Statistical and Causal Analysis of Inbound Supply Chain Inefficiencies

By

Tyler Haley
B.S. Industrial Engineering, California Polytechnic State University, 2006

And

Hossein Nasseri
M.S. Design and Construction Integration, Stanford University, 2010
B.S. Civil and Environmental Engineering, University of California, Los Angeles, 2008

Submitted to the Engineering Systems Division in Partial Fulfillment of the
Requirements for the Degree of

Masters of Engineering in Logistics
at the
Massachusetts Institute of Technology
June 2014

Signature redacted

Signature of Author..
Masters of Engineering in Logistics Program, Engineering Systems Division

Signature redacted
May 9, 2014

Signature of Author..
Masters of Engineering in Logistics Program, Engineering Systems Division
May 9, 2014

Signature redacted

Certified By......................   ....
Dr. Tony Craig
Postdoctoral Associate
MIT Center for Transportation and Logistics
Thesis Supervisor

Signature redacted

Accepted By..............   ....
Dr. Yossi Sheffi
Director, MIT Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering Department

ACKNOWLEDGEMENT

Statistical and Causal Analysis of Inbound Supply Chain Inefficiencies

By

Tyler Haley and Hossein Nasseri

Submitted to the Engineering Systems Division in Partial Fulfillment of the
Requirements for the Degree of

Masters of Engineering in Logistics
at the
Massachusetts Institute of Technology

ABSTRACT

Given the importance of operational inefficiencies and their negative impact on the bottom line in today's competitive economy, CVS/pharmacy is very interested in implementing operational improvement initiatives across its inbound supply chain to minimize the number of non-value-added activities. Undertaking such efforts requires collaboration amongst all trade partners and a systematic approach in measuring the important performance metrics. Currently there is not a single procedure that defines the necessary metrics and the analytical tools necessary for identifying improvement opportunities. Leveraging research from the manufacturing industry, specifically supplier certification and statistical process control, this thesis aims to develop a comprehensive methodology for analyzing, monitoring and improving the operational performance of the retail industry supply chain.

In this thesis, through an innovative approach to perfect order performance measurement combined with the practical application of statistical analysis methods, a complete supplier evaluation process is established. Further, by utilizing statistical sampling and based on the evaluation results, an inspection plan is provided that allows for accurate monitoring of ongoing processes with a reduction in inspection efforts. Finally through introduction of statistical process control models and root cause analysis, a complete procedure is developed for continuous evaluation and improvement, leading to efficiency gains and cost savings across the entire inbound supply chain.

Thesis supervisor: Dr. Tony Craig
Title: Postdoctoral Associate

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This thesis establishes the baseline requirements for an inbound supply chain certification program and evaluates the current state of affairs at CVS/pharmacy's inbound supply chain network. The following section discusses the overall topic, research question, underlying motivations and the thesis structure and provides a background on our sponsor, CVS/Pharmacy.

1.1 Topic Overview

Supply chain improvements offer unique economic and operational competitive advantages to all the parties involved. However functional barriers and lack of a systematic approach in realizing and sharing these benefits often prevent the business-to-business (B2B) trade partners from capitalizing on these advantages. We believe that eliminating these inefficiencies in inbound supply chain, the B2B part of the chain, is more achievable than in the business-to-customer (B2C) part of supply chain due to the relatively higher stability and greater control power. A certification program can enable the B2B partners to realize these benefits.

Supplier certification/evaluation and statistical process control (SPC) historically have been used to improve manufacturing quality control. The improvement is achieved through a systematic quality assurance approach utilizing statistical control charts and historical performance record as predictor of near future state of the affairs. In addition, a certification program establishes a level of trust and cooperation among trade partners that allow effective data sharing and constructive discussion. Applying SPC in retail environment in conjunction with a certification program can eliminate non-value-add activities and increase efficiency.

7

1.2 Motivation

The inbound operations encompass all the electronic and manual activities from ordering and procurement to receiving and quality control at CVS/pharmacy distribution centers (DCs) which involve the suppliers, the carriers and the retailers. As errors occur in the process, additional non-value-add activities are required for documentation and correction of the errors; these activities usually require collaboration across the entire chain and interfere with the routine procedures. Reducing the number of supply chain activities and increasing the quality standards result in the elimination of non-value-add activities and efficiency improvement. This thesis contributes to the operations efficiency and delivery accuracy at the retailer's distribution centers (DC) by developing the foundations of a supplier certification program and a statistical model that accurately measures the relevant metrics and enables the stakeholders to have intelligent discussion about improvements.

1.3 Company Background

CVS/pharmacy is a $120 billion pharmacy company with a network of 18 distribution centers serving pharmacy and front store items to over 7400 CVS/pharmacy/Pharmacy stores in 42 states plus Puerto Rico. The changes in the healthcare and retail industries are driving the needs for new solutions to service the customers which demand improvement on supply chain processes, systems and infrastructure. CVS/pharmacy has undertaken a strategy to ensure that end to end supply chain infrastructure and technology are aligned to support and enable evolving growth and service strategies. In the past decade, due to the acquisitions, supply chain strategy has been focused on building the right infrastructure and developing a standardized process and technology platform to handle the growth and integration of new stores and distribution centers.

1.4   Research Question

This thesis answers the question "How can CVS/pharmacy utilize SPC in developing a

certification program in order to eliminate non-value-add activities in its inbound supply chain?"

To answer, we started by reviewing the existing research on SPC and certification programs in

retail industry and analyzing the current performance of suppliers in CVS/pharmacy's network.

Based on the literature and interviews with CVS/pharmacy's personnel and trade partners, we

performed the initial steps in statistical analysis of available data. These steps consisted of

identifying the key performance metrics and developing a statistical measurement method for

accurately monitoring these metrics. Using the statistical analysis, we then developed a

sampling method and a SPC model. The sampling method, in combination with the SPC model,

allows for accurate monitoring with reduced inspection efforts. These two tools enable the

management to monitor and control the process while continuously seeking improvements. At

the conclusion of this thesis, we make recommendations about benefits and shortcomings of this

strategy with the current trade and information platform.

1.5   Thesis Structure

We have organized the content of this thesis to provide a logical progression of ideas in each

section. Most topics are introduced and discussed individually and at the end of each section we

explain how each idea fits into the overall scope of the thesis. Section 2 provides an overview of

the existing literature as it relates to SPC, supplier certification program, statistical analysis and

the methodology that we have offered in developing the certification program. Section 3

explains the nature of the data provided by CVS/pharmacy and our data classification approach.

Section 4 discusses the performance measurement method and establishes the assessment metrics

required for certification; this section reviews the suppliers' performance and provides different methods to identify statistically significant operational differences. Section 5 gives an overview of statistical sampling and offers a method for reduction of inspection efforts without compromising quality. Section 6 evaluates utilization of SPC in retail and analyzes the benefits of such initiative. Section 7 provides an overview of the potential causes of CVS/pharmacy's inbound network's inefficiencies and investigates the root causes of these inefficiencies and section 8 recaptures the findings of this thesis. We conclude this section by offering recommendations for future work on this topic.

2 LITERATURE REVIEW

The existing research in supply chain quality control using statistical and causal analysis mainly focuses on the manufacturing process specific to product quality and information exchange during production. The other topic studied in supply chain using statistical analysis is supplier-customer and/or retailer-customer interaction (sales) aimed at increasing efficiency through accurate demand forecasts and increased transparency. These researches address the issues at two ends of the supply chain with little emphasis on the linking processes in between. "One fallacy of the traditional management is that it is the firm that delivers the goods and services to the customer. In reality it is the channel (a process itself) that is the actual production and delivery system." (Novack, Grenoble and Goodbread p.49) Our research is focused on the inbound operations required for distribution of the manufactured goods to the end users. This process is different from the mentioned research topics (manufacturing and sales) in two ways: 1. the supply chain and distribution operations' extensive reliance on human resources results in high number of "touches" and a high degree of variation, 2. these operations' lack of direct link to the bottom line. Due to these reasons, management's control and supervision are limited on inbound processes and a higher degree of variability is tolerated due to human errors.

For the rest of this section, we review the existing literature on concepts and tools that we have used in developing our methodology as well as the practical/realized benefits of certification. These concepts include supplier certification/evaluation, perfect order management, statistical analysis, statistical sampling, SPC and root cause analysis. The literature reviews are not inherently interrelated and in the upcoming sections of this thesis, we explain how all of these concepts fit into our overall research work.

2.1 Supplier Certification and Evaluation Process

Different guidelines and strategies are available to assess and improve the supply chain. Amongst these a certification program is both simple and practical enough to be used in addressing the supply chain's inefficiencies. Maass, Brown and Bossert offer methods and insights for establishing such a framework "[to] draw a process map ... a sequential flow of each step of the process [to] identify the area where control mechanisms should be placed." (Maass, et al p.35) They suggest dealing with issues, especially at the data exchange level, requires proper documentation of methods on a process and control details sheet, broken down step by step, with detailed information summarized. "The type of data can be either [measurable] variables or [discrete] attributes determined as a pass/fail based on some gage or methodology." (Maass, et al p.37) J. Boyer has developed a one-page format that details a measurement plan for a specific metric using a definition, goal, accountability, and frequency of review as a means of direct and explicit communication with the supplier.

Brown further discusses the certification process and lays out the ground rules for the program's success. "Supplier certification works well in many industries: the more technical specifications, the more precise the definitions and the easier it is to create a supplier certification system." (Brown p.35) He also offers a high level view of the necessary steps in establishing supplier certification as they relate to the service industry: "separate out the measurable factors," then "meet with service supplier [to] set performance parameters" and take "data for some period of time [with] expectations of future performance. Once a period of time has passed with both parties satisfied with the progress being made, one can think about granting a certification status," using a "total cost model to understand the gains made" (Brown p.38)

Corum highlights the required characteristics of the relationship between the parties for a successful certification program. The fundamental attribute is the trust in the supplier relationship that must exist in various aspects; specifically "deterrence-based trust" and "processed-based trust." Corum notes the problems that may arise with a lack of trust in the supply chain relationship include localized optimization, opportunism, duplicated efforts, and lack of information sharing. Without trust, the transaction costs, including inspection, certification, and verification of the supply chain partners increase. Boyer reiterates these points noting "certification is achieved when the measurement goals have been achieved and maintained ... [but] does not necessarily mean that inspection will cease or that problems will not occur. It simply helps minimize cost-adding activities and problems" (Boyer p.9).

Boyer identifies tactics and objectives for a proper supplier certification program. "Never ask a supplier to perform to a standard that your company cannot achieve; [the] objective of a measurement process is to drive improvement to a level that eliminates cost, improves delivery and improves quality" (Boyer p.5). Corum suggests to "segment the supplier base, identify problems and benefits of supplier evaluation, collect data feeds from disparate systems while ensuring data integrity and punitively use scorecards" to overcome the challenges of certification.

We believe that the first step in development of a certification program should be mapping the process and defining the performance measurements most valued by the supply chain stakeholders. In the next section we review the literature associated with supply chain and logistics performance measurements in the discussion of a perfect order.

13

2.2 Perfect Order Management

Logistics and supply chain management has long been a key function of a successful business; however the methods of performance measurement have evolved over time "from a focus on single measures of functionally oriented distribution costs to an emphasis on multiple measures of process-oriented logistics cost and service outputs" (Novack, Thomas, p.6). In the past, suppliers, retailers, and carriers assumed that single performance measures could be taken independently; however, many measurements are interrelated. A definition that is mentioned throughout the supply chain and logistics industry is the Seven R's of Logistics, where delivery consists of the "*right* product, the *right* quantity, and the *right* condition, at the *right* place, at the *right* time, for the *right* customer, at the *right* cost" (Coyle, Bardi, Langley, p.5). While this definition is qualitative and makes intuitive sense, it does provide the baseline for the measurement of a perfect order.

In simple form, the perfect order is described as being "on time, complete, damage free and having correct documentation" (Vitasek, Manrodt, p.1); however this becomes challenging when each trade partner interprets and measures these attributes in a way that directly translates to customer-satisfaction. In some industries, the supermarket industry for example, the perfect order has been expanded to include on-time delivery, delivery to correct address, complete order, correct paperwork, product undamaged and high quality (C. Morgan, p.878).

Novack and Thomas acknowledge that the perfect order is not the same for all companies, as each company has a different definition of customer-satisfaction; therefore they have described the following seven challenges to overcome to define, implement, analyze and improve a company's perfect order measurement:

14

<u>Defining a Perfect Order</u>

1. Deciding which items to include in the perfect order

2. Internal perfect order vs. external perfect order

3. Impact of substitution/backorder on order fill

4. Impact of buyer request date/appointment date on on-time delivery

<u>Implementing Perfect Order Measurement</u>

5. Linking the appropriate information systems to the purchase order

6. The challenge of lagging variables

<u>Analyzing and Improving Perfect Order Measurement</u>

7. Impacts or correlation among measures in the perfect order

While responding to the above challenges provides a sound methodology that any company can use, it can be difficult to achieve a high score in perfect order metric, as an order must be "perfect" in several different performance measurement to be deemed a perfect order. This is why Vitasek and Monrodt, and Bowersox and LaHowchic, amongst others, have introduced the concept known as Perfect Order Index (POI), where instead of deeming a perfect order when all criteria are met perfectly, the perfect order measures are instead calculated as a percentage individually, then multiplied together to achieve a percentage POI, as shown in the Table 2:

Table 1 - Perfect Order Index (POI) Definition

% on Time	X	% Complete	X	% Damage Free	X	% Accurate Documentation	=	POI
95%	X	95%	X	95%	X	95%	=	84.1%

The POI measurement provides a single measurement that combines several important measures of customer-satisfaction into one, which is the intent of a logistics and supply chain

15

performance measurement system. Once the criteria of a POI is established, it allows for statistical analysis of the individual measurements.

2.3   Perfect Order Statistical Analysis

Statistical analysis is most often performed on a sample of the total population, and therefore that sample holds a certain level of reliability as to how well it represents the population. Confidence interval estimation, as Albright, Winston and Zappe describe, calculates an interval (or statistical range) "that is very likely to contain the true value of the population parameter" (Albright, Winston, Zappe, p.369). They offer several techniques for confidence interval and comparison analysis. The first is the confidence interval for a proportion to determine, with a defined sample size, what statistical range can be calculated using a point estimate for the performance, a standard error of that point estimate, and a corresponding normal distribution z-multiple value derived from the level of confidence desired, which typically is 95%. By performing this test at 95% confidence, we are confident that 19 out of 20 samples of equal size taken of the same population would fall within this interval; therefore producing a reliable estimate of the population performance.

The second technique follows similar logic, however intended to be used for a comparison of two independent proportions, using a point estimate of the difference between the two proportions, a standard error of the difference, and a corresponding normal distribution z-multiple value derived from the level of confidence desired that typically is 95%. While the first confidence interval estimation provides an indicator of overall population performance, this test, using the same statistical background, produces a reliable and confident comparison between two proportions to determine if one has a statistically significant differing performance level.

McGinnis and Vallopra, introduce a new test using the chi-square test statistic. First using contingency tables, McGinnis and Vallopra are able to display and compare the actual numerical results of the two outcomes of performance (i.e. on-time or not on-time). The contingency tables display the data in a 2x2, with explanatory variables on one axis and response variables on the other. Next, using the contingency tables, they perform a test of independence using the chi-square test statistic to determine if there are actual statistical differences between the two samples. Jung, Kang and Ahn also describe the usage of contingency tables for 2x2 analysis using Pearson's Chi-Square Test of Independence; however they also expand the analysis to clustered data with contingency table more than 2x2. From the contingency tables, Jung, Kang and Ahn go on to describe the chi-square analysis defining the null hypothesis that the two sample categories (C) are independent of their respective outcomes (R); therefore they will have the same probability and respective frequencies of the RC possible outcomes. The chi-square test then calculates a chi-square test statistic with a number of degrees of freedom to determine whether the sample observed frequencies are independent.

The literature on statistical analysis discusses the use of samples to describe the attributes of a population hence it is not necessary to inspect every product to gather the required measurements; statistical sampling offers a method for accurate measurements with reduced level of efforts.

2.4   Statistical Sampling

Statistical sampling involves the random inspection of a number of items as a subset of a population for the purpose of making inferences to the whole population that consists of a large number of items and a detailed examination of all items is not possible or costly.

As Gunning, Horgan, and Yancey explain "the application of [sampling is] to obtain and evaluate evidence about characteristic of the items selected in order to form or assist in forming a conclusion concerning the population" (Gunning et al p.2). A. Barnett adds that "in statistical sampling, each member has the same calculable chance of being selected" (Barnett p.28).

A commonly held misconception about statistical sampling is that it removes the need for the use of the professional judgement. While it is true that statistical sampling uses statistical methods to determine the sample size to evaluate, it is the responsibility of the inspector to consider and specify in advance factors such as the expected error rate, the risk of over-reliance or the risk of incorrect acceptance, inherent risk, standard deviation and population size, before the sample size can be determined. Statistical sampling methods don't remove the need for professional judgement, but rather that they allow elements of the evaluation process to be quantified, measured and controlled (Gunning, et al p4). Gunning and Barnett explain advantages of statistical sampling as follow:

1. The sample result is objective and defensible since the processes are based on demonstrable statistical principles.

2. The method provides an estimate of error and the results may be validated in terms of how far the sample projection might deviate from the value that could be obtained by a 100% check.

3. The method provides a means of advance estimation of sample size on an objective basis. The sample size is determined by a statistical method rather than guesswork.

4. The entire test operation has an objective and scientific basis which allows for individuals to participate independently in the same test and combine the results.

5. Objective evaluation of test results is possible. Thus, all parties performing this test would be able to reach the same conclusion about the numerical extent of error in the population. While the impact of these errors might be interpreted differently, there can be no question as to the facts obtained, since the method of determining their frequency in the population is objective.

Gunning et al. suggest using two types of testing in evaluating the statistical samples: compliance and acceptance. Compliance testing is typically concerned with qualitative characteristics or attributes and statistical sampling is used to estimate the proportion of violations associated with a particular set of controls. Acceptance sampling enables the inspector to reject or accept the population under certain conditions. A sample of a given size is drawn and if fewer than a certain number of errors is found, the system is accepted; otherwise it is rejected. The inspector using acceptance sampling seeks to balance out the risks of rejecting "satisfactory" systems and of accepting "unsatisfactory" populations.

A statistical sampling plan requires a certain level of confidence in the ongoing process performance. SPC models can be used for monitoring and controlling a process to recalibrate the confidence level used in statistical sampling plan.

2.5 Statistical Process Control

To analyze and control a process, Novack, Grenoble and Goodbread describe the use of statistical process control (SPC), where "the underlying principle is that to control any process, its variability must be understood ... and more importantly is the ability to identify variances and their causes and to suggest corrective actions to prevent any further occurrences" (Novack, et al,

p.51). Tague goes into more detail and describes a detailed procedure on how to create a SPC chart, while also providing an example of one completed.

To purposefully capture and utilize the collected data, a SPC model can be developed. S. Wang, Y. Wang and X. Jiang demonstrate the usefulness of such method, explaining that "the process quality control charts represent the information of process quality [and the personnel responsible for quality control] can find the reasons for out-of-control process ... by efficient recognition of abnormal mode of control charts." Abnormal modes can include a process that either ascends or descends over time, or takes an abrupt step either up or down, or has a seasonality where performance changes on a seasonal rate. Wang et al. further claim that the intelligent process quality control system in a supply chain provides a new way for enterprises to realize supply chain management.

Morgan takes the statistical approach one step further into the retail supermarket supply chain. From a high-level view, a supermarket supply chain and CVS/pharmacy supply chain are quite similar, in that both have a large supplier base and a large distribution network to support the large number of retail stores. Morgan explains that "in contrast to manufacturing, retailing (in many ways a more dynamic and complex environment in which change is endemic) has tended to be transaction driven. This condition can lead to a short-term view of the buyer/supplier relationship driven by short-term data analysis." (Morgan, p.875) Therefore Morgan discusses the use of SPC charts to track improvement initiatives through consistency, process capability and continuous improvement using methods such as root cause analysis.

2.6 Root Cause Analysis

To analyze the underlying root causes that affect the outcome of a process, Novack, Grenoble and Goodbread describe the use of a cause and effect diagram, otherwise known as an Isikawa or fishbone diagram. "The methodical manner in which this diagram must be constructed helps identify what can impact the output of a process" (Novack et al, p.52). Tague goes into more detail and describes a detailed procedure on how to create a fishbone diagram, while also providing an example of one completed. A Fishbone diagram is not an analysis tool but rather facilitates the brainstorming and analysis process by providing an overview of the current state and potential causal links.

2.7 Practical Application and Realized Benefits

Considering the stakeholders' reliance on electronic data exchange in the consumer packaged goods industry, information accuracy and quality (IQ) are as important as physical products delivery reliability. Accenture investigated this topic in collaboration with Grocery Manufacturers Association and other industry partners and published a report after analyzing the data from leading CPG manufacturers and retailers. As noted in this report, low IQ leads to "order accuracy issues, invoice errors, and ballooning freight and warehouse costs; all of these chronic industry problems have multiple causes. The lack of accurate and synchronized data was pegged as one of the root causes of such inefficiencies" (Accenture p.5). The study observed that solving these issues allows for improvement in other activities such as "accounting administration [and the retailer] due to improved information have reduced reconciliation of invoice by 5 percent. Fewer item data errors [has allowed] the accounting department to spend more time on value added activities." (Accenture p.11) Ge and Helfert found three significant

positive effects of improving IQ on decision quality in general: accuracy, completeness, and consistency. Ge and Helfert suggest that enhanced IQ is an essential aspect in organizational culture since poor IQ may lead to organizational losses such as lost customers, missed opportunities, and incorrect decisions.

In addition to the benefits from high IQ, additional improvements are realized in operations by utilizing a supplier evaluation system. In a two year study sponsored by Vendor Compliance Foundation (VCF), Manrodt and Vitasek used Perfect Order Index (POI) to analyze performance of about five thousand suppliers furnishing over 170,000 POs with five million line items. What they found was that it is practically achievable for all products and lines to have a 100% fill rate. In fact during this study 149 suppliers (~3% of population) achieved perfect order status. It should be noted that most of these suppliers had relatively small operations and simple POs, but nonetheless their achievement attests to the feasibility of perfect order. In two years, performance against complete orders was slightly better than on-time delivery – with the average across the sample falling at 59.9% for the entire sample of shipment data. On-time delivery showed the greatest improvement over two years, improving by 17.8%. "While it might be easy to blame the transportation provider for misses in on-time delivery – suppliers have far more control over complete orders. It is for this reason that we believe suppliers were able to make the greatest performance improvements." (Manrodt and Vitasek p.4)

In this study researchers also found out that best in class suppliers do perform better in each component of the perfect order, from 10% to 27% better than the average. "These suppliers will benefit from lower operating costs and improved service levels with the retailer." (Manrodt and Vitasek p.6) The data over the span of two years showed significant improvement but the

22

average POI score of 66.8% at the end of the study is still far from perfect performance. Most, if not all, retailers would consider this POI score unacceptable.

Using scorecards to track and communicate the performance of suppliers is a cheap but effective method. One large electronics retailer said this about their scorecard project, "Scorecards drive an interactive and collaborative dialog around performance... something we never had when we focused solely on compliance..." (Manrodt and Vitasek p.8) Retailers are recognizing that a robust supplier performance management program with scorecards reinforces what matters most and leads to lower costs and improved customer service.

In addition to scorecards, Morgan discusses the use of statistical process control (SPC) charts to track improvement initiatives through consistency, process capability and continuous improvement. As Novack, Grenoble and Goodbread discuss, SPC is a means to control a process by understanding its underlying variability and using it to develop process limits to both judge the current process and measure the success of process improvement initiatives.

2.8   Summary of Literature Review

Collectively the existing literature acknowledges the potential benefits achieved through a certification program but the offered means and methods vary and we were not able to identify a single comprehensive method that addresses the requirements for a retail certification program. We utilized all the ideas and concepts discussed in this literature review to develop a unique performance measurement system and certification program for CVS/pharmacy.

3 DATA COLLECTION AND CLASSIFICATION

The purpose of this research project was to develop a model that CVS/pharmacy could employ to improve accuracy and efficiency of the inbound operations. Given the large number of suppliers and DCs that CVS/pharmacy operates, we decided to focus on only three suppliers that represented a range of suppliers' performance in three DC locations. Using POI as a guideline we looked for improvement opportunities in accuracy of delivery time and quantity as well as compliance with CVS/pharmacy requirements. We evaluated four performance metrics: 1. line item fill rate, 2. line item quantity fill rate, 3. on time delivery, 4. compliance. As commonly understood in the industry, line item fill rate refers to the accuracy of product type and size received while line item quantity fill rate refers to the received quantity. Compliance issues are based on a set of internal controls pertaining to the packaging and stacking of pallets. This metric and the on time delivery are binary variables; the other metrics were measured in percentage accuracy level. The calculation methods and results of these variables are explained extensively in section 4.

This section provides an overview of the steps in gathering and analyzing data, developing a process map and classifying the data.

3.1 Data Collection

Data collection, including site visits and interviews with employees in the field, took place over three months. The process flow diagram intentionally is kept at a high level.

3.1.1 Site Visits and Phone Conferences

To better understand the inbound operations at CVS/pharmacy and observe the entire receiving shipment process, we visited a large DC in October and a supplier's DC as well as another CVS/pharmacy DC in December. These visits provided us with the knowledge of inbound operations and allowed us to engage stakeholders in the analysis of current performance and expected improvements. We held weekly conference calls with CVS/pharmacy representatives from operations and ordering departments to discuss the scope of the project and presented our incremental findings; occasionally suppliers' representatives joined these calls. The continuous dialogue was instrumental in successful sharing of insights and aligning of the expectations.

3.1.2 Process Flow

Based on the gathered information, a number of operations were identified as components of the inbound supply chain. The process begins with issuance of a Purchase Order (PO), which consists of multiple products (line items) each with their own respective quantities (line item quantities) to specific suppliers. Some items are backordered when the products are temporarily not available while some are eliminated or procured through other sources; however this process imperfection is transparent to CVS/pharmacy and the ultimate grading point and performance measurement is their successful fulfillment of confirmed items and quantities that they are able to deliver. In other words, item unavailability was not considered as a negative factor as long as it was communicated with CVS/pharmacy. Once a supplier has acknowledged the PO, the products in each PO are typically aggregated into a single shipment. Depending on the PO volumes and frequencies multiple POs could be aggregated into a single shipment or a

PO could be delivered in more than one shipment. Deliveries are made either by supplier's own carrier or a preferred third party logistics carrier (3PL).

Approximately 48 hours before the shipment is delivered, the supplier provides an Advanced Shipment Notice (ASN) that informs CVS/pharmacy about the content of the shipment. Prior to this point CVS/pharmacy is not certain about when or in what quantities the accepted POs will be delivered. While the original POs indicate a delivery date, the actual delivery date is determined by the supplier at the time of ASN; however CVS/pharmacy has a general expectation about the delivery of the products routinely ordered. The carrier is usually responsible for booking a delivery appointment from CVS/pharmacy. Some suppliers have regular standing appointments, but others must book appointments with CVS/pharmacy as needed. Figure 1 shows the sequence of actions by each party.
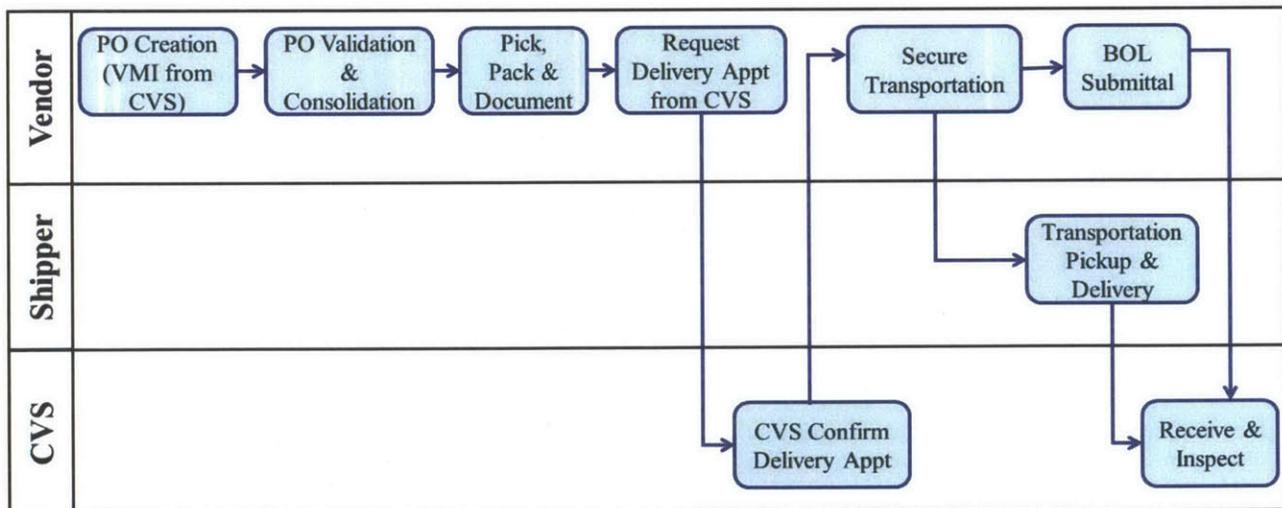

Figure 1 – Inbound Supply Chain Operations Process Flow Diagram

3.2   Data Classification

Given that CVS/pharmacy operates across many geographical regions and trades with various suppliers, it was important to analyze the receiving records for different suppliers in

different regions. The other benefit of cross examining data across different geographies was gaining insights about internal differences in operations at CVS/pharmacy's DCs. We were provided with three months of data for three suppliers' delivery records at three DCs. The three suppliers chosen by CVS/pharmacy representatives, referred to as suppliers 1, 2 and 3, represented different classes of suppliers as it relates to volume, accuracy and other characteristics. The three DCs, referred to as N, O and Y (used the same code names as the raw data), also represented a range of new work force to experienced employees with different geographical conditions.

The first data set consisted of receiving data totaling over 29,000 line items associated with 2,877 POs and a quantity of over a million stock kept units (SKU). Each line of data had values for supplier, DC location, PO number, order date, order quantity, appointment request date, appointment date(s), delivery date, ASN quantity, received quantity and item description. Other information such as supplier's pay ID, item number, etc. were also available in the data, but not used in the analysis. The line item information was grouped into POs and most of the analysis was conducted at PO level.

The second data set consisted of the compliance records associated with the receiving data set through unique PO and ASN numbers. The compliance records noted issues such as damaged pallets, missing shrink wrap, missing shipping lists, and unscannable labels. These attributes all pertain to the physical condition of the delivered shipments and were recorded by receivers responsible for unpacking and sorting the shipped products. Table 1 summarizes the original data set volume at each level.

Table 2 – Supplier-DC Receiving Volume Levels

Supplier	DC	Total PO	Total ASN	Total Del	Total LI	Total LI Qty
1	N	81	41	33	2,944	141,271
	O	58	14	25	1,366	37,274
	Y	65	17	20	1,739	49,204
	All	204	72	78	6,049	227,749
2	N	591	90	49	5,035	268,141
	O	509	25	41	3,926	67,770
	Y	500	24	37	4,705	86,301
	All	1,600	139	127	13,666	422,212
3	N	434	75	47	3,569	315,548
	O	267	22	36	2,304	77,878
	Y	372	32	35	3,588	125,098
	All	1,073	129	118	9,461	518,524
Grand Total		2,877	340	323	29,176	1,168,485

In this section we discussed and displayed the data that was provided to us without any interpretation nor analysis. The following section explains the metrics that were developed for analyzing this data and establishing the basis of a unique performance measurement system.

4   PERFORMANCE MEASUREMENT ANALYSIS

4.1   Perfect Order Measurement

In this section we build upon the literature on POI, discussed in section 2.2 and 2.3, and

define the performance metrics analyzed to construct a POI and the respective analysis methods.

The results of analyzing the suppliers on these metrics are presented at the end of the section.

4.1.1   Perfect Order Measurements and Index Definition

For our analysis of CVS/pharmacy's inbound supply chain, we developed four perfect

order performance measurements for on-time delivery, line item fill rate, line item quantity fill

rate and physical compliance. For explanatory purposes as we introduce and describe the

calculations behind these measurements, the example data in Table 3 will be used.

Table 3 - Sample PO Summary for Performance Measurement Explanation

PO	Line Item SKU	Appt Date	Delivery	ASN Qty	Delivery Qty	Compliance Codes
A	123	03/24	03/24	50	49	U
	134	03/24	03/24	35	35	
B	234	04/15	04/19	80	81	-
C	567	04/26	04/28	15	15	
	578	04/26	-	30	-	-
	589	04/26	04/28	45	45	

- *On-Time Delivery (OTD)* – A PO is considered on-time if the initial delivery date is equal

  to any scheduled appointment date (+/- 2 days to account for CVS/pharmacy processing

  time in database). Shipments are not always delivered on the initial scheduled

  appointments however the supplier and/or their respective carriers can schedule a new

  appointment; therefore there are multiple opportunities for being "on-time."

- o PO A on-time – delivered and processed on appointment date

- o PO B not on-time – delivered and processed 4 days after appointment date

- o PO C on-time – delivered and processed 2 days after appointment date (only items which have been delivered are graded)

- o OTD = 2 on-time POs / 3 delivered POs = 66.6%

- *Line Item Fill Rate (LIFR)* – The LIFR is calculated as the number of line items delivered divided by the ASN committed line item deliveries. Every PO filled by a supplier has one or more line items consisting of multiple SKUs.

  - o Line items 123, 134, 234, 567 & 589 all delivered

  - o Line item 578 not delivered

  - o LIFR = 5 delivered line items / 6 committed line items = 83.3%

- *Line Item Quantity Fill Rate (LIQFR)* – The LIQFR is calculated as the line item quantity delivered divided by the ASN committed line item quantities. Like the LIFR, every PO not only has one or more line items, but those line items also have their own respective quantities from one to many.

  - o Line item 123 delivered 1 short = 49 / 50 = 98% fill rate

  - o Line items 134, 567 & 589 all delivered to correct quantity = 100% fill rate

  - o Line item 234 delivered 1 extra = 1 − (81 − 80 / 80) = 98.8% fill rate

  - o LIQFR = (98% + 100% + 100% + 100% + 98.8) / 5 = 99.4%

- *Physical Compliance (PC)* – A PO is compliant if it does not violate any of CVS/pharmacy's compliance requirements (CVS/pharmacy requested that we omit one code "W – Ops Decision" as it was a test code for internal operations)

  - o PO A non-compliant – assigned compliance code U

o   PO B & C compliant – assigned no compliance codes

o   PC = 2 compliant POs / 3 delivered POs = 66.6%

Rather than simply multiplying these performance measurements together to get the POI, we instead take the weighted average of these four performance measurements with respective weights of 20%, 30%, 30%, and 20%. The weights reflect the significance of each metric based on CVS/pharmacy's perspective. LIFR and LIQFR are weighted higher than the other two due to the efforts associated with inspecting, unpacking and sorting of the received shipments. While the weights are not set and can be changed per management's decision, they still provide a POI metric that all suppliers can be equally graded on. As added research and analysis is performed, the weights can be updated to continue to fine-tune the metric for supplier evaluation.

o   POI = (20%)(66.6%) + (30%)(83.3%) + (30%)(99.4%) + (20%)(66.6%) = 81.5%

The measurements' definition and their respective weights were developed corresponding to CVS/pharmacy's business operations; however as the literature states the development of such metrics for other organizations may be different, but the same methodology can be applied.

4.1.2   Calculated Performance Measurements

The methodology in 4.1.1 was applied to the complete data set for each supplier at each DC, while also calculating total values for each supplier at each of the respective DC locations for the entire three month span, as well as total for each supplier and a grand total overall. As can be seen in the Table 4, each network has different performance levels for each metric, both intra-supplier and intra-DC.

Table 4 – Perfect Order Performance Measurement Summary

Supplier	DC	OTD	LIFR	LIQFR	PC	POI
1	N	100.0%	99.8%	98.9%	82.7%	96.2%
	O	91.4%	99.6%	98.2%	98.3%	97.3%
	Y	100.0%	99.7%	97.2%	100.0%	99.1%
	Total	97.5%	99.7%	98.1%	92.6%	97.4%
2	N	89.7%	99.4%	99.7%	99.7%	96.9%
	O	35.6%	96.5%	99.2%	99.2%	83.0%
	Y	39.8%	99.3%	98.8%	98.6%	87.1%
	Total	56.9%	98.4%	99.3%	93.7%	89.4%
3	N	94.5%	99.3%	99.3%	96.1%	97.7%
	O	49.8%	98.6%	97.8%	97.8%	88.4%
	Y	48.9%	97.0%	97.3%	100.0%	88.1%
	Total	67.6%	98.3%	98.2%	97.9%	92.2%
Grand Total		63.7%	98.6%	98.8%	95.2%	91.0%

These three month average values for each performance measurement allowed us to understand the current performance, however this period is only a small sample size of the total population of POs (OTD, PC), line items (LIFR) and line item quantities (LIQFR) considering CVS/pharmacy's overall ordering volume. In the next section statistical tests will be used to better interpret the supplier performance measurements.

4.2   Perfect Order Statistical Analysis

Building on the concepts covered in literature review sections 2.3, in this section we explain the methods used in supplier performance analysis and share the results. We will introduce three different statistical analyses, one for individual measurements and two for measurement comparisons. For the first two statistical analysis techniques, we will walk through

the calculations for Supplier 2's LIFR performance at DC N and DC O (represented as 2N and 2O), using the actual data in Table 5.

Table 5 – Statistical Analysis Explanatory Sample Data

Data Element	2N	2O
Total POs Delivered	591	509
Total Line Items Confirmed on ASN	4768	4258
Total Line Items Delivered	4741	4109

Confidence Interval for Proportion

For each metric, their relative performance was calculated as a discrete distribution, as opposed to continuous. Based on the data we were presented, the sum of all successes (x) can be divided by the number of samples (n) to get a point estimate of the population probability of success, which is the same calculation that was performed resulting in Table 4.

$$Point\ Estimate: p = \frac{x}{n} \tag{1}$$

$$p_{2N} = \frac{4741}{4768} = 99.4\%$$

$$p_{2O} = \frac{4109}{4258} = 96.5\%$$

Next, the confidence interval is calculated to provide a reliable estimate of the population performance that is represented by the sample. Due to the fact that all sample sizes analyzed in this paper are relatively large (n > 30), according to the Central Theorem the sample distributions can be considered approximately normal, with a mean of p and a standard error (SE) as described below. As the sample size n increases, the SE will decrease, which is intuitive because as more samples are measured, the resulting probability will more accurately represent the population distribution.

33

$$SE(p) = \sqrt{\frac{p(1-p)}{n}} \qquad (2)$$

$$SE(p_{2N}) = \sqrt{\frac{.994(1-.994)}{4768}} = 0.11\%$$

$$SE(p_{2O}) = \sqrt{\frac{.965(1-.965)}{4258}} = 0.28\%$$

Using mean p and standard error SE(p), a statistical confidence interval can be calculated

using a z-multiple based on the level of confidence specified for the analysis. For the purposes

of this analysis, we will be performing the confidence interval based on two tailed distribution

with a significance level á = 0.05, meaning that the confidence interval will extend to within

0.025 of each side of the bell curve. The z-multiple value is then:

$$Z_{1-á/2} = z_{0.975} = 1.96 \qquad (3)$$

Using p, SE(p) and the z-multiple, a confidence interval (CI) for the sample distribution

can be calculated using the below formula. The CI provides a range of values to which the

actual population's performance adhere 95% of the time for the specific supplier and DC.

$$CI = p +/- z - multiple * SE(p) \qquad (4)$$

$$CI_{2N} = 99.4\% +/- 1.96 * .11\% = 99.2\% \ to \ 99.6\%$$

$$CI_{2O} = 96.5\% +/- 1.96 * .28\% = 95.9\% \ to \ 97.1\%$$

CIs were calculated for all nine of the supplier-DC networks using the same formulas to

develop the LIFR graph in Figure 2. We performed the same calculations for the other metrics

and developed similar graphics that can be found in the appendix. In this graphical

representation, the CIs with similar values can be said to have similar performance, whereas the

CIs with drastically different values can be said to have statistically significant differences in

performance; however the next statistical analysis will formalize the comparison analysis.
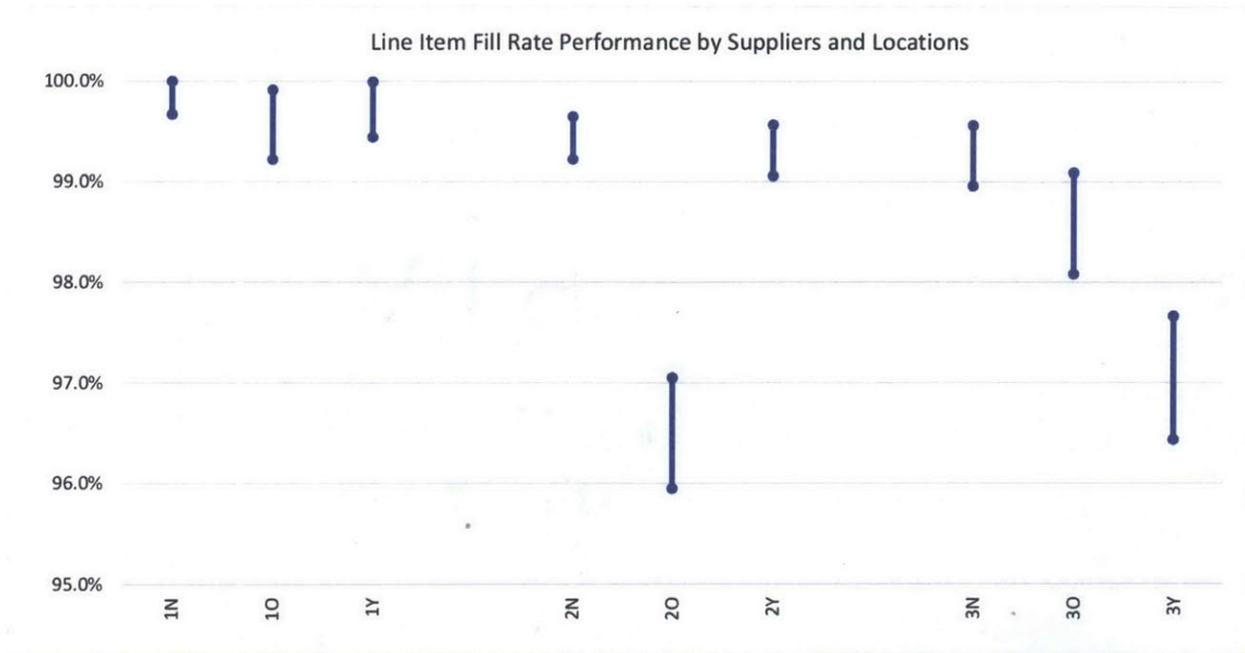
Figure 2 - Line Item Fill Rate Performance Confidence Intervals

Confidence Interval for the Difference Between Two Proportions

While understanding the confidence interval of a single sample is important, it is also important to understand the level of comparison between two separate samples. An important purpose of our analysis was to determine if differences between different networks' performances were statistically significant. To do so, we used confidence interval analysis; however this time as a comparison of two proportions. In this analysis, the point estimates were developed as the difference between two individual sample proportion point estimates.

$$Comparison\ Point\ Estimate: p1 - p2 \qquad (5)$$

$$Comparison\ Point\ Estimate: p_{2N} - p_{2O} = 99.4\% - 96.5\% = 2.9\%$$

Next, similar to the single sample analysis, a standard error must be calculated using both sample distributions:

$$SE(p1 - p2) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \qquad (6)$$

$$SE(p1 - p2) = \sqrt{\frac{.994(1-.994)}{4768} + \frac{.965(1-.965)}{4258}} = 0.30\%$$

Finally a CI can be calculated using the p, SE and z-multiple (which is calculated the same as in the single sample analysis), as follows:

$$CI = p1 - p2 +/- z - multiple * SE(p1 - p2) \tag{7}$$

$$CI = 2.9\% +/- 1.96 * 0.30\% = 2.3\% \ to \ 3.5\%$$

The resulting confidence interval is the statistical range of the difference between the two independent proportions, which provides insight into the significance of whether one sample is a better or worse performer than the other. If the confidence interval includes 0 (goes from positive to negative, or vice versa), it cannot be said that one sample is different than the other at that level of statistical confidence (95%). In these cases, it could be a factor of too small a sample size (larger standard error) or the fact that one is truly not better or worse than the other from a statistical perspective for the inspected period (3 months).

Since the CI comparison of 2N and 2O does not include 0, then it can be said that 2N is a better LIFR performer than 2O with 95% confidence. Comparison CIs were calculated for all 18 supplier-DC network comparisons (intra-supplier and intra-DC) using the same formulas to develop the LIFR graph in Figure 3. We performed the same calculations for the other metrics and developed similar graphics that can be found in the appendix.
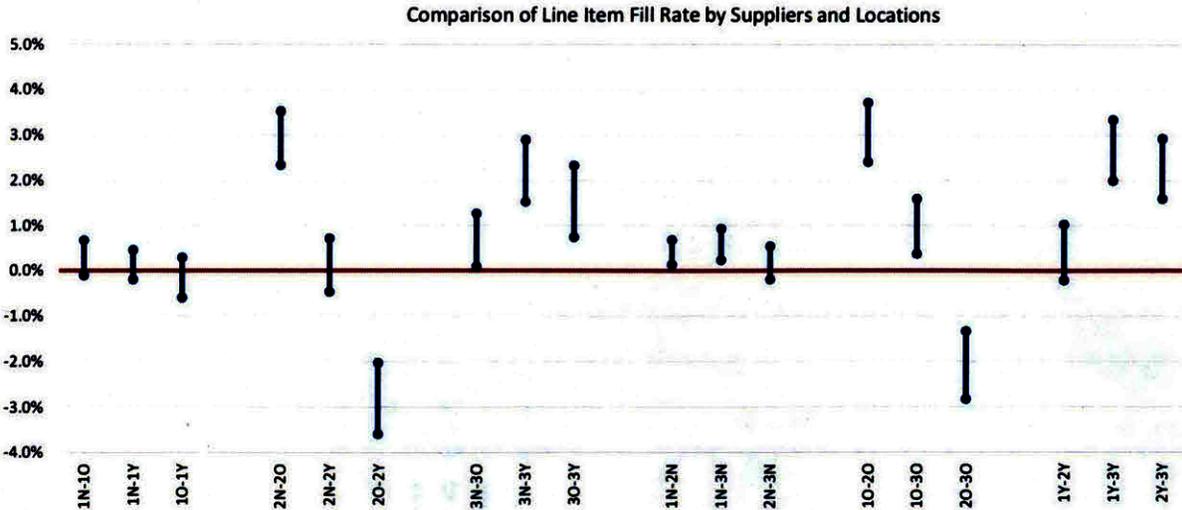
Figure 3 - Line Item Fill Rate Performance Comparison Confidence Intervals

Pearson's Chi-Square Test of Independence

Similar to the confidence interval for comparison of two proportions, Pearson's Chi-Square Test of Independence is used to statistically analyze whether one sample is a better or worse performer than the other, at a certain level of statistical confidence. This test first uses contingency tables, where numerical values are assigned to different categories with one axis of explanatory variables (known category, i.e. black/white, supplier A/supplier B, etc.) and the other response variables (outcome variable, i.e. salary, performance, etc.). A theoretical and the LIFR comparison example contingency tables are shown in table 6 and table 7:

Table 6 – Contingency Table Example with Variables

	Response 1	Response 2	Total
Category 1	a	b	a + b
Category 2	c	d	c + d
Total	a + c	b + d	n (a + b + c + d)

Table 7 – Contingency Table Example with Sample Values

	Delivered	Not Delivered	Total
2N	4741	27	4768
2Y	4109	149	4258
Total	8850	176	9026

The test that is performed in this analysis has a null and an alternative hypothesis, where the null hypothesis (Ho) is that the explanatory variables and the response variables are independent; whereas the alternative hypothesis (H₁) is that the explanatory variables and response variables are not independent, meaning that one could predict the value of response variable by if it was known when explanatory variable category of a sample fell under.

The next step in the analysis is to determine the significance level of the analysis, which in our case, following that of the confidence intervals, is ȧ = 0.05. In addition the degrees of freedom of the analysis must be calculated using the number of rows (r) and columns (c).

$$(r-1)*(c-1) \tag{8}$$

$$(2-1)*(2-1) = 1$$

Initially, the contingency table is populated with the observed frequencies for each explanatory and response variable; however the Chi-Square analysis compares the expected frequencies to those observed frequencies to make a statistical assessment on the samples' independence. The following equation is used to calculate the expected frequency for each field.

$$Expected\ Frequency: E_{r,c} = \frac{(n_r * n_c)}{n} \tag{9}$$

$$Expected\ Frequency: E_{1,1} = \frac{(4768*8850)}{9026} = 4675$$

$$Expected\ Frequency: E_{1,2} = \frac{(4768*176)}{9026} = 93$$

$$Expected\ Frequency: E_{2,1} = \frac{(4109*8850)}{9026} = 4175$$

38

$$Expected\ Frequency: E_{2,2} = \frac{(4109*176)}{9026} = 83$$

Once all expected frequencies are calculated, they can then be statistically compared to the observed frequencies to determine the chi-square test statistic x^2 from the equations below.

$$x^2 = \sum_{i=1}^{n} \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad\quad (10)$$

$$x^2 = \frac{(4741 - 4675)^2}{4675} + \frac{(27 - 93)^2}{93} + \frac{(4109 - 4175)^2}{4175} + \frac{(149 - 83)^2}{83} = 101.21$$

From the calculated chi-square test statistic and the degrees of freedom, a p-value can be derived using the chi-square distribution, where the analysis is focused on the upper region of the distribution, providing the probability that a more extreme outcome would happen, given the null hypothesis (explanatory and response variables are independent). If the p-value is less than the level of significance ($\acute{a} = 0.05$), then the null hypothesis cannot be accepted, thus concluding that there is a relationship between the explanatory variable and response variable. If the p-value is greater than the level of significance ($\acute{a} = 0.05$), then the null hypothesis is accepted, thus concluding that the outcome shown could have happened by chance, given that the explanatory and response variables are independent.

The p-value for the 2N to 2Y comparison of χ^2 of 101.21 and d.f. of 1 is 0.00 at two significant digits, which is less than 0.05; therefore the null hypothesis is rejected and there is a statistically significant difference between the two, which is the same result as the comparison confidence interval. The same analysis was performed on all 18 network comparisons for all 4 performance measurements and the results are shown in table 8, where red values indicate where the test resulting in accepting the null hypothesis (i.e. no statistically significant difference):

Table 8 – Pearson's Chi-Square Test of Independence Results

Compare	OTD χ^2	OTD p-value	LIFR χ^2	LIFR p-value	LIQFR χ^2	LIQFR p-value	PC χ^2	PC p-value
1N-1O	7.24	0.007	2.35	0.125	108.88	0.000	8.50	0.004
1N-1Y	Perfect 100%		0.65	0.421	675.21	0.000	12.43	0.000
1O-1Y	5.84	0.016	0.46	0.499	92.69	0.000	1.13	0.288
2N-2O	350.41	0.000	101.21	0.000	307.94	0.000	38.01	0.000
2N-2Y	303.93	0.000	0.54	0.462	844.86	0.000	5.65	0.017
2O-2Y	1.93	0.165	78.72	0.000	42.25	0.000	56.78	0.000
3N-3O	188.82	0.000	5.71	0.017	1341.70	0.000	1.45	0.228
3N-3Y	213.07	0.000	41.29	0.000	2827.07	0.000	14.89	0.000
3O-3Y	0.05	0.825	12.85	0.000	56.78	0.000	8.44	0.004
1N-2N	9.20	0.002	5.26	0.022	1009.58	0.000	25.84	0.000
1N-3N	4.70	0.030	8.03	0.005	223.89	0.000	21.56	0.000
2N-3N	7.55	0.006	0.91	0.339	387.00	0.000	0.03	0.872
1O-2O	66.93	0.000	36.57	0.000	208.32	0.000	7.16	0.007
1O-3O	33.63	0.000	7.89	0.005	13.81	0.000	0.06	0.804
2O-3O	14.77	0.000	22.49	0.000	427.79	0.000	27.42	0.000
1Y-2Y	83.74	0.000	3.00	0.083	511.67	0.000	0.92	0.337
1Y-3Y	58.74	0.000	33.38	0.000	2.78	0.095	Perfect 100%	
2Y-3Y	7.22	0.007	53.98	0.000	597.74	0.000	5.25	0.022

Comparison of Confidence Interval and Pearson's Tests

As the same inputs were used for both tests, we expected that both tests result in the same comparison statistics regarding the measurements' independence. This is true for 71 of the 72 comparisons analyzed. The only case where the test results differ is 2N-2Y for physical compliance (PC) where the confidence interval test results in no statistically significant difference between the two, whereas Pearson's test results in rejecting the null hypothesis of independence, meaning there is a statistically significant difference. These tests use different statistical distribution tables and while they approximately overlap for most values, there are different results for borderline values due to the nature of the distributions.

5 Statistical Sampling

Once the performance metrics are defined and statistically analyzed, to continue the monitoring process, statistical sampling can be adopted to reduce the inspection efforts. The point estimates calculated in section 4 provide a picture of the historical performance. Following the method discussed in the literature review section 2.4, these same point estimates can be used as the basis of statistical sampling plan for inspection of future shipments.

5.1 Statistical Sampling Procedure

When possible we prefer to inspect only a portion of the received products rather than the entire shipment without impacting the degree of quality control. This is possible only after a high level of performance is constantly achieved and the supplier's shipments volume per period exceeds the required sample size. There is a trade-off between sampling and confidence in accuracy of received product; by inspecting only a portion of the shipments, CVS/pharmacy is accepting the chance of false alarms or errors going unnoticed.

After investigating the point estimate of multiple performance metrics and developing a framework for comparison across CVS/pharmacy's network, we focused on the required amount of inspections to maintain a level of confidence in each supplier's performance. Statistical sampling allows CVS/pharmacy to reduce the resources spent on inspection while having a high level of confidence in the received shipments' accuracy and quality.

The first step is to identify the suppliers suitable for sample inspection; performance below a certain level requires a large sample size and not inspecting the entire shipment for some suppliers does not seem feasible. Three inputs are required to determine the sample size for the desired level of confidence. The first input is the point estimate of the current performance (p) and is the only input controlled by the supplier. We used the LIQFR and LIFR point estimates in determining the required sample sizes. Given the data was only for a three months period, further data collection is required to establish an accurate estimate of current affairs. Due to the significance of this input in determining the sample size, it is important to start with an accurate point estimate. The other two inputs are set by CVS/pharmacy: "confidence interval" (CI) and "acceptable margin of error" (Δ). CI establishes the degree of certainty that CVS/pharmacy wants to achieve by inspecting a portion of the shipments. In other words if CVS/pharmacy chooses a 95% confidence interval then it would be acceptable if once out of every 20 random inspections, the accuracy level is below the agreed upon level. Acceptable margin of error on the other hand reflects the tolerance that CVS/pharmacy is willing to accept. In other words if a 1% margin of error is chosen, then it would be acceptable if the accuracy of an inspected item is within 1% of the agreed upon accuracy level. (e.g. 99% +/- 1% implies a 98% or above is acceptable).

We can find the required sample size using the inputs and the equations below. For any given CI there is a corresponding Z value available from a normal distribution table. Standard deviation of accuracy metric (σ) is defined as

$$\sigma = \frac{\sqrt{p(1-p)}}{\sqrt{s}} \tag{11}$$

42

p is the point estimate of the chosen metric, LIQFR in our case. s is the required sample size. Statistical sampling requires:

$$Z * \sigma \leq \Delta \tag{12}$$

Combining (11) and (12) we can solve for s

$$s = \frac{Z^2 \, p(1-p)}{\Delta^2} \tag{13}$$

As part of our analysis, we calculated the required sample size based on the three input values. Considering that CVS/pharmacy evaluates the suppliers on different metrics, it is important to choose the right metric for the inspection sample size calculation if multiple metrics are measured in each sample. Usually the chosen metric should have the lowest point estimate among all performance metrics since that will result in the largest required sample size for inspection. Once the sample size satisfies the requirement for that metric, CVS/pharmacy safely can assume the same level of certainty for the other metrics.

5.2   Sample Inspection Discussion

Out of the four metrics associated with POI, only two are appropriate for sample inspection: LIFR and LIQFR. The other two metrics, OTD and PC are inevitably recorded for all incoming shipments given the current information system at CVS/pharmacy. The input fields for calculating the appropriate sample size is manual data entry unless CVS/pharmacy chooses to use the same standards across all suppliers which would require a close to uniform performance from all suppliers. Table 9 shows the required inputs in the top three rows and the resulting sample size in the bottom row. The values shown below serve just as an example.

Table 9 – Input table for sample size calculation

Current Performance	97.50%
CI Interval	95.0%
Acceptable Deviation	2%
Required sample size	234

In the industry 95% confidence interval is commonly used hence we calculated the required sample sizes to evaluate the two metrics that can be verified through sample inspections using the current performance point estimates from section 4.1 and tolerating 1% deviation.

Table 10 – Required Inspection Sample Sizes for Each Metric

Supplier	DC	LIFR	LIQFR
1	N	77	418
	O	153	679
	Y	115	1046
2	N	229	115
	O	1298	305
	Y	267	455
3	N	267	267
	O	530	827
	Y	1118	1009

The units for LIFR sample size is line item and the unit for LIQFR is individual product count (SKU). As observed in table 10 the sample sizes can vary significantly with slight changes in performance metric and in general the samples tend to be large. The values chosen for CI and Δ are not too strict and as these values become higher the sample sizes will further grow. Hence a careful review of performance metrics and expectations is required before statistical sampling method can be adopted.

6 STATISTICAL PROCESS CONTROL

Performing sample inspection to measure the performance metrics defined in section 4 allows the management to capture an accurate snap shot of the current affairs. However to use this data in predicting inefficiencies and identifying improvement opportunities a SPC model is used that captures the historical performance of each supplier with respect to time. These two statistical methods should be used in conjunction to ensure the supply chain is operating efficiently and at its full potential.

6.1 Model Development

The SPC analysis is very similar to that of the individual confidence interval analysis, where a point estimate p and standard error SE are used to calculate process control limits, however rather than analyzing the data set as a whole, it is broken out into time periods for historical analysis. For example and explanatory purposes, the LIFR performance of Supplier 2 and DC Y will be used to describe the creation of a SPC chart, where 4058 line item SKUs were committed and 28 were not actually delivered. The first step is to record the data into period buckets over time. Our analysis spans the weeks of 2013-49 to 2014-09:

Table 11 – Supplier-DC Network 2Y Weekly LIFR Performance

Weeks	49	50	51	52	53	01	02
SKUs Committed	179	133	658	17	86	124	464
SKUs Not Delivered	0	0	7	0	0	2	1
LIFR Performance	100.0%	100.0%	98.9%	100.0%	100.0%	98.4%	99.8%

Weeks	03	04	05	06	07	08	09
SKUs Committed	1	570	351	170	696	284	325
SKUs Not Delivered	0	8	2	0	5	3	0
LIFR Performance	100.0%	98.6%	99.4%	100.0%	99.3%	98.9%	100.0%

45

Since the data that we are using to measure and control is binary and the sample sizes change, the appropriate SPC chart is the P-Control Chart, which is thoroughly explained by Tague. The first step is to calculate the point estimate of performance by dividing the number of successes (x) of the total number of samples (n).

$$Point\ Estimate: p\ =\ \frac{x}{n} \tag{14}$$

$$p_{2Y}\ =\ \frac{4030}{4058} = 99.3\%$$

Next, to calculate the control limits, the standard error must be calculated but unlike the CI, the sample size n is not the full sample size, it is a sample size representative of each period. For our analysis, we calculated this in two ways: 1. using the average weekly sample size (290 for 2Y), 2. using the maximum weekly sample size (696 for 2Y). Using the maximum would increase n, thus resulting in a smaller SE and tighter control limits.

$$SE(p)\ =\ \sqrt{\frac{p(1-p)}{n}} \tag{15}$$

$$SE_{avg}(p)\ =\ \sqrt{\frac{.993(1-.993)}{290}} = 0.49\%$$

$$SE_{max}(p)\ =\ \sqrt{\frac{.993(1-.993)}{696}} = 0.31\%$$

In literature, the control limits are typically positioned at three SEs from the mean, and that is how we performed our analysis, using both the average and the maximum sample size SEs.

$$Upper\ Control\ Limit\ (UCL) = p + 3 * SE(p) \tag{16}$$

$$Lower\ Control\ Limit\ (LCL) = p - 3 * SE(p) \tag{17}$$

$$UCL_{avg}\ = 99.3\% + 3 * 0.49 = 100.7\%$$

$$LCL_{avg} = 99.3\% - 3 * 0.49 = 97.9\%$$

$$UCL_{max} = 99.3\% + 3 * 0.31 = 100.2\%$$

$$LCL_{max} = 99.3\% - 3 * 0.31 = 98.3\%$$

In this case and many others, the UCL falls above 100%, which is an impossible performance measurement to achieve; therefore the most important control limit is the LCL. Additionally, even if the UCL fell below 100%, the more important control limit is the LCL; however the UCL is still important in that it would alert that something had changed in the process. Another feature that can be added to a control chart is a target limit for which an organization sets as "target" performance level for the purpose of certification or other scoring criteria. While CVS/pharmacy did not share with us a target limit, for analysis purposes we have chosen the level of 98.5%. The resulting SPC chart is shown in Figure 4.
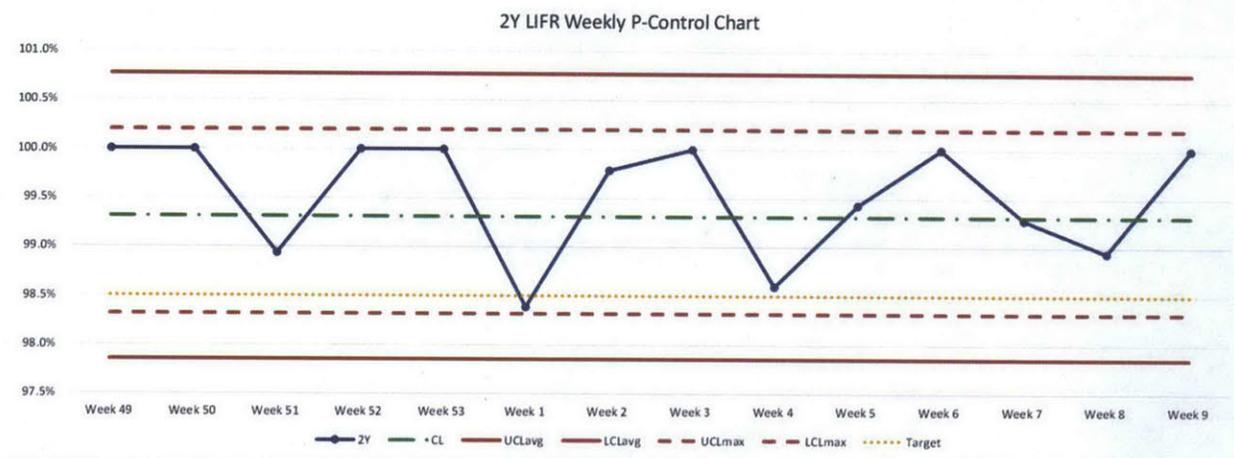
Figure 4 – 2Y LIFR Statistical Process Control P-Chart

6.2 Model Analysis and Discussion

The resulting SPC P-Control Chart for the LIFR performance of Supplier 2 at DC Y shows that the process is in control over that specified span of 14 weekly periods for both control limits

calculated from the average sample size and the maximum sample size; only once does the process fall below the target limit of 98.5% in week 1. In our analysis of other performance measures and networks, this wasn't always the case, especially when analyzing OTD and PC, where some performance levels varied from 0% to 100% (as shown in Figure 5). One of the main reasons for this is the actual amount of data that is available, in that the weekly buckets can make the weekly sample sizes very small, especially for OTD and PC where the sample size unit is not line item but instead PO. If data had been kept for longer periods of time where monthly or quarterly period buckets could be created, the results would have less performance variance leading to tighter control limits more accurately representing the process performance.
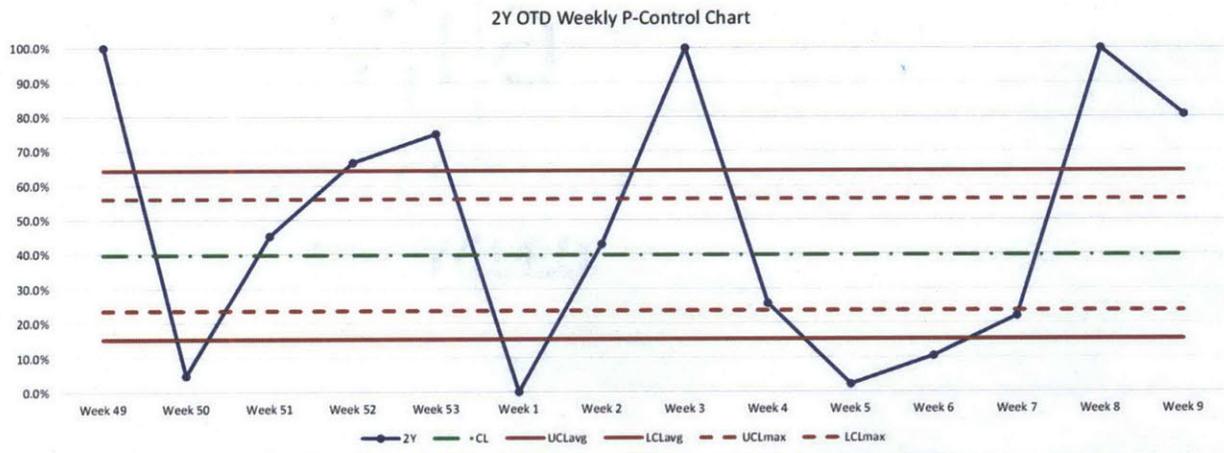
Figure 5 – 2Y OTD Statistical Process Control P-Chart

Additionally, Tague states that data should be collected for at least 20 samples, where we only used a total of 14 samples. The methodology behind creating the control charts would have been better if more data was available where control limits could be calculated for the initial samples and remained at that level for all resulting samples, which would not factor into the control limit calculation. The only time the control limits would need to be changed is if there was a large shift in performance either up or down.

7 ROOT CAUSE ANALYSIS

Sampling plans and SPC provide the necessary data for management to identify

performance inefficiencies. To eliminate these inefficiencies, one must first investigate the root

cause of them. Root cause analysis offers a systematic approach in understanding the problems

and drawing the connections between the causes and the inefficiencies. Once the true causes are

identified, corrective action can be put in place to prevent reoccurrence.

7.1 Methodology

Following Tague's "Quality Toolbox" as a guide, the initial step of a fishbone diagram is

to agree on a problem statement that is both understandable and direct to the point. By doing

this, the resulting fishbone diagram will better describe and list the correct most probable and

potential causes. Next, the major causes should be brainstormed and placed on the diagram, and

common ones are "methods, machines (equipment), people (manpower), materials, measurement

and environment" (Tague, p.247-248). Next, add more causes "bones" to the diagram by asking

the question "why does this happen?" (Tague, p.248) and continue to do this until all potential

causes are listed on the diagram. When creating the fishbone diagram, any and all potential

causes should be brainstormed and listed, then they can all be investigated for actual causation

and/or correlation to the problem statement. As items are verified to not be the cause, they can

remain on the diagram; however colored in a manner that displays their status as a potential

cause.

7.2 Fishbone Analysis and Discussion

Since there are four performance measurements, less than perfect performance in each should be researched and analyzed for root cause and corrective action implementation to improve the process and performance. One fishbone diagram would not be appropriate for all performance measurements, as all the potential causes of not being on-time will not be the same as not delivering the correct quantity. We did however group LIFR and LIQFR fishbone diagrams together, as they are ultimately the same issue of count errors, just at different levels. The resulting fishbone diagrams are followed below.

The fishbone diagrams developed as part of this thesis are in the brainstorming phase. The causes listed on different branches were discussed with CVS/pharmacy personnel and were verified as valid potential issues. As part of the future development of the certification program, CVS/pharmacy team can use these diagrams to analyze the relationship of these potential causes to their respective operational inefficiencies.
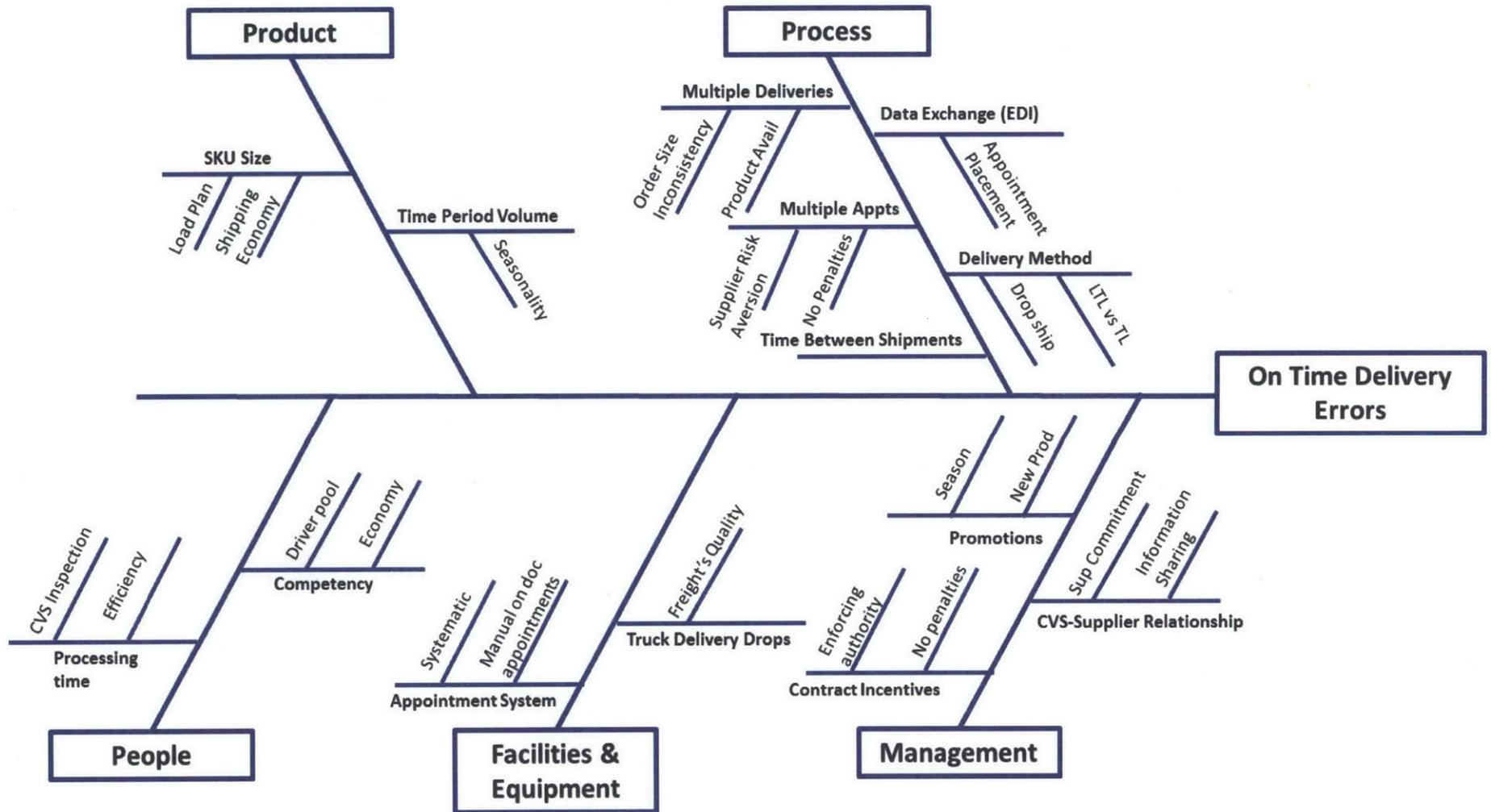
Figure 6 – On-Time Delivery Error Fishbone Diagram
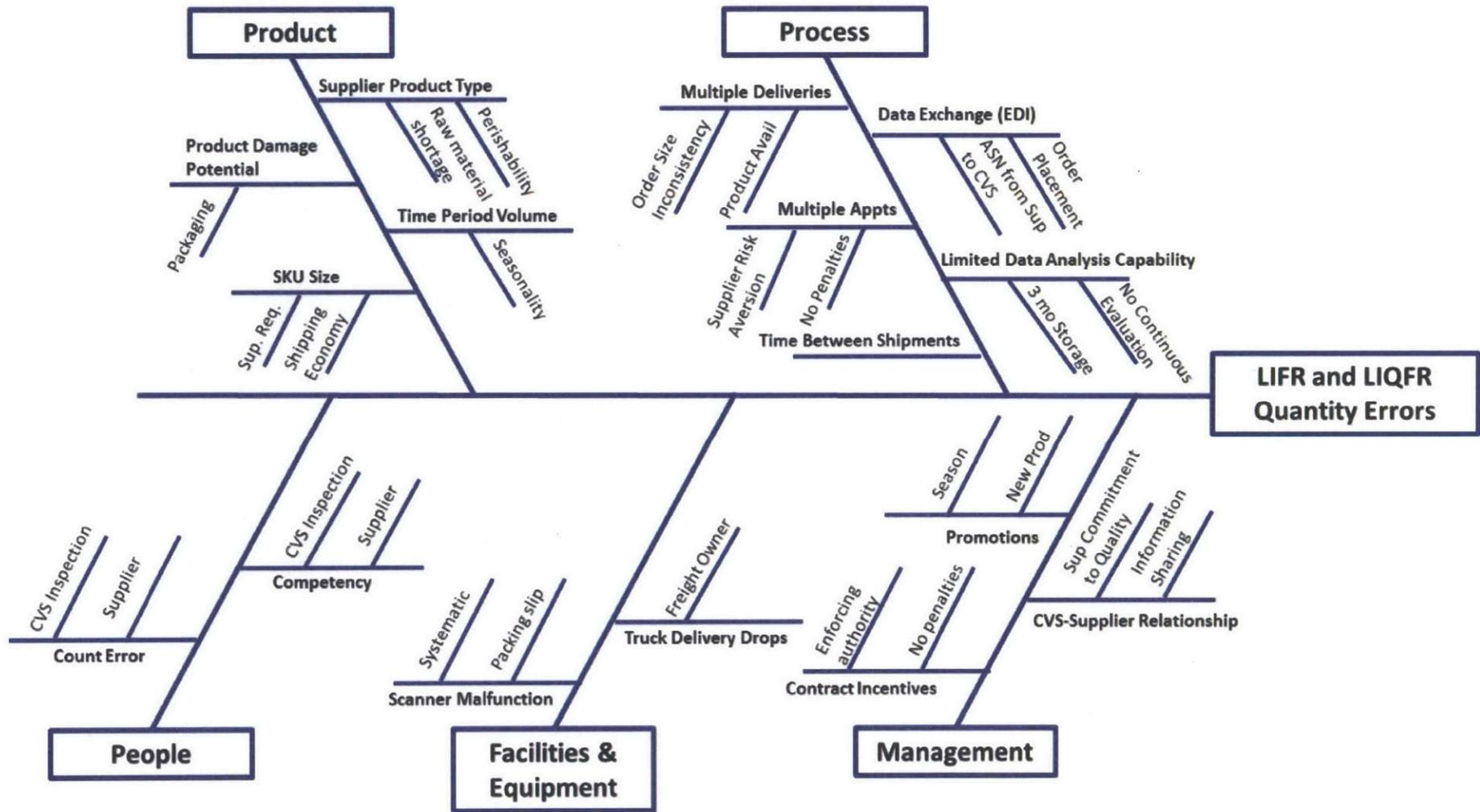
Figure 7 – LIFR/LIQFR Quantity Error Fishbone Diagram
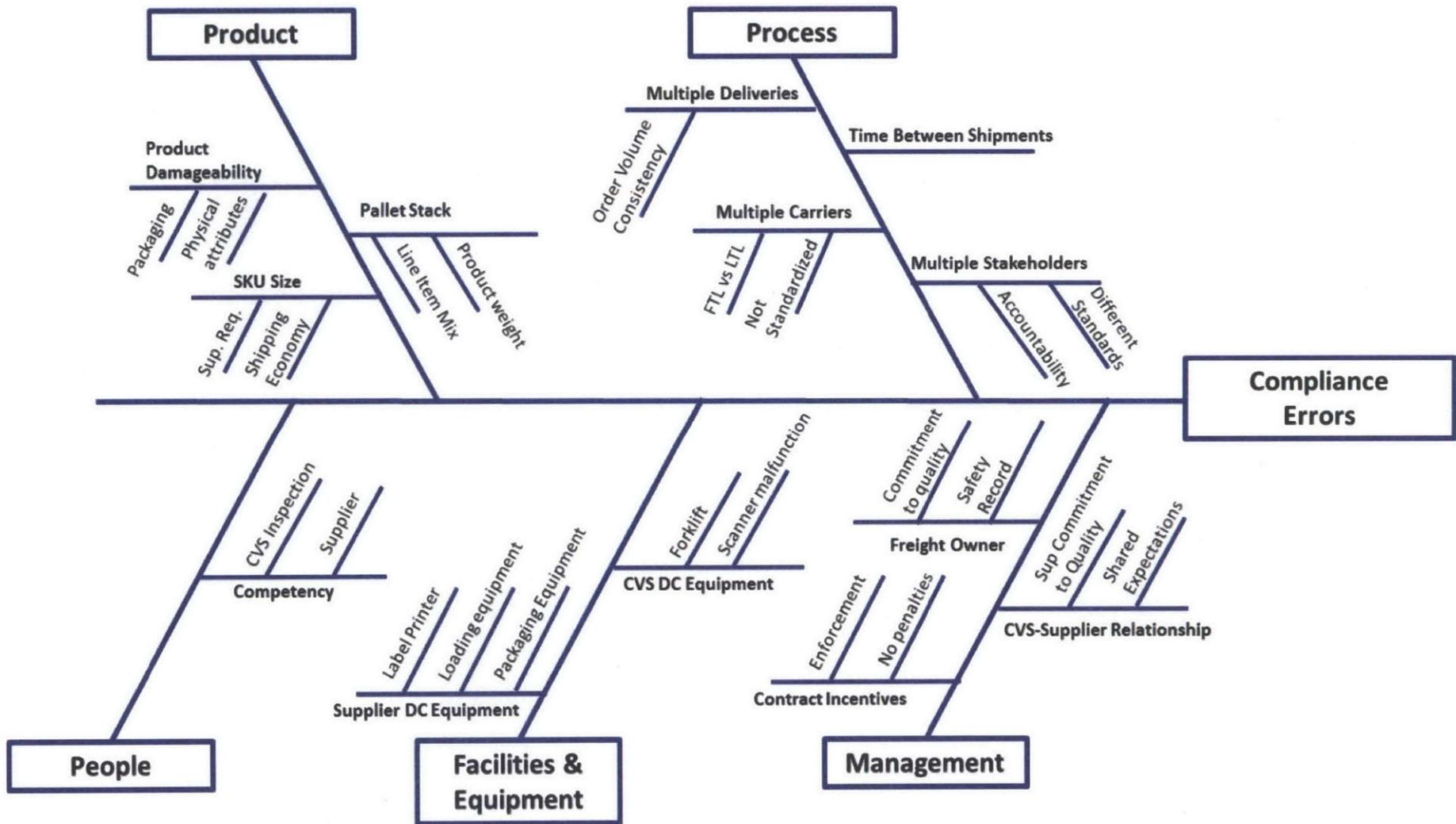
Figure 8 – Compliance Error Fishbone Diagram

8 DISCUSSION

8.1 Results

In the outset of this thesis project, CVS/pharmacy was interested in understanding and quantifying the performance of their suppliers as the basis of a certification program, where benefits and incentives could be introduced to improve efficiency and reduce the cost of the inbound supply chain activities. While CVS/pharmacy did not have a quantitative measurement system in place, they did have an assumed qualitative and intuitive understanding of which suppliers were their best performers. In our analysis of performance based on the three month sample set of data, we expected to find one supplier outperforming; however that proved not to be the case.

According to CVS/pharmacy's assumed intuition, Supplier 1 was believed to be the worst performing supplier, Supplier 2 was believed to be the best performing supplier and Supplier 3 fell somewhere in between. The results of our analysis contradict this belief in two of the developed performance measurements, specifically that Supplier 1 outperforms Supplier 2 in OTD and LIFR at each of the three DCs. While Supplier 2 does outperform Supplier 1 in LIQFR and PC in the majority of cases (at each DC), there still remains one case where Supplier 1 outperforms in PC and another case where the test proved no statistical difference in PC.

Further results show that the suppliers' performance in relationship to one another and even itself between DCs differs throughout the four performance measurements. Specifically, while a Supplier in one DC may be the best performer in one measurement, the same supplier in another DC may be the worst. For example, Supplier 1 consistently has the best performance in

OTD, LIFR and LIQFR in DC N; however its worst performance in PC occurs in that same DC. Similarly, Supplier 1 is the top performer at DC N in OTD and LIFR, but is the worst of the three suppliers at DC N in LIQFR and PC.

These results are important to understand prior to implementing a certification program, as it sheds light into the operations of all parties (supplier, carrier and CVS/pharmacy themselves). These differing levels of performance would need to be investigated using root cause analysis techniques starting with the fishbone analysis in an effort to implement improvements where necessary.

Building on the development and calculations of confidence interval estimates, we were able to develop reliable and statistically sound sampling plans; however with the low sample sizes and high variability, the sample sizes required in many cases exceeded the typical size of a PO shipment, meaning that the sample inspection would result in a 100% inspection. While the actual sampling sizes developed using the three months data may not provide significant operational efficiency improvements, the methodology can be used with a larger data size effectively.

Lastly, using similar statistical analysis techniques, we developed a SPC P-Control charts to measure how historically controlled the performance has been. We found that for LIFR and LIQFR measurements, the SPC charts provided an effective indicator of past performance, even with the limited sample size. However, with OTD and PC, with the sample size being the count of POs, it was too small to gain any statistical understanding of control. In all four performance measurements, we recommend that more data be kept and the period length be increased to result in increased overall sample sizes, increased number of periods and reduced variability.

Although the analysis was conducted on a data set spanning only three months, the results provide a quantitative analysis of three selected suppliers at three DCs that enables CVS/pharmacy to systematically analyze operational performances on a small scale. These analysis results can be openly discussed with the three suppliers and the three DC operations managers to form a collaborative pilot project team for operational improvement initiatives.

8.2 Conclusions

In our literature review we were not able to identify a process that provided a continuous and sustainable supplier evaluation method to use as a baseline for improvement initiatives or certification programs. Hence we developed a process of performance measurement in the retail industry that can be used to quantitatively assess the supply chain's performance and enable an intelligent dialogue between the trade partners.

In this section we summarize our findings and discuss the challenges we faced during our analysis, the benefits that were identified and the risks that could arise of such applications in retail industry. We make further recommendations for both the future research work as well as practical applications.

8.2.1 Benefits

This thesis highlights the current quantitative state of a segment of CVS/pharmacy's inbound supply chain that may have been previously known but mainly through anecdotal evidence. These results were developed through use of statistical methods which are applicable to a wide range of suppliers' performance metrics and could be utilized across any CVS/pharmacy supplier-DC network. The systematic approach enables managers to compare

processes side by side in different locations and across different suppliers and reach comprehensive conclusions regarding the differences and similarities in the system.

8.2.2 Risks and Challenges

The main challenges we encountered in this thesis were 1. limited data period, 2. single point of view, and 3. lack of direct causal investigation. Our analysis was performed on only three months of data, which is the current maximum time span of data storage in CVS/pharmacy's information system. While the methodology used to analyze and compare the performance measures and develop sample sizes hold true for any data, the level of accuracy is questionable with the amount of available data. In addition, no seasonal variability or historical performance data were captured due to the limited time span.

Additionally, all performed analysis reflected only the final state of the shipments at CVS/pharmacy DCs without access to similar quality performance records at the supplier's or the carrier's point of view. Moreover, the final step of the process is performed by CVS/pharmacy personnel where some errors might have been introduced to the records, but our analysis did not distinguish between such errors, and therefore these errors could be misidentified as suppliers' deficiencies.

Finally the quality metrics were developed per CVS/pharmacy's requirements without considering quality control processes in place by the trade partners that contribute to the amount of non-value-add activities. This factor along with lack of differentiation between CVS/pharmacy and their trade partners' operations prevented our ability to pinpoint the causal relationships of deficiencies, except for a qualitative analysis based on interviews and discussions.

These risks and challenges are inherent to this type of analysis for any retailer. It is difficult to have visibility to all aspects of the entire supply chain at the same time as well as a complete historical record of operations; hence it is important to maintain a global perspective and tolerate an acceptable level of uncertainty.

8.3 Recommendations

In this section we provide recommendations based on our analysis and models for the initial stages of a supplier certification program. Additional research and planning are required to develop an executable and sustainable procedure for continuous improvement.

8.3.1 Recommendation for Implementation

Prior to implementation of a certification program using SPC, it is necessary for CVS/pharmacy to gather more data samples to refine and validate the statistical models developed in this thesis based on the three months of performance measurements. In addition, operational and financial impacts of the current processes must be studied more comprehensively to define the performance expectations, such as the SPC target level or required sample size confidence interval and margin of error, to assess the benefits of the suggested models for the entire inbound supply chain.

Once a performance model is defined, CVS/pharmacy must maintain a continuous record of all metrics and reevaluate the target levels periodically to ensure the model's soundness and relevance. Moreover, the data should be used for causal investigation and developing improvement strategies in collaboration with trade partners based on the causal relationships.

8.3.2 Recommendations for Further Analysis

Our analysis investigated only the receiving and compliance data however further analysis can be performed using other variables such as order frequency, carrier type, promotions, seasonality and trade volume that may contribute to the supply chain's efficiency and quality. In addition, investigating the operations of trade partners and carriers may provide additional insights that can benefit both CVS/pharmacy and the supply chain as a whole.

9 APPENDIX

9.1 Individual Confidence Interval Graphs
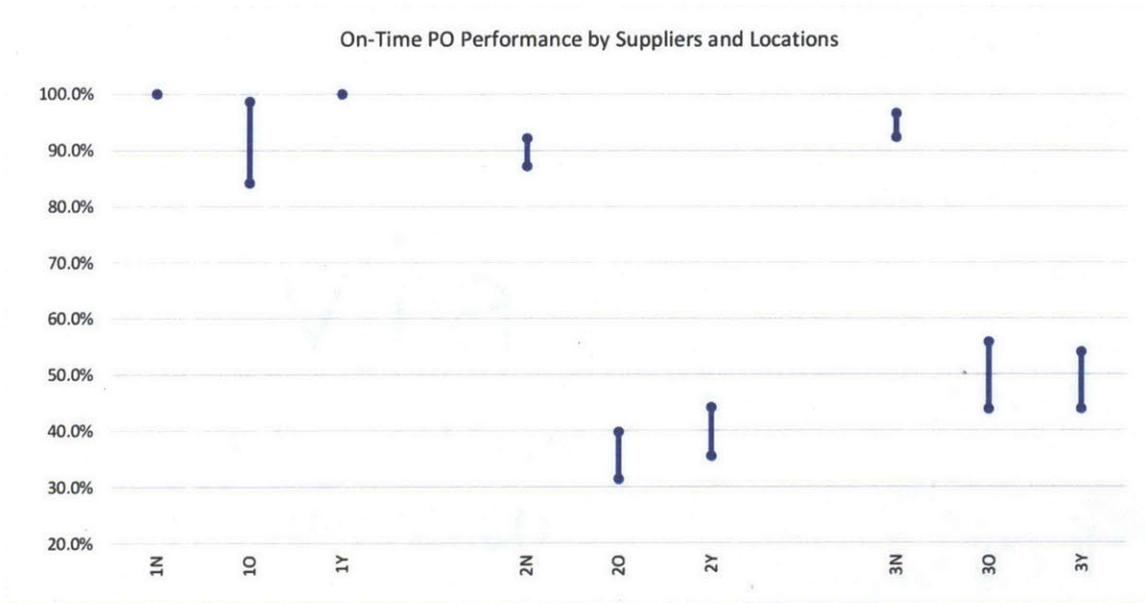
Figure 9 - OTD Individual Confidence Interval
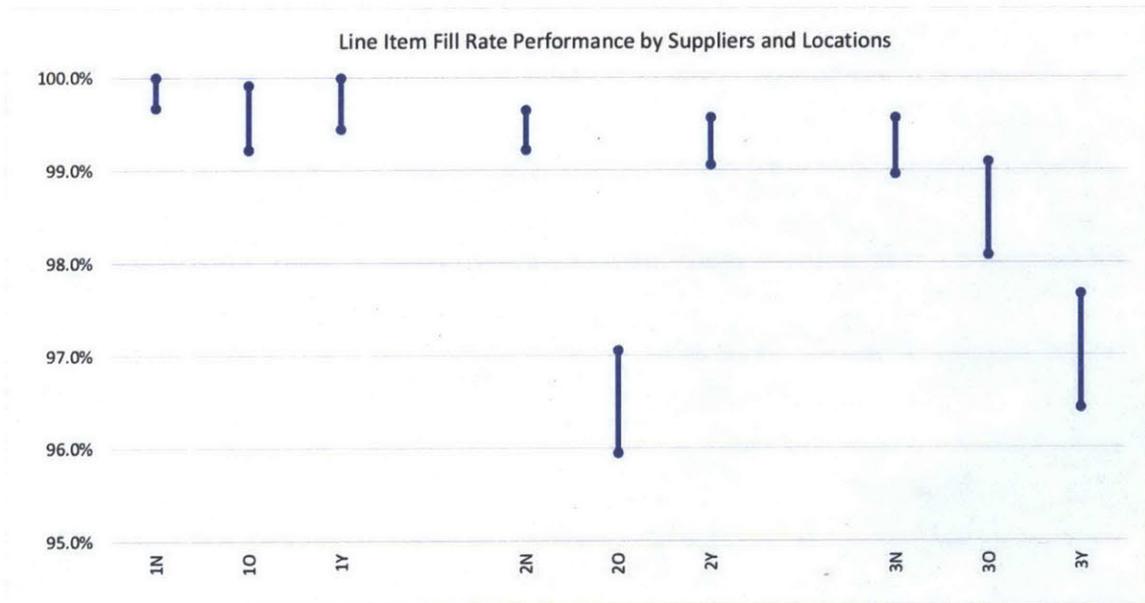
Figure 10 - LIFR Individual Confidence Interval
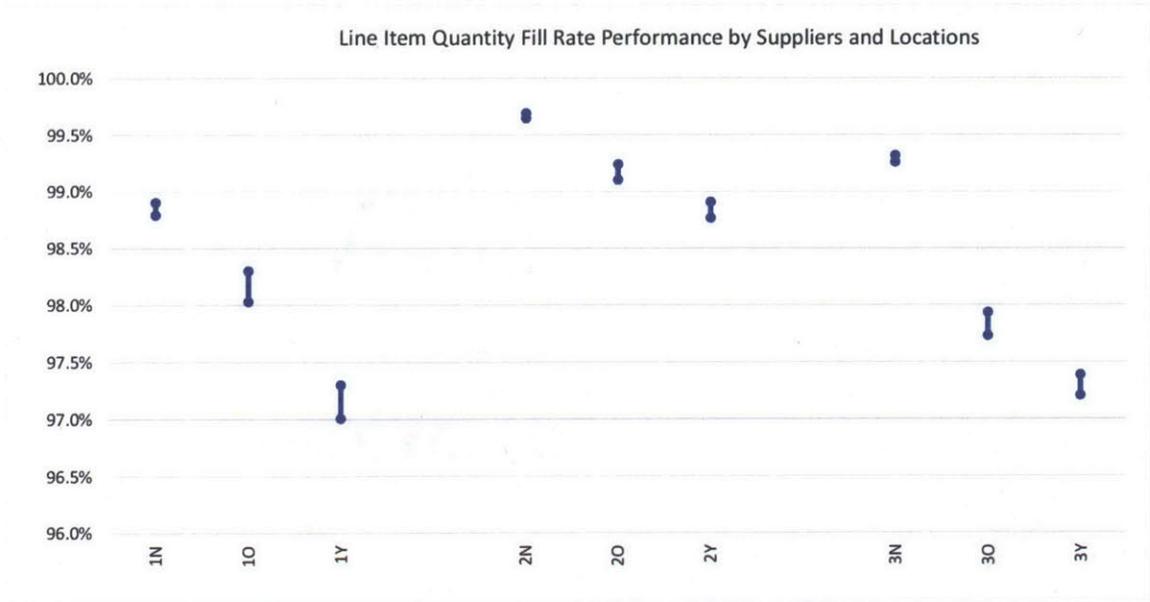
Figure 11 - LIQFR Individual Confidence Interval
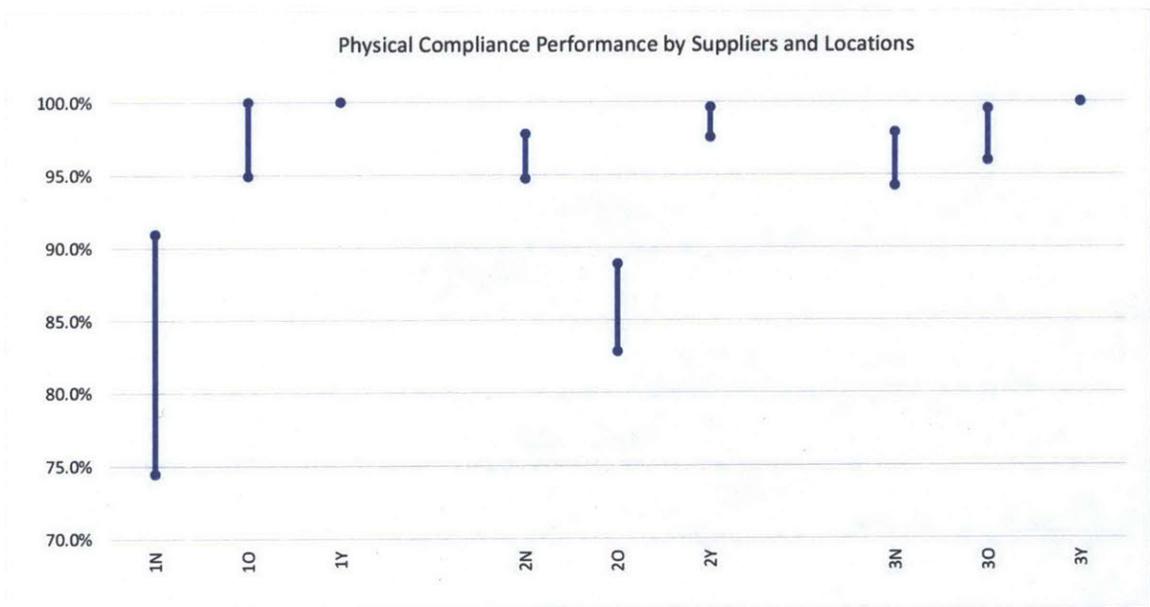
Figure 12 - PC Individual Confidence Interval

61

9.2   Comparison Confidence Interval Graphs
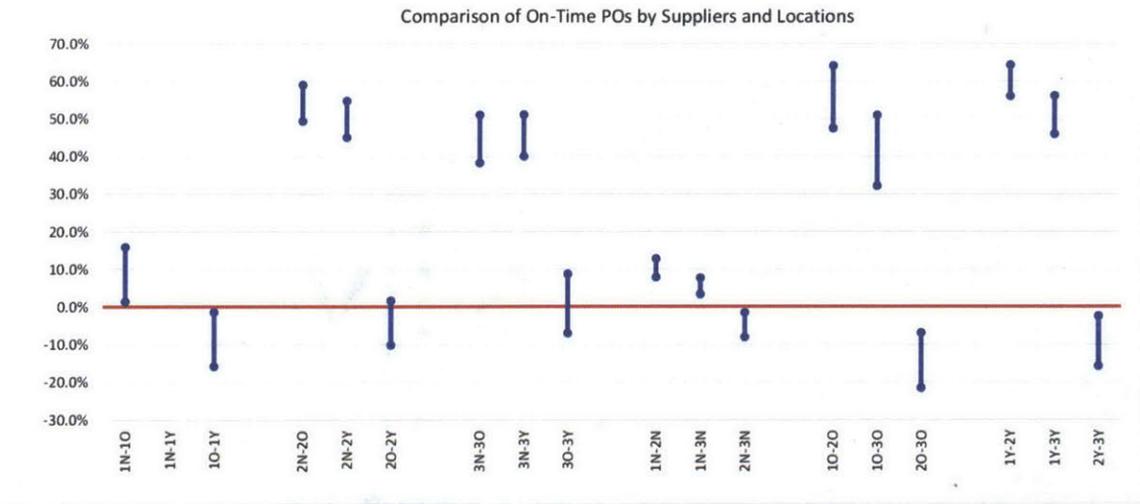

Figure 13 - OTD Comparison Confidence Interval
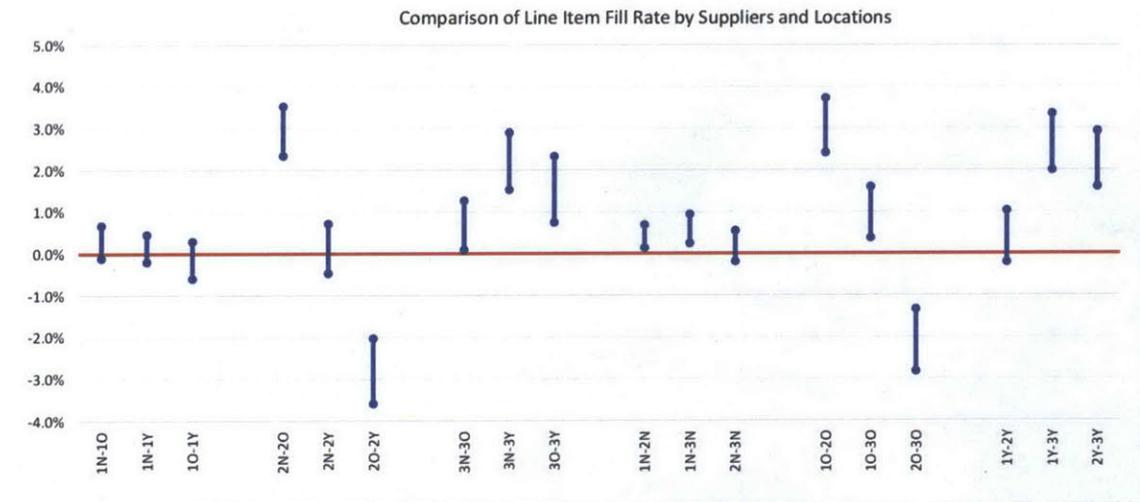

Figure 14 - LIFR Comparison Confidence Interval
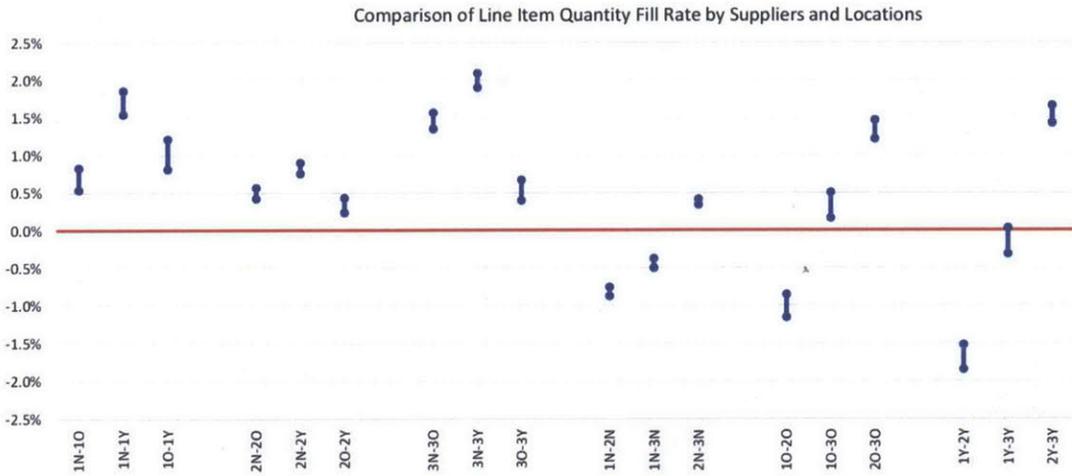
Figure 15 - LIQFR Comparison Confidence Interval
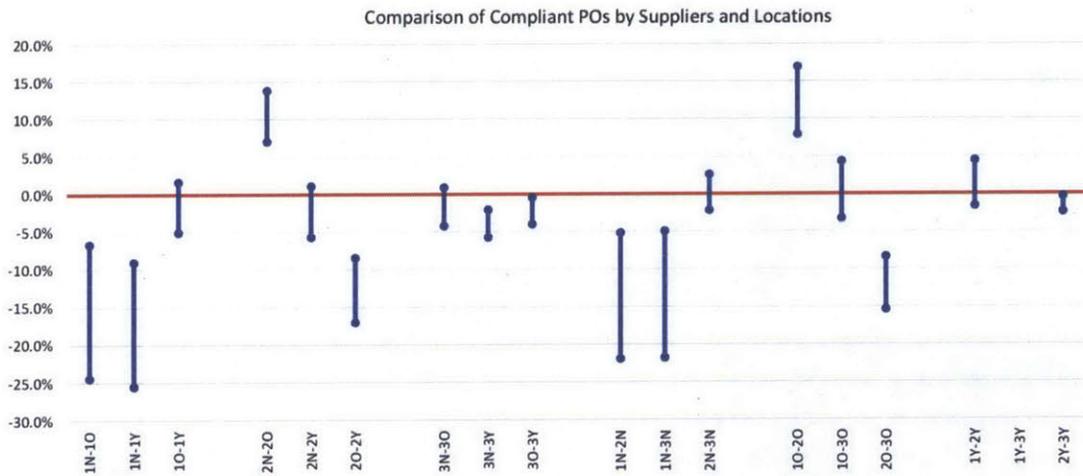
Figure 16 - PC Comparison Confidence Interval

10 REFERENCES

Accenture LLP. (2006). *Synchronization – The Next Generation of Business Partnering.*

Albright, S.C., Winston, W.L., & Zappe, C. (2011). *Data Analysis and Decision Making.* Mason, Ohio: South-Western / Cengage Learning

Bowersox, D.J. & LaHowchic, N.J. How Perfect Is a Perfect Order? (2008). *Industry Week/IW, 257*(11), 58–58.

Boyer Jr., J. E. (1999). The basics of high-performance supplier certification. *Hospital Material Management Quarterly, 20*(3), 4.

Brown, J. O. (1998). A practical approach to service-supplier certification. *Quality Progress, 31*(1), 35.

Corum, A. (Andrew R.). (2009). *Design and development of a supplier evaluation process* (Thesis). Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/49779

Coyle, J.J., Bardi, C., Langley, J.J. (1988). *The Management of Business Logistics.* St. Paul: West Pub. Co.

Mouzhi Ge, & Helfert, M. (2013). Impact of Information Quality on Supply Chain Decisions. *Journal of Computer Information Systems, 53*(4), 59–67.

Gunning, P., Horgan, J.M. and Yancey, W. (2004). Statistical Sampling in Auditing: Geometric Stratification of Accounting Data, *Contaduria y Administration*

Jung, S.-H., Kang, S.-H., & Ahn, C. (2003). Chi-Square Test for R×C Contingency Tables with Clustered Data. *Journal of Biopharmaceutical Statistics, 13*(2), 241.

Maass, R. A., Brown, J. O., & Bossert, J. L. (1990). *Supplier Certification: A Continuous Improvement Strategy.* Milwaukee, WI: Quality Press, American Society for Quality Control.

McGinnis, M. A., & Vallopra, R. M. (1999). Purchasing and Supplier Involvement in Process Improvement: A Source of Competitive Advantage. *Journal of Supply Chain Management, 35*(4), 42–50.

Morgan, C., & Dewhurst, A. (2007). Using SPC to measure a national supermarket chain's suppliers' performance. *International Journal of Operations & Production Management, 27*(8), 874–900.

Novack, R. A., Grenoble, W. L., & Goodbread, N. J. (1993). Teaching Quality in Logistics. *Journal of Business Logistics, 14*(2), 41–70.

Novack, R. A., & Thomas, D. J. (2004). The Challenges of Implementing the Perfect Order Concept. *Transportation Journal (American Society of Transportation & Logistics Inc), 43*(1), 5–16.

Tague, N. R. (2005). *Quality Toolbox.* American Society for Quality (ASQ).

Vitasek, K. & Manrodt, K. (2008). Benchmarking the Perfect Order: A Comparative Analysis of the Perfect Order in the Retail Industry. Retrieved from http://www.compliancenetworks.com/jdownloads/2008_perfect_order_index_study.pdf

Wang, S., Jiang, X., & Wang, Y. (2009). Intelligent process quality control system into supply chain (Vol. 7500, p. 75000C–75000C–9). doi:10.1117/12.858446