data set, the mean, μ, the standard deviation, σ, the user calculated the skewness (referred to by Burr, 1942, as α_3) and the kurtosis (referred to by Burr as α_4). Using the values of α_3 and α_4, find the values of c and k. Burr's Table 1 gave the adjusted μ and σ. With these values, the final calculation uses *Equation 48* (Burr and Cislak, p. 629),

$$\frac{(X - \overline{X})}{s_X} = \frac{(x - \mu)}{\sigma}, \qquad (48)$$

to find x from X.

The first application of Burr's method to capability indices appears to be Castagliola (1996) who used it in evaluating *CPL*, *CPU*, C_p, and C_{pk} for both normal and uniform distributions. He noted in his conclusions that to assess its performance for other nonnormal distributions would require further research. A succinct explanation of the mechanics of Burr's method to capability indices is in Liu and Chen (2006). They found that this method offered superior results to those using the Pearson curves for the calculation of capability indices. Their application processed data simulated by the beta, gamma, and Weibull distributions.

Weighted variance methods. Control charts and PCIs share common roots. The most common control chart, the $\overline{X} - R$, is based on the mean of the mean of the output of a process. It also uses the range covered by the samples taken to monitor the process. The weighted variance approach is somewhat similar to the utilization of the mean and standard deviation in capability indices that, at their simplest, quantify the number of standard deviations between the specification limits.

Using earlier research (Choobineh & Branting, 1986), Choobineh and Ballard (1987) proposed a method for constructing control charts for skewed, that is, nonnormal, distributions. Given a sample mean, one counted the number of observations above the mean and divides by the number of total observations to derive a value *P*, the probability of the observation falling above the mean. Then, the probability of an observation falling below the mean is $1 - P$.

The standard deviation of the entire distribution, σ_x, is broken into two components, σ_a and σ_b, located above and below the mean respectively. Choobineh and Ballard (1987) further indicated that (p.475),

$$\sigma_x^2 = \sigma_a^2 + \sigma_b^2, \tag{49}$$

$$\sigma_a^2 \approx P\sigma_x^2, \tag{50}$$

$$\sigma_b^2 \approx (1-P)\sigma_x^2. \tag{51}$$

The upper control limit factor is $\sqrt{2P}$, and the lower control limit factor $\sqrt{2(1-P)}$. These, taken with a correction factor (p. 475), $A = 3/\sqrt{n}$, generated the upper and lower control limits for the mean and the range.

Abel (1989) was critical of Choobineh and Ballard (1987). Among other objections, he indicated that the calculations of the standard deviations were incorrect and the use of the factor *A*, was not accurate. Shewhart control charts use the factors A_2, D_3, and D_4 (Montgomery, 2013). These contain anti-biasing corrections for the distribution of the standard deviations. Choobineh and Ballard's (1987) factors do not. For information regarding the mathematics behind Abel's objection, see NIST/SEMATECH (2015).

Bai and Choi (1995) and Chang and Bai (2001) refined the weighted variance method of Choobineh and Ballard (1987) for control charts. Later, Chang, Choi, and Bai (2002) applied weighted variance to the construction of PCIs involving skewed distributions. Chang and Bai (2001) split a skewed distribution, $f(x)$, at the mean, μ, and derived two probability density functions using the same mean as *f(x)*. They reflected the distribution around the mean, incorporating the probability, P, of the value being to the left of the mean. These distributions will have different standard deviations because they have differing shapes (p. 398),

$$g_U(y) = \begin{cases} \dfrac{1}{2(1-P)} f(2\mu - y), y \le \mu \\ \dfrac{1}{2(1-P)} f(y), y > \mu \end{cases}, \tag{52}$$

$$g_L(y) = \begin{cases} \dfrac{1}{2P} f(y), y \le \mu \\ \dfrac{1}{2P} f(2\mu - y), y > \mu \end{cases}, \tag{53}$$

Chang and Bai (2001) derived the standard deviations from these equations. They used the semivariance expression from Choobineh and Branting (1986). They calculated weighted standard deviations for the upper and lower standard deviations (p. 399),

$$\sigma_U^W = P\sigma, \tag{54}$$

$$\sigma_L^W = (1-P)\sigma. \tag{55}$$

Using these two σ values in the equation for the normal distribution, with mean μ, the results gave the two probability distribution functions arising from the rotations. Use the values in *Equations 54* and *55* to determine the control limits for the control chart.

Chang et al. (2002) extended this work to PCIs. Building on the work of Chang and Bai (2001), they derived expressions for a C_p and C_{pk} equivalent indices (p.365),

$$C_p^{WSD} = \min\left\{\frac{USL-LSL}{6(2\sigma_U^W)}, \frac{USL-LSL}{6(2\sigma_L^W)}\right\} = \frac{USL-LSL}{6\sigma_x}\min\left\{\frac{1}{2P_x}, \frac{1}{2(1-P_x)}\right\}. \quad (56)$$

If $D_x = 1+\left|1-2P_x\right|$, then,

$$C_p^{WSD} = \frac{C_p}{D_x} \quad (57)$$

$$C_{pk}^{WSD} = \min\left\{C_{pku}^{WSD}, C_{pkl}^{WSD}\right\} = \min\left\{\frac{USL-\mu_x}{6P_x\sigma_x}, \frac{\mu_x-LSL}{6(1-P_x)\sigma_x}\right\}. \quad (58)$$

This weighted value is unlike the simple C_p, and variants, considered previously. Those calculations give only one value for the index. C_p^{WSD} gives a value for each of the distributions generated by rotation around the mean and selects the smaller one.

Building on earlier research, Wu (1998), and Wu, Swain, Farrington, and Messimer (1999) took a different approach to the development of an index based on weighted variance. The approach was unlike that later taken by Chang et al. (2002). Similar to the other methods, they identified the number of observations below the mean as n_1, and above the mean as n_2. The sample standard deviations below and above the mean are (p. 399),

$$S_1^2 = \frac{2\sum_{i=1}^{n_1}\left(X_i - \bar{X}\right)^2}{2n_1 - 1}, \quad (59)$$

$$S_2^2 = \frac{2\sum_{i=1}^{n_2}\left(X_i - \overline{X}\right)^2}{2n_2 - 1}. \tag{60}$$

For indices involving a target value, *T*, they used (p. 399),

$$S_{T1}^2 = \frac{2\sum_{i=1}^{n_1}\left(X_i - T\right)^2}{2n_1 - 1} = \frac{2n_1 - 1}{2n_1} S_1^2 + \left(\overline{X} - T\right)^2, \tag{61}$$

$$S_{T2}^2 = \frac{2\sum_{i=1}^{n_2}\left(X_i - T\right)^2}{2n_2 - 1} = \frac{2n_2 - 1}{2n_2} S_2^2 + \left(\overline{X} - T\right)^2. \tag{62}$$

For all of these equations, the X_is would be those corresponding to those counted by n_1 and n_2 respectively.

Using these *S* values for the standard deviations, the PCIs became (p. 399),

$$\hat{C}_p(WV) = \frac{USL - LSL}{3(S_1 + S_2)}, \tag{63}$$

$$\hat{C}_{pk}(WV) = \min\left[\frac{USL - \overline{X}}{3S_2}, \frac{\overline{X} - LSL}{3S_1}\right], \tag{64}$$

$$\hat{C}_{pm}(WV) = \min\left[\frac{USL - T}{3S_{T2}}, \frac{T - LSL}{3S_{T1}}\right], \tag{65}$$

$$\hat{C}_{pmk}(WV) = \min\left[\frac{USL - \overline{X}}{3S_{T2}}, \frac{\overline{X} - LSL}{3S_{T1}}\right]. \tag{66}$$

The Bootstrap Method

Another method for the evaluation of nonnormal data, not explored in depth in this dissertation, is the bootstrap. Efron and Tibshirani (1993) provided a thorough

description of the technique. Franklin and Gary (1991) applied the method to simulated data from the normal, the *t*, and the chi-squared distributions. They calculated 96% confidence intervals for C_p, C_{pk}, and C_{pm} from the data. Price and Price (1993) used the method to examine quality data from a Ford Motor Company engine casting plant. In this study, they used the method to construct 95% confidence intervals for $\hat{C}_{pk}$.

Pearn et al. (2005) implemented the method to examine asymmetric tolerance intervals for nonnormal data. Their efforts produced a new PCI for this application, $C''_{Np(u,v)}$. Pearn, Tai, Hsiao, and Ao (2014) applied the method to simulated data to develop a confidence interval, and a new, unbiased estimator for C_{Npk} for nonnormal data.

Tong, Chen, and Tai (2008) used the technique to compare confidence intervals from different bootstrap samples from the same distributions. Dharmasena, Zeephongsekul, and Castagliola (2010) implemented the method to calculate fixed-width confidence intervals for C_{pm}. They used simulated data from normal and lognormal distributions.

Robust Methods

Like the weighted variance techniques, the development of a median absolute deviation (MAD) approach to process capability began with the application of the method to control charts. Abu-Shawiesh (2008) proposed the substitution of the average MAD, with an appropriate adjustment factor he provides, for the adjusted standard deviation in the construction of an *s*-like control chart. The calculation of the sample standard deviation used the provided factors and the standard *c4* bias correction factors. The *s*-chart represents a plot of the movement of the standard deviation of samples taken from

the process compared to the average standard deviation. In a simulation study, Abu-Shawiesh showed that the MAD based control chart performed better for heavy-tailed distributions than did the s-chart. This advantage applied to contaminated distributions because they may have heavier tails than a pure normal distribution does.

Adekeye and Azbuike (2012) extended the work of Abu-Shawiesh (2008) to allow the creation of $\overline{\overline{X}}$ control charts. $\overline{\overline{X}}$ refers to the average of the averages of X. Adekeye and Azbuike modified the correction factors developed by Abu-Shawiesh. The new charts use the $\overline{\overline{X}}$ as the centerline with the control limits derived from the MAD using the new correction factors. Adekeye (2012) further refined the correction factors to improve the performance of these control charts.

Adekeye (2013) extended the concepts behind the MAD control charts to include PCIs and developed variants of C_p, C_{pk}, C_{pm}, and C_{pmk}. In the new indices, he applied appropriate correction factors from Abu-Shawiesh (2008) to the mean MAD and the resultant expression substituted for the value of σ in the equation. Adekeye used the technique on four data sets, two real-world, one using data simulated with an exponential distribution, and another using data simulated with a Weibull distribution. He compared these with comparable indices calculated using a percentile method.

For the first real-world set, the MAD C_p and C_{pm} were lower than the equivalent indices, C_{pk} was equal, and C_{pmk} was higher by roughly a third. For the second real-world data set, the MAD indices were all lower by 40 to 24 percent. For the simulated data, the MAD indices were higher than the percentile indices, in one case by 150 percent. While

Adekeye considered the PCI higher values in the latter two cases to be an advantage, his work showed that the use of these indices requires caution.

Mondal, Ray, and Maiti (2014) offered general guidelines for the incorporation of robustness in manufacturing that goes beyond the step of calculating a PCI. Besseris (2014) proposed a modification of existing indices to further improve the robustness of the calculations. Salazar-Alvares and Temblador (2012) provided a general review of PCIs and nonnormal processes.

The Performance of the Different Indices and Methodologies

Throughout their development, researchers have subjected PCIs to testing and scrutiny. These activities led to the development of new indices and methodologies. Much of this activity has centered on the performance of the indices when the data is nonnormal, including data from industry.

English and Taylor (1993) explored the performance of C_p and C_{pk} for nonnormal simulated data from the triangular, uniform and truncated exponential distributions. As a control, they also generated simulated data using the normal distribution. They ran 20 different models with sample sizes varying from small to large. They concluded that the indices were sensitive to the normality assumption. C_{pk} is the more sensitive of the two. They cautioned those who might use these indices to be very careful when working with nonnormal data.

Rivera et al. (1995) examined the performance of C_{pk} for transformed data generated by simulation from the gamma, lognormal, and Weibull distributions. They used logarithmic, square root, inverse, inverse square root, asinh, and power

transformations. The authors emphasized that they examined all transformed data for normality before any testing. Keselman, Othman, and Wilcox (2013) reinforced the importance of this procedure. Rivera et al. (1995) reached the conclusion that a power transform performed best, although problems arose with insufficient shortening of the tail resulting in overly conservative C_{pk} values. This observation lends validity to the sampling effect concern first voiced by Gunter (1998b).

Tang and Than (1999) studied several different methods of overcoming nonnormality. They applied probability plots, weighted variance, Clements's method, and the Box-Cox and Johnson transformations. They used these methods on data simulated from lognormal and Weibull distributions. They concluded that the Box-Cox transformation, a power transformation, performed best. This performance differential was especially apparent for the heavy-tailed lognormal data.

Pal (2005) used the Generalized Lambda Distribution (GLD) method to examine nonnormal data of the length of bolts produced by a process under study. The GLD method is similar to the curve fitting of the Johnson transformation and the percentile calculation of Pearson curves. He concluded that this method was computationally simpler than the other two approaches. He calculated $\hat{C}_{pk}$ using his method and it showed that the process did not meet the requirement of $C_{pk} \geq 1.33$. Unfortunately, he did not compute C_{pk} using the traditional approach for comparison.

As previously mentioned, Liu and Chen (2006) applied a method using the Burr XII distribution to data simulated using the beta, gamma, and Weibull distributions. They compared the results to those reached using Clements's method. They concluded that,

though both methods overestimate *CPU*, especially for heavily skewed distributions, the distortion was less using his modification of the Burr method. They deduced that the Burr method would offer more satisfactory results than Clements' for practitioners.

Han (2006) used simulated data and evaluated it using the Shapiro-Wilk test for normality. His conclusion was that if the data passes the test for normality use the standard PCI based on the estimated standard deviation. If the test indicates nonnormality, he recommended the use of a percentile-based method. The focus of his article is more on the importance of the accuracy of the test for normality, and the selection of the correct significance level, than the performance of the capability indices.

Czarski (2008) compared the results of calculating C_{pk} using percentiles calculated from a nonnormal distribution with the results reached through the application of Clements's method. The object of his research was the thickness of rolled steel plate to which he had empirically fit a Weibull distribution. He found that there was very little difference between the results of the two methods.

Czarski (2008) concluded that, in the case of nonnormal data, it is a mistake to apply methods based on the normality assumption. Instead, fit a distribution if possible, and carry out the calculations based on that distribution. Alternatively, the use of Clements' method may offer close enough results for use as a viable alternative.

Hosseinifard, Abbasi, Ahmad, and Abdollahian (2009) performed a study with two components that compared the results of different methods of dealing with nonnormality first on simulated data, and then in a real-world application. They simulated the data using the gamma, Weibull, and beta distributions. They evaluated four methods

of coping with nonnormality, the root transform, the Box-Cox method, and two percentile methods based on Burr and Clements. Their root transform was different from that previously discussed in Rivera et al. (1995). In Hosseinifard et al.'s (2009) use, a fraction power transformation minimizes the value of the skewness. The concept behind this is that an adjusted normal distribution has a skewness of zero.

Hosseinifard et al. (2009) achieved the best results for the simulated data using the root transformation. They then applied the different to the real-world data, which dealt with contact area in the semiconductor manufacturing industry, and achieved similar results. Their recommendation was that practitioners consider using the root method when dealing with nonnormal data, claiming that it not only gave better results, but also was easier to use.

Kenyon and Sale (2010) developed an index, C_{py}, based entirely on the yield of a process. Their claim was that the index's base is the hard yield. Hard yield is the amount of product produced between the specification limits, compared to the total produced. This comparison made the underlying shape of the distribution describing the process irrelevant. While they stated that a weakness of traditional capability indices is their reliance that the process is under control, others could make the same objection toward their approach. Maiti, Saha, and Nanda (2010) took a similar approach, with the denominator representing the expected process yield.

Goswami and Dutta (2013) compared the results for the calculation of C_p and C_{pk} for data from a chemical manufacturing process fitted to gamma and Weibull distributions. Methods used were the Box-Cox and Johnson transformations, an "ISO"

method apparently based on the percentiles from the two fitted distributions, and Clements method. The "ISO" method yielded the lowest indices that may have been due to the poor fit of the distributions. The other methods produced index results that were similar to each other and considerably higher than those generated by the "ISO" method were.

Kovářík and Sarga (2014) used simulated data from Weibull and lognormal distributions to evaluate the performance of nine different methods, broken into what the authors described as nontransformation and transformation, to overcome nonnormality. The nontransform approaches are a probability graphing method developed by the authors, a tolerance interval/graphical approach from Chan et al. (1988b), weighted variance, and Wright's index (1995). The methods classified as transforms include Clements and Burr-based percentile methods, and several other transformation techniques including Box-Cox and Johnson. The authors studied the index $\hat{C}_{pu}$.

Kovářík and Sarga (2014) concluded that, while non-transformational methods may offer computational simplicity, they do not perform as well as transform methods unless the data are close to normal. Transformation methods, particularly Box-Cox, generally did better than non-transformational methods. The authors found that their accuracy was sensitive to sample size, with better results coming from larger samples. If there was a requirement for small sample sizes, methods using Burr distributions yielded more accurate index estimates.

Safdar and Ahmed (2014) examined the effects of the shape parameter of the Weibull distribution on C_p, C_{pk}, C_{pm}, and C_{pmk}. As expected, they found that it did have an

impact on the estimation of the indices but found that sample size had little or no effect on the values. This lack of influence was in contrast to the findings of several other authors, for example, Kovářík and Sarga (2014) who found such an effect.

Contaminated distributions. Contaminated distributions are combinations of distributions. Tukey (1960, p. 454), showed the form of the probability density function for a mixed normal distribution as,

$$n_{\gamma,h}(z)dz = (1-\gamma)\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz + \gamma\frac{1}{h\sqrt{2\pi}}e^{-\frac{z^2}{2h^2}}dz\,. \tag{67}$$

In this equation, γ is the fraction of the total from the wider distribution that has a scale h times broader than the other. A thorough review of the literature has shown that little research relating this condition to PCIs. Bothe (1999) did consider the selection of samples from multiple process streams and developed a method based on a weighted average of nonconforming parts across the streams. His work did assume a known number of elements from the different streams. Although the contaminated distribution represented above indicates known probabilities for each distribution, in a mixture of lots from an upstream supplier this might not be the case.

Summary and Conclusions

This chapter included a discussion of the development of several approaches to the measurement of process capability. PCIs began as a simple indicator developed by Kane (1986). Since then, PCIs have grown to encompass a broad range of process conditions., and are still considered to be an important part of quality and continuous

improvement activities (Jaca, Viles, Mateo, & Santo, 2012; Kenett & Zacks, 2012; Mariappan, Gaonkar, Sakhardande, & Dhawalikar, 2012).

Many of the uses are in very specialized applications. There was no discussion of many of these specialized indices due to space and, more importantly, scope considerations. Because PCIs rely on the normality of data, this chapter contained discussions of several methods for coping with nonnormality, including transformations and robust methods, as well as weighted variance and bootstraps techniques.

The literature study revealed that many researchers used simulated data to measure the performance of the various indicators. The use of simulation is a critical consideration because the actual nature of ill-behaved data in the real-world is seldom, if ever, known with certainty. Using simulated data provides confidence in the capabilities of the process and allows the use of that knowledge to evaluate the performance of an indicator with certainty. Hence, the use of this approach to generate the data used to assess the PCIs on data from particular contaminated distributions as described in the next chapter.

A gap in the literature exists in the application of PCIs to medical manufacturing specifically. The FDA lot traceability requirement offers a leverage point to make the application of the indices more accurate. This requirement forms the basis for the methods employed in this study. The next chapter contains a description of these methods.

Chapter 3: Research Method

The purpose of this empirical quantitative study was to develop a framework that evaluated the ability of a PCI to accurately measure medical device test data under a scenario where output data combines the effects of mixed production lots of components. The study was comparative in nature. It included an examination of the performance of the most commonly used PCI, C_{pk}, using simulated process data by calculating precise capabilities and then compare these values with the results generated from nonnormal data adjusted indices. Data consisted of combinations of data from different distributions representing the situation where a lot of raw material used in a process consists of material from several supplier lots.

The simulated data represented test values from some production test and are the independent variables. Using simulated data negates the influence of, or need to control for, any independent variables, because only the final value of a test is simulated. The value of the calculated PCI was the dependent variable of interest in this study. Successful completion of this study offers evidence of the applicability of PCIs to results that may come from the combination of distributions.

The research involved an examination of the performance of the most commonly used PCI, C_{pk}, using simulated process data, as well as a description of the population data generated through simulation and then analyzed. A subsequent discussion includes an examination of sample size determination and a description of confidence intervals and hypothesis tests. The chapter ends with a discussion of the validity of the study and overcoming the potential threats to the validity.

Research Design and Rationale

This research was semiexperimental because the data used to answer the research question were simulated, that is, there was an underlying experiment executed to generate the samples. The experimental design itself was quasiexperimental because, even though the simulated data generated to test the hypothesis were random, parameter selection provided control. This random control qualifies it as a quasiexperimental design.

Device components or finished devices testing requires preset criteria. The simulated data represented test results and were the independent variable in this study. These test results might be quantifiable factors, for example, tensile strength, length, electrical resistance, or weight. The nature of the test process itself is not important. What is important is that it generates quantitative, rather than qualitative, data measurable on a scale to ascertain if a datum falls within the specification limits defined by engineering. The PCIs, variances, and standard deviations calculated from the data are the dependent variables.

Simulation runs consist of 10,000 data points. This is a commonly used number in simulation studies; for example, it is the number used by English and Taylor (1993), Han (2006), and Hosseinifard and Abbasi (2009). The appropriateness of this number would depend on the medical device manufactured. A discussion in greater detail appears later in this chapter. However, 10,000 was large enough to determine if the overall approach taken in this study was valid, but might require some adjustment under specific, large scale application of the methodology.

Several factors informed the decision to use simulated data. The first, and most important, was the ability to control the nature of the data. Because of the examination of defined distributions during the study, the ability to create test data that would follow specific distributions when tested would require a great deal of luck. By using simulated data, with the parameters of the generating distributions known, it was possible to determine the proportion of the data that should lie above or below a set point, for example, a specification limit.

Another consideration was the time and expense that would be involved in generating 10,000 data points for 12 different distributions. Even if the process to make the parts to be tested was simple, creating 120,000 parts would require considerable time and expense. After the completion of part manufacturing testing requires additional time.

Methodology

This section contains a description of the methodology used, including the mathematical model construction, data generated for the study, sampling techniques, and statistical tests used. I generated the data for this study through computer simulation. The data represent manufacturing test results applicable to all goods produced through manufacturing processes worldwide where the raw materials, or components, for the process, originate in different lots with subsequent mixing in the production process.

While the FDA requires that the lots used in the creation of a subsequent lot be traceable, a mixture of the lots could occur in production (Identification and Traceability, 2016). See Figure 4 for an illustration.

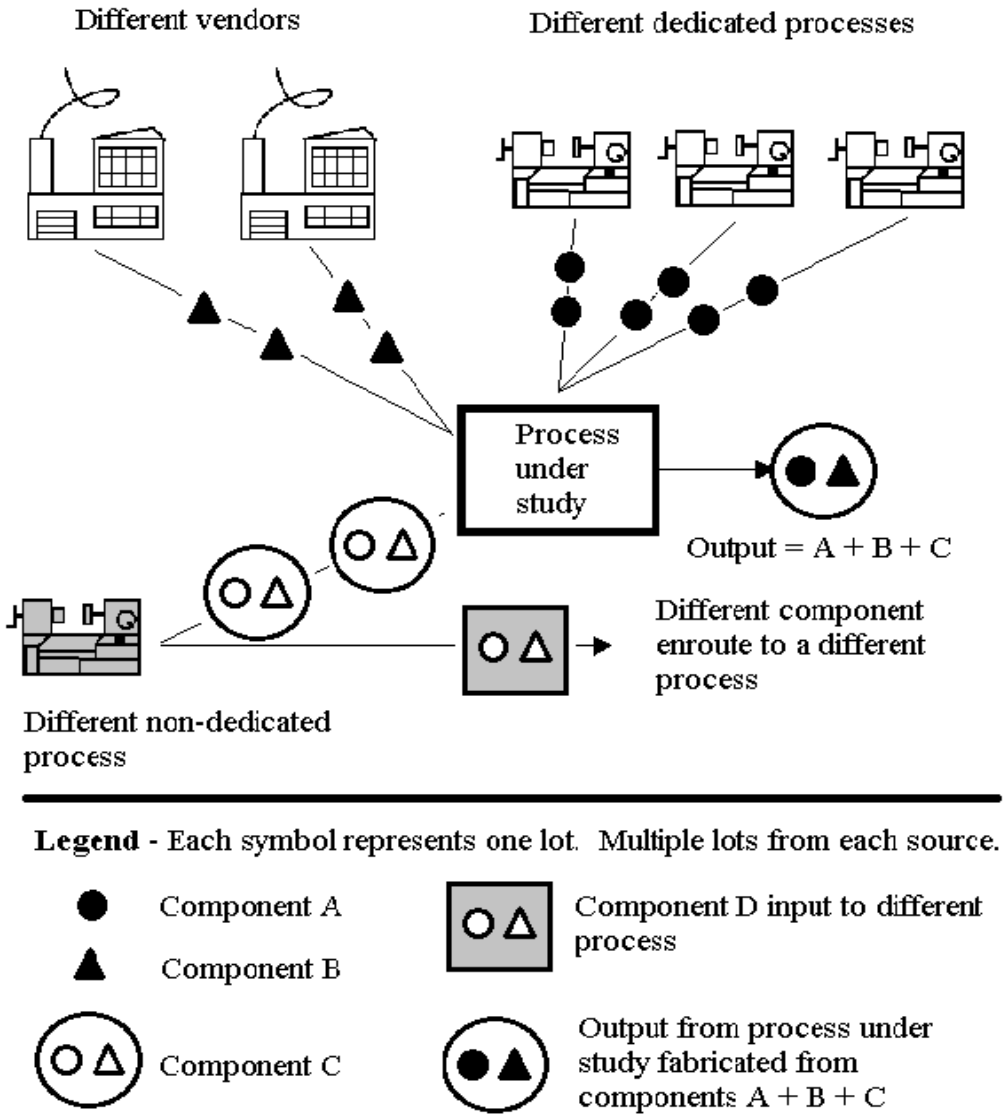

Figure 4. Illustration of how components from different lots come together in a manufacturing process.

This mixing of components can quickly become very complicated. Figure 5 contains an illustration of this lot mixing activity. Each substep uses three lots for illustration purposes. The number of lots could easily be larger depending on the complexity of the component manufactured.

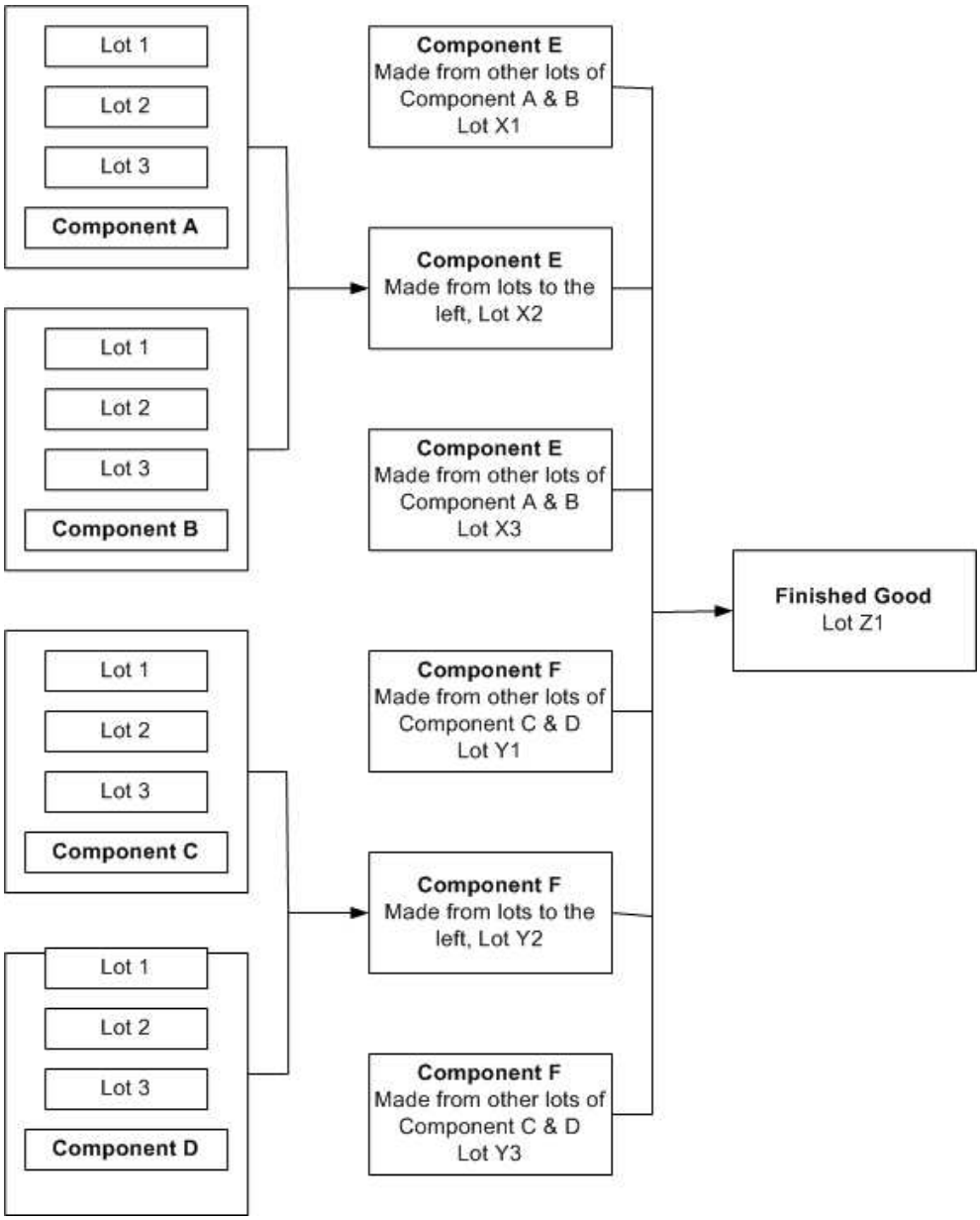

Figure 5. An example of lot mixing in production for a device with four third level components, two second level components, and o top level finished good.

Sampling and Sampling Procedures

I used R's (2016) sample function to choose samples from the simulated data. Four different sample sizes are used: 10, 30, 59, and other values dependent on the execution of the simulation. A sample size of 10 imitates a convenient number possibly

selected by an engineer who is statistically naïve. A sample size of 30 represents the situation where statistics students correctly learn that this is the number at which a *t* distribution approximates a normal distribution and use it as a default sample size. However, this fact does not make it suitable as a sample size in all situations. Cohen (1990) and Mathews (2010) expressed caution concerning its indiscriminate use as a sample size. A sample size of 59 is the value that achieves a 95% level of confidence and a 95% reliability level. Based on the success run theorem (Lipson & Sheth, 1973), the medical device industry often uses this sample size as a convenience.

The values of the mean and standard deviation generated from these different sample sizes depend on calculations from samples. The requirement was values that would represent the expected results generated from samples of these different sizes. Therefore, sample size calculations are also required to arrive at the values of the mean and standard deviation for the calculation of the PCIs.

To calculate the sample size for the mean, Mathews's (2010, p. 8) Equation 1.12 was used under the assumption that the test was being constructed as if it is a two-sided hypothesis test. The known mean of the underlying standard deviation, 100, is the test standard, and five is the known standard deviation for this distribution. The difference between a C_{pk} of 1.00 and 1.33 is one standard deviation, so the effect size is taken as δ = 5.00. To achieve 95% confidence, and power, π, of 0.90, the sample size calculation is,

$$n = \left(\frac{\left(z_{\alpha/2} + z_{\beta} \right) \sigma_x}{\delta} \right)^2 = \left(\frac{(1.96 + 1.282) 5.00}{5.00} \right)^2 = 10.51. \tag{68}$$

This rounded result is 11. Next, a comparison of the calculated sample size for the standard deviation with that required for the mean determines if it is smaller or larger than that value. Mathews's (2010, p. 59) is an appropriate approximate method. For a 95% confidence interval with a confidence interval, δ, of 0.10 of the standard deviation, the calculation is,

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{\delta}\right)^2 = \frac{1}{2}\left(\frac{1.96}{0.1}\right)^2 = 192.08 \tag{69}$$

A rounded result of 200 agrees with the table of exact results in Mathews (2010). Because this value of n is greater than that calculated for the mean, it becomes the sample size for the calculation of the mean and standard deviation of the applied sample sizes. All further reference to values derived from the different sample sizes refer to the mean and standard deviation calculated from a sample size of 200. The next section contains an elaboration on the determination of sample sizes during the simulation.

Sample size determination. The sample size intimately relates to the testing performed on the PCIs calculated for this study. The comparison of variances is the principle behind hypothesis testing of PCIs and careful analysis is critical to accurate results (Álvarez, Moya-Férnandez, Blanco-Encomienda, & Muñoz, 2015). Consider *Equations* 1 and 2 (repeated for convenience),

$$C_p = \frac{USL - LSL}{6\sigma},$$

$$C_{pk} = Min\left\{\frac{\mu - LSL}{3\sigma}, \frac{USL - \mu}{3\sigma}\right\}.$$

The value of σ, the square root of the variance, provides the base for both equations. Mathews (2010) pointed out that these PCIs are arithmetic transformations of the standard deviations of the output of the process under study. The values of the numerator of these fractions, combined with a value for the index, define the maximum σ allowable to meet these conditions. The results for the σ would be directly applicable to the variance.

The comparison of variances forms the basis for the hypothesis test using the chi-squared distribution. The relationship in *Equation (*70) forms the basis for the test,

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}. \tag{70}$$

Wackerly, Mendenhall, and Scheaffer (2002) sketched out a proof of this equation, and Penn State University (2015) provided more details of the proof. In this equation, n represents the sample size and s the sample standard deviation. The chi-squared distribution parameter is the degrees of freedom, $n - 1$. Because the definition of sample variance is,

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}, \tag{71}$$

Equation 68 simplifies to,

$$\chi^2_{n-1} = \frac{\sum_i (x_i - \bar{x})^2}{\sigma^2} = \frac{SS}{\sigma^2}. \tag{72}$$

For a given sample size and probability, this value of χ^2_{n-1} can be used to calculate the effect size which would be the difference between the variance needed to meet the PCI

requirement and the maximum value from the sample that yielded a χ^2 larger than the statistic based on the desired significance and the value of the sample mean.

Before addressing power, the implications of Type I and Type II errors need clarification for medical devices. It is important to remember that PCIs are calculated using the inverse of the standard deviation. Smaller values of σ result in higher index values. In comparing variances, three different hypothesis tests are possible.

Table 1

Hypothesis Tests for the Variance

Test Type	Hypotheses
Two-tailed	$H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 \neq \sigma_0^2$
One-tailed, lower	$H_0 : \sigma^2 \leq \sigma_0^2, H_1 : \sigma^2 > \sigma_0^2$
One-tailed, upper	$H_0 : \sigma^2 \geq \sigma_0^2, H_1 : \sigma^2 < \sigma_0^2$

In the case of a Type I error for the two-tailed test, resource waste and delays in the introduction of a device could occur while time is spent fixing a process that is not broken. A Type II error could result in actions that would endanger patient wellbeing by allowing a product that does not meet a specification to enter the marketplace. For the one-tailed, lower test, the same error ramifications would hold. For the one-tailed, upper test, reverse the results. In this case, a Type I error could endanger patient safety, and a Type II error could lead to resource waste and unnecessary delays.

Much hypothesis testing of capability indices in the medical field examines the one-tailed, lower case for the capability index that is the index less than a predetermined

value. Because the basis of the index calculation is the reciprocal of the standard deviation, the corresponding variance test is the one-tailed upper test. The emphasis is on maximizing patient safety, so the testing focuses on whether or not the standard deviation or variance exceeds that for an acceptable value of the capability index. If the standard deviation exceeds the limit, the value of the index will be smaller than the target. To achieve a balance between safety and wasted resources, α and β in this study use the same value, 0.05.

Mathew's (2010, p. 60) equation was used to determine the sample size. The value of the sample variance at the critical χ^2 point is derivable from *Equation 68*,

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}. \tag{73}$$

If $\sigma = \sigma_0$ represents the variance of a sample distribution of s^2 for H_0, and σ_1 represents the values of σ under H_1, then at the point where the chi-squared values are equal,

$$\frac{\chi^2_{1-\alpha,n-1}\sigma_0^2}{n-1} = \frac{\chi^2_{\beta,n-1}\sigma_1^2}{n-1}. \tag{74}$$

Rearranging this expression yields,

$$\frac{\chi^2_{1-\alpha,n-1}}{\chi^2_{\beta,n-1}} = \frac{\sigma_1^2}{\sigma_0^2}. \tag{75}$$

One solves *Equation 73* iteratively to find the value of n, the sample size, for which it is true. Iteration is one of the methods suggested by Guenther (1965), Mathews (2010), and Zar (2014). It does make sample size calculations impossible until the values of σ_1 are calculated using simulation or available data. Confidence intervals for the variance come directly from *Equation 74* (Mathews, 2010, p 58)

$$P\left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2}}\right) = 1-\alpha\,. \tag{76}$$

The contents of this section included specifics addressing effect size, power calculations, and sample size through the examination of hypothesis testing for variances. These are closely related to hypothesis testing for capability indices but are more direct because the values of the direct comparison of the variances. The section contains an explanation of the theory for the comparison of PCIs while avoiding the additional complexity that inverses and numerators unequal to one would introduce. A later section contains an explanation of the application of these methods to the indices.

Procedures for Simulated Data Generation

I generated 12 different sets of data with the random number generators in R (2016). Each set of data has three components. Fifty percent of the data, 5,000 points, come from a normal distribution. Twenty-five percent of the data, 2,500 points, come from an alternative distribution with its median located above the mean of the normal distribution. The other 25% of the data come from an alternative distribution with its median located below the mean of the normal distribution. The alternative distributions are the gamma, lognormal, and Weibull.

I selected the gamma distribution because of its use in several studies on PCIs. Ahmad, Abdollhian, and Zeephongsekul (2008) chose the gamma distribution to expand the work of previous authors who used it (Liu & Chen, 2006; Tang & Than, 1999). Chang and Bai (2001) selected it for their development of weighted variance control charts. Derya and Canan (2012) also used it in control chart development.

George and Ramachandran (2011) used it to explore the Johnson transform. Hosseinifard et al. (2009) also used it to examine transformations. Pal (2005) generated a nonnormal data set with the gamma distribution to evaluate PCI development using the generalized lambda distribution. Rivera et al. (1995) selected it to generate data for evaluation of several transformation methods.

I selected the Weibull distribution because of its widespread use in reliability studies, and the lognormal distribution because it often results when output characterized by several different distributions is combined (Ott, Schilling, & Neubauer, 2000). This distribution is appropriate because one of the goals for this study was to examine the effect on PCIs of a mixture of components from several distributions representing the same mixing that might occur in upstream processes. The inclusion of the lognormal distribution allows for this possibility.

Table 2 is an illustration of the generation of the component distributions. Table 3 is an illustration of the method of combining them.

Table 2
Composition of Study Simulated Data

ID	Title	Description
		Central Distributions
A	nAtT	Normal distribution, μ at target = 100, σ = 5
B	nAtO	Normal distribution, μ offset left 1σ, σ = 5

(continued)

ID	Title	Description
		Lateral Gamma Distributions
C	gammaUpperT	Gamma distribution, shape = 100, rate = 1, median transposed 1.5 σ to the right of the target.
D	gammaLowerT	Gamma distribution, shape =100, rate =1, reversed, median transposed 1.5 σ to the left of the target.
E	gammaUpperO	Gamma distribution, shape = 100, rate = 1, median transposed 2.5σ to the right of the target.
F	gammaLowerO	Gamma distribution, shape =100, rate =1, reversed, median transposed 2.5 σ to the left of the target.
		Lateral Lognormal Distributions
G	lognormalUpperT	Lognormal distribution, generated as indicated in note, median transposed 1.5 σ to the right of the target.
H	lognormalLowerT	Lognormal distribution, generated as indicated in note, reversed, median transposed 1.5 σ to the left of the target.
I	lognormalUpperO	Lognormal distribution, generated as indicated in note, median transposed 2.5σ to the right of the target.
J	lognormalLowerO	Lognormal distribution, generated as indicated in note, reversed, median transposed 2.5 σ to the left of the target.
		Lateral Weibull Distributions
K	weibUpperT	Weibull distribution, shape = 10, scale = 10, reversed, median transposed 1.5 σ to the right of the target.
L	weibLowerT	Weibull distribution, shape = 10, scale = 10, median transposed 1.5 σ to the left of the target.

(continued)

ID	Title	Description
M	weibUpperO	Weibull distribution, shape = 10, scale = 10, reversed, median transposed 2.5σ to the right of the target.
N	weibLowerO	Weibull distribution, shape = 10, scale = 10, median transposed 2.5 σ to the left of the target.

Note. All data trimmed at the *LSL* and *USL*. Trimming occurs after transposition of the lateral distributions. Trimmed data replaced by sampling with replacement from the trimmed distribution until the number of data points equals the pre-trimmed number. For gamma distribution random number generation, R (2016) uses rate as a simulation parameter. Rate defined as 1/Scale. Lognormal values generated by taking μ = 92.5, and σ = [log (107.5) – log (77.5)]/6, that is, determining a value based on the premise that *USL* – *LSL* = 10σ. The Weibull is a left-tailed distribution, therefore, unlike the gamma and lognormal, the upper distributions, rather than the lower, have reversed positions.

Table 3
Formation of Combined Distributions from Underlying Normal and Lateral Distributions.

Formation of Offset Distributions
offsetDistributions1 = gammaUpperT + gammaLowerT
offsetDistributions2 = gammaUpperO + gammaLowerO
offsetDistributions3 = lognormalUpperT + lognormalLowerT
offsetDistributions4 = lognormalUpperO + lognormalLowerO
offsetDistributions5 = weibUpperT + weibLowerT
offsetDistributions6 = weibUpperO + weibLowerO
Formation of Combination Distributions
combinationDistribution1 = nAtT + offsetDistributions1
combinationDistribution2 = nAtO + offsetDistributions1

(continued)

combinationDistribution3 = nAtT + offsetDistributions2

combinationDistribution4 = nAtO + offsetDistributions2

combinationDistribution5 = nAtT + offsetDistributions3

combinationDistribution6 = nAtO + offsetDistributions3

combinationDistribution7 = nAtT + offsetDistributions4

combinationDistribution8 = nAtO + offsetDistributions4

combinationDistribution9 = nAtT + offsetDistributions5

combinationDistribution10 = nAtO + offsetDistributions5

combinationDistribution11 = nAtT + offsetDistributions6

combinationDistribution12 = nAtO + offsetDistributions6

The first group of data consists of a normal distribution with its mean centered at the target value with specification limits located five standard deviations on either side of the mean. Using their medians, alternative distributions are located first at 1.5 standard deviations on either side of the mean, followed by locations at 2.5 standard deviations on either side of the mean. This relocation of data is a standard method in simulation studies (Law, 2007).

The application of a simple linear transformation reversed the values to the left of the target for the gamma and lognormal distributions, and to the right of the target for the Weibull distribution, so that the tail of the distribution exhibits skew away from the mean. This reversal imitates the real-world situation of distribution tails often pointing in the least favorable direction and the scarcity of left tailed probability distributions (Sleeper, 2007). The application of a custom R (2016) function truncated the values from

the alternate distributions at a value equal to the *LSL* and *USL*. Such truncation is common in simulation studies (Law, 2007). For example, truncated data results if, upon inspection of process output, the quality department rejected all units with a value above or below a quality limit (Nadarajah & Kotz, 2006; Plansky, Chou, & Mason, 1998). Figure 6 shows the process of generating a combined distribution.

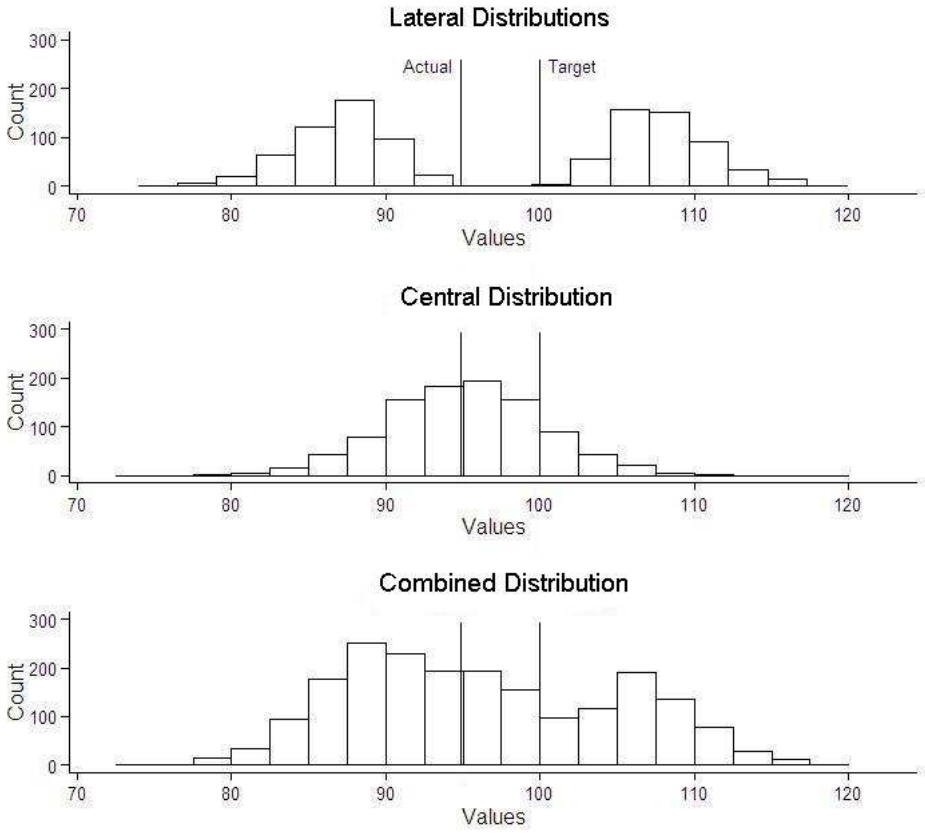

Figure 6. An illustration of combined distribution generation.

The second sample group is similar to the first, but with the normal and alternate distributions locations shifted one standard deviation to the left of the target value. This shift results in a less capable process, and the intent was to see how accurately the process capability calculations reflects this. Using weighted probability values, accurate

calculations yielded the equivalent capability indices. These are the standards of comparison for the calculated values.

Along with multimodality, each distribution exhibits fat tails. This characteristic may affect the required sample size needed to make accurate projections about the features of the distribution (Wilcox, 2012). Each combination distribution is obviously multimodal. Application of R (2016) functions transformed each distribution with the Box-Cox, square root, inverse, inverse square root, and asinh (Rivera, 1995) transforms. The application of normality tests determined if the results of the transformation are beneficial. Results from the transformations indicated that normality is unachievable for two distributions and evaluation required an alternative method.

I collected data from each distribution with the R *sample* function. Sampling occurred without replacement. This replicated a situation in which destructive testing occurs, for example, conducting a pull test to a joint failure.

Instrumentation and Operationalization of Constructs

Data collection in this study does not require instrument use. Computer simulation generated all data. The data represents the value of a quality test performed on the output of a process. The distributions of this test data, rather than the data themselves are the focus of this study. Because the test values come directly from the simulated data, no operationalization of constructs was required.

While variables in this study are unitless, because the indices are unitless, in actual applications of the methods used, the user would directly measure the variables. For example, some breaking force measures are pounds, ounces, newtons, or dynes.

Potential length measures are inches, feet, millimeters, centimeters, or meters. Dimensionally, because a division occurs with a unit of measure divided by the same unit of measure in the calculation of an index, the indices are unitless. Measurement directly yields the units of interest, so no operationalization was required.

Data Analysis Plan

This section contains an identification of the software used for analysis, the data, the research questions, and the analysis plan. The description of the analysis plan includes the statistical tests used to test the hypotheses. It also contains a section on the interpretation of test results.

Software

The software used for the analysis was R (2016). In addition to base R, the analysis required the use of several R packages: AID (Dag, Asar, & Ilk, 2015), ggplot2 (Wickham, 2009), gridExtra (Auguie, 2016), MASS (Venables & Ripley, 2002), moments (Komsta & Novomestky, 2015), and nortest (Gross & Ligges, 2015).

Data Cleaning and Screening

Computer simulation generates the data used in this study. The careful selection of the distribution parameters eliminates the need for data cleaning or screening. This negates the need for any further data modification prior to analysis.

Research Questions and Hypotheses

The quantitative research question of this research was how accurately does the calculated value of C_{pk}, under the assumption of normality, reflect the actual probabilities of nonconformance from simulated distributions representing the mixture of components

from different upstream production batches in a subsequent process? This question reflects the real-world situation in which a production line uses components from several different lots. A PCI, as used in industry, is primarily a point value (Porter and Oakland, 1991). A quality manual or protocol may state that the C_{pk} value must be greater than 1.33, 1.50, or some other value. This question leads to three formal research questions.

The first series of tests involve the comparison of calculated values of the PCIs based on samples taken from each distribution set to a required value.

Research Question 1: Do the PCIs calculated from samples of the combined distributions meet the industry standard?

$H_0 1$: $PCI_C \geq 1.33$,

$H_a 1$: $PCI_C < 1.33$.

Peña-Rodríguez (2013) suggests the value of 1.33. PCI_C uses the calculated C_{pk} from sample sizes at levels of 10, 30, 59, and a value determined from each of the 12 distributions using the method described by Mathews (2010). The data in all research questions represents the test results from a test conducted after the testing of a component assembled from parts. I tested the data for normality using the Anderson-Darling method. When the test results indicate nonnomality, I transformed and retested it. The second series of tests involve a comparison of the calculated positions on the *x*-axis of 4 standard deviations, equivalent to a C_{pk} of 1.33 to the value calculated from the parameters of the 12 combined distributions and their transformed values when required.

Research Question 2: Do the values calculated from samples taken from a combined distribution exceed the actual values required to meet the standard?

$H_0 2$: $x_{pci} \geq x_{pdf}$,

$H_a 2$: $x_{pci} < x_{pdf}$.

x_{pci} is the x-axis value calculated from the required value of the PCI, and x_{pdf} is the value calculated directly from the combined probability density functions. The results of this test should mirror those of research question 1, but show the percentage difference between the actual and calculated values. The test schema included only raw data because of the inability to transform to normality.

Previous tests have used the combined distributions. The third series of tests tested the components of the combined distributions individually and comparing the PCI calculated from these distributions with the standard C_{pk} of 1.33.

Research Question 3: Do the data values from the lateral distributions, isolated from the underlying normal distributions, meet the industry standards?

$H_0 3$: $PC_{gamma} \geq 1.33$,

$H_a 3$: $PC_{gamma} < 1.33$,

$H_0 3$: $PC_{lognormal} \geq 1.33$,

$H_a 3$: $PC_{lognormal} < 1.33$,

$H_0 3$: $PC_{Weibull} \geq 1.33$,

$H_a 3$: $PC_{Weibull} < 1.33$.

In these hypotheses, PC_{gamma} is the C_{pk} calculated from the gamma distributions, $PC_{lognormal}$ is the C_{pk} calculated from the lognormal distributions, and $PC_{Weibull}$ is the C_{pk} calculated from the Weibull distributions. These tests evaluate the suggestion that lots constructed using different lots of components require individual tests. This individual

testing might compensate for the effect of a fattened tail on the value of the standard deviation used to calculate the capability index. Test input consists of both raw and transformed data.

The operationalization of the above RQ into a null and alternative hypothesis required the calculation of the PCIs from sample sizes of 10, 30, 59, and a sample size needed to achieve an a priori specified power 1-beta. The value of 10 represents a low convenience value. The value of 30 is key because it is the value at which a *t*-distribution begins to correspond to a normal distribution. The sample size of 59 achieves a 95% level of confidence (Lipson & Sheth, 1973). The calculated sample size based on power was indeterminate until the generation of simulated results. The calculated sample size based on power was indeterminate until the generation of simulated results.

Next, data from these samples permitted the calculation of a 95% confidence interval for the PCIs. Because simulated data generates the entire populations, it was possible to compute the true value of all of the capability indices. Finally, if PCI_C represents the calculated PCI, and PCI_T represents the true index, then the hypothesis tests operationalizing the above RQ was:

H_0: $PCI_C \geq PCI_T$ or required

H_a: $PCI_C < PCI_T$ or required

My testing used methods for the variance/standard deviation and the chi-square distributions and determined if the calculated PCI is greater than or equal to the true PCI. This testing examined if a Type II error was occurring in the calculation of the PCI. The definitional equation for C_{pk} was the basis for calculating the values for the sample PCIs.

Next, does the calculation of actual values of the indices agree with the probabilities as defined by the generating cumulative distribution function? From the definitional equation for C_{pk}, it is readily apparent that the value of the index is directly proportional to the distance between the mean of the samples and the specification limit, and inversely proportional to the standard deviation of the sample. This establishes cause and effect of the data on the value of the PCI.

Data Analysis Procedures

Structure of analysis. Hypothesis testing for variances and standard deviations to give theoretical foundation for the tests outlined in this section. Those methods apply to the testing of capability indices, but require an adjustment because capability indices rely on the inverse of the standard deviation. Thus, if $A < B < C$ is true, then, when inverses are considered, the expression becomes, $1/C < 1/B < 1/A$. Given the inverse relations of the definitional equations, the larger the variances/standard deviations, the smaller the capability indices are.

Figure 7 is a schematic representation of the analysis plan.

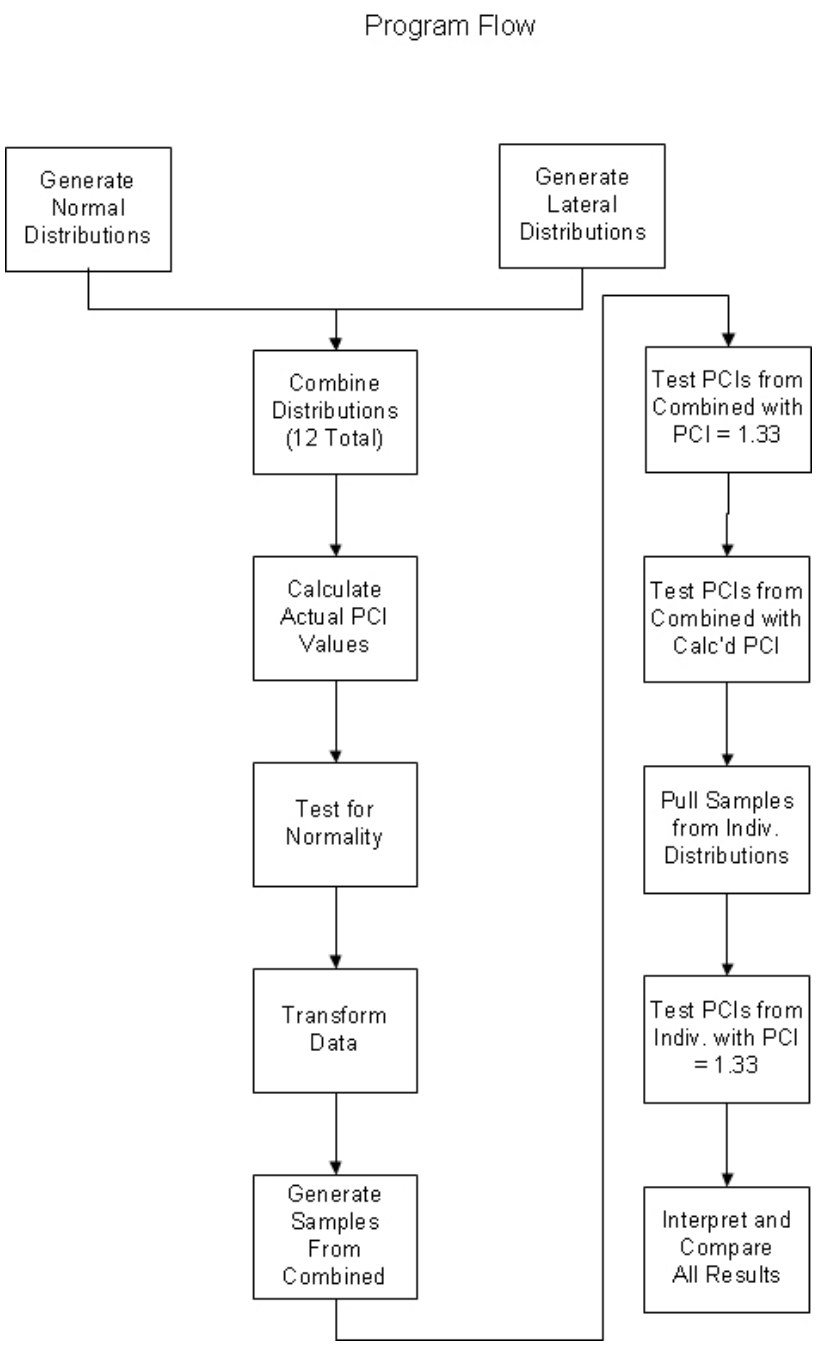

Figure 7. Schematic representation of the data analysis plan.

Confidence intervals. This study requires the calculation of PCIs from samples drawn from the simulated output of a process described earlier. The underlying assumption is that the process is under control and that whatever test performed on the output from that process, for example, measuring fill weight of a container, will produce normally distributed results. The assumption of normality is critical to hypothesis testing of capability indices. In the case of the simulated data, testing confirmed or denied the assumption of normality, but the calculation of the indices proceeded even if the results from the raw data indicated nonnormality. This compromise allowed a comparison of the conclusions drawn from an unwarranted assumption of normality to those based on reality.

Consider first *Equation 1* for C_p (repeated unnumbered for convenience),

$$C_p = \frac{USL - LSL}{6\sigma}.$$

If given a required value of C_p, and values for *USL* and *LSL*, the maximum value of σ that will result in this value of C_p can be calculated from this expression as,

$$\sigma = \frac{USL - LSL}{6C_p}. \tag{77}$$

If σ is larger than this value, the value of C_p is less than required.

Now, consider *Equation 2* (repeated unnumbered for convenience),

$$C_{pk} = Min\left\{\frac{\mu - LSL}{3\sigma}, \frac{USL - \mu}{3\sigma}\right\}.$$

In this expression, for a given σ, the calculation of the value of C_{pk} uses the smaller of the two values, $\mu - LSL$ or $USL - \mu$. The smaller of the two uses the nearest specification limit or *NSL*. The expression for C_{pk} then becomes,

$$C_{pk} = \frac{NSL}{3\sigma}. \tag{78}$$

As was done for C_p, the maximum value of σ that gives the required value of C_{pk} is,

$$\sigma = \frac{NSL}{3C_{pk}}. \tag{79}$$

The comparison of these values of the standard deviations to those calculated for each simulated data set, for both raw and transformed data, uses the confidence intervals calculated from the sample data.

Hypothesis testing, general. The research question in this study was, given that the value of a process capability, calculated from a set of data composed of the combination of values from three different distributions, does that index adequately reflect the overall capability of the process? Answering the research question requires three different sets of hypothesis tests.

Hypothesis tests against the requirement. This series of tests calculated the values of the PCIs using samples taken from each distribution set and tested them against a desired value. The value is 1.33 as suggested by Peña-Rodríguez (2013). If PCI_T represents the calculated capability index, that is, either C_p or C_{pk}, the hypothesis test was:

$H_0 1$: $PCI_C \geq 1.33$,

$H_a 1$: $PCI_C < 1.33$.

These tests included data from all 12 distributions.

Hypothesis tests against the calculated value. Because the combined distributions were constructed using probability distributions with known parameters, a capability index of 1.33 determines an exact *x*-axis value. This series of tests compared the *x*-axis values of the PCIs calculated using samples taken from each distribution set against that value. If the value calculated directly from the probability density functions is x_{pdf}, and the value for the capability indices is x_{pci}, then the hypothesis test becomes,

$H_0 2$: $x_{pci} \geq x_{pdf}$,

$H_a 2$: $x_{pci} < x_{pdf}$.

Tests included data from all 12 distributions.

Hypothesis tests with components from each distribution taken separately. Combining the three distributions fattened the tails of the majority normally distributed data and increased the value of the standard deviation. This series of tests individually addressed data for the gamma, lognormal, and Weibull distribution placed to the left of the mean of the normal distribution. It was not necessary to test the normal distribution or the other distribution located to the right of the mean of the normal distribution. Because of their construction, these distributions would yield a capable process. The testing concludes with comparisons of the calculated capability index values to the standard of 1.33. Representing these capability indices by PC_{gamma}, $PC_{lognormal}$, and $PC_{Weibull}$, the hypothesis tests become,

$H_0 3$: $PC_{gamma} \geq 1.33$,

$H_a 3$: $PC_{gamma} < 1.33$,

$H_0 3$: $PC_{lognormal} \geq 1.33$,

$H_a 3$: $PC_{lognormal} < 1.33$,

$H_0 3$: $PC_{Weibull} \geq 1.33$,

$H_a 3$: $PC_{Weibull} < 1.33$.

These tests evaluate the suggestion that lots constructed using different lots of components require individual tests. This individual testing might compensate for the effect of a fattened tail on the value of the standard deviation used to calculate the capability index.

Threats to Validity

This section contains a discussion of external, internal, construct validity, and ethical considerations. In the extensive literature review conducted for this study, the mathematical nature of the research was paramount, and no concerns about validity arose except those arising from the possible lack of process control or nonnormality of the data. This nonnormality condition was the focus of this study.

External Validity

All analysis in this study uses definitional equations and simulated data. Because of this, the results do not tie to any particular period, product, or process. The FDA requirement for lot traceability of components used in the manufacture of pharmaceuticals and medical devices led to the choice of the medical device manufacturing segment as the setting for this study. This requirement guarantees that the information to separate components by production lot is available. The results of this

study would be generalizable to any scenario in which information from component lots is accessible. There are no anticipated external threats to validity.

Simulations consisted of 10,000 data points. This is a commonly used number in simulation studies; for example, it is the number used by English and Taylor (1993), Han (2006), and Hosseinifard and Abbasi (2009). Whether or not this number is appropriate would depend on the medical device manufactured. In the case of hypodermic syringes, where worldwide daily production would number in the millions of units per day, it might be small. In the case of M.R.I. machines, where the total number of machines in existence worldwide is only 36,000 (Rinck, 2016), it might be too large. However, it was large enough to determine if the overall approach taken in this study is valid, but might require some adjustment under specific, large-scale application of the methodology.

This study was very specific in its limitations, that is, cases where traceable lots from upstream processes are mixed. The test results of the testable product made from those components are separable by lot. In other cases, the method may not be applicable.

Internal Validity

All analysis in this study uses definitional equations and simulated data. There are no human or animal subjects involved. Because of the lack of subjects who can react or adjust to the research, there are no anticipated internal threats to validity.

Construct Validity

Construct validity refers to whether an instrument actually measures what it was constructed to measure. Historically, it was a general category for all validity (Warner, 2008). It is in this sense that the following discussion takes place.

One area of this study that might raise questions is trimming the simulated data to eliminate any points laying outside of 5 standard deviations and replacing the trimmed data with new points drawn from the remaining distribution using the sampling procedure in R (2015). Sleeper (2007) cautioned "Truncation is usually the wrong way to solve a modeling problem" (p. 50). His point is valid if the raw process is considered. In contrast, this study is examining components that have already gone through a sorting process by a quality control department and the truncation and replacement that would take place to achieve a full lot size reflects reality (Nadarajah & Kotz, 2006; Plansky et al., 1998).

Truncating a distribution does change the nature of the distribution. Consider the standard normal distribution for simplicity's sake. Figure 8 contains an illustration of a comparison of the CDF of a truncated distribution (assuming upper and lower control limits of ±1.64, with the upper and lower 5% eliminated).

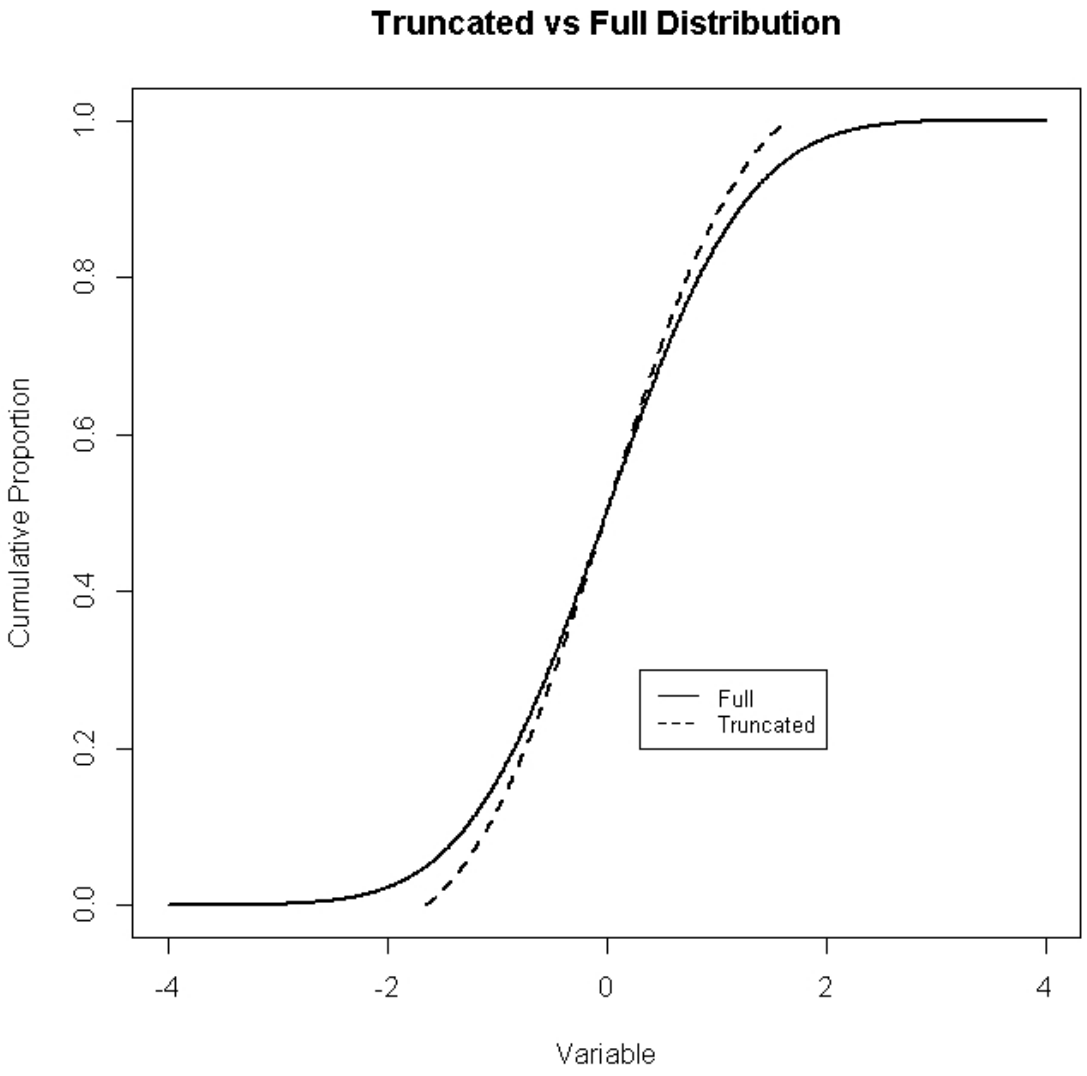

Figure 8. Comparison of a trimmed and non-trimmed CDF of a normal function.

The trimmed distribution begins later than the non-trimmed distribution because of the truncation, crosses at the mean, and then increases more rapidly. In addition, the truncated distribution boundaries are ± 1.64, while the nontruncated distribution's domain is ± ∞.

Another area of construct validity that might arise was the use of C_{pk} itself as a standard for judging the performance of a process. This also could overlap into the realm

of face validity. As several authors cited in this study have pointed out, this is a valid question. However, C_{pk} has become the standard and a goal of this study was to help evaluate its suitability for that task.

The validity of the conclusions of this study might be subject to some scrutiny under the new examination of the use of *p*-values in research by the American Statistical Association (Wasserstein & Lazar, 2016). Any judgment as to the suitability of *p*-values in research lies far out of the scope of this study.

Ethical Procedures

This study was a purely quantitative comparative research in mathematical modeling and simulation. Computer simulation generates all data specifically for this study. Values for the initial target value are set as 100 for ease of computation and to aid in the understanding and visualization of differences in the values. Setting this value at 100 makes it easier to think in terms of the target being 100%. This resemblance may also aid understanding. The upper and lower specification limits are set as being equivalent to five standard deviations because this would correspond to plus or minus 25% of the target value while also allowing a five-unit transposition of the mean to still fall within the specification limits. Parameter values of the lateral distributions are set as typical values that would give these distributions their characteristic shapes.

No human or animal participants are involved. There are no anticipated requirements for the use of confidential information; therefore, there are no storage, anonymity, or privacy issues. This study used no corporate data or equipment. The

Walden Institutional Review Board (IRB) approval number for this study is 11-16-16-0029746.

Summary

This chapter contains a summary of the methodology, including the population studied, mathematical model construction, sampling techniques, and statistical tests. Model construction combines a normal central distribution, and two noncentral distributions. The noncentral distributions were the gamma, the lognormal, and the Weibull distributions. The location of the median of the noncentral distributions is one and a half standard deviations to the left of the mean of the central distribution. Reversal of the left noncentral distribution yields left pointing tails. Combining the distributions fattened the tails of the overall distribution.

The research included two cases of the combined distributions. The first cases assumed that the target value of the distribution is the mean of the central distribution. The second case offset the central distribution and the two noncentral distributions to the left of the mean of the original central distribution by one standard deviation. In both instances, the mean of the original, non-offset, central distribution was the target of the distribution.

There is an infinite number of distributions, combinations of distributions, and offsets possible. The methods of analysis developed for this study are not unique to those selected, and are adaptable to other combinations of distributions. The concern was to simulate subcomponents, with different physical characteristics, from different lots combined in a downstream production process in the manufacture of medical devices or

pharmaceutical products. The research requires the application of a series of hypothesis tests, designed to measure the effects of lot mixing on the most commonly used measures of the process capability of the final process.

The next chapter contains a description of the individual and combination distributions developed for this study, as well as the results of the hypothesis tests performed on a series of samples taken from the distributions.

Chapter 4: Results

Introduction

The purpose of this empirical quantitative study was to develop a framework that evaluates the ability of a PCI to accurately measure medical device test data under a scenario where output data combines the effects of mixed production lots of components in the medical device industry. Computer simulation generated twelve different sets of data representing mixed production test results where subcomponents came from different production lots. Testing of the sampled data used an industry standard PCI, C_{pk}, to determine if the test results reflected the actual process capability by answering several research questions.

Research Question 1: Do the PCIs calculated from samples of the combined distributions meet or exceed the industry standard of 1.33?

Research Question 2: Do the values calculated from samples taken from a combined distribution exceed the actual values required to meet the industry standard?

Research Question 3: Do the data values from the lateral distributions, isolated from the underlying normal distributions, meet the industry standard?

This chapter begins with a description of the equivalent of a pilot study for this type of research. Next, it moves to the timing and methods used to generate the data. Following this is a description of the data generated, the rationale behind the statistical methods used for analysis, the results of the data analysis, including power and effect sizes, and the outcomes of the hypothesis tests formulated to answer the research

questions. The chapter concludes with summary of the results from the analysis and the research questions.

Data Generation

Background

From the conceptual stage of the development of this study, the intent was to use computer simulated data. There are three key reasons for this. First, gaining approval for the use of actual data would be difficult because the purpose of the study is to examine data on the boundaries of those which would yield acceptable results for examining the production of a medical device. Given the litigious nature of our society, securing permission to use questionable data would be difficult.

The second reason is the difficulty in producing data specifically for use in this study. The research investigates 12 different data sets with 10,000 points in each set. The time and expense involved in generating this quantity of information is prohibitive.

Last, the simulated data allow the exact characterization of each of the frequency distributions generated. This precision permits the precise comparison of the results of the statistical analyses with known information for each distribution. The result is an exact answer to RQ 3.

Development of the programs used to analyze the data began prior the time that the prospectus received approval. Final generation of the computer simulated data did not occur until after IRB approval of this study. Through the use of a different random seed, the data for this study is different than that used to develop the program.

Data Generation Procedure

The generation of the simulated data followed the procedures listed in Tables 2 and 3. The *rgamma*, *rnorm*, and *rweibull* random number generators in the base R (2016) package generated the random numbers using the parameters listed in Table 2. Generating the random numbers using these functions is straightforward.

Generating the lognormal random numbers requires first creating a series of random numbers using *rnorm*. These random numbers corresponded to the logarithms centered at the value of the mean of the distribution with a standard deviation equivalent to the value of the log of the upper tolerance limit minus the lower tolerance limit divided by 10. The subtraction operation is not division in this case, even though it involves logarithms. Nor does the division by 10 represent finding a root. Instead these operations divide up the *x*-axis to find the standard deviation. In the lognormal distribution, it is the logarithms of the data values that are normally distributed. The actual data values come from raising *e* to the powers generated with these steps.

Combinations of data generated from the R (2016) random number generators formed 12 different distributions. The details of the combinations are in Table 3. Application of the R *sample* function provided sample sizes of 10, 30, and 59 data values with an additional value calculated for each distribution based on the standard deviation of the distribution sampled.

The data represent the results of a quantifiable test applied to a component made up of subcomponents from three identifiable lots. This situation occurs in medical device manufacturing because of the FDA requirement for lot traceability. While an infinite

number of distributions of test results would arise in the real world, the methods developed in this research are applicable to any fat tailed distributions and are not limited to the data generated for this research.

Histograms of the data from the 12 combination distributions are in Appendix A. In Figure A1, *nAtT* is the underlying normal distribution used with the lateral distributions. Figure A2, *nAtO*, is the offset normal distribution used with the lateral distributions. The R (2016) code to generate the data is in the file *DCode03_Simulations.r* located in Appendix B.

Study Results

Descriptive Statistics

Table 4 provides a summary of the descriptive statistics for the 12 combination distributions. Each distribution contains 10,000 data points. The column labeled Sample Size is the calculated sample size used in the data analysis along with the values of 10, 30, and 59. I used the method found in Mathews (2010) for this calculation. Skew and kurtosis functions are not part of the base R (2016) package. The *moments* package (Komsta & Novomestky, 2015) furnishes these functions.

Note that two of the calculated sample sizes are size 10. The analysis results from values that duplicate those of the chosen sample size of 10. While elimination of the results from the calculated sample size of 10 avoids repetition, keeping them facilitates any comparisons of the results of the calculated sizes.

Table 4

Summary Statistics of Combination Distributions

Distribution	Mean	Median	Standard Deviation	Skew	Kurtosis	Calculated Sample Size
comboDist1	100.01	100.05	8.68	-0.02	3.42	16
comboDist2	97.51	96.52	9.00	0.50	3.36	14
comboDist3	99.19	99.68	10.80	-0.08	2.77	9
comboDist4	96.69	95.54	10.87	0.45	2.88	9
comboDist5	100.06	100.09	6.77	-0.02	2.07	48
comboDist6	97.56	95.90	7.20	0.40	2.25	34
comboDist7	97.57	100.05	10.11	-0.52	2.05	10
comboDist8	95.07	95.09	9.78	-0.01	1.93	11
comboDist9	99.04	99.44	9.14	-0.07	3.27	13
comboDist10	96.54	95.70	9.18	0.48	3.45	13
comboDist11	97.56	98.97	10.69	-0.18	2.63	9
comboDist12	95.06	94.76	10.39	0.36	2.95	10

Preliminary Steps

I drew samples of each sample size for analysis using the *sample* function in R (2016) and used the *ad.test* function from the *nortest* package to test for normality. The *ad.test* uses the Anderson Darling test (Gross & Ligges, 2015). This test is particularly sensitive to the tails of a distribution. Because the distributions created for this study purposely have heavy tails, the Anderson Darling test would provide the strictest results.

Table 5 contains the results of the Anderson Darling test for all of the samples from the different distributions. The column labeled AD Statistic has the Anderson

Darling statistics for the distributions, and *p* Value is the associated *p* value. In the case of the Anderson Darling statistic, smaller is better. Values of *p* less than .05 indicate that the test failed to show normality. The *p* values have 3 decimal places to clearly illustrate the failures.

Table 5

Anderson Darling Test Results with Failures in Bold Font

Distribution	AD Statistic	*p* Value
Dist1SampSize10	0.36	.368
Dist1SampSize30	0.25	.725
Dist1SampSizeC16	0.31	.527
Dist1SampSize59	0.16	.946
Dist2SampSize10	0.24	.689
Dist2SampSize30	0.28	.605
Dist2SampSizeC14	0.26	.638
Dist2SampSize59	0.31	.546
Dist3SampSize10	0.26	.624
Dist3SampSize30	0.42	.306
Dist3SampSizeC9	0.27	.581
Dist3SampSize59	0.27	.668
Dist4SampSize10	0.24	.684
Dist4SampSize30	0.64	.084
Dist4SampSizeC9	0.22	.756
Dist4SampSize59	0.32	.529

(continued)

Distribution	AD Statistic	*p* Value
Dist5SampSize10	0.22	.762
Dist5SampSize30	0.21	.853
Dist5SampSizeC48	0.57	.134
Dist5SampSize59	0.58	.128
Dist6SampSize10	0.12	.985
Dist6SampSize30	0.39	.365
Dist6SampSizeC34	0.64	.087
Dist6SampSize59	0.71	.062
Dist7SampSize10	0.42	.259
Dist7SampSize30	0.59	.114
Dist7SampSizeC10	0.42	.259
Dist7SampSize59	**1.86**	**.000**
Dist8SampSize10	0.25	.669
Dist8SampSize30	0.26	.679
Dist8SampSizeC11	0.26	.631
Dist8SampSize59	**0.75**	**.048**
Dist9SampSize10	0.19	.868
Dist9SampSize30	0.25	.719
Dist9SampSizeC13	0.22	.794
Dist9SampSize59	0.32	.517

(continued)

Distribution	AD Statistic	*p* Value
Dist10SampSize10	0.26	.616
Dist10SampSize30	0.37	.400
Dist10SampSizeC13	0.40	.306
Dist10SampSize59	0.22	.833
Dist11SampSize10	0.25	.651
Dist11SampSize30	0.41	.320
Dist11SampSizeC9	0.33	.435
Dist11SampSize59	0.68	.074
Dist12SampSize10	0.30	.524
Dist12SampSize30	0.21	.856
Dist12SampSizeC10	0.30	.524
Dist12SampSize59	0.36	.439

Only two datasets failed the normality test; both data sets had the largest sample size, 59, and the lognormal lateral distribution. I next transformed the data identified as nonnormal using the Box-Cox, square root, inverse, inverse square root, and asinh transforms. A Box-Cox transformation function is not part of the base R (2016) package but is in the *AID* package (Dag, Asar, & Ilk, 2015). The *boxcoxnc* function applies a sequence of lambda value to the transform and then tests for normality using seven different normality tests, including the Anderson Darling test.

Table 6 contains the transformation results. The *p* values have 3 decimal places to clearly illustrate the failures. None of the transformation methods succeeded. Because one of the purposes of this study to examine the results a statistically unsophisticated

engineer might generate, the nonnormal data will remain part of the analysis. Statistically, this is incorrect practice. The engineer should adjust the process to achieve normal output, or output transformable to normality.

Table 6

Anderson Darling Normality Test Results of Transformed Nonnormal Data

	Dist7SampSize59		Dist8SampSize59	
Method	AD Statistic	*p* Value	AD Statistic	*p* Value
Box-Cox	0.96	.015	0.75	.047
Square Root	2.03	.000	0.76	.045
Inverse	2.61	.000	0.89	.021
Inverse Square Root	2.41	.000	0.83	.030
Asinh	2.21	.000	0.79	.039

Statistical Assumptions

Normality of data was the underlying assumption behind this study. PCI calculation depends on this assumption. The analytical tool used to answer RQ1, using the chi squared distribution to compare two variances, also depends on normality of data. The testing for normality using the Anderson Darling test showed the validity of this assumption for all but two of the distributions. Attempts to transform these distributions to normality failed, but the analysis continued as if normality was present for reasons explained in the previous section.

RQ2 and RQ3 were deterministic with results based on calculated numbers from the distributions rather than through statistical analysis. Answering these questions involved no statistical assumptions. The lack of statistical inference restricts the answers

to these questions to the generated distributions, so generalization of the conclusions is more difficult. Generalizing these answers requires careful qualification of the applicability conditions.

Research Question 1

Research Question 1 was, do the PCIs calculated from samples of the combined distributions meet the industry standard of 1.33.

$H_0 1$: $PCI_C \geq 1.33$,

$H_a 1$: $PCI_C < 1.33$.

Answering RQ1 was a multiple stage process beginning with comparisons of variance. Common tests for equality of variance do not apply because they depend on values calculated from individual data points from two samples. These common tests include the F-test (Sheskin, 2000), Levene's (1960) test, etc. Sampling distributions of variances follow a chi squared distribution. This forms the basis for answering this research question.

The calculations for *Dist1SampSize30* form an example. The mean of the data from this distribution is 99.9460 and the pooled standard deviation is 8.6647. The required standard deviation for a C_{pk} of 1.33 is the minimum of (99.9653 – 75) and (125 – 99.9653), where 75 and 125 are the lower and upper tolerance limits, divided by 4. For this distribution, the required standard deviation is 6.2413. The standard deviations of 8.6637 and 6.24 correspond to variances of 75.0600 and 38.9376 respectively.

The variances provide the basis for calculating the confidence intervals for the respective standard deviations and C_{pk} values. The relationship used to do the calculations is found in Zar (2014, p. 127),

$$\chi^2 = \frac{vs^2}{\sigma^2}, \tag{80}$$

where v is the degrees of freedom, s is the sample standard deviation, and σ is the standard deviation under study. The process begins with calculating the chi squared values corresponding to the probability density function values. Figure 9 is a plot of chi squared values with the value at the 95% confidence value indicated by a solid line.

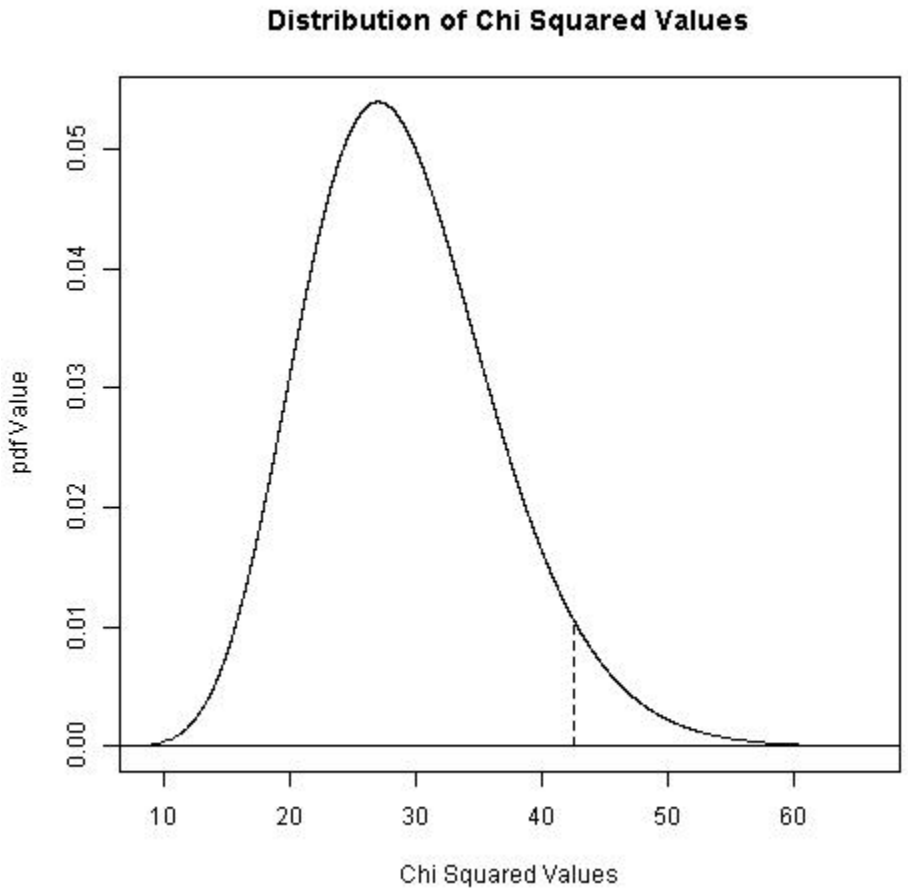

Figure 9. Chi squared values for a distribution with nine degrees of freedom.

Algebraic manipulation of *Equation* 80 yields the relation needed to calculate confidence intervals for the variance. The relationship for the variances is in *Equation* 81. Figure 10 contains a plot of the variance values with confidence interval indicated. The plot is for a standard deviation of 6.2413. This is the standard deviation needed to result in a C_{pk} of 1.33 when the mean is 99.9653 for *Dist1SampSize30*, that is, the value of the mean for *Dist1SampSize30*.

$$\frac{vs^2}{\chi^2_{\alpha/2,v}} < \sigma^2 < \frac{vs^2}{\chi^2_{1-\alpha/2,v}}. \tag{81}$$

Proper understanding of Figure 10 is critical to the understanding of the analysis in this study. Unless a variance lies to the right of the 95% boundary, it cannot be concluded that it is larger than the variance it is being compared to, If the test variance lies to the left of the dotted line, the null hypothesis that it is less than or equal to the comparison variance is accepted. If the formulation of the null hypothesis was the C_{pk} values were less than 1.33, then the lower 5% boundary is applicable.

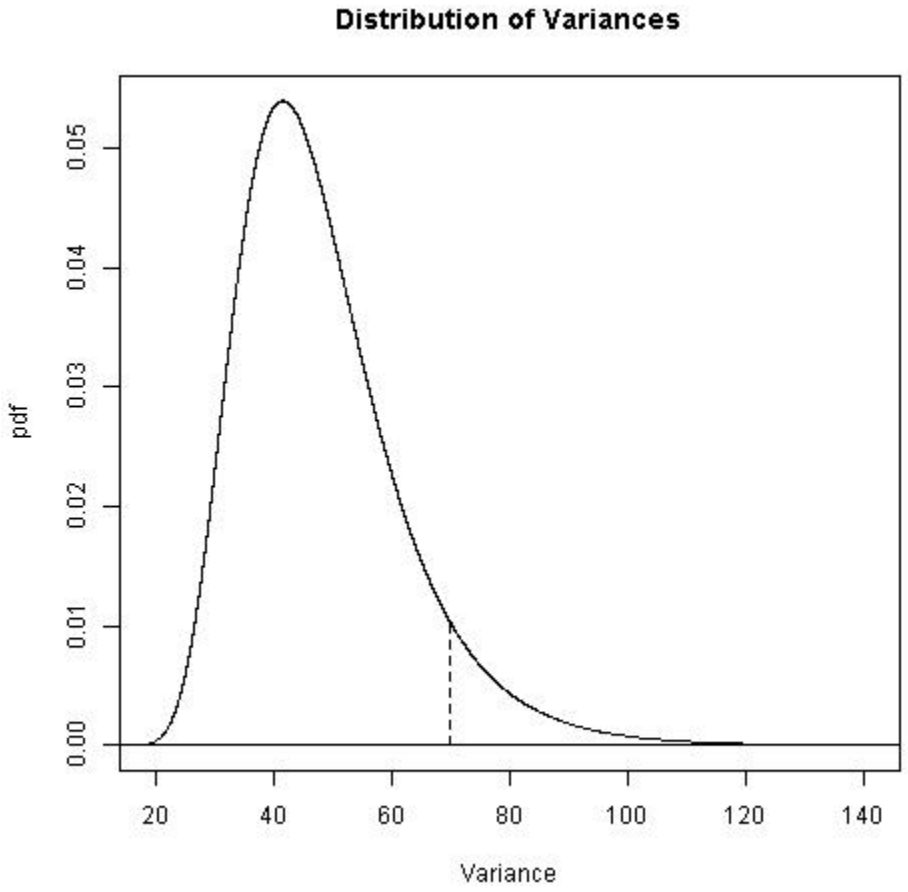

Figure 10. Distribution of the variance for Dist1SampSize30.

Taking the square root of *Equation* 81 gives the relationship for the standard deviation. Figure 11 contains an illustration of the transformation from variance to standard deviation with the 95% confidence value indicated by a dotted line.

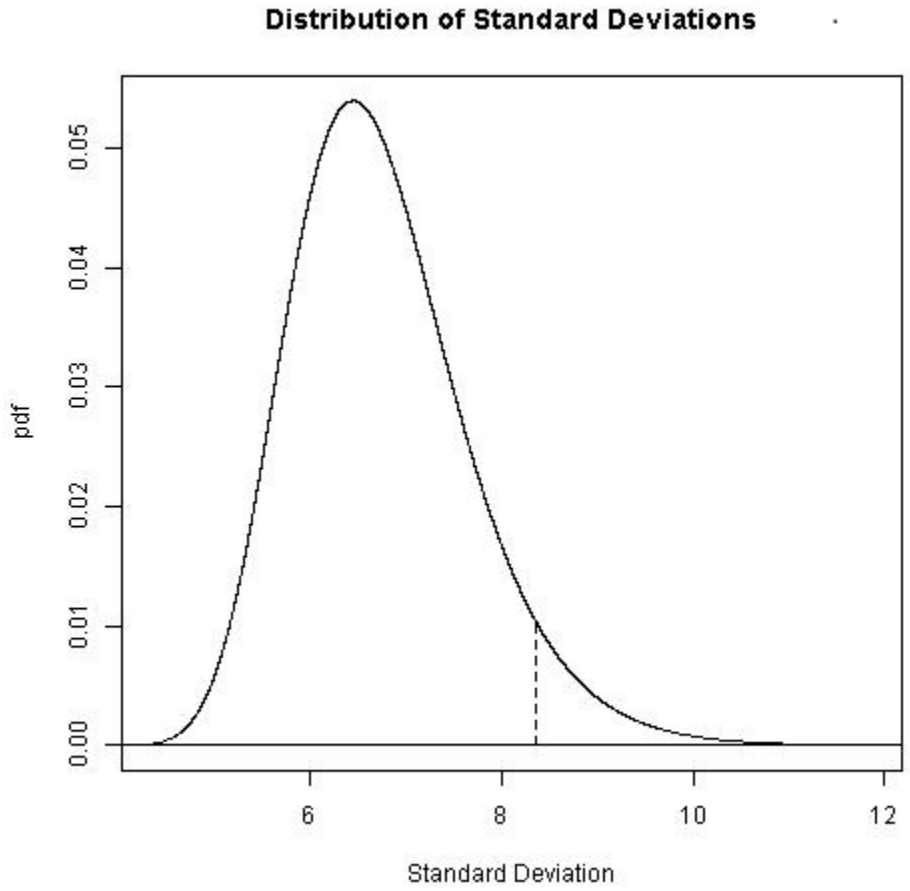

Figure 11. Distribution of the standard deviation.

The value of C_{pk} is a transformation of the value of the standard deviation. Further manipulation of *Equation* 81 gives the confidence limits for C_{pk}.

$$\frac{Interval}{3\sqrt{\frac{vs^2}{\chi^2_{\alpha/2,v}}}} < \frac{Interval}{3\sigma} < \frac{Interval}{3\sqrt{\frac{vs^2}{\chi^2_{1-\alpha/2,v}}}}, \tag{82}$$

where *Interval* is the smaller of the mean minus LTL, and UTL minus the mean. Figure 12 contains the results of the application of this transformation to the distribution of standard deviations. This figure represents the distribution of C_{pk} values if *Dist1SampSize30* did have a C_{pk} value of 1.33. The 5% boundary moves from the left to the right of the curve because the calculation of C_{pk} requires the use of the inverse of the standard deviation.

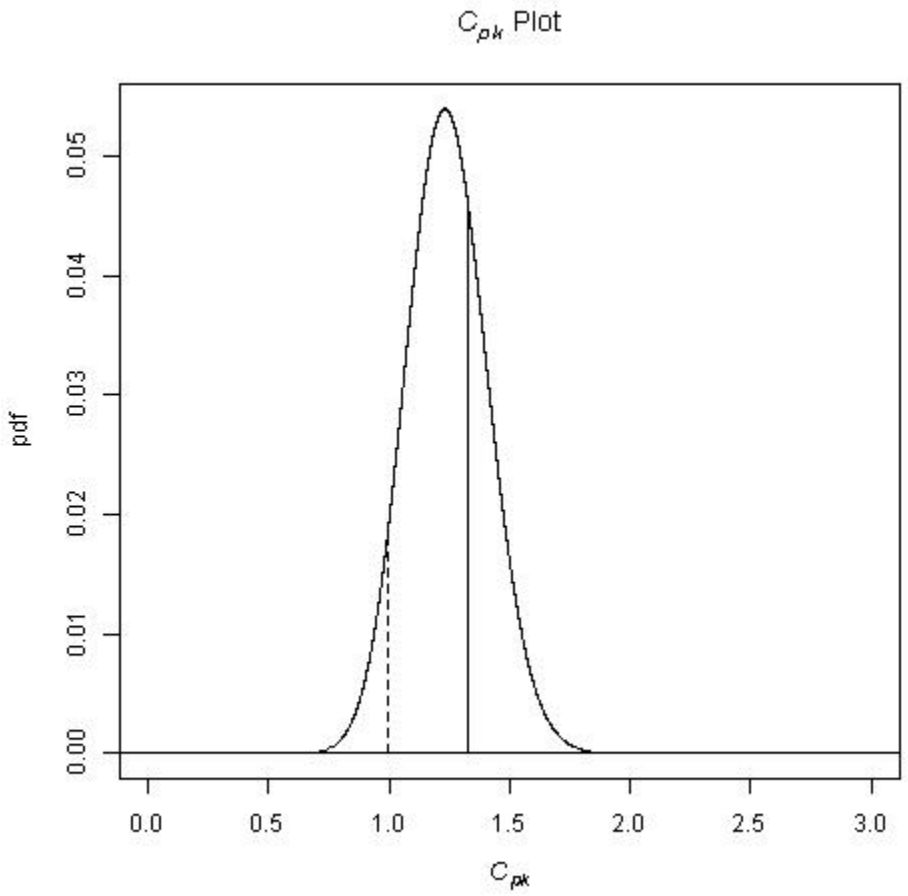

Figure 12. Distribution of the C_{pk} values.

With the relationship between chi squared values and C_{pk} values established, this relationship furnished the tool needed to evaluate if the C_{pk} values from the distribution

differed from the 1.33 value at a statistically significant level. The program used to generate the results for these calculations is in the file *DCode21_RQ1.r* in Appendix B. The power of the result is determined by using *Equation* 83 (Zar, 2014, p. 130),

$$1-\beta = P\left(\chi^2 \geq \chi^2_{\alpha,v}\frac{\sigma_0^2}{s^2}\right), \tag{83}$$

where v is the degrees of freedom, σ_0^2 is the variance tested against, and s^2 is the variance being tested.

The use of the term effect size may be misleading in this study because of the lack of any treatment. Assuming that whatever difference in parameters present in the simulated test results and those need to achieve a C_{pk} of 1.33 constitute the treatment, then the effect size can be calculated by adapting an equation from Grissom and Kim (2012, p. 63),

$$\Delta_G = \frac{C_{pk_R} - C_{pk_T}}{\sigma_T}. \tag{83}$$

In this equation, the C_{pk} values substitute for the mean values. The subscript, R, is the C_{pk} value compared to the standard of 1.33 based on the standard deviation σ_T.

I calculated the one tailed 95% confidence interval for the C_{pk} s that would result from the conditions necessary to achieve a value of 1.33. The last calculation was for the p value for the results from the simulated data. Table 7 contains the results.

Table 7

Results for Research Question 1. Nontransformable, Nonnormal Data in Bold Font

Distribution	Variance Test Statistic Chi Value	Tested Variance Chi Value	Variance 95% Confidence Limit	Equivalent C_{pk}
Dist1SampSize10	16.92	17.34	[39.87, ∞]	1.32
Dist1SampSize30	42.56	55.70	[50.98, ∞]	1.17
Dist1SampSizeC16	25.00	28.87	[45.03, ∞]	1.24
Dist1SampSize59	76.78	110.42	[56.06, ∞]	1.11
Dist2SampSize10	16.92	22.82	[42.44, ∞]	1.15
Dist2SampSize30	42.56	73.90	[54.72, ∞]	1.01
Dist2SampSizeC14	22.36	33.41	[47.00, ∞]	1.09
Dist2SampSize59	76.78	146.63	[60.61, ∞]	0.96
Dist3SampSize10	16.92	28.13	[60.50, ∞]	1.03
Dist3SampSize30	42.56	90.28	[78.02, ∞]	0.92
Dist3SampSizeC9	15.51	25.28	[59.26, ∞]	1.04
Dist3SampSize59	76.78	183.08	[87.38, ∞]	0.86
Dist4SampSize10	16.92	35.25	[60.89, ∞]	0.92
Dist4SampSize30	42.56	114.30	[79.38, ∞]	0.81
Dist4SampSizeC9	15.51	32.00	[60.05, ∞]	0.93
Dist4SampSize59	76.78	231.44	[88.91, ∞]	0.77

(continued)

Distribution	Variance Test Statistic Chi Value	Tested Variance Chi Value	Variance 95% Confidence Limit	Equivalent C_{pk}
Dist5SampSize10	16.92	10.32	[23.73, ∞]	1.71
Dist5SampSize30	42.56	33.94	[30.92, ∞]	1.49
Dist5SampSizeC48	64.00	54.65	[33.21, ∞]	1.44
Dist5SampSize59	76.78	68.08	[34.48, ∞]	1.42
Dist6SampSize10	16.92	14.18	[26.63, ∞]	1.46
Dist6SampSize30	42.56	46.91	[35.14, ∞]	1.27
Dist6SampSizeC34	47.40	53.60	[35.83, ∞]	1.25
Dist6SampSize59	76.78	94.49	[39.17, ∞]	1.20
Dist7SampSize10	16.92	28.17	[53.15, ∞]	1.03
Dist7SampSize30	42.56	90.84	[68.40, ∞]	0.91
Dist7SampSizeC10	16.92	27.94	[52.84, ∞]	1.04
Dist7SampSize59	**76.78**	**186.18**	**[77.15, ∞]**	**0.86**
Dist8SampSize10	16.92	33.17	[49.48, ∞]	0.95
Dist8SampSize30	42.56	107.95	[64.26, ∞]	0.84
Dist8SampSizeC11	18.31	37.02	[51.00, ∞]	0.94
Dist8SampSize59	**76.78**	**220.85**	**[72.43, ∞]**	**0.79**
Dist9SampSize10	16.92	20.19	[43.01, ∞]	1.22
Dist9SampSize30	42.56	65.09	[55.53, ∞]	1.08
Dist9SampSizeC13	21.03	26.88	[46.03, ∞]	1.18
Dist9SampSize59	76.78	133.93	[62.78, ∞]	1.01

(continued)

Distribution	Variance Test Statistic Chi Value	Tested Variance Chi Value	Variance 95% Confidence Limit	Equivalent C_{pk}
Dist10SampSize10	16.92	25.31	[43.24, ∞]	1.09
Dist10SampSize30	42.56	82.27	[56.33, ∞]	0.96
Dist10SampSizeC13	21.03	34.02	[46.57, ∞]	1.05
Dist10SampSize59	76.78	168.51	[63.44, ∞]	0.90
Dist11SampSize10	16.92	31.70	[59.90, ∞]	0.97
Dist11SampSize30	42.56	102.18	[76.89, ∞]	0.86
Dist11SampSizeC9	15.51	28.29	[58.05, ∞]	0.99
Dist11SampSize59	76.78	209.85	[86.46, ∞]	0.81
Dist12SampSize10	16.92	37.67	[56.27, ∞]	0.89
Dist12SampSize30	42.56	122.12	[72.64, ∞]	0.79
Dist12SampSizeC10	16.92	38.12	[56.47, ∞]	0.89
Dist12SampSize59	76.78	250.05	81.47	0.74

Table 8

Additional Results for Research Question 1. Nontransformable, Nonnormal Data in Bold Font

Distribution	Power	*p* Value	Effect Size
Dist1SampSize10	0.46	0.04	-0.06
Dist1SampSize30	0.81	0.00	-0.06
Dist1SampSizeC16	0.60	0.02	-0.06
Dist1SampSize59	0.96	0.00	-0.06

(continued)

Distribution	Power	*p* Value	Effect Size
Dist2SampSize10	0.67	0.01	-0.09
Dist2SampSize30	0.97	0.00	-0.09
Dist2SampSizeC14	0.80	0.00	-0.09
Dist2SampSize59	1.00	0.00	-0.09
Dist3SampSize10	0.80	0.00	-0.10
Dist3SampSize30	0.99	0.00	-0.09
Dist3SampSizeC9	0.77	0.00	-0.10
Dist3SampSize59	1.00	0.00	-0.10
Dist4SampSize10	0.89	0.00	-0.12
Dist4SampSize30	1.00	0.00	-0.12
Dist4SampSizeC9	0.87	0.00	-0.12
Dist4SampSize59	1.00	0.00	-0.12
Dist5SampSize10	0.10	0.33	-0.01
Dist5SampSize30	0.16	0.24	-0.02
Dist5SampSizeC48	0.20	0.21	-0.02
Dist5SampSize59	0.24	0.17	-0.02
Dist6SampSize10	0.29	0.12	-0.05
Dist6SampSize30	0.61	0.02	-0.05
Dist6SampSizeC34	0.66	0.01	-0.05
Dist6SampSize59	0.85	0.00	-0.05

(continued)

Distribution	Power	*p* Value	Effect Size
Dist7SampSize10	0.80	0.00	-0.10
Dist7SampSize30	0.99	0.00	-0.10
Dist7SampSizeC10	0.79	0.00	-0.10
Dist7SampSize59	**1.00**	**0.00**	**-0.10**
Dist8SampSize10	0.87	0.00	-0.13
Dist8SampSize30	1.00	0.00	-0.13
Dist8SampSizeC11	0.89	0.00	-0.13
Dist8SampSize59	**1.00**	**0.00**	**-0.13**
Dist9SampSize10	0.58	0.02	-0.07
Dist9SampSize30	0.92	0.00	-0.07
Dist9SampSizeC13	0.67	0.01	-0.07
Dist9SampSize59	1.00	0.00	-0.08
Dist10SampSize10	0.74	0.00	-0.10
Dist10SampSize30	0.99	0.00	-0.10
Dist10SampSizeC13	0.83	0.00	-0.10
Dist10SampSize59	1.00	0.00	-0.10
Dist11SampSize10	0.85	0.00	-0.11
Dist11SampSize30	1.00	0.00	-0.11
Dist11SampSizeC9	0.82	0.00	-0.11
Dist11SampSize59	1.00	0.00	-0.11

(continued)

Distribution	Power	*p* Value	Effect Size
Dist12SampSize10	0.91	0.00	-0.13
Dist12SampSize30	1.00	0.00	-0.14
Dist12SampSizeC10	0.91	0.00	-0.14
Dist12SampSize59	1.00	0.00	-0.14

Based on these results, I cannot reject the alternate hypothesis for the distributions listed in Table 8. I cannot reject the null hypothesis for the distributions listed in Table 9.

Table 9

Distribution Failing to Meet the Null Hypothesis Condition.

Dist1SampSize30	Dist4SampSize59	Dist7SampSize30	Dist10SampSize59
Dist1SampSize59	Dist5SampSize30	Dist7SampSize59	Dist11SampSize30
(Dist2SampSize30	Dist5SampSizeC48	Dist8SampSize30	Dist11SampSize59
Dist2SampSize59	Dist5SampSize59	Dist8SampSize59	Dist12SampSize30
Dist3SampSize30	Dist6SampSize30	Dist9SampSize30	Dist12SampSize59
Dist3SampSize59	Dist6SampSizeC34	Dist9SampSize59	
Dist4SampSize30	Dist6SampSize59	Dist10SampSize30	

Note. Reject null hypothesis for listed distributions.

Table 10

Distributions Meeting the Null Hypothesis Condition.

Dist1SampSize10	Dist4SampSize10	Dist8SampSize10	Dist11SampSize10
Dist1SampSizeC16	Dist4SampSizeC9	Dist8SampSizeC11	Dist11SampSizeC9
Dist2SampSize10	Dist5SampSize10	Dist9SampSize10	Dist12SampSize10
Dist2SampSizeC14	Dist6SampSize10	Dist9SampSizeC13	Dist12SampSizeC10
Dist3SampSize10	Dist7SampSize10	Dist10SampSize10	
Dist3SampSizeC9	Dist7SampSizeC10	Dist10SampSizeC13	

Note. Cannot reject null hypothesis for listed distributions.

Research Question 2

Research Question 2 was, do the values calculated from samples taken from a combined distribution exceed the actual values required to meet the standard. While stated as a hypothesis test, the answer comes from deterministic mathematical calculations. The premise of the calculation of C_{pk} is that the possibility of a process yielding a value outside of the tolerance limit is equal to the probability of a point falling more than four standard deviations from the mean in a standard normal distribution. Answering Research Question 2 determines how many points from the combination distributions meet this criteria.

Determining the answer to this question begins by calculating the limits of the distribution. The limits are the points laying beyond the intervals formed by taking the mean plus four standard deviations, and the mean minus four standard deviations. The number of points laying beyond these boundaries provides the answer to the question.

Verification of the results consisted of determining the empirical cumulative distribution function (ecdf) for each of the 12 combination distributions. Determining if any points of the ecdf fall outside of the probabilities of laying more than four standard deviations from the mean answers the question. If no values lie outside of the tolerance limits, then the process is still producing the desired output even if the C_{pk} value is below 1.33.

Table 10 contains the results of this analysis. The “0” and “-Inf” entries indicate that no values lay outside of the boundaries. Calculations are to four decimal points to increase the clarity of the results.

Table 11

Results from Research Question 2

Distribution	Low Points	High Points	Max Low Value	Min High Value	Proportion Low Value	Proportion High Value
comboDist1	0	0	-Inf	-Inf	0	0
comboDist2	0	0	-Inf	-Inf	0	0
comboDist3	0	0	-Inf	-Inf	0	0
comboDist4	0	0	-Inf	-Inf	0	0
comboDist5	0	0	-Inf	-Inf	0	0
comboDist6	0	0	-Inf	-Inf	0	0
comboDist7	0	0	-Inf	-Inf	0	0
comboDist8	0	0	-Inf	-Inf	0	0
comboDist9	0	0	-Inf	-Inf	0	0
comboDist10	0	0	-Inf	-Inf	0	0
comboDist11	0	0	-Inf	-Inf	0	0
comboDist12	0	0	-Inf	-Inf	0	0

Table 12

Additional Results from Research Question 2.

Distribution	Min Value	Max Value
comboDist1	75.0205	124.8790
comboDist2	75.0205	124.8790
comboDist3	75.0228	124.9458
comboDist4	75.0228	124.9458
comboDist5	81.0589	118.9411
comboDist6	76.7805	118.9411
comboDist7	75.0253	118.9411
comboDist8	75.0253	118.9411
comboDist9	75.0007	124.9886
comboDist10	75.0007	124.9886
comboDist11	75.0003	124.9886
comboDist12	75.0003	124.9886

The calculations indicate that the process is producing no output beyond the tolerance limits. Because the lateral distributions were trimmed, and the tolerance limit is at least four standard deviations from the mean of the underlying distributions, this is not an unexpected result. While not ideal results because of the lack of normality, engineering time spent "fixing" these processes could take second place to higher priorities.

Research Question 3

Research Question 3 was, do the data values from the lateral distributions, isolated from the underlying normal distributions, meet the industry standard. While stated as a hypothesis test, the answer comes from deterministic mathematical

calculations. Input to answer this question included the mean of each distribution, the population standard deviation, the standard deviation required to meet the 1.33 value (for comparison), and the C_{pk} calculated using the population standard deviation. Table 11 contains the results including those of the underlying normal distributions, nAtT and nAtO.

Table 13

Results from Research Question 3

Distribution	mean	Population Standard Deviation	Required Standard Deviation	C_{pk}
nAtT	100.11	5.05	6.22	1.64
nAtO	95.11	5.05	5.03	1.64
gammaLowerT0	92.13	9.98	4.28	0.83
gammaUpperT0	107.64	9.98	4.34	0.83
gammaLowerT	93.18	8.91	4.55	0.93
gammaUpperT	106.64	8.97	4.59	0.92
gammaLowerO0	82.13	9.98	1.78	0.83
gammaUpperO0	112.64	9.98	3.09	0.83
gammaLowerO	86.15	7.13	2.79	1.16
gammaUpperO	110.39	8.19	3.65	1.01
lognormalLowerT0	92.45	3.04	4.36	2.73
lognormalUpperT0	107.55	3.04	4.36	2.73
lognormalLowerT	92.45	3.04	4.36	2.73
lognormalUpperT	107.55	3.04	4.36	2.73
lognormalLowerO0	82.45	3.04	1.86	2.73
lognormalUpperO0	107.55	3.04	4.36	2.73
lognormalLowerO	82.50	2.97	1.88	2.79
lognormalUpperO	107.55	3.04	4.36	2.73
weibUpperT0	108.72	10.42	4.07	0.80
weibLowerT	89.21	7.36	3.55	1.13
weibUpperT	106.73	8.39	4.57	0.99

(continued)

Distribution	mean	Population Standard Deviation	Required Standard Deviation	C_{pk}
weibLowerO0	76.03	10.42	0.26	0.80
weibUpperO0	108.72	10.42	4.07	0.80
weibLowerO	83.12	5.20	2.03	1.60
weibUpperO	106.90	8.39	4.52	0.99
weibLowerT0	86.03	10.42	2.76	0.80

As expected, the underlying normal distributions exceeded the C_{pk} value of 1.33. Only the distribution based on the upper offset Weibull distribution, and those based on the lognormal distribution also exceeded the desired C_{pk} value of 1.33. The results from this research question and Research Question 2 require joint analysis because of the relationships between the data.

Summary

The purpose of this study was to develop a framework that evaluates the ability of C_{pk} to accurately measure medical device test data under a scenario where output data combines the effects of mixed production lots of components. Simulation with the R (2016) programing language generated 12 different fat tailed distributions. These distributions represented the test results of some product characteristic related to the performance or safety of a medical device. I formulated three different research questions to evaluate the performance of C_{pk} as a measure of process capability under the fat tailed distribution scenario.

The first research question examined the overall performance of C_{pk} as a tool to evaluate these processes. The analysis provided mixed results for this question. In 46% of

the cases, the null hypothesis that the C_{pk} value met or exceed the desired 1.33 value was not rejectable. In the other 54%, it was.

The second research question complimented the first. It provided a count of the number of data points, as a number and as a percentage, that laid outside of the boundaries needed for a C_{pk} of 1.33. Using separate methods for the count and the percentage calculation provided an additional level of validity for the results. The results showed that no points were outside of these boundaries.

The third research question, also a compliment to Research Question 1, was to compare the results of the calculated C_{pk} for each component distribution with the desired value of 1.33. The outcome evidenced mixed results. Both underlying normal distributions met the criterion as expected. One distribution, based on the Weibull lateral distribution, met the criterion. All of the distributions based on lognormal lateral distributions also met the criterion.

The results were more varied than expected at the outset of this research. Chapter 5 contains more analysis and further discussion of the results given in this chapter.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this empirical quantitative study was to develop a framework that evaluates the ability of a PCI to accurately measure medical device test data in a scenario where output data combines the effects of mixed production lots of components. The study was comparative in nature, involving an examination of the performance of the most commonly used PCI, C_{pk}, using simulated process data by calculating precise capabilities and then comparing these values with the results generated from nonnormal data adjusted indices.

The reasons for conducting the study are the ongoing FDA and ISO requirements to prove that processes to manufacture medical devices are under control. Often the tools used to determine control are PCIs, primarily C_{pk}. This index depends on the normality of the data under test, and statistically unsophisticated engineers may often use it in the absence of normality and unknowingly accept inaccurate results. Kane (1986) warned against this misuse in the article in which he first introduced this index.

The results of this study validated the concerns of Kane (1986) and Gunter (1989a, 1989b, 1989c, 1989d) regarding the application of this index to nonnormal data. Although the 12 different data sets generated through simulation for this study contained no points outside of the upper or lower tolerance limits, the calculated C_{pk} values generally failed to meet the industry standard of a value of 1.33 or greater. However, proper interpretation of the values would lead to a good understanding of the output of the process and the direction any corrective action should take. The research findings also contribute to filling a significant gap in the literature regarding the application of PCIs to

the medical device manufacturing field where mistakes can be life threatening if a procedure relies on a potentially defective device.

Interpretation of Findings

The design of the research questions in this study reflects an attempt to leverage the FDA requirement for lot traceability of the subcomponents used in the manufacture of a medical device to possibly overcome the limitations that nonnormal test results might impose on the ability of the capability index C_{pk} to monitor a medical device manufacturing process. The three research questions complement each other. The purpose of the first question was to determine the control status of several different processes as indicated by the calculation of C_{pk} for the processes calculated for different sample sizes. The goal of the second question was to evaluate the actual output from the processes to determine if defective output resulted. The objective of the last question was to evaluate the possibility that decomposing the output of a process by lot could yield a more accurate determination of process capability.

The results of this research both confirm and extend earlier findings. In the article introducing C_{pk}, Kane (1986) cautioned that a prerequisite for its application was normality of data. An early criticism of PCIs (Gunter, 1989a, a989b, 1989c, 1989d) held that this dependency on normality is a weakness of the technique. The results of this study generally show the process as being out of control, judged by the C_{pk} values calculated in the first research question, while the results of the second research question show that no defective product results from the processes. Relying on C_{pk} in this instance

could result in time consuming and expensive efforts to "fix" a process that is producing good product.

The probability of this situation occurring is increased by the fact that almost all of the distributions examined in this study would be accepted as normally distributed by the Anderson Darling test commonly used in normality testing. This result could lead a practitioner to believe that there was no problem with the data. Examination of the plots in Appendix A, or of the kurtosis results in Table 4, could indicate the soundness of this conclusion with the possible exception of combination distributions five, six, seven, and eight. These distributions, constructed using lognormal lateral distributions, are platykurtic.

An unexpected and potentially significant result of this research was that, in spite of their kurtosis values, combination distributions five and six had the highest C_{pk} values and are the only distributions that yielded a C_{pk} value higher than 1.33. Three of the four distributions with a value greater than 1.33 came from combination distribution 5 composed of the underlying normal and the offset lognormal distributions centered at the target of 100. Combination distribution 5 appears to exhibit some degree of bimodality, a characteristic even more pronounced in combination distribution 6. The values of the standard deviation distributed around two modes symmetrically distant from the mean could contribute to this result.

The only two combination distributions that were nonnormal by the Anderson Darling test, and nontransformable, had lognormal lateral distributions with sample sizes of 59. Figure 13 illustrates C_{pk} values by distribution. Because of the range of values for

the calculated sample sizes, this analysis only includes the standard sample sizes of 10, 30, and 59. Including the calculated samples sizes that ranged from seven to 48 could mask any pattern related to sample size.

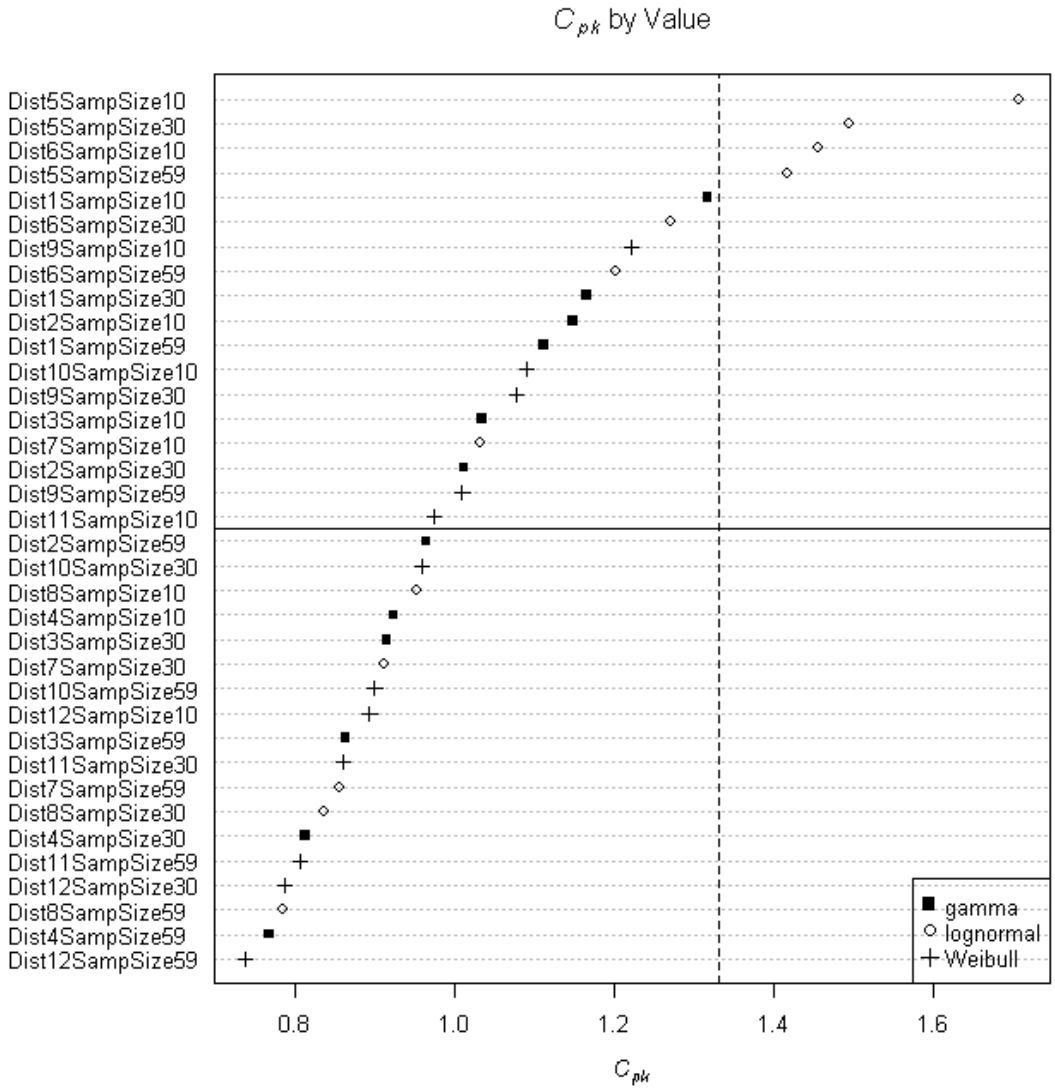

Figure 13. Combination distributions sorted by C_{pk} values.

What is particularly disturbing is that these values come from the lognormal lateral distributions. Ott (2000) noted that lognormal distributions often represent data

mixed from several distributions. This is the very problem this study addresses. Although the data in the study came from simulation, these results indicate that if mixed data comes from upstream processes, it may lead to deceptive results requiring that the steps in the study need application throughout the process rather than just where a clear mixing of lots occurs.

A significant finding follows from the theory of sampling and statistical quality control (Deming, 1950/1966, 1960/1990; Ott, 2000; Shewhart, 1931/2015, 1939/1986) is the critical importance of choosing an appropriate sample size when evaluating a process using C_{pk}. Figure 14 illustrates how the value of C_{pk} decreases with sample size. The panels in this plot have the C_{pk} values arranged from the smallest sample size, 10, on the bottom, to the larges sample size, 59, on the top. The C_{pk} values all migrate from right to left with increasing sample size.

This effect could be due to the presence of n - 1, the unbiased sample size, in the denominator of the equation used to calculate the sample standard deviation. As the value of n increases, the sample standard deviation would decrease. To balance this effect, the sample size calculation, based on the "… smallest value that is still considered to be practically significant …" (Mathews, 2010, p.16), assumes more importance.

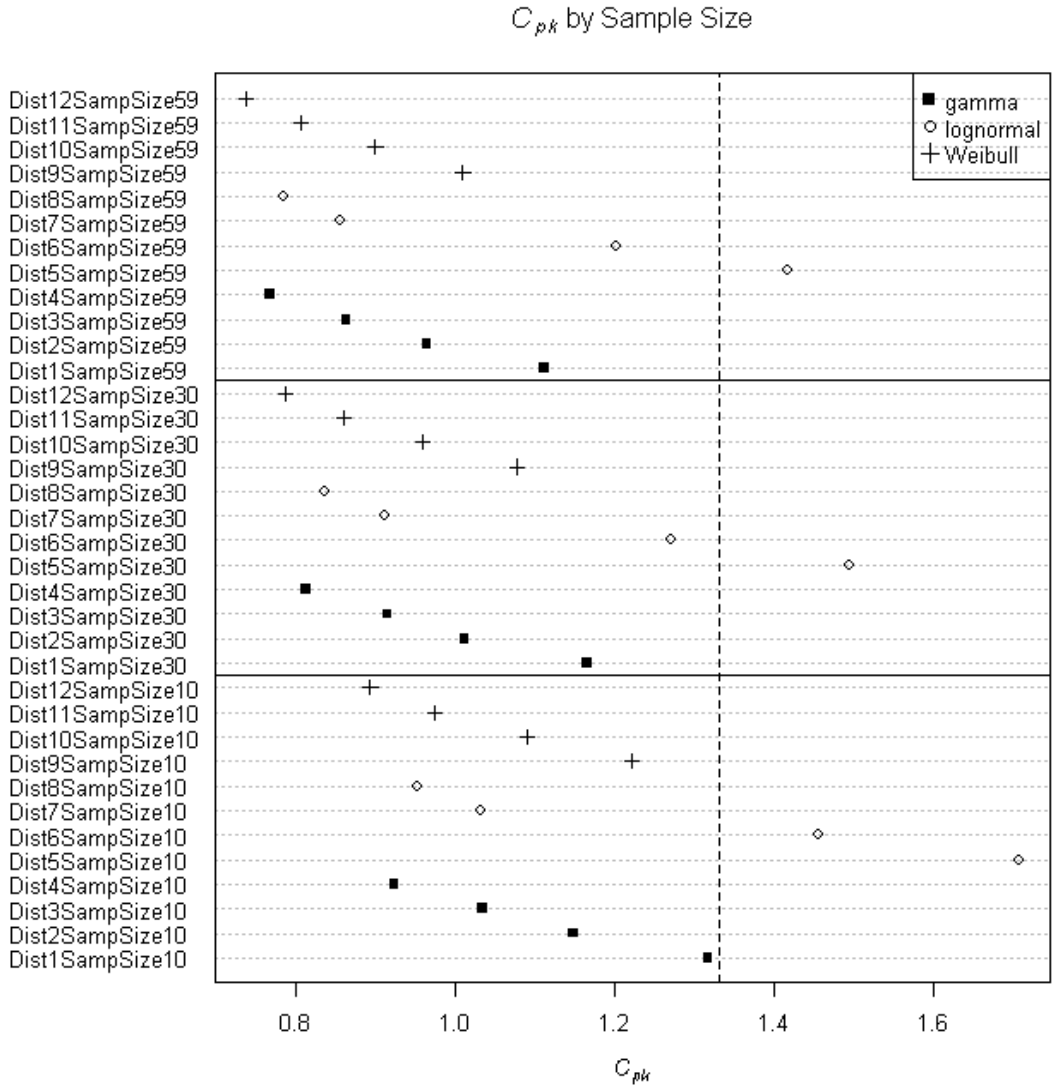

Figure 14. C_{pk} variation by sample size.

This research extends current theory in two different ways. First, other researchers have previously examined the behavior of C_{pk} using distributions other than normal distributions. English and Taylor (1993) used the triangular, uniform and truncated exponential distributions. Rivera et al. (1995) used the gamma, lognormal, and Weibull distributions. Tang and Than (1999) used the lognormal and Weibull distributions. Pal (2005) used the generalized lambda distribution, and Liu and Chen (2006) used the beta,

gamma, and Weibull distributions. A thorough search of the research could find more examples. This study considered the component distributions together rather than separately as might be the case with mixing of production lots as might be the case in real world production environments.

Another extension to theory came from research question three, with the different distributions considered individually rather than in combination. The focus in this research question was to look at the variance of the individual distributions. Other researchers introduced penalties or other considerations in their modifications of C_{pk} for missing target values. Kane's (1986) original work addressed target values. Chan et al. (1988a), and Taguchi (as cited in Boyles, 1991) followed this path. Pearn, Kotz, and Johnson (1992) with their introduction of C_{pm}.

In spite of the focus on the target, there are actually two factors that determine the value of $C_{pk,}$ the interval between the mean and the closest specification limit, and the standard deviation of the distribution of the samples. In some cases, the latter might be a more appropriate place to focus quality improvement efforts than the former. While the values of the C_{pk}s calculated from the overall combination distributions generally failed to meet the standards, the results from research question 3 indicated that many of the variances were sufficient to meet the 1.33 standard with the component distributions considered separately. In light of the results from Research Question 2, showing that no defective output resulted from any of the processes, this might also be a significant finding.

Limitations of the Study

This study had many limitations. First, the study design incorporated a total of four different, but specific, frequency distributions, an underlying normal distribution, a gamma distribution, a lognormal distribution, and a Weibull distribution. When considering the different parameters that define these distributions, it is obvious that there is an infinite number of possible distributions. Any variation in the parameter set used in R's (2016) random number generator would result in different distributions and possible different results. A practitioner applying the methods developed in this study would have to replicate many of the study steps to generate results. The study did not, nor was it intended to, generate a generalizable formula to evaluate process output.

Second, the overall symmetry of the combination distributions may have affected the results of the Anderson Darling normality tests. The creation of fat tailed distributions by doing this was by design to increase the likelihood of Anderson Darling failures. In actuality, the symmetry created the opposite effect. Nonsymmetrical distributions may have better suited this design goal.

Third, this study takes advantage of the lot traceability present in the manufacture of medical devices. While this closes a gap in the literature, it also limits the applicability of this study to industries where this condition is present, either formally, as is the case in FDA regulated industries, or informally where management has made the decision to incorporate such traceability on their own. Attempts to apply the methods used in this study to situations where lot traceability is not present would be difficult or perhaps impossible.

Recommendations

Recommendations for further study begin with addressing the limitations identified in the previous section. Rather than distributing the lateral distributions around the target mean of the underlying normal distribution by identical distances on either side, asymmetric placement represents an alternative. For example, if the left lateral distribution is one standard deviation from the target, the right distribution offset could equal two standard deviations.

With symmetric placement of the lateral distributions, the mean of the combination distribution stays centered around the target value. Each point to the left of the mean has a balancing value to the right of the mean. Consider the placement of the right lateral distribution two standard deviations rightward while the left lateral distribution remains offset left by only one standard deviation. Because of the lack of balance of values around the target, the mean would also shift to the right.

Along with the mean shift, the standard deviation would also grow larger due to the increase in dispersion around the mean. The interval between the mean and the UTL would decrease. It could be valuable to quantify the effect of the shortened interval paired with the larger standard deviation upon the value of C_{pk}. Would C_{pk} increase or decrease monotonically, or does it reach a maximum or minimum and then reverse direction?

A more extreme extension to this approach is eliminating one of the lateral distributions altogether. Again, the mean would shift in the direction of the lateral distribution, and the standard deviation would increase. Like the previous case, quantification and determining the characteristics of the effect would be interesting.

Both lateral distributions in this study had values above and below the tolerance limits trimmed. This represented a normal screening process of subcomponents occurring before the process combining the components from different lots. The trimming operation eliminated the more extreme values and decreased the value of C_{pk} compared to untrimmed components. A further extension to this research would be eliminating the trimming and measuring the effect this would have on the C_{pk} values. This approach could also use the asymmetric placement of the lateral values described previously.

In this research, I used the gamma, lognormal, and Weibull distributions as the lateral distributions. Other researchers have studied C_{pk} behavior using other distributions than these, although never in combination like the current study did. A follow-on study could substitute the triangular or uniform distributions (English & Taylor, 1993), the Generalized Lambda Distribution (Pal, 2005), or the beta distribution (Hosseinifard et al., 2009) for the lateral distributions used in this study. This could also use asymmetric placement, or only one lateral distribution.

In all of the extensions so far, I assumed the presence of the underlying normal distribution. How could the elimination of this distribution affect the results? For example, if two lateral distributions based on the gamma were studied, one tailing right by transformation, and the other naturally tailing right, would the effect be close enough to a normal distribution that a valid value of C_{pk} would result?

One distribution that was absent from the literature is the Gumbel distribution formulated to address extreme or rare cases (Gumbel, 1958/2004). While capability and control are goals for production processes, exceptions do occur. Studying the effect of

rare malfunctions in a process could be interesting. While rare, the effects from such an event could be very serious.

This study, by design, focused on one PCI, C_{pk}, because of its acceptance as the industry standard (Peña-Rodríguez, 2013). It is possible that one of the many other indices, developed since the introduction of C_{pk}, but not yet widely accepted, could provide more accurate results under this study's parameters. Further research could investigate the performance of these other indices under the original design or under one of the alternatives proposed in this section.

The current study addressed fat tailed processes. A last branch from the current study would examine the opposite situation, that is, thin or nonexistent tails. These often arise from a lack of granularity in the measurement process and can be difficult to interpret properly (Sleeper, 2007). Further research using the comparison of variance techniques used in this study might furnish a method to better analyze these situations.

Implications

The implications of this study affect three segments of society, the individual patient, the manufacturers of medical devices, and the regulatory agencies responsible for insuring that only safe products enter the marketplace. The individual patient relies on the knowledge and professionalism of medical device manufacturers when undergoing a treatment regimen that requires the use of a medical device. Patients enter treatment expecting to have to fight a disease or other medical condition and they rely on the quality of the devices used by medical professionals to help them in this effort. The results of this study could help insure that the patients receive treatment with devices that

perform as expected by making the device manufacturers more aware of the potential shortcomings or the tools they rely on to insure quality.

Manufacturers face a constant balancing act between cost, quality, and availability. A perfect device for the treatment of a condition with absolute reliability might be possible to make. But, if that device costs 5 million dollars to manufacture, and takes 100 skilled workers a year to make each one, it probably is not viable as a product (note that single use devices, unlike MRI machines or something similar that can help many patients over a long period of time, are the topic of this discussion). In a large scale production environment, manufacturers have come to rely on statistical methods to insure meeting quality and cost objectives. The regulatory agencies impose these requirements. Failing to meet these requirements can have serious consequences for both the patients and the manufacturers.

In January of 2017, the Department of Justice announced that Baxter Healthcare had agreed to pay more than $18 million to resolve issues that arose from its failure to follow Good Manufacturing Practices in one of its plants (U.S. Department of Justice, 2017). That sum paid to the government represents money that will not be invested in research and development, plant expansion, or passed on to medical facilities or stockholders. In addition, it is highly likely that the cost of this fine will eventually fall on the healthcare consumers, contributing to ever increasing healthcare costs. While this case involved pharmaceuticals rather than devices, it nevertheless indicates the possible consequences companies face for failing to control their processes.

By pointing out the weakness of a commonly used index, this study could help healthcare companies, and their engineers, better understand the implications of reading too much into the results of the applications of theses indices. It is not enough to apply a formula to results without a good understanding of the data to which the index is applied. Such understanding requires meaningful training.

Disseminating knowledge that one index alone cannot provide all of the needed answers regarding process capability could be an important outcome of this research. Additionally, a requirement that engineers in the medical industry have enough statistical training to ask the right questions and to know when to seek expert help when confronted with a situation beyond their statistical ability. Regulatory and standards organizations impose many requirements on medical manufacturers. Recognizing and requiring statistical literacy from those responsible for medical devices could be a worthy addition to those requirements.

Conclusions

From the beginning, practitioners and theoreticians knew the weaknesses of PCIs in dealing with nonnormal data. This study provided another verification of that weakness, but also extended that finding to more complicated cases consisting of multiple distributions combined into one. Perhaps the most significant finding of this study is that a low value of C_{pk} did not necessarily indicate the production of failing product. Had the data generated for this study represented an actual production situation, misinterpretation of the low values of C_{pk} could have prevented the release of a new product while the engineers fixed a product that met specifications.

Related to this was the performance of the Anderson Darling test statistic to detect nonnormality in fat tailed distributions. A conscientious, but statistically unsophisticated, engineer might accept the results of this test without question or further examination of the data. The conclusion could be that the process is under control when it exhibits an undesirable degree of spread.

I began this study with the hope of simplifying the use of C_{pk} to evaluate process capability by decomposing distributions by lot identity. Based on the conclusions of this study, it is apparent that a single capability index is unreliable when used alone to judge if a process is capable. Instead, the index must be part of a system that combines both statistical tools, for example, other tests for normality, variance comparison methods, and so forth, and intimate process knowledge to evaluate the output of a process. Until that occurs, overall process capability will still be difficult to evaluate properly.

References

Abel, V. (1989). Comment on: Control-limits of QC charts for skewed distributions using weighted-variance. *IEEE Transactions on Reliability*, *38*, 265. doi:10.1109/24.31118

Abu-Shawiesh, M. O. A. (2008). A simple robust control chart based on MAD. *Journal of Mathematics and Statistics*, *4*, 102-107. doi:10.3844/jmssp.2008.102.107

Adekeye, K. S. (2012). Modified simple robust control chart based on median absolute deviation. *International Journal of Statistics and Probability*, *1*, 91-95. doi:10.5539/ijsp.v1n2p91

Adekeye, K. S. (2013). PCIs based on median absolute deviation. *International Journal of Applied Science and Technology*, *3*, 43-48. Retrieved from http://www.ijastnet.com

Adekeye, K. S., & Azubuike, P. I. (2012). Derivation of the limits for control chart using the median absolute deviation for monitoring non-normal process. *Journal of Mathematics and Statistics*, *8*, 37-41. doi:10.3844/jmssp.2012.37.41

Ahmad, S., Abdollhian, M., & Zeephongsekul, P. (2008). Process capability estimation for non-normal quality characteristics: A comparison of Clements, Burr and Box-Cox methods. *Australian and New Zealand Industrial and Applied Mathematics*, *49*, C642-C665. Retrieved from http://journal.austms.org.au

Álvarez, E., Moya-Férnandez, P. J., Blanco-Encomienda, F. J., & Muñoz, J. F. (2015). Methodological insights for industrial quality control management: The impact of various estimators of the standard deviation on the process capability index.

Journal of King Saud University-Science, *27*, 271-277. doi:10.1016/j.jksus.2015.02.002

Auguie, B. (2012). gridExtra: Miscellaneous functions for "Grid" (R package version 2.21). Retrieved from http://CRAN.R-project.org/package=gridExtra

Bai, D. S., & Choi, I. S. (1995). X and R control charts for skewed populations. *Journal of Quality Technology*, *27*, 120-131. Retrieved from http:// asq.org/pub/jqt

Benson, E. D. (1994). *Statistical properties of a system of fourth-generation process capability indices* $C_{psk}(u, v, w)$. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9514495)

Besseris, G. (2014). Robust process capability performance: An interpretation of key indices from a nonparametric viewpoint. *TQM Journal*, *26*, 445-462. doi:10.1108/TQM-03-2013-0036

Bothe, D. R. (2001). *Measuring process capability*. Cedarburg, WI: Landmark Publishing.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*, 211-252. doi:10.2307/2984418

Boyles, R. A. (1991). The Taguchi capability index. *Journal of Quality Technology*, 23, 17-26. Retrieved from http://asq.org/pub/jqt

Brook, Q. (2006). *Six sigma and Minitab* (2nd ed.). Hampshire, UK: QSB Consulting.

Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, *13*, 215-232. doi:10.1214/aoms/1177731607

Burr, I. W. (1973). Parameters for a general system of distributions to match a grid of α_3 and α_4. *Communications in Statistics*, *2*, 1-21. doi:10.1080/03610927308827052

Burr, I. W., & Cislak, P. J. (1968). On a general system of distributions I. Its curve-shape characteristics II. The sample median. *Journal of the American Statistical Association*, *63*, 627-635. doi:10.1080/01621459.1968.11009281

Castagliola, P. (1996). Evaluation of non-normal process capability indices using Burr's distributions. *Quality Engineering*, *8*, 587-593. doi:10.1080/08982119608904669

Chan, L. K., Cheng, S. W., & Spiring, F. A. (1988a). A new measure of process capability: C_{pm}. *Journal of Quality Technology*, *20*, 160-175. Retrieved from http://asq.org

Chan, L. K., Cheng, S. W., & Spiring, F. A. (1988b, May). *A graphical technique for process capability*. Paper presented at the ASQC Quality Congress, Dallas, TX. Retrieved from http://asq.org

Chang, Y. S., & Bai, D. S. (2001). Control charts for positively-skewed populations with weighted standard deviations. *Quality and Reliability Engineering International*, *17*, 397-406. doi:10.100/qr2.427

Chang, Y. S., Choi, I. S., & Bai, D. S. (2002). Process capability indices for skewed populations. *Quality and Reliability Engineering International*, *18*, 383-393. doi:10.1002/qre.489

Choobineh, F., & Ballard, J. L. (1987). Control-limits of QC charts for skewed distributions using weighted-variance. *IEEE Transactions on Reliability*, *R-36*, 473-477. doi:10.1109/TR.1987.5222442

Choobineh, F., & Branting, D. (1986). A simple approximation for semivariance. *European Journal of Operational Research*, *27*, 364-370. doi:10.1016/0377-2217(86)90332-2

Clements, J. A. (1989). Process capability calculations for non-normal distributions. *Quality Progress*, 22 (9), 95-100. Retrieved from http://asq.org/qualityprogress

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312. doi:10.1037/0003-066X.45.12.1304

Cullen, C. G. (1972). *Matrices and linear transformations* (2nd ed.). Reading, MA: Addison-Wesley.

Czarski, A. (2008). Estimation of process capability indices in case of distribution unlike the normal one. *Archives of Materials Science and Engineering*, *34*, 39-42. Retrieved from http://www.archivesmse.org/

Dag, O., Asar, O., & Ilk, O. (2015). AID: Estimation of Box-Cox power transformation (R package version 1.5) [Computer software]. Retrieved from http://CRAN.R-project.org/package=AID

Deming, W. E. (1966). *Some theory of sampling*. New York, NY: Wiley (Original work published 1950).

Deming, W. E. (1990). *Sample design in business research* (Wiley Classics Library ed.). New York, NY: Wiley (Original work published 1960).

Department of Statistics, University of Minnesota, Morris (n.d.). *Karl Pearson*. Retrieved from http://mnstats.morris.umn.edu/introstat/history/w98/Pearson.html

Derya, K., & Canan, H. (2012). Control charts for skewed distributions: Weibull, gamma, and lognormal. *Metodolški Zvezki*, *9*, 95-106. Retrieved from http://mrvar.fdv.uni-lj.si

Dharmasena, L. S., Zeephongsekul, P., & Castagliola, P. (2010). Estimating process capability index Cpm using a bootstrap sequential sampling procedure. *Communications in Statistics: Simulation and Computation*, *39*, 1097-1110. doi:10.1080/03610911003790114

Dodson, B. (2006). *Weibull analysis handbook* (2nd ed.). Milwaukee, WI: Quality Press.

FDA, Maximum Penalty Amounts, 21 CFR § 17.2, (2011a [Revised 2014]).

FDA. (2009). Medical device quality systems manual A small entity compliance guide. Retrieved from http://www.fda.gov/medicaldevices/deviceregulationandguidance/postmarketrequirements/qualitysystemsregulations/medicaldevicequalitysystemsmanual/default.htm

FDA, Process Validation CFR 21 § 820, 75, (2011b [Revised 2013])

FDA. (2017). *Regulatory procedures manual*. Washington, DC: GPO Retrieved from https://www.fda.gov/ICECI/ComplianceManuals/RegulatoryProceduresManual/default.htm

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the bootstrap*. New York, NY: Chapman & Hall.

English, J. R., & Taylor, G. D. (1993). Process capability analysis—A robustness study. *International Journal of Production Research*, *31*, 1621-1635. doi:10.1080/00207549308956813

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, England: Sage.

Fisher, R. A. (1935/1971). *The design of experiments* (8th ed.). New York, NY: Hafner Publishing.

Franklin, L. A., & Gary, W. (1991). Bootstrap confidence interval estimates of Cpk: An introduction. *Communications in Statistics—Simulation and Computation*, *20*, 231-242. doi:10.1080/03610919108812950

French, T. E., & Vierck, C. J. (1972). *Engineering drawing and graphic technology* (11th ed.). New York, NY: McGraw-Hill.

George, F., & Ramachandran, K. M. (2011). Estimation of parameters of Johnson's system of distributions. *Journal of Modern Applied Statistical Methods*, *10*, 494-504. Retrieved from http://digitalcommons.wayne.edu/jmasm

Global Harmonization Task Force. (2004). Quality management systems—Process validation guidance (2nd ed.): Global Harmonization Task Force.

Goswami, A., & Dutta, H. N. (2013). Some studies on normal and non-normal process capability indices. *International Journal of Mathematics and Statistics Invention 1*, 31-40. Retrieved from http://www.ijmsi.org

Grant, E. L., & Leavenworth, R. S. (1980). *Statistical quality control* (5th ed.). New York, NY: McGraw-Hill.

Grau, D. (2011). Process yield and capability indices. *Communications in Statistics: Theory and Methods*, *40*, 2751-2772. doi:10.1080/03610926.2010.491594

Grau, D. (2012). Process yield, process centering, and capability indices for one-sided tolerance processes. *Quality Technology & Quantitative Management*, *9*, 153-170. doi:10.1504/IJQET.2011.043169

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York, NY: Routledge.

Gross, J., & Ligges, U. (2015). nortest: Tests for normality (R package version 1.0-4) [Computer software]. Retrieved from http://CRAN.R-project.org/package=nortest

Gruska, G. F., Mirkhani, K., & Lamberson, L. R. (1979). *Nonnormal data analysis*. Dearborn Heights, MI: Multiface Publishing Company.

Gryna, F. M. (1988). Manufacturing planning. In J. M. Juran & F. M. Gryna (Eds.), Juran's quality control handbook (4th ed., pp. 16.11-16.59). New York, NY: McGraw-Hill.

Gumbel, E. J. (2004). *Statistics of extremes*. Mineola, NY: Dover (Original work published 1958).

Guenther, W. C. (1965). *Concepts of statistical inference*. New York, NY: McGraw-Hill.

Gunter, B. H. (1989a). The use and abuse of Cpk. *Quality Progress*, *22*, 72-73. Retrieved from http://asq.org/qualityprogress

Gunter, B. H. (1989b). The use and abuse of Cpk, part 2. *Quality Progress*, *22*, 108-109. Retrieved from http://asq.org/qualityprogress

Gunter, B. H. (1989c). The use and abuse of Cpk, part 3. *Quality Progress*, *22*, 79-80. Retrieved from http://asq.org/qualityprogress

Gunter, B. H. (1989d). The use and abuse of Cpk, part 4. *Quality Progress*, *22*, 86-87. Retrieved from http://asq.org/qualityprogress

Halberstam, D. (1986). *The reckoning*. New York, NY: William Morrow and Company.

Han, C. P. (2007). Effect of testing normality on estimating process capability indices. *Quality Engineering*, *18*, 391-395. doi:10.1080/08982110600719514

Hogg, R. V. (1985). Statistical education for engineers: An initial task force report. *American Statistician, 39*, 168-175. doi:10.1080/00031305.1985.10479422

Hosseinifard, S. Z., Abbasi, B., Ahmad, S., & Abdollahian, M. (2009). A transformation technique to estimate the process capability index for non-normal processes. *International Journal of Advanced Manufacturing Technology*, *40*, 512-517. doi:10.1007/s00170-008-1376-x

Identification and Traceability, 21 C.F.R, § 820 (2016)

Industry standard. (n.d.) In *Collins English dictionary*. Retrieved from http://www.collinsdictionary.com/dictionary/english/industry-standard

International Organization for Standardization (2016). ISO 13485:2016 Medical devices—Quality management systems—Requirements for regulatory purposes. Geneva, Switzerland: ISO.

Jaca, C., Viles, E., Mateo, R., & Santo, J. (2012). Components of sustainable improvement systems: Theory and practice. *TQM Journal*, *24*, 142-154. doi:10.1108/17542731211215080

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*, 149-176. doi:10.1093/biomet/36.1-2.149

Juran, J. M. (1988). *Juran's quality control handbook* (4th ed.). New York, NY: McGraw-Hill.

Kabacoff, R. I. (2015). *R in action* (2nd ed.). Shelter Island, NY: Manning.

Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology, 18*, 41-52. Retrieved from http:// asq.org/pub/jqt

Kenett, R. S., & Zacks, S. (2012). Industrial statistics applications in the semiconductor industry: Some examples. *Applied Stochastic Models in Business and Industry*, *29*, 319-333. doi:10.1002/asmb.1959

Kenyon, G. N., & Sale, R. S. (2010, May). Calculating process capability index with limited information. Paper presented at the *POMS 21st Annual Conference*, Vancouver, Canada. Retrieved from http://www.pomsmeetings.org/ConfPapers /015/015-0330.pdf

Keselman, H. J., Othman, A. R., & Wilcox, R. R. (2013). Preliminary testing for Normality: Is this a good practice? *Journal of Applied Statistical Methods*, *12,* 1-19. Retrieved from http://digitalcommons.wayne.edu/jmasm/

Kettenring, J. R. (1995). What industry needs. *American Statistician, 49*, 2-4. doi:10.1080/00031305.1995.10476100

Komsta, L., & Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurtosis and related tests [Computer software] (R package version 0.14). Retrieved from https://CRAN.R-project.org/package=moments

Kotz, S., & Johnson, N. L. (1993). *Process capability indices*. London, UK: Chapman & Hall.

Kotz, S., Pearn, W. L., & Johnson, N. L. (1993). Some process capability indices are more reliable than one might think. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 42 55-62. doi:10.2307/2347409

Kotz, S., Johnson, N. L., Hubele, N. F., Spring, F., Cheng, S., Yeung, A., … Ramberg, J.S. (2002). Process capability indices—A review, 1992-2000. *Journal of Quality Technology, 34*, 2-53. Retrieved from http:// asq.org/pub/jqt

Kotz, S., Pearn, W. L., & Johnson, N. L. (1993). Some process capability indices are more reliable than one might think. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *42*, 55-62. doi:10.2307/2347409

Kotz, S., & Lovelace, C. R. (1998). *Process capability indices in theory and practice*. New York, NY: Arnold.

Kovářík, M., & Sarga, L. (2014). Process capability indices for non-normal data. *WSEAS Transactions on Business and Economic*s, *11*, 419-429. doi:10.1080/08982110008962614

Law, A. M. (2007). *Simulation modeling and analysis* (4th ed.). New York, NY: McGraw-Hill.

Lawson, J. (2014). *Design and analysis of experiments with R*. Boca Raton, FL: CRC Press.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. C. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics* (pp. 278-292). Stanford, CA: Stanford University Press.

Lipson, C., & Sheth, N. J. (1973). *Statistical design and analysis of engineering experiments*. New York, NY: McGraw-Hill.

Liu, P. H., & Chen, F. L. (2006). Process capability analysis of non-normal process data using the Burr XII distribution. *International Journal of Advanced Manufacturing Technology*, *27*, 975-984. doi:10.1007/s00170-004-2263-8

Maiti, S. S., Saha, M., & Nanda, A. K. (2010). On generalizing process capability indices. *Quality Technology & Quantitative Management*, *7*, 279-300. Retrieved from http://web.it.nctu.edu.tw

Mariappan, V., Gaonkar, R. S. P., Sakhardande, M., & Dhawalikar, M. (2012). An integrated statistical analysis for process improvement. *International Journal of System Assurance Engineering and Management*, *3*, 184-193. doi:10.1007/s13198-012-0109-6

Mathews, P. (2010). *Sample size calculations: Practical methods for engineers and scientists*. Fairport Harbor, OH: Mathews, Malnar and Bailey.

Mays, P. (2008). *Seal strength models for medical device trays.* (Doctoral dissertation, Texas A&M University). Retrieved from http://repository.tamu.edu/handle/1969.1/ETD-TAMU-2756

Minitab. (2010). *Minitab 16 statistical software* [Computer software]. State College, PA: Minitab.

Mondal, S., Ray, P., & Maiti, J. (2014). Modelling robustness for manufacturing processes: a critical review. *International Journal of Production Research*, *52*, 521-538. doi:10.1080/00207543.2013.837588

Montgomery, D. C. (2001). *Design and analysis of experiments* (5th ed.). New York, NY: John Wiley & Sons.

Montgomery, D. C. (2013). *Introduction to statistical quality control* (7th ed.). New York, NY: John Wiley and Sons.

Nadarajah, S., & Kotz, S. (2006). R programs for truncated distributions. *Journal of Statistical Software*, 16, 1-8. doi:10.18637/jss.v016.c02

Nagreha, N., & Parmar, M. (2011). Study of medical device recalls by FDA over last four years (2008-2011). *Applied Clinical Trials Online,* (Sep 20, 2011). Retrieved from http://www.appliedclinicaltrialsonline.com

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, *8*, 343-366. Retrieved from http://www.math.jhu.edu/

Niavarani, M. R., Noorossana, R., & Abbasi, B. (2012). Three new multivariate process capability indices. *Communications in Statistics: Theory and Methods*, *41*, 341-356. doi:10.1080/03610926.2010.526743

NIST/SEMATECH. (2015). *e-handbook of statistical methods* Retrieved from http://www.itl.nist.gov/div898/handbook/

Ott, E. R., Schilling, E. G., & Neubauer, D. V. (2000). *Process quality control* (3rd ed.). New York, NY: McGraw-Hill.

Pal, S. (2005). Evaluation of nonnormal process capability indices using generalized lambda distribution. *Quality Engineering*, *17*, 77-85. doi:10.1081/QEN-200028709

Pearn, W. L., & Chen, K. S. (1995). Estimating process capability indices for non-normal Pearsonian populations. *Quality and Reliability Engineering International*, *11*, 386-388. doi:10.1002/qre.4680110510

Pearn, W. L., & Chen, K. S. (1997). Capability indices for non-normal distributions with an application in electrolytic capacitor manufacturing. *Microelectronics Reliability*, *37*, 1853-1858. doi:10.1016/S0026-2714(97)00023-1

Pearn, W. L., Chen, K. S., & Lin, G. H. (1999). A generalization of Clements' method for non-normal Pearsonian processes with asymmetric tolerances. *International Journal of Quality & Reliability Management*, *15*, 507-521. doi:10.1108 /02656719910243618?journalCode=ijqrm

Pearn, W. L., & Kotz, S. (1994). Applications of Clements' method for calculating second- and third-generation process capability indices for non-normal Pearsonian populations. *Quality Engineering*, *7*, 139-145. doi:10.1080/08982119408918772

Pearn, W. L., & Kotz, S. (2006). *Encyclopedia and handbook of process capability indices*. Singapore: World Scientific Publishing.

Pearn, W. L., Kotz, S., & Johnson, N. L. (1992). Distributional and inferential properties of process capability indices. *Journal of Quality Technology*, *24*, 216-231. doi:10.1080/0266476042000280364

Pearn, W. L., Tai, Y. T., Hsiao, I. G., & Ao, Y. P. (2014). Approximately unbiased estimator for non-normal process capability index C_{Npk}. *Journal of Testing and Evaluation*, *42*, 1-10. doi:10.1520/JTE20130125

Pearn, W. L., Wu, C. W., & Wang, K. H. (2005). Capability measure for asymmetric tolerance non-normal processes applied to speaker driver manufacturing. *International Journal of Advanced Manufacturing Technology*, *25*, 506-515. doi:10.1007/s00170-003-1858-9

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London, Series A*, *185*, 71-110. doi:10.1098/rsta.1894.0003

Pearson, K. (1901). Mathematical contributions to the theory of evolution, X: Supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London, Series A*, *197*, 443-459. doi:10.1098/rsta.1901.0023

Pearson, K. (1916). Mathematical contribution to the theory of evolution, XIX. Second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, *216*, 429-457. doi:10.1098/rsta.1916.0009

Peña-Rodríguez, M. E. (2013). *Statistical process control for the FDA-regulated industry*. Milwaukee, WI: American Society for Quality.

Penn State University. (2015). Sampling distribution of sample variance, Retrieved from https://onlinecourses.science.psu.edu/stat414/print/book/export/html/174

Pluta, P. L. (2014). Statistical tools for development and control of pharmaceutical processes: Statistics in the FDA process validation guidance | IVT. *Journal of Validation Technology*, *20*. Retrieved from http://www.ivtnetwork.com

Polansky, A. M., Chou, Y.-M., & Mason, R. L. (1998). Estimating process capability indices for a truncated distribution. *Quality Engineering*, *11*, 257-265. doi:10.1080/08982119808919237

Porter, L. J., & Oakland, J. S. (1991). Process capability indices—An overview of theory and practice. *Quality and Reliability Engineering International*, *7*, 437-448. doi:10.1002/qre.4680070602

Price, B., & Price, K. (1993). A methodology to estimate the sampling variability of the capability index, C*pk*. *Quality Engineering*, *5*, 527-544. doi:10.1080/08982119308918995

Pyzdek, T. (1992). Process capability analysis using personal computers. *Quality Engineering*, *4*, 419-440. doi:10.1080/08982119208918923

R Development Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.0) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rinck, P. A. (2016). *Magnetic resonance in medicine. The basic textbook of the European Magnetic Resonance Forum*. Retrieved from http://www.magnetic-resonance.org/

Rivera, L. A. R., Hubele, N. F., & Lawrence, F. P. (1995). Cpk index estimation using data transformation. *Computers & Industrial Engineering*, *29*, 55-58. doi:10.1016/0360-8352(95)00045-3

Romeu, J. L. (2013). Professional organizations and the learning of statistics after college. *Revista Empresarial Inter Metro / Inter Metro Business Journal, 9*, 20-37. Retrieved from http://ceajournal.metro.inter.edu

Safdar, S., & Ahmed, E. (2014). Process capability indices for shape parameter of Weibull distribution. *Open Journal of Statistics*, *4*, 207-219. doi:10.4236/ojs.2014.43020

Salazar-Alvarez, M.-I., & Temblador, C. (2012, May). Review of tools for capability analysis for non normal process behavior. Paper presented at the *Industrial and Systems Engineering Research Conference*, Orlando, FL.

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Boca Raton, FL: Chapman & Hall.

Shewhart, W. A. (2015). *Economic control of quality of manufactured product*. Mansfield, CT: Martino Publishing. (Original work published 1931).

Shewhart, W. A. (1986). *Statistical method from the viewpoint of quality control*. New York, NY: Dover (Original work published 1939).

Sleeper, A. (2007). *Six sigma distribution modeling*. New York, NY: McGraw-Hill.

Slifker, J. F., & Shapiro, S. S. (1980). The Johnson system: Selection and parameter estimation. *Technometrics*, *22*, 239-246. doi:10.2307/1268463

Spiring, F., Leung, B., Cheng, S., & Yeung, A. (2003). A bibliography of process capability papers. *Quality and Reliability Engineering International*, *19*, 445-460. doi:10.1002/qre.538

Sullivan, L. P. (1984). Reducing variability: A new approach to quality. *Quality Progress*, *17*, 15-21. Retrieved from http://asq.org

Tang, L. C., & Than, S. E. (1999). Computing process capability indices for non-normal data: A review and comparative study. *Quality and Reliability Engineering International*, *15*, 339-353. doi:10.1002/(SICI)1099-1638(199909/10)15:5 <339::AID-QRE259>3.0.CO;2-A

Taylor, F. W. (1998). *The principles of scientific management*. Mineola, NY: Dover. (Original work published 1911).

Tobias, P. A., & Trindade, D. C. (2012). *Applied reliability* (3rd ed.). Boca Raton, FL: CRC Press.

Tong, L. I., Chen, H. T., & Tai, Y. F. (2008). Constructing BCa bootstrap confidence interval for the difference between two non-normal process capability indices C_{Npmk}. *Quality Engineering*, *20*, 209-220. doi:10.1080/08982110701724348

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. C. Madow & H. B. Mann (Eds.), *Contributions to probability and statistics* (pp. 448-485). Stanford, CA: Stanford University Press.

U.S. Department of Health and Human Services, Food and Drug Administration [FDA]. (2011). *Process validation: General principles and practices*. Washington, DC: U.S. Department of Health and Human Services. Retrieved from http://www.fda.gov/downloads/Drugs/Guidances/UCM070335.pdf

U.S. Department of Health and Human Services, Food and Drug Administration [FDA]. (2013). *Medical device recall report: FY2003 to FY2012*. Washington, DC: U.S. Department of Health and Human Services. Retrieved from http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHTransparency/UCM388442.pdf

U.S. Department of Justice. (2017). *Baxter Healthcare Corporation to pay more than $18 million to resolve criminal and civil liability relating to sterile products*. Retrieved from https://www.justice.gov/opa/pr/baxter-healthcare-corporation-pay-more-18-million-resolve-criminal-and-civil-liability

University of Bristol, School of Mathematics, Institute of Pure Mathematics. Retrieved from Institute of Pure Mathematics website: http://www.bristol.ac.uk/maths/research/pure/

von Bertalanffy, L. (1969). *General systems theory*. New York, NY: George Braziller.

Vännman, K. (1995). A unified approach to capability indices. *Statistica Sinica*, *5*, 805-820. Retrieved from http://www3.stat.sinica.edu.tw

Vännman, K. (1997). A general class of capability indices in the case of asymmetric tolerances. *Communications in Statistics—Theory and Method*, *26*, 2049-2072. doi:10.1080/03610929708832031

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer.

Wackerly, D. D., Mendenhall III, W., & Schaeffer, R. L. (2002). *Mathematical statistics with applications* (6th ed.). Pacific Cove, CA: Duxbury.

Warner, R. M. (2008). *Applied statistics*. Thousand Oaks, CA: Sage.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 00-00. doi:10.1080/00031305.2016.1154108

Wheeler, D. J. (2000a). *Beyond capability confusion* (2nd ed.). Knoxville, TN: SPC Press.

Wheeler, D. J. (2000b). *Normality and the process behavior chart*. Knoxville, TN: SPC Press.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.

Wilcox, R. R. (2012). *Introduction to robust estimation & hypothesis testing*. Waltham, MA: Elsevier.

Wright, P. A. (1995). A process capability index sensitive to skewness. *Journal of Statistical Computation and Simulation*, *52*, 195-203. doi:10.1080/00949659508811673

Wu, C. H., Lin, S. J., Yang, D. L., & Pearn, W. L. (2014). Box-Cox transformation approach for evaluating non-normal processes capability based on the C_{pk} index. *Journal of Testing and Evaluation*, *42*, 949-961. doi:10.1520/JTE20130080

Wu, H. H. (1998). *Development of a weighted variance approach to modify current process capability indices*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9834325).

Wu, H. H. (2001). Performance problems of families of non-normal process capability indices. *Quality Engineering*, *13*, 383-388. doi:10.1080/08982110108918666

Wu, H. H., Swain, J. J., Farrington, P. A., & Messimer, S. L. (1999). A weighted variance capability index for general non-normal processes. *Quality and Reliability Engineering International*, *15*, 397-402. doi:10.1002/(SICI)1099-1638(199909/10)15:5<397::AID-QRE274>3.0.CO;2-N

Ye, R., Ma, T., & Wang, S. (2011). Generalized confidence intervals for the process capability indices in general random effect model with balanced data. *Statistical Papers*, *52*, 153-169. doi:10.1007/s00362-009-0216-x

Yum, B. J., & Kim, K. W. (2011). A bibliography of the literature on process capability indices: 2000-2009. *Quality and Reliability Engineering International*, *27*, 251-268. doi:10.1002/qre.1115

Zar, J. H. (2014). *Biostatistical analysis* (5th ed.). Noida, India: Pearson Indian Education Services.

Zimmer, W. J., & Burr, I. W. (1963). Variables sampling plans based on nonnormal populations. *Industrial Quality Control*, *20*, 18-36. Retrieved from http://asq.org/qic

Appendix A: Histograms of Combination Distributions

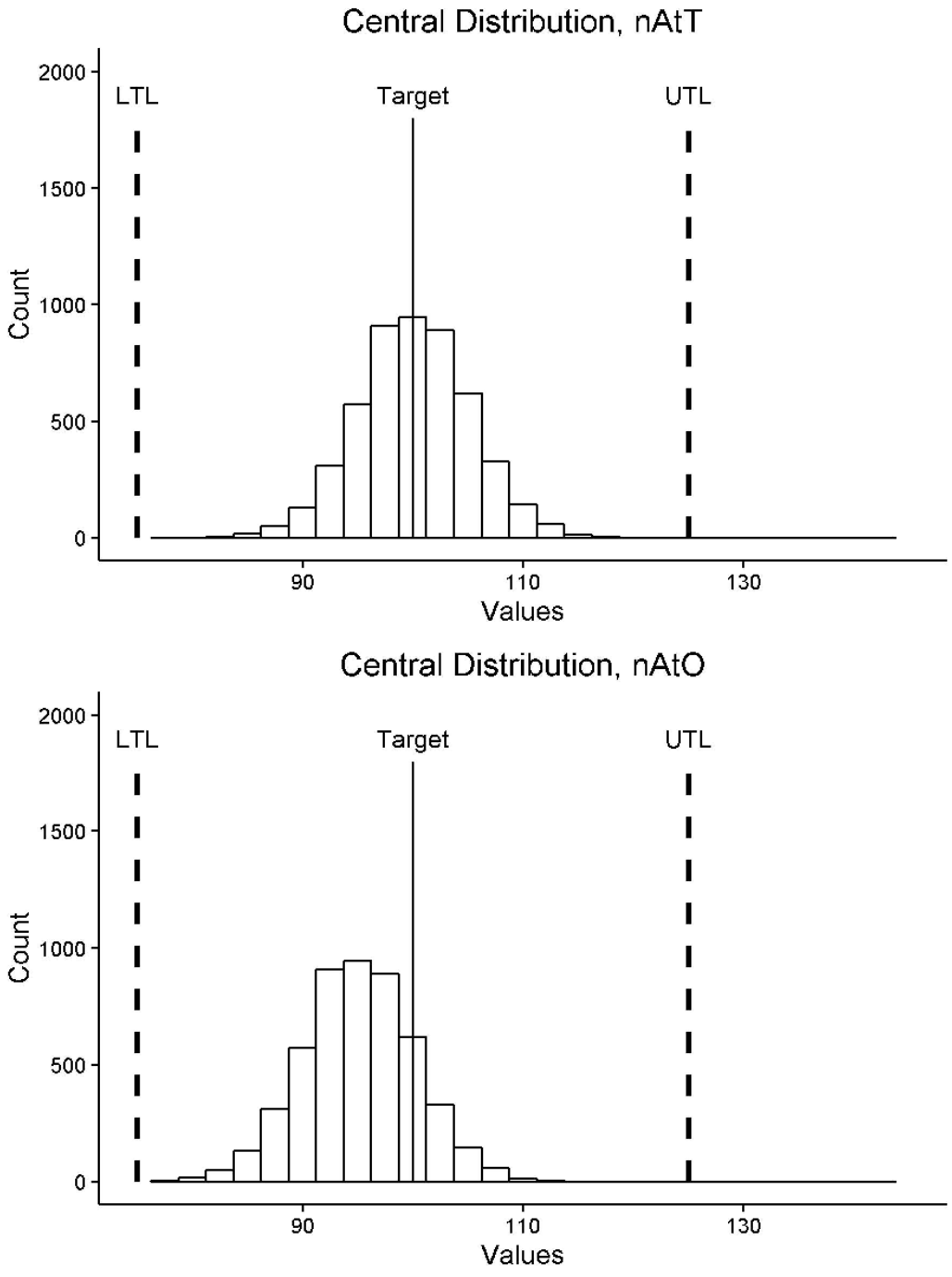

Figure A1. Underlying normal distributions.

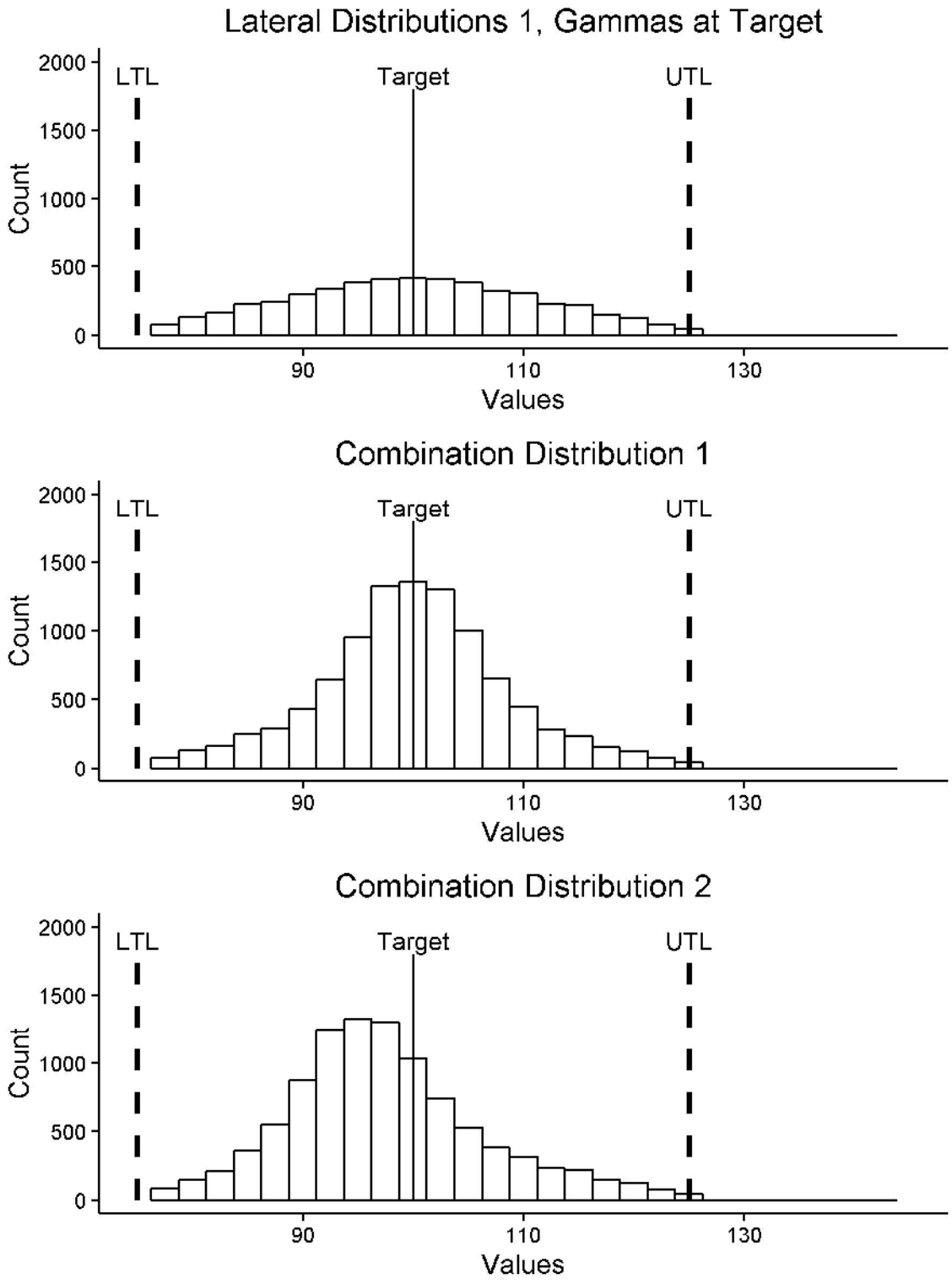

Figure A2. Combination distributions formed with centered gamma distributions.

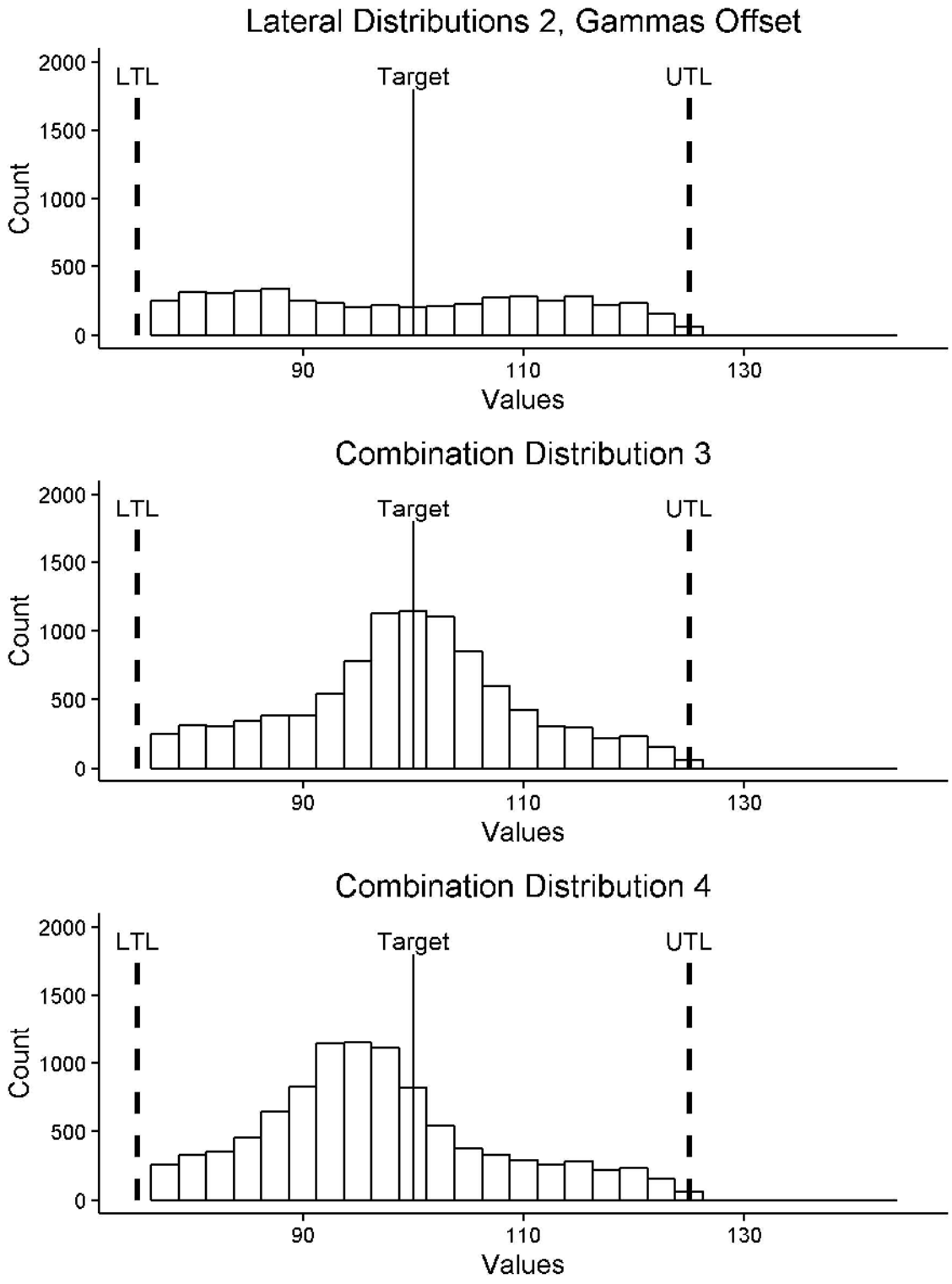

Figure A3. Combination distributions formed with offset gamma distributions.

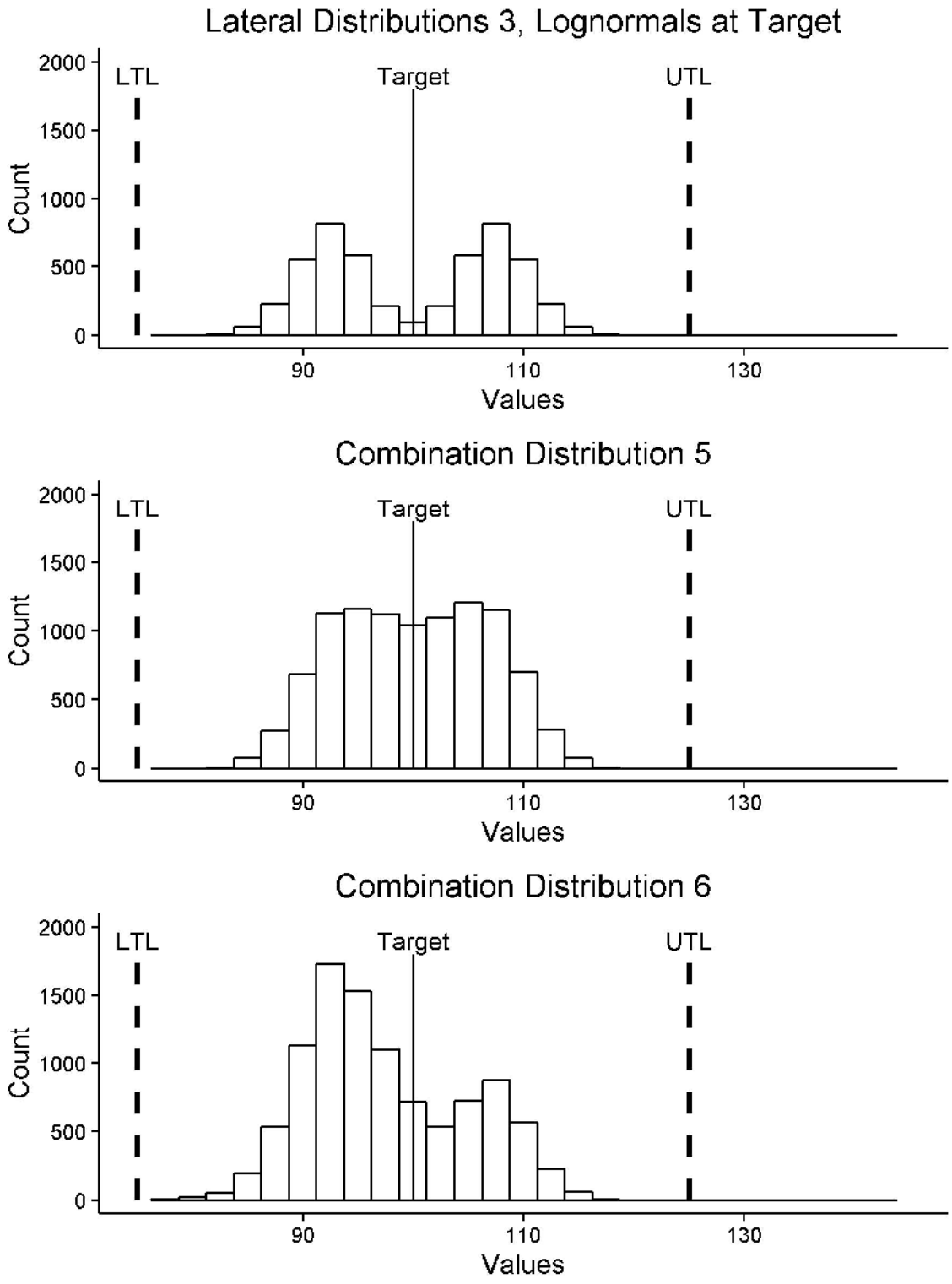

Figure A4. Combination distributions formed with centered lognormal distributions.

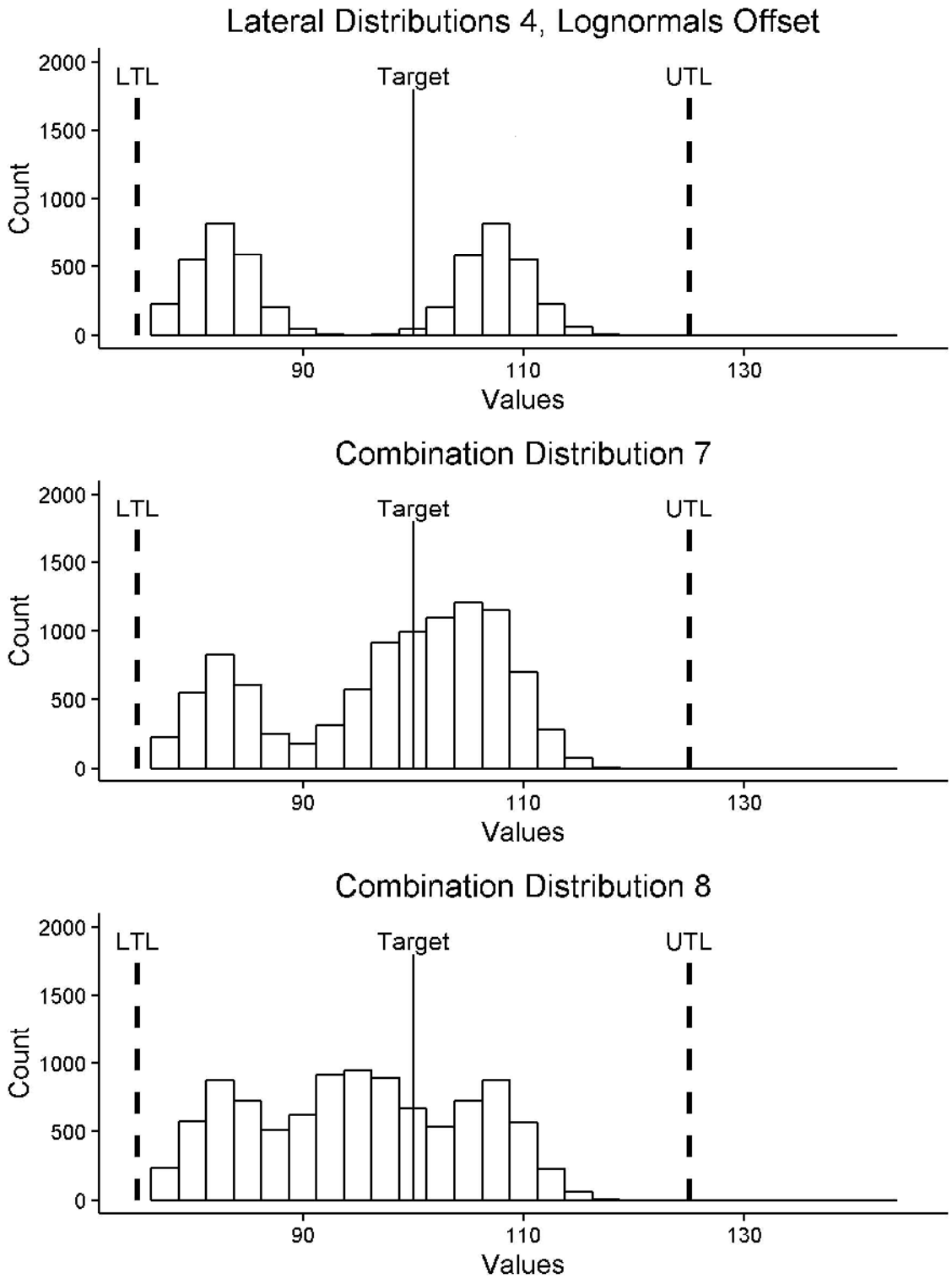

Figure A5. Combination distributions formed with offset lognormal distributions.

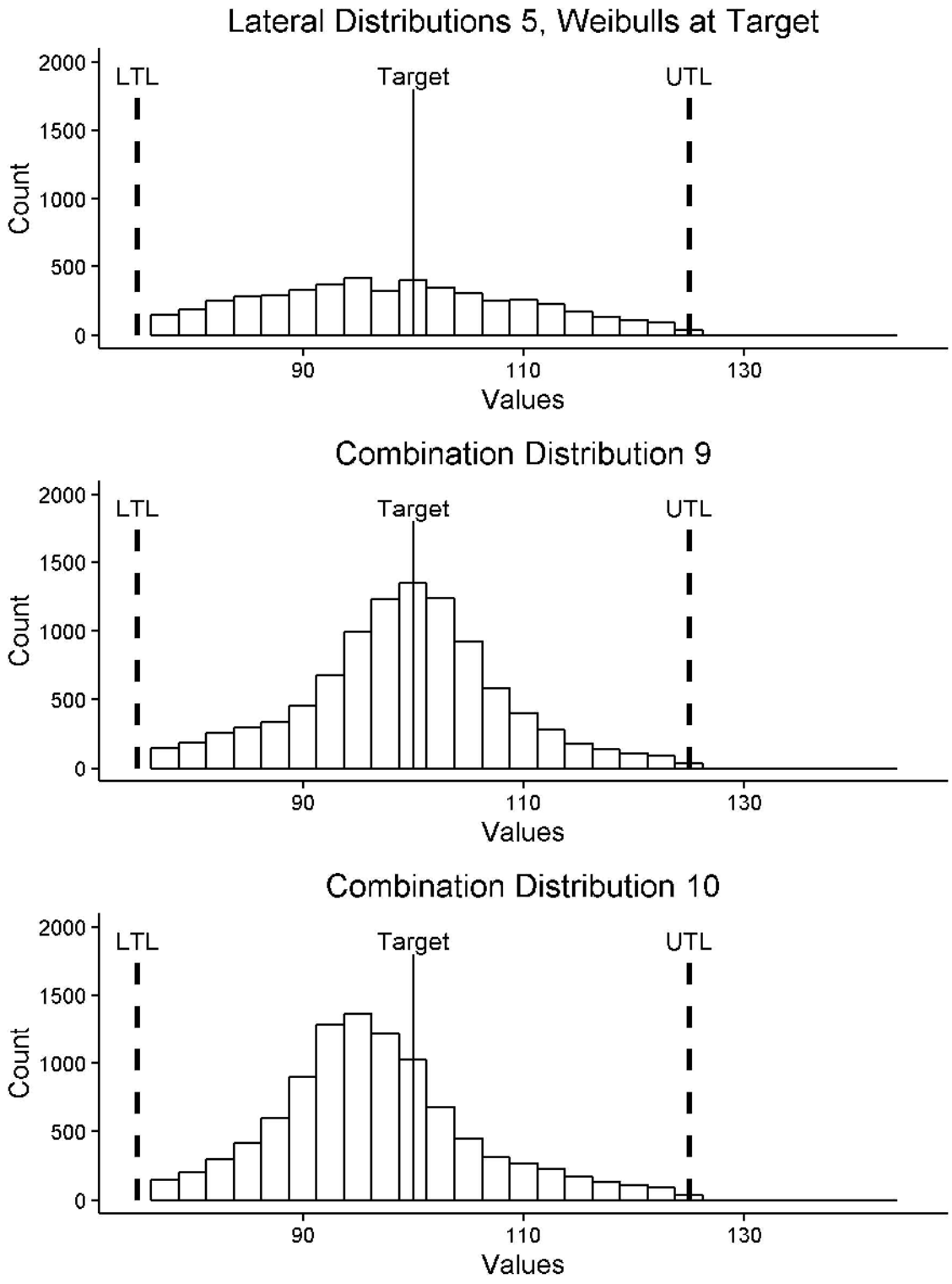

Figure A6. Combination distributions formed with centered Weibull distributions.

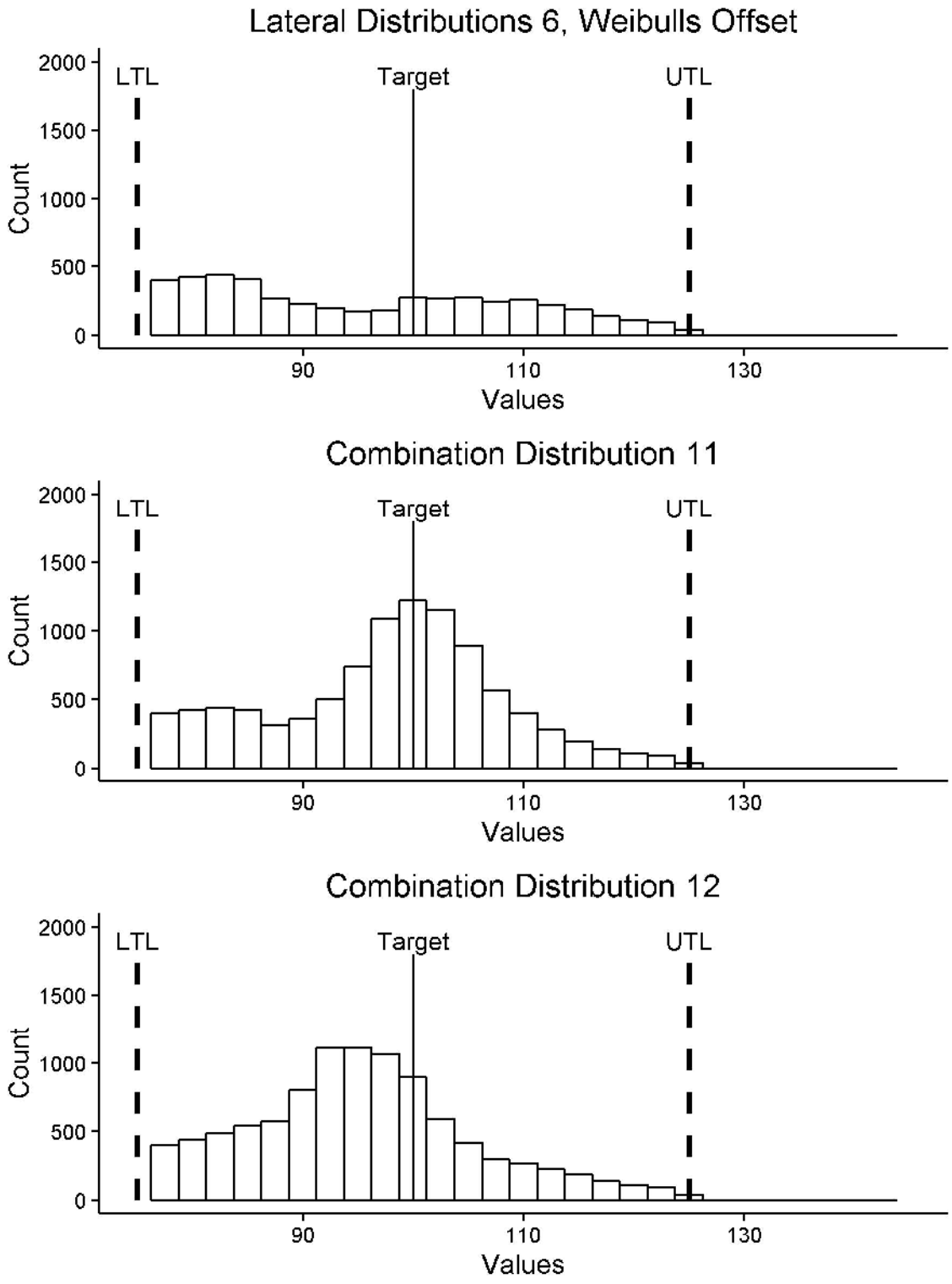

Figure A7. Combination distributions formed with centered Weibull distributions.

Appendix B: R Code

The embedded object on this page contains the R code used to create the figures used and do the analysis described in this study.

KwiecienJW_R_Code_02262017.docx

I will also place the R code on GitHub at https://github.com/JimKw1091/JWK-Dissert.git.