Better Contract – Better 3PL Performance

by

Mykola Oleksyn
Bachelor of Science in Project Management, Wentworth Institute of Technology, 2020

and

Tony S. Kook
Bachelor of Arts in Economics and International Relations, University of Southern California, 2013

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Signature of Author: _____
Department of Supply Chain Management
May 6, 2022

Signature of Author: _____
Department of Supply Chain Management
May 6, 2022

Certified by: _____
Dr. Maria Jesus Saenz
Executive Director, Supply Chain Management Program
Capstone Advisor

Certified by: _____
Dr. Mehdi Farahani
Postdoctoral Associate
Capstone Co-Advisor

Accepted by:

_____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Better Contract – Better 3PL Performance

by

Mykola Oleksyn

and

Tony S. Kook

Submitted to the Program in Supply Chain Management
on May 6, 2022 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Contract design and management are critical for Fast-Moving Consumer Goods (FMCG) companies. Well-designed contracts promote a value-driven relationship between the buyer and the supplier. Under the governance of such contracts, the service provider will proactively lead continuous improvement efforts in the direction that best aligns with the goals of the buyer. Such goals may include cost savings, service level, or context-dependent KPIs. On the other hand, poorly-designed contracts can lead to overspending and compromised performance. In 2017 Acme, a leading North American FMCG, designed a new contract model for managing its Third-Party Logistics (3PL) suppliers. The key differences introduced by the new contract were a stronger emphasis on cost savings, an enhanced review process of unbudgeted spending, and higher levels of risk/reward sharing. Acme is interested in learning whether the new contract created value for itself and the suppliers. To answer this question, we use the econometrics model, Difference-in-Differences (DiD). Our results show that despite the increase in 3PL spending over time, there is no evidence to suggest a causal impact of the new contract on the cost performance of the 3PLs. We then proceed with the discussion of how the new approach to governance provides some benefits to Acme since it promotes better alignment of financial incentives between the parties and protects the buyer against the risks of gross underperformance on the part of the supplier. Additionally, we develop a framework that helps the reader to understand the key structural elements of contracts that define the relationship between an FMCG buyer and a 3PL service provider. Such elements include Performance Measurements, Compensation/Incentive Structure, and Governance Process. We then describe the appropriate functional levers and the trade-offs that apply to each of the three structural elements of value-driven contracts.

Capstone Advisor: Dr. Maria Jesus Saenz
Title: Executive Director, Supply Chain Management Program

Capstone Co-Advisor: Dr. Mehdi Farahani
Title: Postdoctoral Associate

ACKNOWLEDGMENTS

I would like to thank our advisors, Dr. Maria Jesus Saenz and Dr. Mehdi Farahani, for their guidance and support throughout this project. A special thanks to Toby Gooley, for her attention to detail and tremendous support with our writing. Finally, I would also like to thank my wife, Ella, and the family for supporting my journey. Without your support, it would have been an impossible journey.

-Tony Kook

I thank our advisors, Dr. Maria Jesus Saenz and Dr. Mehdi Farahani, for the continuous support they provided to us on the long and winding road to completing this project. As a failed musician, I especially appreciate that they didn't deny my attempts to blend rock-n-roll themes with supply chain management in our poster presentations.

I am grateful to Toby Gooley because she didn't give up on trying to level up my writing skills. Toby saved this paper from the intrusion of far-fetching metaphors that would undoubtedly make me cringe upon revisiting this work in the future.

Finally, I thank my Ukrainian compatriots for the strength they display in the fight that defines a future worth living. I draw strength and inspiration from the unbreakable spirit of the Ukrainian nation. Слава Україні! ПТН ПНХ. Російський воєнний корабель…

-Mykola Oleksyn

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1 Introduction

The fast-moving consumer goods (FMCG) market in North America is served by supply chains that often rely on third-party logistics (3PL) providers. This means that FMCG companies outsource warehousing, distribution, and even certain stages of the manufacturing process. Outsourcing is an example of vertical supply chain collaboration in which the partnering companies are at different supply chain levels. Outsourcing allows companies to focus on their non-logistics core competencies and remain competitive in the industry.

In North America, the 3PL market is estimated at $252 bn. (Thuermer, 2021). On average, FMCG companies spend about 7.5% of sales revenue on 3PL services (Alicke & Lösch, 2010). This is significantly higher than the pharmaceutical industry at 1.8% and retail or high-tech at around 5%. With FMCG being a low-margin and competitive business, supply chains must be cost-effective and reliable in delivering products to the end-consumer. Consequently, 3PL contract design and management are critical for FMCG companies.

Optimal contracts promote efficiencies, increased accountability, continuous improvement, and value for all the parties involved. At the same time, sub-optimal contracts may create misaligned incentives, unfair evaluation, and financial compensation that lead to subpar 3PL performance coupled with an excessive amount of governance required. As a result, suboptimal 3PL contract design and management can cause erosion of value up to 9 percent of customer's annual revenues (Belotserkovskiy et al., 2018).

1.1 Problem Statement and Key Research Question

For this project, we have partnered with Acme – a large FMCG company that operates globally. The company has partnerships with six different 3PLs that manage over 20 facilities to ensure North America product distribution. These partnerships have been previously managed under the

pain-share-gain-share (PSGS) model (PSGS model is commonly used by the supplier and buyer to share cost savings or budget overruns to align the goals of two parties).

However, Acme felt that the PSGS model did not adequately incentivize the 3PL to drive process improvements or cost-saving initiatives. Therefore, in an attempt to improve 3PLs' performance in the two areas, Acme has developed a contract model called Value Partner Award (VPA).

VPA is an open-book (commonly known as cost-plus) contract. Based on this contract, Acme's payments to its partners consist of the expenses of running the facility plus a fixed base management fee (BMF) multiplied by a performance-based coefficient. BMF is roughly 6% of the site's budget and can be inflated (up to 185%) or deflated (down to 15%) depending on the operational excellence and cost-savings achieved by the supplier. Therefore, higher expenses may result in an increased management fee (although unjustified spending leads to losses for the 3PLs due to performance penalties). The process of determining proper spending and required performance is managed using key performance indicators (KPIs).

Historically, the incentives at Acme are intensely focused on cost savings (~65% of weight), while operational KPIs like on-time delivery, safety, and employee retention rates have received less weight (~35%). The VPA assumes a relatively stable business environment, i.e., consistent demand patterns, readily available workforce, typical procurement costs, and reliable supplier lead times. However, the timing of the switch to the new contracts coincided with turbulent times for the FMCG industry.

The implementation of the VPA contracts at the facilities we analyzed was initiated within six months before the COVID-19 pandemic. As the costs of the 3PL services kept rising, Acme decided to investigate whether the new contract was the driving cause of deteriorating

performance. We identified several reasons why the VPA contract could fall short in the conditions of an unstable business environment:

- Occasionally, sharp increases in demand call for a sacrifice in margins for the sake of throughput since the goal is to maximize the overall profit. If the contract overvalues efficiencies, it can promote losing profits to maintain the KPIs. As a result, the contract compensation system may be misaligned with broader business objectives. In these situations, KPIs designed for efficiency may prevent 3PLs from doing what is best for their partner.
- Labor shortages in the market may lead to increased expenses. To account for shortages, Acme may need to offer higher pay to attract new workers or supplement the current workforce through overtime. Thus, the legacy KPIs may fail to distinguish warranted spending from inefficient management.
- Volatile demand may require 3PL partners to carry excess labor capacity for addressing upswings in production loads. The KPI may penalize such extra spending.

The pandemic is not the single source of problems with the VPA model. It only illustrates the broader point – contracts between FMCG companies and 3PLs need to be resilient in a changing business environment. Acme's 3PL partnerships are built around the principles of transparency, shared risk/reward, and win-win solutions. However, the changing business environment results in Acme spending additional effort and resources on governing the contracts. For example, a supplier manages resources according to the forecasts, and when demand overshoots the estimates, more labor spending is required. The contract stipulates strict financial controls and, as a result, an elaborate review of extra spending is conducted. Management of both companies

usually arrives at mutually accepted amendments to compensations and evaluations, but the process can be time-consuming and, thus, expensive.

Acme is currently engaged in a phased process of automation of the facilities used by the 3PL providers. Acme estimates that the transition can take years for each site. Up to 65% of employees may be gradually reduced during such a process. The VPA model may not be suitable to measure the accountability of suppliers during the transition period and after the efficiency gains have been realized.

The VPA model may not be a suitable model given the increasing governing costs from the changing business environments. Therefore, the project's goal is to answer the following key research questions:

1. What is the financial impact of switching from the PSGS model to the VPA model?
2. What are the key elements and trade-offs of designing a value-driven 3PL service contract?

Our research presents a quantitative analysis of the impact of the new contract. We then create a framework of how companies can develop, design, and manage their service contracts and supplement it with discussion and recommendations.

For quantitative analysis, we outline fundamental assumptions, review different causal inference frameworks, and choose the most appropriate modeling method to compare how much more or less effective the new contract (VPA) is compared to the previous one (PSGS). From the literature review, we learn to pay particular attention to measures of uncertainty in the business environment.

For the recommendations section, we explore the results of quantitative research and apply the concepts from the literature review to present compelling arguments for contract-specific attributes that drive the outcomes we observed.

2 Literature Review

The main objective of this paper is to identify how much value (in terms of cost-savings and operational performance measurement) can be gained or lost from improving the design of service contracts. We seek to identify how to improve the strategic alignment of the buyer (Acme) and its service providers. In other words, how can the service contract be designed to better motivate the service provider to not only provide the basic service but also go beyond and act as a strategic partner that can drive process improvements and cost-savings on behalf of its client?

First, we cover the different types of service contracts prevalent in the current 3PL market with different trade-offs between cost (design and governance) and benefits (value creation – cost-savings or performance improvement). Then, we cover the key structural elements of effective contracts.

2.1 Types of Supply Chain Contracts

Various goods, services, and information services activities are vertically integrated through outsourcing and contracts within the wide range of the supply chain spectrum. Raw materials and goods are purchased from upstream suppliers through purchasing contracts. These raw materials and goods can be manufactured or assembled into a finished product and distributed to the downstream customers through service contracts, such as those for warehousing and transportation. In addition, software contracts can be used to purchase software (or licenses) to manage the information and financial flow throughout a supply chain. For example, the

sponsoring company Acme uses the help of 3PL partners for staffing and running warehousing and distribution facilities. This is why, for this study, we will focus on the service contracts that dictate the management and relationship between FMCG companies and 3PLs in the context of warehousing and distribution operations.

2.2 3PL Warehouse Service Contracts

3PL warehousing contracts can be generalized into two types: closed-book and open-book contracts. Closed-book contracts set pre-defined prices for services, and suppliers assume all the risks associated with operations. When open-book contracts are used, suppliers bill clients for the actual costs of delivering the service while adding an agreed-upon profit margin. In certain cases, companies may form hybrid models to use both types (BEROE Advantage Procurement, 2016).

2.2.1 Closed Book Contracts

Closed book contracts have set prices, and associated fees are charged accordingly. The 3PL margins are built into the pricing and hidden from the buyer. This margin is used to protect the supplier from business uncertainties and risks.

Closed book contracts are generally used for standardized services and are built around simple metrics such as cost per transaction, cost per labor hour, or cost per materials used. Due to their simplicity, closed book contracts have lower administration costs for functions like drafting, management, and governance. Therefore, such contracts are suitable for smaller operations with standardized service requirements. On the other hand, closed book contracts can suggest a lack of trust due to their transactional nature. Since the supplier is getting paid per transaction, it may engage in opportunistic behaviors due to misalignment of goals (BEROE Advantage Procurement, 2016).

2.2.2 Open-Book Contracts

Under open-book contracts, known as cost-plus contracts, the buyer remunerates the suppliers for the general operating expenses. The buyer also pays an additional margin to the supplier. Examples of margins include lump-sum agreements, performance-based awards, performance-based incentives (PBI), or pain-share-gain-share (PSGS) (BEROE Advantage Procurement, 2016).

PBI is one of the most widely used types of agreements in open-book contracts. Under PBI, the buyer compensates the supplier based on pre-determined goals or objectives. Some examples of the goals include cost-savings initiatives or on-time shipping performance measurement. The buyer and supplier pre-define the work scope and the performance-based compensation structure. Both parties continuously monitor and review the performance to drive efficiency through their partnership. Under PSGS, the buyer and supplier agree on a target budget. Cost savings by the supplier are shared between both parties. The sharing ratio or the expectations are pre-determined.

Open-book contracts are designed to optimize performance and generate cost savings for the buyer. They incentivize the supplier with monetary compensation that can align the goals of the buyer and the supplier. Such contracts are commonly used in larger operations with high service specificity where per-unit-basis calculations are difficult. However, the effort and costs associated with contract governance are high compared to closed book contracts. Open-book contracts require pre-determined performance metrics and key results or achievements. They are designed for stable business environments and may not deliver desired results when normal operations are disrupted. Acme (our sponsoring FMCG company) is a dominant market player in North America. Open-book contracts are the preferred choice for the governance of large

operations (BEROE Advantage Procurement, 2016) and therefore are the focus of our research. Given the benefits and administration costs of open-book contracts, it was necessary for us to study the core trade-off of contract design. After we understood the factors that tend to drive performance and costs in the process of contract governance, we were better equipped to perform quantitative analysis as part of our research.

2.3 The Core Trade-Off of Contract Governance Optimization

Contract development requires careful consideration of some trade-offs. For example, detail-oriented contracts have their advantages, but the cost to administer and govern such contracts can be high. The following are significant trade-offs related to the costs and benefits of 3PL open-book contract administration.

Cost Structure of Contract Administration. Ideally, long-term supply chain partnership contracts are designed to drive value for all parties involved. However, it can be costly to implement and manage such commercial agreements. One way to look at cost structure is to consider the lifecycle of contracts (Belotserkovskiy et al., 2018):

1. Precontracting
2. Contract writing
3. Implementation and vendor management

The Value of Potential Improvements. Studies show that just in the procurement function alone, poor design, lack of funding, and managerial neglect of contracts can lead to erosion of value estimated at around 9% of the client's annual revenues (Belotserkovskiy et al., 2018). To put it in perspective, the five-year average (2017-2021) net profit margin (net profit divided by total revenues) for S&P 500 companies is 10.8% (Butters, 2021). While it is prudent to recognize that

the above numbers are just averages, the writing is on the wall – the negative financial impact of poor contract governance measures in trillions of USD.

The Trade-Off: Balancing Costs of Governance with Value Derived. Well-coordinated and flexible contracts can be so costly to administer that the contract designer will prefer a simple agreement even if it does not fully optimize the company's performance. Further still, extremely simple contracts can lead to nearly optimal performance (Cachon, 2003). The emerging theme is that the process of design and administration of contracts comes down to finding the balance between the following two performance attributes:

1. Incremental costs of designing and administering the contract as compared to what Cachon (2003) calls naive and simple contracts
2. Incremental value driven by the contract when compared to naive and simple contract alternatives

For example, an extremely simple open-book 3PL contract might be presented as something like "all expenses are covered, and 10% is added as an extra charge for the services provided." We can theorize that such a contract would be easy to initiate but is unlikely to perform well. It is a carte blanche for overspending. If a client wants an effective contract, some degree of managerial oversight is required. Now, the question is whether more oversight is always better. And the answer, unsurprisingly, is: up to a point. This conclusion is not a deep insight that requires extensive studies. Rather, it stems from the fact that 3PL contracts are designed to optimize the efficiency of collaboration between partnering companies. The value contracts create is usually proportional to the waste they help eliminate, and such value is finite. Now, with finite value at stake, there can only be a limited amount of governance resources that can be justified as an

expense to unlock such value. Clearly, there is a point where extra spending on contract administration yields insufficient value to justify the added expense. As a result, we can say that an ideal open-book contract would achieve maximum net value output from the supplier in the context of optimal spending.

Usually, open-book contracts with 3PLs are used when a company outsources one or more organizational processes. Contracts play important roles in governing such processes. In our research, we limit ourselves to analyzing the contract models governing warehousing and finished goods distribution. In the sections that follow, we review the critical components of well-performing contracts.

2.4 Key Elements of Effective 3PL Contracts

Key elements in contracts include an effective performance measurement system, a set of appropriate target goals and expectations, and a sensible incentive structure for the 3PL.

2.4.1 Effective Performance Measurement Systems

Open-book contracts are usually explicitly built around quantifiable measurements of key performance indicators (KPIs). Many experts agree that this is a prudent approach. For example, Belotserkovskiy et al. (2018) claim that based on experience, such a formal agreement with rewards and penalties tied to performance levels is the key to optimizing the entire contracting process.

Performance measurement is critical to contract governance, as it measures the efficiency, effectiveness, and capability of an action, a process, or a system against a given norm or target (Fortuin & Korsten, 1988). According to Drucker (1963), effectiveness is a measure of doing the right job, while efficiency is the measure of doing the job right.

The Evolution of Performance Measurement and Accountability. According to Bourne et al. (2003), modern management accounting (prerequisite of performance measurement systems) started taking shape in the United States between the 1850s and 1920s in the context of the following business trends inside the individual organizations:

- Shift from piece-work payments to wage-based employment
- Activity expansion from single to multiple operations
- Shift from integrated production plants to vertical integration
- Move from individual businesses to multi-divisional firms

As time has shown, the above trends stay relevant and align with the challenges that open-book contracts are designed to tackle. Traditional accounting-based performance measures are financially based, internally focused, backward-looking, and more concerned with local departmental performance than with the overall success of the business (Bourne et al., 2003). In the 1990s, more comprehensive performance measurements systems were developed; the most notable is called Balanced Scorecard.

Balanced Scorecard Overview. Balanced Scorecard puts strategy and vision instead of control at the center (Kaplan & Norton, 1992). It sets goals but leaves behaviors and actions up to the people who are responsible for achieving them. At the same time, effective measurements are an integral part of the management process (Kaplan & Norton, 1993).

Balanced Scorecard establishes four high-level business goals (perspectives to consider) and then links three to five performance measures to each of them:

1. Customer perspective (How do customers see us?)
2. Internal perspective (What must we excel at?)
3. Innovation and learning perspective (Can we continue improving and creating value?)
4. Financial perspective (How do we look to shareholders?) (Kaplan & Norton, 1992)

3PL contract design predominantly focuses on the questions of improving the internal perspective of the business. The internal perspective considers productivity measures and prospects of future competitiveness in the market. Balanced Scorecard puts performance measurement in the context of a larger perspective of sustainable business practice, i.e., balancing short and long-term objectives while considering a broader set of stakeholders. In Kaplan and Norton's view, the Balanced Scorecard should be the cornerstone of how the business operates, and it should be the core of the management system and not the measurement system (Kaplan & Norton, 1993).

Design of Performance Measurement Systems. No individual performance measurement is sufficient in itself, but a comprehensive and balanced combination of measurements, i.e., a performance measurement system, is required. Such a system is at the core of the evaluation and governance of open-book contracts. Hwang (2011), in his analysis and adaptation of the ideas of A. D. Neely (2002), states that effective performance measurement systems are composed of well-defined metrics, reviewal processes of the metrics and behavior changes.

According to Hwang (2011), Balanced Scorecard by itself does not drive the behavioral changes within an organization. He continues that Balanced Scorecard reviews tend to transform into

once-a-month formal rituals that are quickly forgotten until the following monthly meeting. To define an effective metrics set, we need to understand the value and purpose behind every measurement. A. D. Neely (2002) outlines the different purposes of measuring performance. For our context, the measurement should serve to evaluate the performance and motivate the 3PL to improve performance. In addition, it should serve as a guide to better identify areas of improvement.

The above are high-level objectives. Meanwhile, many voices in the field attempt to define the attributes and specific goals of performance measures. Kennerley and Neely (2003) outline the characteristics of good performance measures through a comprehensive list (Kennerley and Neely, 2003). In summary, the performance measures should consist of specific goals that are derived from the business strategy to serve as an informative measure for improvement and provide accurate feedback on the performance.

However, 3PL contracts are redesigned and renegotiated every three years on average (BEROE Advantage Procurement, 2016). Changing business environments can lead to shifts in either the business strategy or the performance measurement. Consequently, the performance measures outlined in the original contract can become outdated long before the contract expiration or renegotiation date.

Seven deadly sins of performance measurement. Hammer (2007) shares the “seven deadly sins of performance measurement.” In summary, an organization should design, measure, and evaluate performance objectively and justly. In the context of 3PL contracts, the managing party measures not itself but an external partner, and the metrics are the basis for periodic evaluation and compensation. In such cases, the managing party should be careful of possible

‘provincialism’ (letting organizational boundaries and concerns dictate performance metrics) or ‘narcissism’ (measuring from one’s own point of view) in the design and measurement of performance.

Temporal Tense of Performance Measurement. According to Melnyk et al. (2004), one of the important attributes of a metric is its temporal tense, i.e., the depiction of historical performance, the present state, or the future. The two tenses usually identified are:

- Lagging indicators – outcome-oriented indicators that analyze the past and ideally provide the data to improve the future
- Leading indicators – predictive indicators that are likely to correlate strongly with the KPIs of interest

3PL contracts are often focused on lagging indicators of actual performance provided. On the one hand, it only makes sense since such contracts are designed for evaluation and compensation, as opposed to hands-on management of the partners. On the other hand, better integration of leading performance indicators could potentially help with navigating the changes in the business environment.

2.4.2 Goals and Expectations

Setting proper goals and expectations is one of the key attributes of an effective 3PL contract. Establishing appropriate performance goals can be extremely challenging. This aspect of contract design is outcome-binding for the 3PL, which means it has to be accepted by both sides. Lengthy negotiations are not uncommon at this stage. Literature in the field of goal-setting tends to agree that effective goals are high enough to encourage elevated performance but low enough so they

appear attainable. Setting too high goals jeopardizes motivation and commitment and may even create a culture of dishonesty and cutting corners (Bennett, 2003).

Already in the early 1990s, professionals in the field started recognizing that performance measures need to be dynamic to remain relevant in the ever-changing business environment (Dixon et al., 1990). Maintaining open feedback channels between management and operations is essential to achieving optimal business performance. Efficiency benchmarking is often used for setting baselines in contracts. This approach may work reasonably well in some situations when there is reliable historical performance data and the changes in the business environment are minimal. Unfortunately, benchmarking turned into a fad, and most companies who tried it achieved very little in return because it is a lot easier to benchmark a process than to benchmark output (Kaplan & Norton, 1993). Still, 3PL open-book contracts tend to set goals in terms of output measures, e.g., unit volume processed, resources consumed, or savings achieved.

2.4.3 Incentive Structures

Why Proper Incentives Are Important. Another critical element for effective 3PL contracts is a sensible incentives structure. Proper performance measurement and goal setting will only work if the incentives and compensation are well designed. According to Eccles (1991), aligning incentives with the measurement system is a crucial component of the successful implementation of a measurement system.

Cachon (2003) states that poor performance in supply chains may not be due to the incompetency of managers or poorly designed policies but rather misaligned incentives. The alignment of interests is one of the keys to creating sensible incentive structures. It is also crucial to predict and manage unintended negative consequences (also known as the “cobra effect” (Hartley, Dale, 2016)) related to contract design incentives and performance requirement aspects.

In fact, renowned economist, Steven Levitt (2010), calls “the law of unintended consequences” to be one of the most powerful laws in the universe that applies to virtually anyone in any industry.

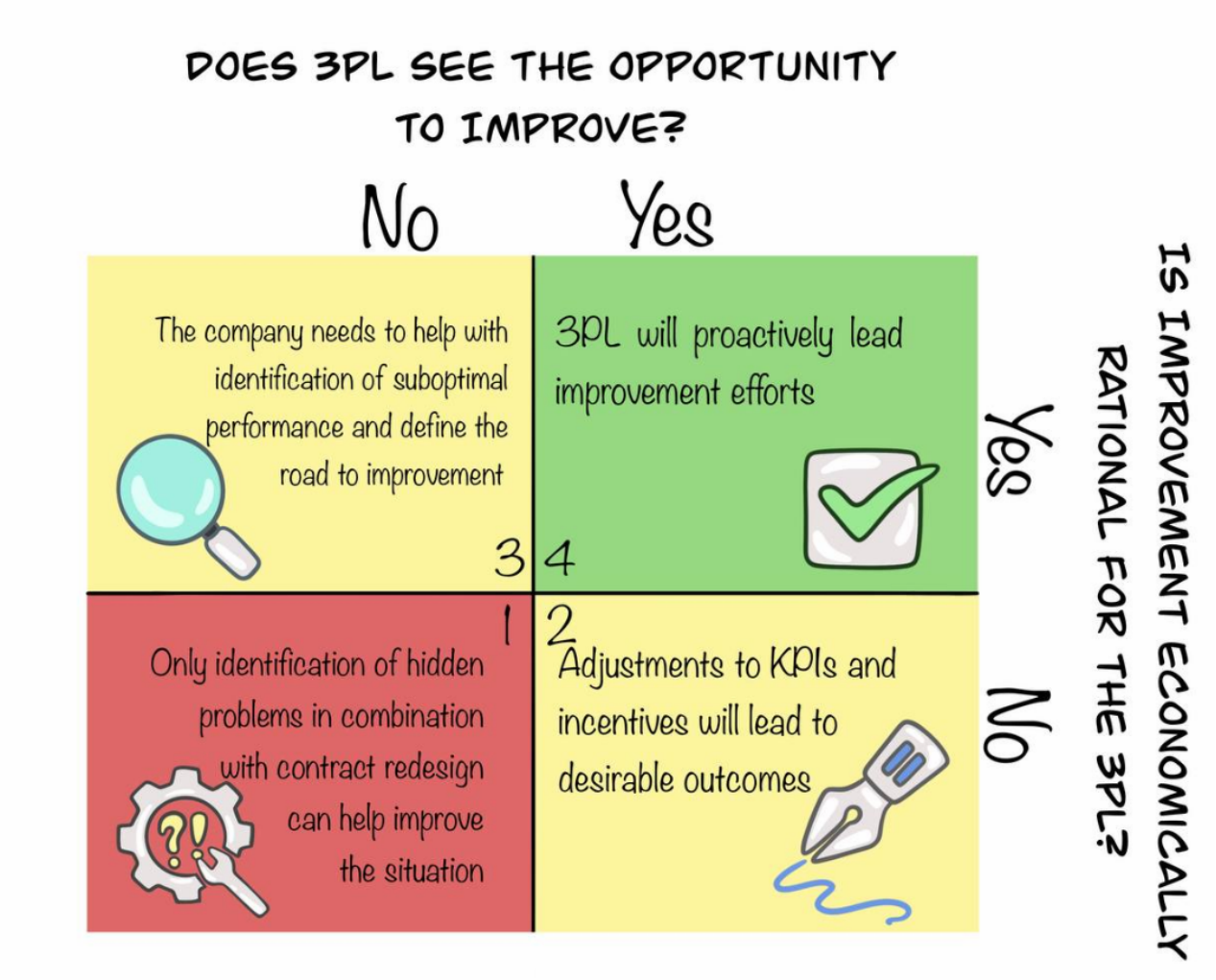
The Four Scenarios of Incentives Alignment and Problem Detection. A 3PL may or may not be aware of suboptimal performance (which in reality is defined not by the contract but by their ability to execute). Also, there might or might not be an economic reason for the 3PL to improve. Based on our review of the performance measurement systems, goal-setting approaches, and incentives structures, we have identified the following four possible scenarios represented in the quadrants in Figure 1:

1. The 3PL is neither aware of the suboptimal performance nor has sufficient economic reason to fix it upon identification. It is a complicated issue that requires identification of the problem and change of the incentives in the contract.
2. The 3PL is aware that they can do better, but the lack of incentives prevents them from taking action. This problem can be solved by adjusting the incentives within the contract and by sharing the benefits of the improvements.
3. The 3PL is unaware that making specific improvements is possible and will positively affect its profitability. This is a problem of integrated management, and to find the solution, the contract designer might need to get involved in the process management procedures. Doing so can be risky, and the potential damage from micromanagement can be significant in terms of governance costs and supplier morale deterioration.

4. The 3PL is aware of the suboptimal performance and has the incentives and the ability to improve. This scenario calls for no action on the part of the contract designer. Instead, the 3PL takes the initiative and leads the improvement efforts.

Figure 1

The Four Scenarios of Process Improvements Initiatives Between a Client and a 3PL



Quadrant 2 represents the low-hanging fruit of improving the performance of the 3PL (that is, when measured relative to quadrant 1 or quadrant 3). It represents a scenario where the 3PL sees the opportunity to make economically rational improvements. Quadrant 3 represents a scenario

where the 3PL does not see the opportunity for improvement, while it is economically rational. In this case, training might be the solution to fixing the problem since, according to Couzens (2006), training can improve the success rate of performance improvement initiatives.

2.5 Supply Chain Resiliency

Even with the well-designed performance measurement systems and incentive structures, supply chain contracts may not perform well in the event of disruptions. Therefore, supply chains should be resilient in a way to withstand and manage the stresses caused by disruptions or variability of the environment (McDonald, 2006). In the context of Acme's warehousing operations, the primary risk factors of the environment are labor shortages and demand volatility, which have been exacerbated by the recent COVID-19 pandemic.

Per Wieland and Wallenburg (2013), there are two dimensions to resiliency: robustness and agility. Robustness refers to how a firm can prevent or resist external shock or disruptions, while agility refers to how a firm can recover after a shock. Through their analysis, Wieland and Wallenburg argue that communication between the buyer (in our case, Acme) and supplier (3PL) positively influences resilience. Through better communication, firms can improve visibility and better anticipate future changes (Wieland & Wallenburg, 2013). Furthermore, if the relationships are based on cooperative and extensive communication, agility can be improved. Better relationships can increase the willingness to make sensitive information visible to partners, which allows responsiveness to disruptions. (Wieland & Wallenburg, 2013).

2.5.1 Resillience through Performance Measurement Systems Design

Per Bühler and Wallenburg (2016), accounting for external turbulence through performance measurement systems (PMS) can improve organizational resilience. By incorporating risk metrics to account for external turbulence, an organization can be more proactive (monitor risks and prepare for disruptions) and reactive (detect and respond faster to disruptions). As a result, improvements in supply chain resilience can indirectly enhance the service performance of an organization (Bühler et al., 2016).

2.5.2 Social Capital - Benefits of Collaboration

Wieland and Wallenburg (2013) claim that maintaining a cooperative relationship with supply chain partners can also positively affect resiliency. The benefits mentioned align with Coleman's (1990) definition of social capital, a valuable asset that can be gained through social relationships. Nahapiet and Ghosal (1998) break down the social capital into three dimensions: cognitive, relational, and structural, as shown below. They highlight how social capital can facilitate the overall performance (efficiency and creativity) in buyer-supplier relationships. Through social capital, firms can gain access to valuable information, reduce monitoring costs, and suppress opportunistic behaviors (Nahapiet, 1998).

Cognitive Social Capital - refers to the shared values, visions, or goals. Through common understanding and approach, firms can promote harmony of interests

Relational Social Capital – refers to trust, obligations, respect, and friendships. Through engagement in open communication and transparency, firms can promote cooperative behaviors among stakeholders.

Structural Social Capital – refers to a pattern of connections between parties. Through the creation of a structure with dense interactions and multiple connections, firms can exchange reliable and diverse information.

2.5.3 Social Capital - Dark Side of Collaboration

Collaboration does not always guarantee performance improvements. Per Villena et al. (2011), there is an inverted curvilinear relationship between social capital and performance. Despite the potential benefits, social capital, if taken to an extreme, can be very costly to administer. As social capital passes the optimal level or the threshold level, "risks of opportunism, loss of objectivity, ineffective decision making, and costly investments may begin to outweigh the benefits" (Villena et al., 2011).

2.6 Literature Review Conclusion

Existing literature has shown that outsourcing core logistics functions can lead to cost reduction and performance improvement inside organizations (Stank et al., 2001). A substantial body of theoretical and empirical research has been developed in the context of performance measurement and evaluation. However, most of the studies cannot be directly applied to the problem of evaluation and cost-benefit analysis of open-book 3PL commercial contracts. It may be due to the specifics of the context in which such contracts are developed and governed:

- Contracts primarily focus on the lagging performance indicators of value output
- Two different parties conduct performance evaluation and day-to-day management
- Impact of resiliency is difficult to measure and is not captured in the contracts
    - ~~Collaborative relationships and elements of social capital are not captured in contracts~~

- It is hard to evaluate the effectiveness of contracts because:
    - The cost of governance is often difficult to estimate. There usually are no records of management costs specifically attributable to contract governance efforts. Several stakeholders with different roles at a client company are involved in periodic review and management of contracts. Over the years, key employees change their roles, and it is incredibly challenging to get an accurate measure of how much more or less governance effort is spent on the current contract compared to the one that was in place a few years ago.
    - Causal relationship between the design and effectiveness of contracts is hard to establish. We should recognize that contract is just one of many things that could impact the KPIs of effectiveness that we measure. Even with large amounts of historically extensive data, one has to use some strong assumptions to develop a model where, in theory, the impact of the contract is isolated and presented as causal.

Additionally, the extent of the impact of contract design on supply chain resilience requires more research. Wieland and Wallenburg (2012) believe that the agility and robustness of a supply chain (which are integral to resilience) have non-negligible effects on business performance. Contract design that promotes the robustness of the supply chain can deliver additional financial benefits because it mitigates the impact of negative outcomes that could have happened. There is a literature gap in empirical research of such dynamics in 3PL service contracts.

Moreover, we identified a literature gap in empirical research regarding the causal impact of open-book 3PL service contracts on the cost-effectiveness of collaboration. In the following

chapters, our research and analysis have attempted to address that gap by applying the methodology described in Chapter 3.

3 Methodology

Our project aims to evaluate the effectiveness of the new contract model (VPA) against the previous model (PSGS). We approached the project in four phases. We first started with a qualitative investigation by conducting interviews with stakeholders and reviewing relevant literature in the field. Second, we collected (received) the warehouse operations and costs data. Third, we cleaned and analyzed the data and developed a model from the analysis. Finally, in the fourth phase, we explore the results of quantitative research and apply the concepts of the qualitative reviews to share business insights with our sponsor.

3.1 Operational Context

Acme has two different types of facilities: mixing centers and supply warehouses. Our research was focused on the mixing centers. Mixing centers receive products from different supply warehouses or manufacturing sites and consolidate customers' orders.

Below are definitions that will be used throughout this chapter:

- Approved Budget – the operating budget that includes but is not limited to all labor costs, equipment maintenance, repair, and rental costs divided into 12 months
- Actual Cost – the actual cost of operating the site.
- Should-Cost or Flex Budget – the budget as adjusted each month by mutual agreement of both parties for changes in throughput volume and other factors that are beyond 3PLs' control.

- Labor Index – indicates the employment and wage information for workers in the region to determine the wage percentile. Acme uses this to determine the hourly wage of its warehouse employees. It also could help indicate the changes in the labor workforce in the industry.
- Base Management Fee (BMF) – indicates the constant dollar amount based on the percentage of the approved budget

3.2 Data Gathering

Acme's Canada site first transitioned from PSGS to VPA in 2017. Other sites have asynchronously transitioned from the PSGS model to the VPA from 2017 through 2020. We originally requested the data from 2016 through 2021 for all 20 warehouses that Acme operated in North America. However, due to the differing nature of the warehouses, we narrowed our focus to the three of the mixing center warehouses located in Texas, Illinois, and Pennsylvania. Furthermore, Acme's internal and external stakeholders have validated that these locations were the most representative of the projects at hand.

3.2.1 Stakeholder Interviews

From our interviews with Acme's internal and external stakeholders, we learned more about the operational context of its mixing centers. For Acme, these centers are used to consolidate partial pallets or cases from different inbound plants to ship to different retail facilities. Due to its operating nature, mixing centers' main handling units are pallets, and the majority of its workforce is forklift operators.

3.2.2 Data Retrieval

We received the monthly summary of the following data for the three warehouses, as shown in Table 1. We also received an additional timeline detailing the month and the year of the

transition from PSGS to VPA. Furthermore, we acquired the state-level unemployment rate from the U.S. Bureau of Labor Statistics (U.S. Bureau of Labor Statistics Data).

Table 1

List of Provided Data

Data Group	Variables	
Cost	Actual Cost	Should-Cost
	Budgeted Cost	
	Base Management Fee	Cost Savings
Volume	Forecast Pallet Qty	Forecasted Cases
	Actual Pallet Qty	Actual Cases
Labor	Employee (Headcount)	Unemployment Rate
	Hourly Wages	Labor Cost (Wage_Emp)
KPIs	KPI 1	KPI4
	KPI 2	KPI5
	KPI 3	KPI6
VPA (Treatment)	Binary Value	

Note: KPI descriptions have been removed to maintain Acme's confidentiality

3.2.3 Data Preparation

The cost was the dependent variable in our analysis. We wanted to determine the impact of the VPA contract, volume, labor, and KPI on the cost-per-pallet. Instead of the Should-Cost, we used the Actual cost because it represents the total actual spend. We have also excluded the usage of BMF and cost savings in our analysis because the calculation of BMF is inconsistent throughout the timeline of the two contracts.

Based on our stakeholder interviews, we have identified that the main driver of volume at the mixing centers was pallets. Therefore, we have used pallets, instead of cases, as the main variable for our volume data. In addition, we have added an additional variable, CPP_SC (Should-Cost per pallet), by dividing the Should-Cost by the actual pallet quantity. Furthermore,

we have added a new variable, FE_ABS, for forecast error by taking the absolute value of the difference between the actual and forecasted pallet quantity.

For labor information, we have used all four variables: employee, hourly wages, labor cost, and the unemployment rate specific to the location. The labor cost information was the product of hourly wages and the number of employees in a given month. For hourly wages, we used the historical wages (location specific) of the forklift drivers, as the majority of the workers at the mixing centers were forklift operators.

For operational KPI information, we needed to remove five of the KPIs as they were inconsistent from the two periods. Furthermore, from our stakeholder interviews, we have learned that while the Should-Cost review process can adjust for fair compensation when an event is beyond the 3PL's control, it may not always adjust the KPIs of recorded performance. Therefore, even using the remaining KPI was questionable.

3.2.4 Feature Selection

Through stakeholder interviews, we determined that the four features would be important in estimating the costs: actual pallets, forecast error, labor cost information, seasonality impact, and unemployment rate.

Contrary to our expectations, certain features showed little to no predictive power. For example, the impact of seasonality on cost-per-pallet performance was consistently insignificant for all facilities. Also, unemployment rate measured as a state-level average percentage for a given month (U.S. Bureau of Labor Statistics Data) showed no predictive power in our modeling.

In addition, we used the Least Absolute Shrinkage and Selection Operator (LASSO) method to select the features for our model. The LASSO method was used to remove redundant variables

and reduce overfitting (Fonti, V., & Belitser, E. (2017). From this method, the LASSO method selected the following parameters: actual pallets, labor cost, and forecast error. We then tested for multicollinearity between independent variables in the data for both facilities. Kutner et al. (2005) describe two methods of diagnosing multicollinearity: informal and formal.

The informal method consists of analyzing absolute values of Pearson correlation coefficients between the pairs of independent variables. The informal method is not rigorous and has significant limitations. We include the results of multicollinearity reports in Table 2 and Table 3.

Table 2

Multicollinearity Report for Independent Variables at the Texas Facility

	VPA	Wage_Emp	Actual_pallets	Forecast_AE
VPA	1.000000	0.775158	0.273558	0.434652
Wage_Emp	0.775158	1.000000	0.614758	0.438411
Actual_pallets	0.273558	0.614758	1.000000	0.104078
Forecast_AE	0.434652	0.438411	0.104078	1.000000

Table 3

Multicollinearity Report for Independent Variables at the Illinois Facility

	VPA	Wage_Emp	Actual_pallets	Forecast_AE
VPA	1.000000	0.738524	-0.497463	0.698199
Wage_Emp	0.738524	1.000000	-0.143859	0.452629
Actual_pallets	-0.497463	-0.143859	1.000000	-0.643822
Forecast_AE	0.698199	0.452629	-0.643822	1.000000

We observe that some of the correlation values are relatively high (above 0.6). For our model, it is essential to ensure a lack of multicollinearity between the independent variables. Since the

results of the informal analysis are inconclusive, we use a more robust and formal approach of calculating the variance inflation factors (VIF) for each independent variable in the dataset we use for modeling. Kutner et al. (2005) suggest that VIF values over 10 are usually interpreted as signs of multicollinearity. Table 4 includes the VIF report of independent variables.

Table 4

Variance Inflation Factors (VIF) Report for Independent Variables at the IL and TX Facilities

Variables	Texas VIF	Illinois VIF
VPA	3	3.9
Wage_Emp	4.8	2.6
Actual_pallets	2	2
FE_ABS	1.3	2.5

We observe that all the VIF values are well below 10, so we can assume that there is no multicollinearity between the independent variables we use.

3.3 General experimental design

Our primary goal is not to deconstruct VPA and analyze its elements but rather to assess its impact on the performance of suppliers. Such an approach provides us with a natural experiment setting in new policy intervention. The defining feature of natural experiments is that group assignment is random or seemingly random (Messer, 2016). In our case, the assignment of VPA or no-VPA governance is random. This condition of random assignment holds since the implementation of VPA contracts was timed depending on the expiration dates of existing agreements. Historically, the process of negotiation and 3PL contract initiation depends on a variety of factors not directly within Acme's control. Conventionally, the expiration time of a

contract is set every three years. This makes the timings of contract expirations (and subsequently VPA initiations) essentially random.

Below are the assumptions that have enabled our analysis:

- The Should-Cost assessments of 3PL performance were conducted fairly across the facilities in the dataset.
- The accuracy of data provided was consistent throughout the period used in the analysis. Such data includes cost measurements (budgets, actual spending, and Should-Cost assessments), volume measures (forecasts and actuals), and the timeline of the contract changes.
- The effect of various exogenous factors was relatively the same across the 3PL site locations (the impact of COVID-19 is addressed separately).

3.4 Exogenous Factors and Preparation for Causal Analysis

Even though the reliability and integrity of the collected data were high, we have to approach causal inference analysis with caution. The facilities we analyzed have distinct locations that could be subject to different exogenous factors, the most notable of which is COVID-19. From our interviews with the stakeholders, from the point of view of operations management, we determined that the impact of COVID-19 meant the following:

- Sharp increase in market demand for a number of product categories
- Increased difficulties with human resource recruiting and management.

Failure to properly account for the above impacts can lead to faulty attribution of causal effects within our model. Moreover, both spikes in product demand and labor-management problems can have cascading effects on operational efficiencies.

Elevated demand translates into higher output, which consumes more operational expenditures. It would be incredibly difficult for an outsider to make sense of the cost-appropriate relationships between operational KPIs and the budget. However, Acme has a budget review and adjustment process, or the Should-Cost review. This means that demand volatility can be viewed as part of normal operations.

Labor shortages, high turnover rates, and declines in levels of experienced senior workforce members were all associated with COVID-19. We incorporated region-appropriate data provided by Acme's Labor Purchase team to control such effects. The team investigates and identifies the percentile range of hourly wages for similar warehouse workers in the local market. We also derived data from the U.S. Bureau of Labor Statistics. These measurements helped our model account for non-uniform regional effects of the pandemic and overall labor trends.

3.5 Choosing the Method of Causal Analysis

When choosing the statistical framework for causal inference, we considered the following models:

- Rubin's Causal Inference can be an excellent tool for clinical research and economic studies. It implies developing potential outcomes or "counterfactuals" (Rubin, 2005). Doing so can work well in large-scale randomized experiments. In our case, making speculations about potential outcomes of KPIs given alternative contracts is unwarranted.
- Structural Equation Modelling (SEM) is one of the most popular and (if properly applied) sophisticated frameworks for causal analysis. SEM used for exploratory research where theory is less developed is called Partial Least Squares Structural Equation Modeling (PLS-SEM) (Sarstedt et al., 2017). Its counterpart, Covariance-based Structural Equation Modeling (CB-SEM), is used to test pre-existing theories (Hair et al., 2017). Initially, we

were eager to apply PLS-SEM to developing our model since it doesn't require a preconceived theory and doesn't need a large sample size (Sarstedt et al., 2017). However, upon further reflection, we realized that PLS-SEM is best suited for establishing causal relationships between parameters with continuous ranges of outcomes (or at least multi-tier categorical). In our case, the variable of causal interest is binary, i.e., either presence or absence of the VPA contract of governance.

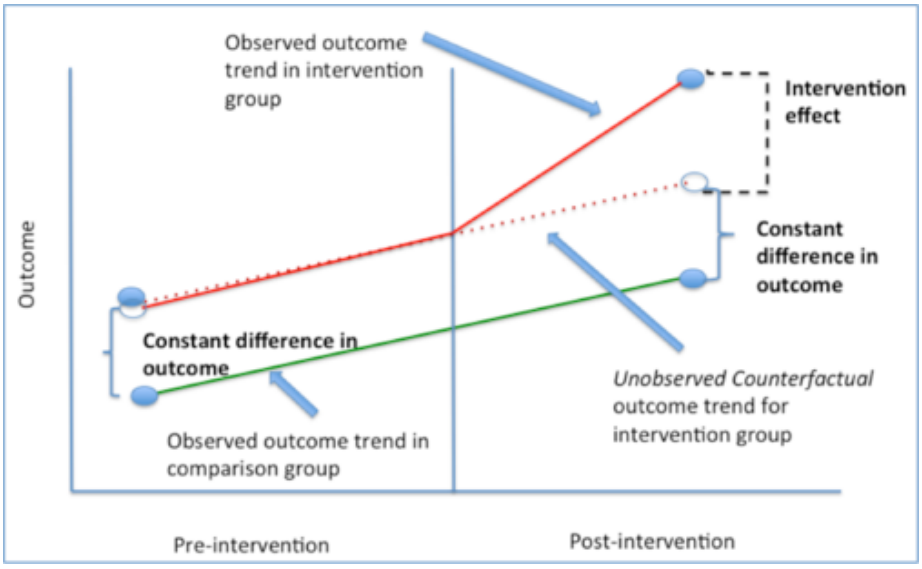
- Difference in Differences (DiD) is one of the most widely adopted methods of evaluation of causal effects of programs or policies (Jiménez & Perdiguero, 2019). Since the implementation of the VPA contract essentially means a new policy, the DiD was chosen as the most suitable framework and was used for further analysis.

3.6 Difference-in-Differences (DiD) Method Overview

The idea of DiD is simple. We need two similar groups (subjects of research; in our case, 3PL sites) in two different time periods. During the second period, one of the groups is affected (treated) by a new policy (VPA implementation). The goal is to assess how the treated group's performance changes when compared to the control group. We control double difference, i.e., changes over time and over the control group. This control aspect of double differencing removes biases in the second period (when we compare treatment and control groups) that could result from permanent differences between sites and the effects of time trends (Imbens & Wooldridge, 2009). Figure 2 provides a visual representation of the DiD method (Difference-in-Difference Estimation | Columbia Public Health, n.d.)

Figure 2

Visual Representation of the DiD Method (Difference-in-Differences Estimation) | Columbia Public Health, n.d.)



3.6.1 DiD Structure

As shown in Table 5, we will be measuring an outcome of interest denoted as "Y" and will populate a table similar to the one below (Jiménez & Perdiguero, 2019). Superscripts "TRT" and "CRT" above the variable of interest "Y" denote treatment and control groups. Subscripts "1" and "0" indicate whether the new contract (VPA) was in place during a given period.

Table 5

Group and Time Effects

		Time effect		
		Before (Be)	After (Af)	Difference (Af−Be)
Group effect	Control (Co)	Y_0^{TRT}	Y_1^{TRT}	$Y_1^{TRT} - Y_0^{TRT}$
	Treated (Tr)	Y_0^{CRT}	Y_1^{CRT}	$Y_1^{CRT} - Y_0^{CRT}$
Difference (Tr−Co)		$Y_0^{TRT} - Y_0^{CRT}$	$Y_1^{TRT} - Y_1^{CRT}$	$Y_1^{TRT} - Y_1^{CRT} - (Y_0^{TRT} - Y_0^{CRT})$

If we only consider the difference between before and after, we are ignoring the differences within each group (last column). At the same time, if we only consider the difference between the treated and control group, we are ignoring the time effects (last row). This is why we need to

evaluate the above differences simultaneously, i.e., assessing how the treatment group changes over time in relation to the control group.

Our estimation must include three binary variables: after (treatment has been implemented within the treatment group), treated (to denote that the group is the treatment group), and the interaction of the two (denoting the treatment group at the time when the treatment was implemented). As seen in (1), the mathematical foundation on which the DiD framework is built.

Equation 1

DiD Framework

$$Y = \beta_0 + \beta_1 After + \beta_2 Treated + \beta_3 After{*}Treated + \varepsilon_{it} \quad (1)$$

The explanation of the coefficients in (1) is presented in Table 6 (Jiménez & Perdiguero, 2019).

Table 6

Coefficients of Group and Time Effects

		Time effect		
		Before (Be)	After (Af)	Difference (Af−Be)
Group effect	Control (Co)	β_0	$\beta_0 + \beta_1$	β_1
	Treated (Tr)	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Difference (Tr−Co)		β_2	$\beta_2 + \beta_3$	β_3

The mathematical goal of the DiD method is the evaluation of the value and statistical significance of the coefficient β_3, which represents the impact of policy intervention on the outcome of interest.

3.6.2 DiD Assumptions

Like any causal inference method, DiD relies on several assumptions. We would like to reiterate the insightful words of Emeritus Professor of Statistics at Harvard University Donald Rubin:

“Causal inference is impossible without making assumptions, and they are the strands that link statistics to science. Assumptions are always needed and it is imperative that they be explicated and justified.” (Rubin, 2005, p. 324).

The following are our assumptions and justifications for the assumptions in the context of our research setting:

- Exogeneity assumption: As described by Lafontaine and Slade (2008), the policy or change has to be an exogenous change or what is also called a natural experiment, which is defined by three aspects:
    - Exogenous policy change (imposed externally)
    - Group observations that are affected by the change
    - Group observations that are not affected by the change

  The new policy intervention was implemented on a virtually random basis, and the data for both affected and unaffected groups were provided. Such a setting justifies all three of the above requirements, and we can consider the VPA implementation process a natural experiment.

- Identifying assumption: The temporal effect in the two groups is the same in the absence of the exogenous change. In other words, the changes between the groups during the pretreatment period have equal averages (also known as parallel trend assumption). To ensure this, we do an endogeneity test. This step is important but often overlooked in applied literature on DiD (Albalate, 2007). While looking only at the pretreatment period in both groups, we test whether the treatment and control groups have the same trends over time. If we fail to reject the statistical hypothesis of the two trends being the same,

we reaffirm the identifying assumption(Albalate, 2007) and thus have conducted a successful endogeneity test.

- Stable unit treatment value assumption (SUTVA). This comprises two sub-assumptions. First, it assumes no interference between groups, i.e., neither group is affected by what treatment the other has received. The facilities are managed according to established sets of KPIs and Standard Operating Procedures (SOPs). Implementation of VPAs at one of the facilities will not affect the operations of a different site.

  Second, SUTVA assumes that there are no hidden versions (variations) of treatments. The VPA contract was designed with the goal of consistency of governance across the facilities. After analyzing the contracts and interviewing the stakeholders, we have confirmed that there is no significant variation between the implementation of the VPA contract across different facilities.

4 Results

The objective of our DiD model is to find out if there was an improvement or deterioration in operations of the supplier attributable to a switch in the contract governance framework – and if so, by how much.

4.1 Primary Unit of Measuring Performance

In our case, the most suitable way of measuring the volume of work performed by the supplier is the actual amount of pallets they process. Processed pallets represent expected revenues downstream in the supply chain. Also, pallets of output are cost drivers at the 3PL facility. As a result, we chose Cost Per Pallet as the primary measurement unit of efficiency of the 3PL facility.

4.2 Relevant Data and Preliminary Analysis

Acme initiated the research project to learn whether the new contract promoted more effective relationships with the suppliers. The VPA contract model started getting implemented at Acme's partner 3PL facilities around 2017, but the exact timing depended on the expiration date of the existing contracts. Our data includes fiscal performance for the period of 2016-2021. In addition, we have detailed data for two mixing centers of Acme, at both of which the VPA contract was implemented throughout 2019.

When comparing pre-VPA and VPA contract periods, both facilities show a significant increase in cost-per-pallet metrics. We are interested in learning whether the new contract is causing such a trend or whether the relationship is correlational. For example, a significant increase in MAPE of forecast accuracy over time could be one of the reasons for the gradual increase in cost-per-pallet performance. In Table 7, we illustrate the increase in cost-per-pallet performance and decrease in forecast accuracy under old and new contract governance at the two facilities.

Table 7

Cost Performance and Forecast Accuracy at Acme's Mixing Center

Location	Contract Type	Average Cost-Per-Pallet*	Average MAPE of Forecasts
Texas	PSGS (old contract)	$10.00	8.80%
	VPA (new contract)	$11.76	15.90%
Illinois	PSGS (old contract)	$10.00	3.70%
	VPA (new contract)	$12.59	21.70%

Note: Average Cost-Per-Pallet shown are for illustrative purposes and are not actual values

A decrease in the reliability of forecasts during the VPA period is striking. We know that there was no significant change in forecasting methodology at Acme over the period in question,

which suggests that the differences in forecast errors are likely attributable to unpredictable changes in the business environment.

4.3 Relevant Independent Variables

As discussed in Section 3.2.4, below are the key independent variables listed below:

- "VPA" – a binary indicator of whether the VPA contract was administered during the reported period
- "Forecast_AE" – absolute error value (measured in pallets) of the forecast used for planning the operations for a given month.
- "Actual_pallets" – the number of pallets processed by the supplier in a given month
- "Wage_Emp" – the product of the current wage and the number of employees at the facility at a given month.

4.4 Analysis of Each Facility

In preparation for building the Difference in Differences (DiD) analysis, we constructed linear regression models to predict the value of monthly cost-per-pallet measurements at the Texas and Illinois facilities.

We succeeded in creating simple predictive models with significant explanatory power. Table 8 and Table 9 present the results of the Ordinary Least Squares (OLS) regression model predicting Cost-Per-Pallet (CPP) Performance at the Texas and Illinois facilities.

Table 8

Texas Facility - OLS Regression Result for Predicting Cost-Per-Pallet (CPP) Performance

OLS Regression Results

Dep. Variable:	CPP	R-squared:	0.723
Model:	OLS	Adj. R-squared:	0.703
Method:	Least Squares	F-statistic:	35.84
Date:	Fri, 22 Apr 2022	Prob (F-statistic):	9.97e-15
Time:	10:42:47	Log-Likelihood:	-39.709
No. Observations:	60	AIC:	89.42
Df Residuals:	55	BIC:	99.89
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	[redacted]	0.715	14.471	0.000	8.914	11.780
VPA	0.5752	0.233	2.466	0.017	0.108	1.043
Wage_Emp	0.0007	0.000	5.300	0.000	0.000	0.001
Actual_pallets	-2.204e-05	3.27e-06	-6.741	0.000	-2.86e-05	-1.55e-05
Forecast_AE	-5.253e-06	3.21e-06	-1.638	0.107	-1.17e-05	1.17e-06

Omnibus:	11.535	Durbin-Watson:	1.591
Prob(Omnibus):	0.003	Jarque-Bera (JB):	12.152
Skew:	0.883	Prob(JB):	0.00230
Kurtosis:	4.319	Cond. No.	3.41e+06

Note: Intercept value has been removed to maintain Acme's confidentiality

Table 9

Illinois Facility - OLS Regression Result for Predicting Cost-Per-Pallet (CPP) Performance

OLS Regression Results

Dep. Variable:	CPP	R-squared:	0.884
Model:	OLS	Adj. R-squared:	0.876
Method:	Least Squares	F-statistic:	105.0
Date:	Fri, 22 Apr 2022	Prob (F-statistic):	4.49e-25
Time:	10:50:34	Log-Likelihood:	-20.512
No. Observations:	60	AIC:	51.02
Df Residuals:	55	BIC:	61.49
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	[redacted]	0.838	10.084	0.000	6.769	10.128
VPA	0.0038	0.189	0.020	0.984	-0.375	0.383
Wage_Emp	0.0004	5.28e-05	7.601	0.000	0.000	0.001
Actual_pallets	-1.011e-05	1.41e-06	-7.166	0.000	-1.29e-05	-7.28e-06
Forecast_AE	2.223e-06	7.72e-07	2.881	0.006	6.77e-07	3.77e-06

Omnibus:	20.535	Durbin-Watson:	1.397
Prob(Omnibus):	0.000	Jarque-Bera (JB):	34.493
Skew:	1.150	Prob(JB):	3.24e-08
Kurtosis:	5.917	Cond. No.	1.09e+07

Note: Intercept value has been removed to maintain Acme's confidentiality

4.5 Comparing the Two Facilities

The results of the two models are consistent with each other and with our intuition regarding the forces affecting the costs of operations. The adjusted R squared of 0.876 for the Illinois facility model is so high that we should investigate it further. Upon reflection, we notice that a predictable trend in cost-per-pallet performance for this location persists even despite changes in contract type and a dramatic increase in forecast MAPE (from 4% to 22% when comparing the two periods of contract governance). In fact, the high variance of forecast errors at the Illinois facility provides our model with additional predictive power when compared to the more typical adjusted R squared of 0.7. for the Texas model. Table 10 reports the differences in absolute forecast errors (Forecast_AE) at the two facilities.

Table 10

The Comparison of Historical Absolute Forecast Error Values at the IL and TX Facilities

Facility/Period	Forecast_AE Mean	Forecast_AE Standard Deviation	Forecast_AE Coef. of Variation
Texas Overall	34898	22831	0.65
Texas Pre-VPA	27940	18521	0.66
Texas Post-VPA	48815	24674	0.51
Illinois Overall	70545	95612	1.36
Illinois Pre-VPA	21969	19226	0.88
Illinois Post-VPA	160758	114540	0.71

Forecast error as an independent variable of our OLS models yields a p-value of 0.006 in Illinois compared to 0.1 in Texas. The data suggests that every additional 1000 wrongly forecasted pallets at the Illinois facility add about 0.2 cents to the expected cost-per-pallet performance metric. To put this in perspective, the model suggests that every additional 6% in absolute percent error in forecasting at the Illinois facility is expected to increase cost-per-pallet by about 1%.

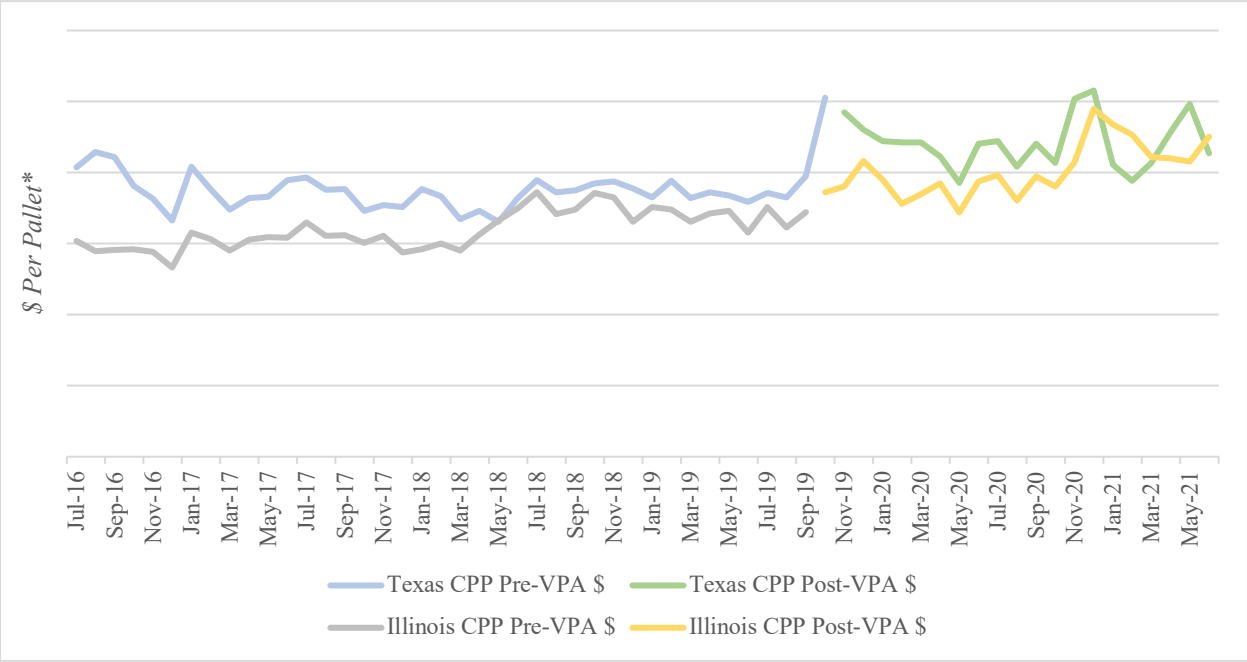
The Texas facility has a less predictable pattern of cost performance over time. The post-VPA time period (an independent variable) is statistically significant at predicting the elevated cost-per-pallet performance of about 5%. Meanwhile, the same time-period variable shows no statistical significance at predicting cost-per-pallet performance in Illinois. Also, when comparing Texas to Illinois, the model suggests a higher level of potential economies of volume

at the Texas facility. Every 3% increase in total pallet volume is expected to produce a 1% decrease in cost per pallet.

Figure 3 illustrates the cost-per-pallet (CPP) performance at the two facilities:

Figure 3

Historical Cost-Per-Pallet (CPP) Performance in TX and IL Facilities



Note: Actual values has been removed to maintain Acme's confidentiality

4.6 Difference in Differences and Causal Analysis

In our case, we used the Illinois facility as the control since cost-per-pallet performance there followed a more predictable pattern. We then re-created the post-VPA performance at the Illinois facility with the calculated parallel trend from the pre-VPA period. The impact of using such a parallel assumption was insignificant (MAPE of cost-per-pallet recalculated according to the modeling assumption compared to the actual cost-per-pallet is 5.2%). Still, such an approach provided additional consistency with the notion of a no-policy-change control mechanism.

Now, having the Illinois facility as control and looking into the effects of VPA implementation at the Texas facility, we can run the Difference in Differences model to measure the effects over time (pre and post-VPA) and over facility (Texas compared to Illinois). As a result, the variable of interest is the product of the Texas facility indicator and post-VPA time period indicator (denoted in our model as "Texas : Post_VPA_Period_at_TX").

The results of the Difference in Differences model are included in Table 11.

Table 11

DiD Result of Assessing the Causal Effect on Cost-Per-Pallet (CPP) of the VPA Contract

OLS Regression Results

Dep. Variable:	CPP	R-squared:	0.832
Model:	OLS	Adj. R-squared:	0.823
Method:	Least Squares	F-statistic:	93.22
Date:	Fri, 22 Apr 2022	Prob (F-statistic):	2.04e-41
Time:	11:07:41	Log-Likelihood:	-62.310
No. Observations:	120	AIC:	138.6
Df Residuals:	113	BIC:	158.1
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	[redacted]	0.768	11.969	0.000	7.673	10.717
Post_VPA_Period_at_TX	-0.0868	0.204	-0.426	0.671	-0.491	0.317
Texas	-0.6807	0.397	-1.715	0.089	-1.467	0.106
Wage_Emp	0.0004	5.13e-05	8.087	0.000	0.000	0.001
Actual_pallets	-1.144e-05	1.37e-06	-8.360	0.000	-1.42e-05	-8.73e-06
Forecast_AE	-1.185e-06	8.56e-07	-1.384	0.169	-2.88e-06	5.11e-07
Texas:Post_VPA_Period_at_TX	0.8499	0.196	4.333	0.000	0.461	1.238

Omnibus:	33.596	Durbin-Watson:	1.326
Prob(Omnibus):	0.000	Jarque-Bera (JB):	79.749
Skew:	1.082	Prob(JB):	4.82e-18
Kurtosis:	6.356	Cond. No.	1.07e+07

Note: Intercept value has been removed to maintain Acme's confidentiality

We observe that the Post-VPA time period is correlated with increased costs at both facilities. Moreover, it seems that the VPA may have increased the cost-per-pallet spending. While the p-value of the result is statistically significant, we should account for the extremely limited amount

of data we are able to use. However, we need to take into account the fact that we only had data for two facilities, which requires a robustness check on the model.

We use the treatment period randomization technique similar to what's known as the falsification test to evaluate whether semi-random dummy assignments of the "placebo treatment" period (given the same performance data) will lead to similar results. In their research, Whaley et al., (2021) suggest that finding associations within falsified scenarios suggests spurious primary results. In the case of a valid and robust model, only the correct assignment of the VPA intervention period would yield statistical significance of the result findings. As shown in Figure 4, we create two alternative dummy assignments of the policy intervention treatment for our Texas dataset. In Table 12, we compare the results of three DiD models, only one of which reflects the true timeline of the VPA contract implementation.

Figure 4

Timeline of Real and Dummy Treatment Assignments at the Texas Facility

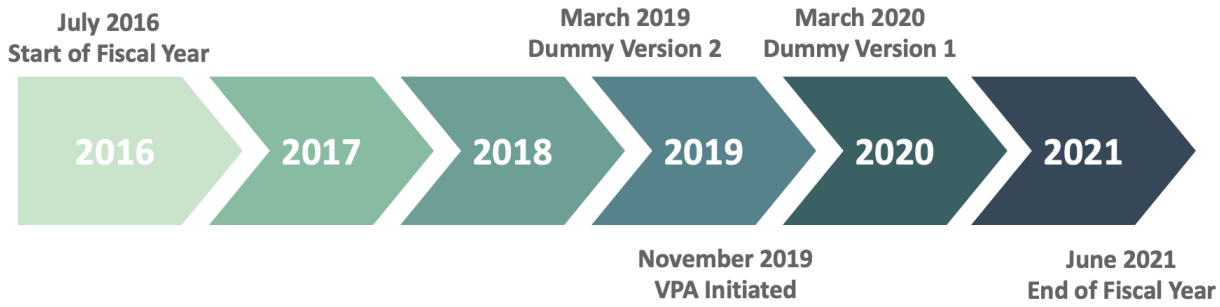


Table 12

Comparison of Models With Real vs. Dummy Treatment Assignment Parameters

Parameter of Comparison	Real VPA Treatment	Dummy Treatment Version 1	Dummy Treatment Version 2
Treatment Effect on CPP	$0.85	$0.66	$0.98
95% Confidence Interval	$0.46-1.24	$0.14-1.19	$0.62-1.34
Standard Error	$0.20	$0.27	$0.18
p-value	0.000	0.014	0.000
Adjusted R Squared	0.823	0.791	0.817

As we can see, our model also suggests similar results when we use dummy assignments of treatment periods. On these grounds, we suggest that there is no evidence of a causal link between the new contract implementation (VPA) and the increased cost-per-pallet performance at the Illinois and Texas facilities.

The finding is valuable because it helps Acme's management properly assess the impact of the VPA contract on cost performance. We lack the data to conclude whether the new governance model improves other essential aspects of the buyer-supplier relationship. But we believe this model better aligns interests between the buyer and the supplier through enhanced risk and

reward sharing. Also, more aligned interests reduce Acme's risk when it comes to possible events of severe financial underperformance on the supplier's part.

5 Discussion and Recommendations

We need to address the fact that our research findings are subject to significant limitations. Also, it is important to consider some explicit as well as implicit costs and benefits associated with Acme switching to the new contract.

5.1 Insights From Quantitative Results

The quantitative part of our research project aimed to assess the financial impact of Acme's switch from the old (PSGS) to the new (VPA) contract model. Acme effectively redefined the concept of "value" in its value-driven relationship with the 3PLs. After the change, much more emphasis was placed on cost-related KPIs. Also, overperformance or underperformance measured against budgeted goals were linked to higher rewards or penalties.

Our difference-in-differences analysis suggests no causal relationship between Acme's contract revision and the financial performance of its suppliers. From a managerial perspective, the result is surprising in two ways:

There was a measurable deterioration in cost performance of the 3PLs shortly after the new contract was put in place. Our analysis of the relationship between contract model change and cost performance suggests correlation (likely induced by exogenous factors like the pandemic) as opposed to causality. Managers rarely conduct rigorous cause-effect analyses of their actions. In the corporate environment, conclusions are often made based on incomplete and selective data. Our research suggests that the VPA model is likely not to be blamed for the deterioration of efficiencies.

In some instances, strong incentives targeting cost savings do not directly lead to cost savings. Placing high rewards and penalties to improve a specific attribute of supplier performance might not always be the most effective way to achieve the goal.

In the course of our literature review and quantitative analysis, we learned to look at contracts more holistically. Contracts consist of a few major components and some features that define crucial trade-offs. For some readers, making sense of contract design and governance might seem daunting, so we designed an easy-to-understand framework that covers such concepts.

5.1 3PL Service Contract Framework

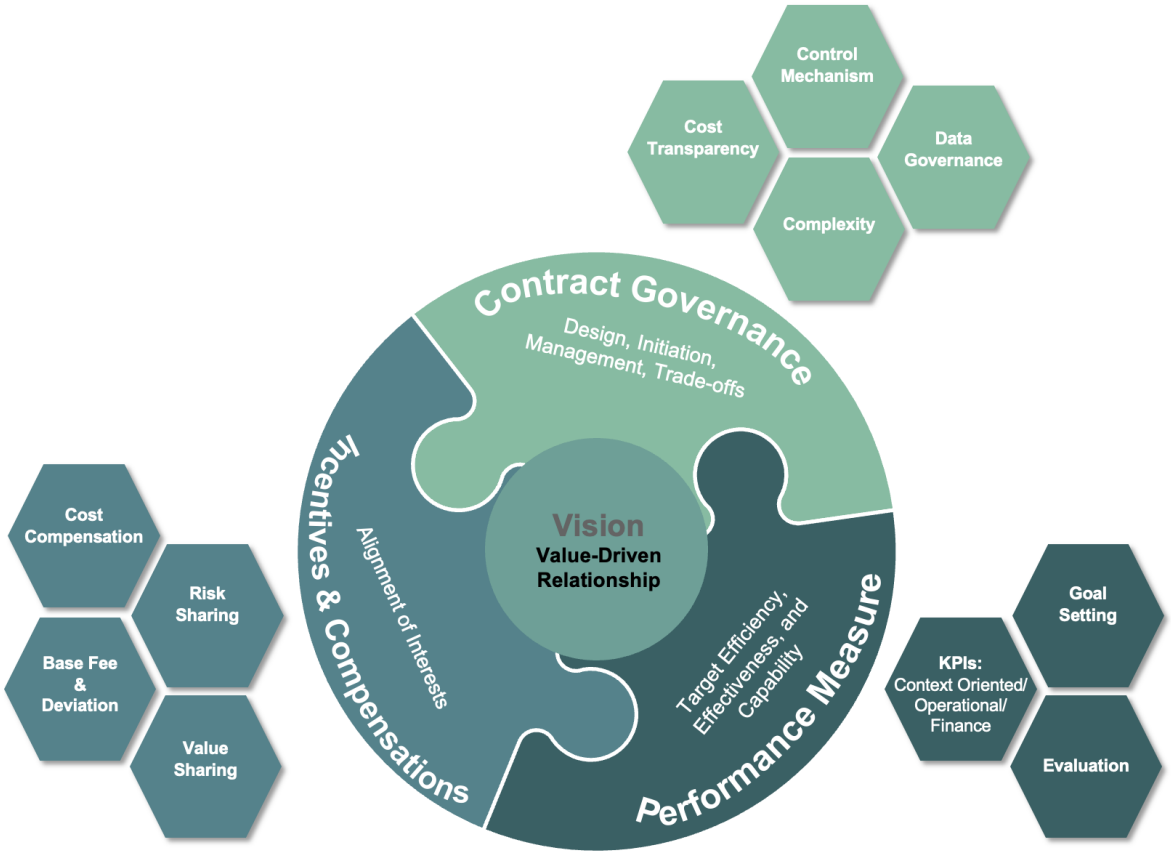
3PL Contract design and management are critical for FMCG companies. Well-designed contracts can bring value to all parties involved by promoting efficiency, performance improvements, and cost-savings. On the other hand, poorly-designed contracts can cause erosion of value, measuring up to 9 percent of customer's annual revenues (Belotserkovskiy et al., 2018). Therefore, by incorporating the learnings from the literature review, we developed a framework that can guide companies when analyzing or designing 3PL service contracts (Figure 5). The framework, however, is not a turnkey solution. Every buyer-supplier relationship is unique, and depending on the circumstances, companies may adopt different approaches.

Our framework suggests that the contract's overall vision and goal rely on mutually-consistent execution of the three primary components: performance measures, compensation structure, and governance processes. In addition, we present key features that define each primary component. We may also think of such features as levers that define key trade-offs inherent in the contract. Changing the parameters of any defining feature of the contract almost always leads to changes in some other attributes, although such dynamics might not be immediately evident.

We will use Acme to illustrate the execution of the framework. Through illustration, we will highlight the areas that Acme is excelling at as well as the areas that Acme can enhance.

Figure 5

3PL Service Contract Framework



5.1.1 Identify the Vision

Based on the company's strategic goal, the buyer (Acme) should identify the vision of the 3PL partnership. For example, the vision for Acme's contract was to increase the level of accountability of its 3PL. Acme wanted the 3PL to have more 'skin in the game' and take ownership to drive cost reductions and performance improvements.

5.1.2 Performance Measure

Based on the vision, the buyer can set explicit, quantifiable performance measurements. To define an effective metrics set, the company should understand the value and purpose behind each metric. After conducting the literature review and analyzing Acme's performance reports, we suggest grouping the performance KPIs into three categories: financial, operational, and context-dependent.

Performance measurements should serve as indicators of efficiency, effectiveness, and capability of an action, a process, or a system against the given norm (Fortuin & Korsten, 1988). These are often captured via the financial, operational, and context-dependent KPIs. The targets for these performance measures should be high enough to encourage elevated performance but low enough, so they appear attainable. Therefore, proper goal-setting and evaluation are crucial for effective performance management.

Furthermore, performance measurement systems (PMS) that account for external turbulence can improve organizational resilience. By incorporating risk metrics and identifying the potential risks, an organization can improve its resilience while becoming more proactive and reactive. In times of disruptions, resilience can be a source of competitive advantage, thereby indirectly improving service performance (Bühler et al., 2016).

For Acme, the operational performance measures are explicitly defined. Although we cannot disclose the details of each KPI, each metric has a clear value and purpose behind them. The target setting has received mixed responses. One of the 3PLs that we interviewed viewed the KPI targets to be attainable with considerable effort. However, another 3PL voiced concerns that some of the KPI targets were unattainable. In this case, the target should be further discussed through open feedback channels between the management and operations to better align on the

target. Setting goals too high can jeopardize motivation and commitment and may even eradicate social capital through dishonesty and corner-cutting (Bennett, 2003).

For Acme, another performance measure that they consider is the year-over-year (YoY) cost-saving. The details are less explicitly defined. Without clear guidance, 3PL is expected to deliver a 3% YoY savings to avoid a decrease in its base-management fee (BMF). This was the common concern for both 3PLs that we interviewed. Acme owns the facility, material handling equipment, and warehouse management systems (WMS), while the 3PL directly controls the labor and labor management. Such an arrangement leads to the fact that 3PLs are often limited in improving labor efficiencies. However, as the site matures, the labor optimization tapers off, and it becomes increasingly difficult to deliver 3% YoY savings. We believe that sooner or later, Acme and its 3PLs should review the targets for cost-savings. The areas in which the 3PLs are expected to drive savings will need to be reviewed, and the target expectations might need to be renegotiated.

5.1.3 Incentives and Compensations

The vision and the performance metrics should be aligned with the incentives and compensations structure. Aligning the performance metrics and incentives is crucial for successfully implementing a service contract. Performance metrics will only work if the incentives and compensations are properly designed. Misaligned incentives, in otherwise well-designed contracts, can lead to unintended negative consequences (also known as “cobra effects” (Hartley, Dale, 2016)) and opportunistic behaviors on the part of the supplier.

As a rule, all open-book contracts compensate the full amount of costs incurred by the supplier during operations. However, the contracts differ in handling deviations from the specified goals. Defining an appropriate level of the base management fee, risk/reward sharing, and

consequences of deviating from the mutually agreed-upon targets is paramount for an effective buyer-supplier relationship.

For Acme, the VPA model has provided a better incentive structure than that of the PSGS model. From our interviews, we learned that there were cases of 'gaming the system' prior to VPA. For example, once a 3PL failed to meet one of the metrics, they would shift focus to other KPIs to maximize their incentives. On the other hand, the VPA model has a tiered system where the 3PL can still recoup some of the incentives even after missing its target.

5.1.4 Governance

Contracts also require governance to manage and administer the performance metrics and incentive structures. There is a trade-off between the incremental cost of governance and the incremental value driven by the contract.

Every buyer of 3PL services desires comprehensive reporting of costs, state-of-the-art data management systems, and effective control mechanisms. Yet, improvements in those directions increase the complexity of contract governance. High governance complexity may create friction and increase the costs of doing business on both sides. As a rule of thumb, the larger and more integrated the partnership is, the higher the degree of justifiable complexity of governance.

From our interviews, we learned that the VPA model required greater governance effort than the PSGS model. The process of cost review has become more elaborate and time-consuming. Yet, no new hires were made to support the VPA governance. Nevertheless, it is possible that the additional effort and time required for governance may have negative long-term effects due to stress, fatigue, and even increased turnover rates of the key employees.

5.1.5 Social Capital

Our framework is built on the three mechanisms: performance measurement, incentive structure, and governance. For the mechanisms within the framework to work effectively, there needs to be a sufficient level of social capital and trust between the buyer and the supplier. (Tate, 1996) supports such a line of reasoning by claiming that successful logistics partnerships require mutual commitment and trust. As described in Section 2.5, social capital is a valuable asset that can be gained through social relationship. Social capital can facilitate the overall performance of the relationship (Nahapiet and Ghosal 1998) and improve resiliency through information sharing. However, collaboration does not always guarantee improvements. As the level of social capital passes the optimal level, it may increase the chance of opportunism and may begin to outweigh the benefits (Villena et al., 2011).

Evaluating the level of social capital for Acme and its 3PL was difficult, as there is no simple way to measure it. However, there seems to be a significant level of social capital based on the length of 3PL executives' tenure on Acme's account. However, Acme's managers are rotated to different sites or roles every three to five years. Although this may prevent the dark side of collaboration, it can lead to lower cognitive or relational social capital, as mentioned in Section 2.5.

5.2 Resiliency and 3PL Contracts

In the FMCG industry, 3PL services account for about 7.5% of overall spending (Alicke & Losch, 2010). As a result, companies are actively looking into ways of decreasing 3PL spending. In fact, Acme set an annual cost savings target of 3% for its service providers. From conversations with Acme's partners, we learned that intense pressure to decrease costs

sometimes causes erosion of relationships and decreases the prospects of long-term business relationships.

Suppliers may give in to the cost pressures of buyers and lower their budgeted costs by reducing the upside (safety) throughput capacity of their operations. Such savings projections are estimated primarily by reducing the number of full-time employees. Employee headcount planning relies on forecasts, which, if correct, facilitate near-100% utilization of resources and very efficient operations. In reality, forecasts are often wrong, and our work with Acme shows that errors can run as high as 40%. In our research, the largest forecast errors usually underestimate the demand. In such a case, the client company has to assume a lot of 3PL employee overtime spending and, even so, is likely to fall short of meeting the market demand. In addition, increasing competition and extraordinary events (e.g., pandemic) can increase employee churn rates and lead to labor shortages. When understaffed, 3PLs experience even more difficulties in satisfying higher-than-forecasted demand.

Projected savings due to a decrease in headcount at 3PL facilities look good on paper. But in reality, the buyer pays a 50% premium (overtime) for processing most of the volume that overshoots the forecasts. For example, a simple retrospective what-if scenario shows that if Acme were to increase the budget-approved employee count at a 3PL facility by 10%, it could result in up to 2-3% savings in total spending on 3PL (based on avoiding overtime spending). In the world of FMCG business, such savings are substantial.

Section 2.4.3 discussed the importance of aligning the company's goals with the goals of a partnering service provider. If priorities of the client change (e.g., service level performance

becomes more critical), the way they measure and compensate a 3PL has to adapt accordingly. Typical contracts make it hard to make such alignments, so in practice, adjustments are made predominantly when a new contract is renegotiated (once every three years on average).

Steven Levitt (2010) claims that incentives always work, even though not always in a predictable way. The "godfather" of the Theory of Constraints, Eliyahu Goldratt (1990), argues that the way we measure the performance of an organization is a good predictor of how it will behave. Adapting means changing and adjusting behavior. Contracts prevent adaptability by fixing the way operations are measured and compensated. Creating an adaptable contract that promotes resiliency is a complex task far beyond the scope of this research. Still, it is worthwhile to outline the problem, and we hope others will conduct further work in this direction.

6 Conclusion

To focus on their non-logistics core competencies and remain competitive in the industry, companies are becoming more vertically integrated with third-party logistics providers. Therefore, it is imperative for companies to understand the impact of contract design in determining the trade-offs between governance costs and incremental value driven through contract design.

Our project sought to empirically analyze the impact of Acme's new contract (VPA) on its 3PL cost performance. Using VPA as the treatment, we developed a Difference-in-Differences model to estimate the quantitative impact of the new contract. We found no statistical evidence that the VPA model impacted the overall cost-performance of Acme's suppliers. In addition, we developed a framework that can help FMCG companies analyze and improve their 3PL open-

book contracts. Acme can apply both the quantitative model and the contract framework to evaluate and improve its warehouses and mixing centers.

6.1 Managerial Implications

It is common wisdom that accurate forecasts improve operations planning and cost performance. In the quantitative part of our research, we explored how pallet volume forecast accuracy is associated with increased spending by Acme's suppliers. For example, we presented estimates of what can be the impact of improving forecast accuracy at the Illinois facility. The mean absolute percent error (MAPE) of forecasts in Illinois rose to 22% after the pandemic. Our research suggests that every 6% of additional forecast error at this facility contributes about 1% in extra cost-per-pallet. Improving forecasting performance to match the pre-pandemic levels (about 4% MAPE) constitutes a 2-3% opportunity to improve the cost performance of this facility.

After looking into the pattern of forecast errors at the Illinois facility, we recommend Acme (and other FMCG companies where the phenomenon exists) works on removing bias from forecasting. An unbiased forecast is a forecast that is expected to produce an error of zero on average. In practical terms, an unbiased forecast is as likely to fall short as it is to exceed the actual demand. A substantial body of literature, including Manary et al. (2019) and Graves & Willems (2000), emphasizes the importance of creating unbiased forecasts in manufacturing and operations. The value of such "neutral" forecasts is in the fact that they enable scientifically sound approaches to managing complex operations in conditions of uncertainty. Instead of "padding" the forecasts, a company should strive to mitigate the impacts of uncertainties by utilizing buffers and safety stocks (this idea applies to both labor and inventory needs).

The magnitude and bias of forecast accuracy errors can be unequal at different facilities. For example, post-COVID-19 at the Illinois facility, we observed a persistent positive bias (over-

forecasting) measuring at about 21% of monthly pallet volume at the facility. The equivalent metric for the Texas facility is only 1.9%. We believe further research regarding the impact and the ways of improving forecast accuracy can be valuable for Acme and the FMCG industry.

The framework we presented in section 5.1 could serve as a valuable tool for managers looking to better understand or even improve their existing 3PL contracts. We would like to reiterate how important it is to achieve mutually-consistent execution of the three primary components of the framework: performance measures, compensation structure, and governance processes. Inevitably, there will be trade-offs as part of any contract design. We emphasize that managers should pay careful attention to how pursuing improvements in one area might lead to sacrifices elsewhere.

Additionally, fine-tuning each component of the contract framework requires a balanced approach to several critical activities, some of which include:

- Setting performance goals that are neither too high nor too low
- Measuring the data points that matter and align with organizational goals
- Fairly sharing rewards and losses based on factors within and outside of the supplier's control

Optimal contracts are not static and require ongoing work on behalf of both buyer and supplier. Contracts help facilitate productive logistics partnerships. According to Tate (1996), such partnerships are like marriages and require open communication, mutual commitment, fairness, and agility.

6.2 Limitations

One of the biggest weaknesses of our research is the limited pool of data. At the end of our data cleaning and review, we were able to use only data from two facilities. Acme works with over a dozen 3PL mixing centers and warehouses. Still, we could only use the data for the two facilities because there were idiosyncratic ways in which the other sites were measured and managed. Some specific examples that warranted disqualification of facility-specific data from our research included:

- Change of a 3PL provider at the facility during the period in question. Acme owns the facilities and can switch service providers at its discretion. When such an event occurs, there are significant losses in short-to-medium term operational productivity at the facility.
- Existence of different management and evaluation processes at the facility. Some sites are run by Acme and can't be directly compared to 3PL-run sites.
- Implementation of a new warehouse management system (WMS) at the facility. Such projects can disrupt normal operations for, on average, six months or more. The longer-term effect of WMS upgrades is an open question and subject to another line of research.

Nevertheless, we were assured by Acme that the two locations we used were the most representative of the project at hand.

Another weakness of our data is that we do not have a proper control facility where no contract changes happened. The Difference in Differences method relies on a parallel trend assumption at the facility where the supposed impact of policy change is being measured. At the same time, we compare both the parallel trend and the actual performance of the facility in question to the

facility where no policy change took place. We used a parallel trend assumption to create the counterfactual of control conditions (no contract change) at the Illinois facility. The data from the control scenario was used to enable our Difference in Differences model.

REFERENCES

Albalate, D. (2007). A. Natural Experiments: Lowering Blood Alcohol Content Levels to Save Lives, The European Experience. *Journal of Policy Analysis and Management*, *27*, 20–39. https://doi.org/10.1002/pam.20305

Alicke, K., & Lösch, M. (2010). How does your supply chain shape up? *McKinsey&Company. https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/operations/pdfs/lean_and_mean-how_does_your_supply_chain_shape_up.pdf*

Belotserkovskiy, R., Lietke, B., Sewak, J., & Teodorian, A. (2018). Contracting for Performance: Unlocking Additional Value. *McKinsey&Company. https://www.mckinsey.com/business-functions/operations/our-insights/contracting-for-performance-unlocking-additional-value*

Bennett, D. (2009, March 15). Ready, aim ... fail. Why setting goals can backfire. *Boston Globe*. Retrieved October 31, 2021, from http://archive.boston.com/bostonglobe/ideas/articles/2009/03/15/ready_aim____fail/

BEROE Advantage Procurement. (2016). *Market Intelligence on Warehousing Contract Structure and Incentives* [Unpublished Internal Report]. BEROE.

Bourne, M., Neely, A., Mills, J., & Platts, K. (2003). Implementing performance measurement systems: A literature review. *International Journal of Business Performance Management*, *5*(5), 1-24. https://doi.org/10.1504/IJBPM.2003.002097

Bühler, A., Wallenburg, C. M., & Wieland, A. (2016). Accounting for external turbulence of logistics organizations via performance measurement systems. *Supply Chain Management: An International Journal*, *21*(6), 694–708. https://doi.org/10.1108/SCM-02-2016-0040

Butters, J. (2021). *S&P 500 Is Reporting Second Highest Net Profit Margin Since 2008 for Q2*. Factset.Com. Retrieved October 27, 2021, from https://insight.factset.com/sp-500-is-reporting-second-highest-net-profit-margin-since-2008-for-q2

Cachon, G. P. (2003). Supply Chain Coordination with Contracts. In *Supply Chain Coordination with Contracts: Vol. Volume 11* (pp. 227–339). Elsevier. https://doi.org/10.1016/S0927-0507(03)11006-7

Couzens, S. A. (2006). *The DNA of a High-Performing Manufacturing Organization: Improving Operations Capability through Performance Measurement*. https://dspace.mit.edu/handle/1721.1/37137

Difference-in-Difference Estimation | Columbia Public Health. (n.d.). Retrieved November 28, 2021, from https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation

Dixon, J. R., Nanni, A. J., & Vollmann, T. E. (1990). *The new performance challenge: Measuring operations for world-class competition*. Dow Jones-Irwin.

Drucker, P. F. (1963, May 1). Managing for Business Effectiveness. *Harvard Business Review*. https://hbr.org/1963/05/managing-for-business-effectiveness

Eccles, R. G. (1991, January 1). The Performance Measurement Manifesto. *Harvard Business Review*. https://hbr.org/1991/01/the-performance-measurement-manifesto

Fortuin, L., & Korsten, A. (1988). Quantitative methods in the field: Two case studies. *European Journal of Operational Research*, *37*, 187–193. https://doi.org/10.1016/0377-2217(88)90328-1

Goldratt, E. (1990). *The Haystack Syndrome: Sifting Information Out of the Data Ocean*. North River Press.

Hair, J. F., Babin, B. J., & Krey, N. (2017). Covariance-Based Structural Equation Modeling in the Journal of Advertising: Review and Recommendations. *Journal of Advertising*, *46*(1), 163–177. https://doi.org/10.1080/00913367.2017.1281777

Hartley, Dale. (2016, October 8). *The Cobra Effect: Good Intentions, Perverse Outcomes*. Psychology Today. https://www.psychologytoday.com/us/blog/machiavellians-gulling-the-rubes/201610/the-cobra-effect-good-intentions-perverse-outcomes

Hwang, D. (2011). *Performance Measurement System Design for Supply Chain Organizations*. https://dspace.mit.edu/handle/1721.1/66076

Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Jiménez, J. L., & Perdiguero, J. (2019). Difference-in-Difference. In *Encyclopedia of Law and Economics* (pp. 551–555). Springer New York. https://doi.org/10.1007/978-1-4614-7753-2_664

Kaplan, R. S., & Norton, D. P. (1992, January). The Balanced Scorecard—Measures that Drive Performance. *Harvard Business Review*.

Kaplan, R. S., & Norton, D. P. (1993, September). Putting the Balanced Scorecard to Work. *Harvard Business Review*.

Kennerley, M., & Neely, A. (2003). Measuring performance in a changing business environment. *International Journal of Operations & Production Management*, *23*(2), 213–229.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed). McGraw-Hill Irwin.

Lafontaine, F., & Slade, M. (2008). Exclusive Contracts and Vertical Restraints: Empirical Evidence and Public Policy1. *Handbook of Antitrust Economics*.

Levitt, S., & Dubner, S. (2010). *SuperFreakonomics: Global Cooling, Patriotic Prostitutes, and Why Suicide Bombers Should Buy Life Insurance*. HarperCollins. https://www.amazon.com/SuperFreakonomics-Illustrated-Patriotic-Prostitutes-Insurance-ebook/dp/B003V1WSVE/

McDonald, N. (2006). Organisational resilience and industrial risk. In E. Hollnagel, D. D. Woods, & N. Leveson, *Resilience engineering: Concepts and precepts* (pp. 155–179). Ashgate.

Melnyk, S., Stewart, D., & Swink, M. (2004). Metrics and Performance Measurement in Operations Management: Dealing With the Metrics Maze. *Journal of Operations Management*, *22*, 209–218. https://doi.org/10.1016/j.jom.2004.01.004

Messer, L. (2016). Natural Experiment. In *Britannica*. Retrieved November 26, 2021, from https://www.britannica.com/science/natural-experiment

Nahapiet, J. (1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, 23, 242-266. https://www.jstor.org/stable/259373

Neely, A. D. (Ed.). (2002). *Getting the measure of your business*. Cambridge University Press.

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, *100*(469), 322–331.

Sarstedt, M., Ringle, C., & Hair, J. (2017). *Partial Least Squares Structural Equation Modeling*. https://doi.org/10.1007/978-3-319-05542-8_15-1

Stank, T., Keller, S., & Closs, D. (2001). Performance Benefit of Supply Chain Logistical Integration. *Transportation Journal*, *41*, 32-46.

Tate, K. (1996). The elements of a successful logistics partnership. *International Journal of Physical Distribution & Logistics Management*, *26*(3), 7–13. https://doi.org/10.1108/09600039610115045

Villena, V. H., Revilla, E., & Choi, T. Y. (2011). The dark side of buyer-supplier relationships: A social capital perspective⋆. *Journal of Operations Management*, *29*(6), 561–576. https://doi.org/10.1016/j.jom.2010.09.001

Whaley, C. M., Cantor, J., Pera, M., & Jena, A. B. (2021). Assessing the Association Between Social Gatherings and COVID-19 Risk Using Birthdays. *JAMA Internal Medicine*, *181*(8), 1090–1099. https://doi.org/10.1001/jamainternmed.2021.2915

Wieland, A., & Wallenburg, C. M. (2012). Dealing with supply chain risks: Linking risk management practices and strategies to performance. *International Journal of Physical Distribution & Logistics Management*, *42*, 887–905. https://doi.org/10.1108/09600031211281411

Wieland, A., & Wallenburg, C. M. (2013). The influence of relational competencies on supply chain resilience: A relational view. *International Journal of Physical Distribution & Logistics Management*, *43*(4), 300–320. https://doi.org/10.1108/IJPDLM-08-2012-0243