Snider Undergraduate Research Journal an Undergraduate Research Journal of Business

The Robert H. Smith School of Business

Volume II

2019
Table of Contents

A Note from the Student Leadership ................................................................. 5
A Note from the SURE Faculty Director .......................................................... 6
Acknowledgements ........................................................................................... 7
Financial Barriers to Asthma Care for Black Families in Urban Areas............ 8
The Impacts of Coastal Erosion and Rising Sea Levels on Louisiana and Other Areas Around the World ................................................................. 26
Potential Losses and Data Value for Cybersecurity Breaches ....................... 46
The Effects of Decriminalization of Drugs: Lessons Learned from Portugal’s Public Health Approach ................................................................. 79
Differences in Media Representation in Low Casualty Terror Attacks .......... 125
Issues with Public Disclosure of Non-GAAP Financial Metrics ................. 159
Evaluating Corporate Self-Governance in Eradicating Forced Labor .......... 208
Evaluating Patients’ Trust in Physicians Through Yelp Reviews ................ 235
Public Pensions: A Crisis of Political Negligence ....................................... 269
Impacts of Data Analysis Pitfalls on Decision-Making .............................. 305
Hydroponics and Vertical Farming: A Study on the Multi-Level Perspective of Transitional Sciences in Global Agriculture .............................................. 339
SEC Rulings on Dark Pool Transparency: A Case Study to Understand Its Effects on Price Discovery ................................................................. 379

Proper Fit: Implications of Personalized Breast Cancer Screening ............. 415

Dispelling the Hype Over Hemp: How “Green” are the Production Opportunities for Hemp in U.S. Agriculture? ......................................................... 442

The Great Recession: Impact on Commercial Real Estate Leasing Markets ..... 487

Do Voices of General Employees Matter? The Importance of Their Involvement and Impact in Designing and Implementing a Human Resource Information System (HRIS) ................................................................. 505
A Note from the Student Leadership

Dear Readers:

The second edition of the Snider Undergraduate Research Journal marks the continued success, growth, and desire of our students to remain curious, think big and ask questions. Parallel to its purpose, of providing a platform for conducting truly independent research, the SURE program itself was an experiment that brought one student’s proposition, of giving undergraduates investigative freedom, to life. In the words of Albert Einstein, “if we knew what it was we were doing, it would not be called research, would it?” What began in 2016 with five fellows, has transformed into a program that is 18-students strong and representative of nine academic disciplines at the University of Maryland. Spanning from a local to global scope, each fellow’s research paper gives an insightful clue into the social, political, and economic topics that are top of mind in today’s bright minds.

All of our fellows demonstrate the attitude and determination necessary to nurture a research project from its question-scoping phase to a body of work that may ultimately be published. Under the guidance of remarkable subject matter experts, the SURE 2018-19 cohort serves as a beacon of hope and declaration from the decision-makers, innovation drivers, and influencers of tomorrow that they are listening to the topics of most relevance today.

Mohammed Alghaffari

Ethan Liu

Mario Menendez
A Note from the SURE Faculty Director

SURE is a unique opportunity to experience research, not as a team or analyzing a small part of a faculty project; rather, SURE fellows are looking for a challenging and messy experience that forces them to dive deep into a well scoped-out question and come up with an answer grounded in data. Since its inception, SURE has been a student run program. Over the years, student leaders have created templates and timelines to facilitate everything from the candidate recruitment and selection process to the research outcomes and final paper. From each class, a few fellows step up to assume the next year’s leadership. It’s a great opportunity for the student leaders to iterate and improve on the previous year’s program. As the faculty director for SURE, it gives me real pleasure to support the fellows’ and leaders’ continued growth and success.

Christina Elson

Associate Research Scholar

Managing Director, Ed Snider Center for Enterprise and Markets
Acknowledgements

Thank you to the faculty advisory board: Dr. Rajshree Agarwal, Robert H. Smith School of Business; Dr. Mircea Raianu, Department of History; and Dr. Jim Purtilo, Department of Computer Science.

During the year, we also call on faculty to provide workshops and we are grateful to Dr. Joe Bailey for the scoping workshop and Dr. Purtillo and Dr. Raianu for the research methods workshop.

Finally, the Snider Undergraduate Research program is very grateful for the participation of the mentors who make time to advise fellows on their papers. This year we thank Hossein Abbasi, William Braniff, Progyan Basu, Mark Flood, Paul Goeringer, Gordon Gao, Virginia Haufler, Jose-Luis Izursa, Hui Liao, Henry Lucas, Vojislav Maksimovic, Amy Pate, Courtney Paulson, David Waguespack and Katie Zafft.
Financial Barriers to Asthma Care for Black Families in Urban Areas

My-Asia Chaplin
University of Maryland- College Park

My-Asia Chaplin is a SURE Fellow in the School of Public Health at the University of Maryland, College Park, MD. Their email address is: myasia.chaplin@gmail.com.

Abstract

Commonly highlighted in the wider literature is that asthma is a chronic respiratory disease that is relatively manageable with quality care. However, while a large proportion of people living with asthma can effectively treat their condition with regular medication, a segment of the population with asthma struggling with poor health outcomes because of financial limitations. This paper will narrow its scope to the analysis of financial barriers to asthma care for black families in urban areas. The paper follows a systematic literature review methodology for data collection and subsequent analysis. The results will show that material deprivation can be used as a complementary indicator of socioeconomic...
status to describe and explain health inequalities among people with asthma.

**Introduction**

Asthma is one of the most common chronic diseases. It is characterized by permanent inflammation of the bronchial mucosa, which leads to hypersensitivity of the respiratory tract. The result is a swelling of the mucous membrane and excessive mucus production, which leads to a narrowing of the bronchi, the main passageway of air to the lungs. In an acute asthma attack, the bronchial muscles become cramped, exhaling becomes more difficult, and shortness of breath occurs. Symptoms include constant coughing, wheezing, chest tightness, viscous mucus, and seizure-related respiratory distress. Many people living with asthma can treat their condition with regular medication. But for minorities and low-income people, especially in black communities in urban areas, there is an unmet need in providing adequate therapies for asthma. The standout challenge for many of them is financial limitation -- they cannot afford or are not insured for diagnosis and proper management of asthma.

This research initiative aims to reveal the financial barriers to needed asthma care for low-income Black families in urban areas; offer a new
perspective toward approaching the problem; enhance awareness of the
consequences of financial barriers to asthma treatment; serve as a call to
action; encourage stakeholders to understand the impact of asthma on this
community and work to provide these populations with asthma care and
treatment they need and deserve.

**Literature Review**

In the wider literature, asthma is classified as a respiratory illness.
The full medical name of is bronchial asthma. The International consensus
report on the diagnosis and treatment of asthma defines it as a chronic,
inflammatory respiratory disease with persistent hypersensitivity.
According to the same report, in persons with a hypersensitive bronchial
system, the inflammation leads to episodic respiratory distress due to an
acute narrowing of the respiratory tract, i.e. bronchial obstruction. This
airway narrowing is caused by increased secretion of mucus, cramping of
the bronchial musculature and formation of edema of the bronchial mucosa,
it is capable of regressive treatment. A variety of stimuli can cause an
increase in airway sensitivity (bronchial hyperreactivity or even bronchial
hyperresponsiveness) and the associated inflammation. Common
complaints associated with asthma include wheezing, dry cough, cough
with tough mucus, tightness in the chest and shortness of breath (Sheffer et al., 1992).

In terms of the prevalence of incidence of the disease, studies show that 5% of adults and 7% to 10% of children suffer from bronchial asthma. Studies show that asthma normally manifests in childhood but often disappears toward adulthood, as the symptoms decrease with good management (Engelkes et al., 2015). Why asthma develops or disappears is not yet fully understood or explained in the available literature. Because medication can relieve the symptoms but cannot cure the disease, asthma is characterized as chronic.

For a person with asthma to maintain the same level of health and quality of life as a person with no asthma, management of the disease is key. Since the diagnosis, treatment and management of asthma usually come with a relatively heavy financial requirement, low-income populations, especially minorities, struggle with negative health outcomes. In her study on “asthma disparities in urban environments,” Bryant-Stephens (2009) found a substantial difference in asthma health outcomes across race/ethnicity groups, income brackets, and language. Whites were more likely to have a usual source of care at 96 percent, while Blacks were scientifically less likely. Ethnic and racial minorities experienced disparities in access to asthma care that lead to greater morbidity and
mortality. Black children who lacked regular care were more likely to have increased morbidity (Bryant-Stephens, 2009).

On the other hand, Mansour et al. 2000 identified financial barrier as a clear challenge limiting effective management of asthma in his black urban community. In his study, parents making an income below the designated national minimum indicated they would not be able to afford medication without insurance or that changing an insurance plan interrupted the continuity of the ongoing care. Because of their low-income status, some parents also said they are forced to live with individuals who smoke, making it difficult to limit their children’s exposure to negative stimuli (Mansour et al. 2000). Smoking or being exposed to smoke increases the seriousness of asthma symptoms. Physicians around the University of Glasgow's Neil Thomson have found that lung function improves by 15 percent six weeks after the last cigarette.

Wasilewski et al. (1996), found that when comparing asthma outcomes across insurance type, the discharge rate for Black children enrolled in Medicaid was nearly twice that of non-enrolled Black children. By the same token, 2.5 times those of Whites in the youngest age group but were nearly five times greater between the age group 15-19 years (Wasilewski et al., 1996). These findings were complemented by Ybarra et al, (2017) which found that Black children were more likely to be on
Medicaid and to seek primary care for asthma than White children and other privately insured Black children (Ybarra et al., 2017).

According to Valerio et al. (2006), barriers to asthma care specific to Medicaid insured patients included difficulty maintaining continuity of care due to physician participation in Medicaid programs and concerns about differences in asthma care from health care providers because of their Medicaid insurance status (Valerio et al., 2006). Rand et al., (2000) showed that the lack of insurance coverage affected the quality of care one received for asthma. The study found that children with no medical care coverage were significantly less likely to identify a primary care provider than children with medical coverage. They also were far more likely than children with medical coverage to receive only emergency care. On the other hand, after adjusting for sociodemographic variables and asthma status, Lieu et al., (2002) found that Black and Latino children were less likely to be using inhaled anti-inflammatory medication than White children (relative risk for Blacks: 0.69; relative risk for Latinos: 0.58). They were more likely to have less effective home nebulizers. a difference possibly attributed to financial burden that comes with seeking professional medical interventions for asthma without insurance.

Studies also have demonstrated that the quality of living conditions affects asthma health outcomes. According to Hughes et al. (2017),
households owned by someone living in the home had 13% lower odds of having a child in the home with asthma, and 37% lower odds of ED visits for asthma. Households with poor quality housing had 50% greater odds of an asthma-related ED visits in the past year. Correspondingly, among unassisted low-income renters, 13.5% reported current asthma, while only 8.7% adults did. Among HUD-assisted adults, 16.3% reported current asthma.

**Theoretical Framework**

For proper care and treatment, persons with asthma must have, or be provided with, the financial capacity to have proper asthma management and environments that reduce the occurrence of asthma triggers. Low income Black families in urban areas do not have the financial capacity that allows them to successfully manage asthma. Many do not live in environmentally safe housing free from inner city air pollution or household smoke. This status quo can be explained with the material deprivation theory.

Material deprivation is the restriction of living conditions by limited financial resources. It is measured by the lack of access to services and goods that reflect the average standard of living. Materially deprived populations occupy the lower levels of socioeconomic status. Studies have
shown that persons with a low socioeconomic status have a much poorer health outcome than those with a higher socioeconomic status. They often cannot afford good healthcare insurance and, as a result, suffer poor health outcomes that lead to health disparity.

These studies are based on the classic indicators of socioeconomic status: education, occupation and income. Since the 1980s, the immediate living conditions of a person have been discussed as indicators of socio-economic conditions (Townsend, 1979). The lack of a good standard of living due to financial constraints is understood in relatively rich industrialized countries as a relative disadvantage to social participation (Marmot, 1999). An inadequate standard of living due to financial constraints is also referred to as relative or material deprivation.

Following Wilkinson's psychosocial comparison hypothesis, material, or relative, deprivation has a direct effect not only on material deprivations, but also indirectly on a relative sense of deprivation of social exclusion (Marmot, 2006; Wilkinson & Pickett, 2010). Corresponding studies have shown a close correlation between material deprivation and psychosocial stress symptoms (Foulds et al., 2014). Similarly, because of lack of access to quality healthcare this phenomenon can also manifest itself in many chronic diseases and disorders, such as asthma.
The reason for this is clear. Income in the market can be exchanged for goods and activities that have a direct impact on health. As a result, income acts as a health resource that increases one’s standard of living, a direct and active effect that can be used more directly at the level of social and health policy.

On the other hand, without a healthy income, a person’s ability to pay for health care, to afford a healthy quality of life, is compromised. Material deprivation, therefore, contributes significantly to a person’s general health outcomes. Thus, any measures to solve the asthma care challenge for Black families in urban areas should be aimed at improving their quality of life (for example, housing) and their income situation overall (Mansour et al. 2000).

**Methodology**

This study’s methodology of data collection and analysis follows a systematic literature review format. The systematic selection and use of literature are essential when exploring a topic that cannot be straightforwardly addressed using primary data. Intensive examination of the secondary literature helps to shed light on the topic, to substantiate or dismiss the hypothesis and arrive at concrete theses based on the analyzed materials. The selection of useful literature is key to focusing on the
research in a thematic way. The systematic literature review process was divided into four steps: literature search; literature selection, literature synthesis; and presentation.

The literature synthesis is about extracting the information needed to answer the research question or address the topic. For this purpose, relevant articles/literature were selected, summarized and assigned to categories. Formation of categories was done using a deductive approach, where categories are formed from the research question, forming themes and defining main concepts. Eventually, the categories were refined, adjusted and narrowed with the progression of literature synthesis. To further limit the list of literature for synthesis, several qualitative approaches also were used.

Analysis and interpretation of literature were included in the literature synthesis category. Different objectives or approaches were used, including mapping subject discourses; statistical analysis (e.g. analysis of relationships and frequencies); working out theories and patterns with the aim of producing typical arguments; and working out new research questions and hypotheses. The final step in the systematic review process was presenting the results or findings.
Data Analysis

Analysis of data on health outcomes of asthma sufferers, especially in Black families in urban areas, fell into the following thematic elements.

i. **Socioeconomic Status (SES) Is A Determinant of Health Access and Outcomes.**

The data points toward the thesis that social economic status is highly consequential in determining whether one will get treatment for asthma and the quality of the treatment. Bryant-Stephens (2009) found that low social economic status is inversely proportional to low health outcomes (health disparities). According to Mansour et al. (2000), socioeconomic status determines the ability to meet the cost of asthma medication. Consequently, low-income inner-city parents find it hard to afford medication and treatment for asthmatic children.

The SES reflects the individual position in the social hierarchy and is mostly covered by the information on the level of education, occupational status, and income situation. The available research broadly suggests that people with low SES are at higher risk for many chronic diseases and disorders than those with higher SES. Low SES individuals also rate their general health as poor and increasingly report health-related problems in coping with everyday life. Socioeconomic differences also show up in the
distribution of behaviorally correlated risk factors, such as smoking, lack of exercise, obesity, high blood pressure and lipid metabolism disorders. The increased prevalence of disease and health effects and underlying risk factors of low SES people ultimately translates into higher premature mortality and 5- to 10-year below the mean life expectancy at birth.

ii. Health Disparities with Respect to Asthma Are Evident Across Race/Ethnicity.

Bryant-Stephens (2009) demonstrated that health disparities were also distributed across ethnic/racial groups. For example, Whites were more likely (96%) than Blacks to access more quality care. Wissow, L S, et al. (1998) also showed that the discharge rates were higher (2.5 times) among Black children than White children, a result, in part, of income disparity between the groups. White children were more likely to be covered by private insurance.

iii. The Type of Insurance Determines the Health Outcomes Among Persons with Asthma.

are more likely to have positive asthma health outcomes than those with Medicaid.

iv. **Housing is an Important Determinant of Health Outcomes Among Persons with Asthma.**

Housing is a key indicator of quality of life and determines the health outcomes of persons with asthma. Asthmatic people living in poor housing situations are more exposed to adverse triggers of asthmatic attacks, including air pollution or a household with smokers (Hughes et al., 2017; Helms et al., 2017).

The four themes highlighted above are underscored by the theoretical framework of material deprivation. Basically, material deprivation can lead to a low socioeconomic standing, force one to opt out of private insurance, and draw a person into poor housing. Overall, this contributes to a quality of life which causes poor health outcomes.

**Conclusion**

This paper explains the financial barriers to adequate asthma care for low-income Black families in urban areas. The concept of material deprivation allowed for an immediate assessment of living conditions and an expanded view of socioeconomic status and how it affects individuals
with asthma. Material deprivation is associated with an exclusion from the generally accepted standard of living and is closely associated with a variety of health indicators that are negatively consequential to people suffering from asthma, including inability to fund quality health care and healthy housing.
References


diagnosis and management of asthma. Allergy, 47(suppl. 13), R1-R61.

The Impacts of Coastal Erosion and Rising Sea Levels on Louisiana and Other Areas Around the World

Madeleine Cheng
Research Mentor: Henry Lucas

Madeleine Cheng is a SURE Fellow in the Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: madeleinecheng28@gmail.com.

Abstract

This study looks at how coastal erosion has impacted different regions of the world and examines predictions of future land loss rate and its implications. While this paper focuses on the state of Louisiana, other coastal regions are examined as examples of what has already happened and what is to come. These case studies provide insight into the aftermath of previously inhabited land lost to rising sea levels and how this will create new problems and obstacles that local, regional and national governments will have to address and solve. Additionally, this paper aims to shed light on the impacts of coastal erosion and hopefully inspire more environmentally conscious behavior by residents and government.
Introduction

Lurking threats to present and future global populations are coastal erosion and rising sea levels. Coastal erosion is both a natural and human-induced process that causes net removal of sediments or bedrock from the shoreline and loss of coastal land. It occurs when the transport of material away from the shoreline is not balanced by new material being deposited back (“Coastal Erosion”). Since the start of time, the repeated crashing of waves against the shore and the pounding of rainfalls on coastal ground and rocks have worn away shoreline and cliffs, resulting in a very gradual receding of the shoreline. Coastal erosion is further accentuated by extreme weather events, such as coastal storms, hurricanes, flooding, monsoon, and tsunamis. Waves and currents from these storms can be so intense they can extend to inland areas that are usually out of normal wave and current reach. Heavy rainfalls saturate the soils and cause them to lose structure and shear strength (“Coastal Erosion”). This reduction in strength leads to a higher chance of soil erosion and landslides. The most vulnerable areas are those composed of unconsolidated sediments, such as beaches, dunes, and sand cliffs on shores of coastal lakes and lagoons and open coasts that experience net long shore drift of sediment (“Coastal Erosion”).
The increased presence of humans and human activities on the environment has compounded the effects of coastal erosion in many areas. The construction of infrastructure and systems by the coast and the actions of dredging or sand-mining are both direct ways that can weaken the coastline (“Coastal Erosion”).

An indirect way is through the burning of fossil fuels and other activities that add pollution into the environment. The issue of climate change has become an increasingly hot-button topic. While people disagree about the legitimacy and causes of climate change, rising sea levels around the globe over the last century should be an undisputed fact based on measurements and scientific observations. One cause of rising sea levels is the melting of the polar ice caps and glaciers, a result of warmer global temperatures and oceans. Carbon is trapped and absorbed by these large bodies of water, raising sea temperatures gradually. Over time, increased saturation of carbon is predicted to contribute to increase overall sea level rise.

There are two primary types of coastal erosion. These include a rapid-onset hazard which occurs, as the name suggests, rapidly, in a period of days to a week (“Coastal Erosion”). The second type is a slow-onset hazard, which takes place over many years, decades, or centuries (“Coastal Erosion”).
Coastal erosion is already starting to affect many regions around the world, as lands that humans inhabit are being submerged and lost. With this loss of land comes a host of economic, social and cultural impacts that will have a negative long-term impact. It is critical to analyze this issue now, so it can be addressed and mitigated before it is too late.

Areas Already Severely Impacted

One place that has experienced significant effects of coastal erosion is Tangier Island, Virginia, a small island in the Chesapeake Bay. Since 1850, Tangier has lost at least two-thirds of its land, partially as a result of a phenomenon called glacial rebound (Kormann), which has caused the island to sink a millimeter or two each year. Storm-driven erosion and sea-level rise, both increasing as a result of climate change, are compounding the losses. (Kormann). Further studies by scientists have suggested that the sea-level rise is tripling or quadrupling the natural normal land loss rate (Kormann).
Figure 1: Tangier Island erosion projections by Dave Schulte and Karin Dridge, Army Corps of Engineers.

Top left: 1851; top right: 2013; bottom left: 2038; bottom right: 2113 (Kormann).

Newtok, Alaska is another place that has been severely impacted by coastal erosion. A small coastal village on the banks of the Ninglick River, Newtok has been attempting to relocate since 1994, because of erosion that has been occurring since the late 1950s (Mandel). The village is losing about 70 feet of land each year to coastal erosion. At that rate, residents fear
their homes and land will soon be overtaken in the next century (Mandel). In response, the village has secured more than $15 million to relocate households and the roughly 350 residents to safer grounds inland (Mandel).

However, this will not be enough to relocate the entire village to Mertarvik, roughly 9 miles away (Mandel). To save on construction costs for new housing units, Newtok plans to transport military barracks from Anchorage to Mertarvik to convert into housing (Mandel).

A third island severely affected by coastal erosion is Isle de Jean Charles, Louisiana, 75 miles south of New Orleans. It has lost 9 percent of its land to the Gulf of Mexico’s waters due to a combination of levee construction, coastal erosion, sinking land, rising seas, and hurricane damage (Mariani). The Biloxi-Chitimacha-Choctaw Tribe have been living on Isle de Jean Charles since the 1830s, when their ancestors were seeking escape from the harsh tribal relocation pressures of the Indian Removal Act (Dermansky). On Isle de Jean Charles, they have created a community that preserves the culture and traditions of their forefathers.

In January 2016, the U.S. Department of Housing and Urban Development gave the tribe $48 million to relocate through the National Disaster Resilience Competition (Mariani). However, in early 2019 the tribe turned down the offer because of the tactless way the State of
Louisiana made known to them that it was planning to purchase their land (Dermansky). Instead of direct communication between the Tribe Chief or Tribe Executive Secretary and the State, the tribe found out about these purchasing plans through an online press release (Dermansky). The tribe fears that the State has no respect for their culture and the history of Isle de Jean. (Dermansky).

These situations set a precedent on how the government and community both should and should not act in solving this population relocation crisis. Governments will have to offer some sort of economic plan and budget to allow communities to move to a new location. At the same time, they have to be sensitive enough about existing communities and their traditions and values. It is important to not just find a place to relocate these people, but to also do it in a way that preserves the cohesiveness of the tribe, the traditions, the relationships, and the cultural practices that these tribe members have.

Coastal erosion usually does not qualify for federal disaster since they are slow-moving issues, but areas like Isle de Jean Charles and Newtok have qualified for funding because they are both suffering from significant and rapid erosion.
Louisiana Coastal Erosion

Louisiana has experienced a rapid loss of land over the past century as the result of coastal erosion, land subsidence, channeling the river, and sea level rise. Projections suggest that in a future without action, the next 50 years could result in the loss of 1,750 additional square miles of land area (Glick).
Coastal land loss will directly affect many homes and businesses along the coast. A recent study found that land loss in Louisiana poses a direct risk to as much as $3.6 billion in assets that support $7.6 billion in economic activity each year--updated to 2015 dollars (Barnes, Stephen R., et al.). More importantly, land loss reduces the storm protection services of coastal wetlands.

This study estimated that land loss in Louisiana may increase storm damage from a single storm by as much as $138 billion and generate an additional $53 billion in lost economic output from storm disruptions updated to 2015 dollars (Glick). Land loss brings more densely populated
and heavily built areas such as the cities of New Orleans, Houma and Lake Charles closer to the Gulf of Mexico and increases the risk of storm damage to those areas. The economic links across the state mean that the impact of land loss and the subsequent increase in storm damage will be felt further inland than the highly publicized land loss maps suggest. A primary goal of this study is to investigate how land loss impacts different regions along the coast considering both the direct damage facing coastal regions and the economic disruptions generated from land loss impacts across the coast.

Figure 4: Land loss costs in Louisiana with data taken from LSU report (Barnes, Stephen R.)

To protect the well-being of its residents and its economic base, the state of Louisiana and coastal parishes plan to invest billions of dollars in
coastal restoration and protection projects. This large-scale investment seeks to reduce the long-term consequences of land loss, but also provides a real business opportunity for firms engaged in coastal protection and restoration-related work. To help policy makers, the business community, and the public better understand the economic opportunity created by implementing the state’s Coastal Master Plan, this report also summarizes the economic activity generated by continued efforts to protect and restore Louisiana’s coast (Glick). The economic implications of coastal erosion in Louisiana will spread throughout the nation due to the state’s importance in shipping, energy production, chemicals, and other sectors. As much as $3.6 billion in Louisiana’s business, residential, and infrastructure assets are at risk due to land loss over the next 50 years (Glick). These assets support an additional $7.6 billion in economic activity throughout the nation each year (Glick).

**Areas at Risk**

In addition to the aforementioned regions, there are countless other areas around the world that are at risk of severe coastal erosion and will need to relocate populations within the next century. EarthTime is an online tool that uses blended satellite Earth images with data from universities, government agencies, and nongovernmental organizations to create
animations that show what will happen with the projected sea level rise in the coming decades (“Sea Level Rise Viewer”). EarthTime allows users to adjust sea levels and see how much land in a region will be inundated (“Sea Level Rise Viewer”). Right now, sea levels are rising at a rate of approximately 3 millimeters per year and average global temperature is about 12.7 degrees Celsius (“Sea Level Rise Viewer”). The animations will allow simulations of what land will look like in a world where temperature rises to 16.7 degrees Celsius (“Sea Level Rise Viewer”). Miami, Florida is one of the top cities in the world that is at risk of significant changes due to rising sea levels and coastal erosion.

Figure 5: Miami landscape with current water levels taken from EarthTime (“Sea Level Rise Viewer”)
Figure 6: Miami landscape at a 3-foot sea level rise according to EarthTime (“Sea Level Rise Viewer”)

Figure 7: Miami landscape at a 6-foot sea level rise according to EarthTime (“Sea Level Rise Viewer”)
In a report done by the Organization for Economic Cooperation and Development (OECD), Risk Management Solutions (RMS), and the University of Southampton, it was found that many areas in Asia, the United States, and the Netherlands are at high risk in terms of the size of the population that will be affected (Nichols, R.D., et al.) The two tables below show their findings of the top 20 cities in terms of population size that will be exposed and the top 20 cities in terms of assets exposed to coastal flooding (Nichols, R.D., et al.).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INDIA</td>
<td>Kolkata (Calcutta)</td>
<td>1,929,000</td>
<td>14,014,000</td>
</tr>
<tr>
<td>2</td>
<td>INDIA</td>
<td>Mumbai (Bombay)</td>
<td>2,787,000</td>
<td>11,418,000</td>
</tr>
<tr>
<td>3</td>
<td>BANGLADESH</td>
<td>Dhaka</td>
<td>844,000</td>
<td>11,135,000</td>
</tr>
<tr>
<td>4</td>
<td>CHINA</td>
<td>Guangzhou</td>
<td>2,718,000</td>
<td>10,333,000</td>
</tr>
<tr>
<td>5</td>
<td>VIETNAM</td>
<td>Ho Chi Minh City</td>
<td>1,931,000</td>
<td>9,216,000</td>
</tr>
<tr>
<td>6</td>
<td>CHINA</td>
<td>Shanghai</td>
<td>2,353,000</td>
<td>5,451,000</td>
</tr>
<tr>
<td>7</td>
<td>THAILAND</td>
<td>Bangkok</td>
<td>907,000</td>
<td>5,138,000</td>
</tr>
<tr>
<td>8</td>
<td>MYANMAR</td>
<td>Rangoon</td>
<td>510,000</td>
<td>4,965,000</td>
</tr>
<tr>
<td>9</td>
<td>USA</td>
<td>Miami</td>
<td>2,003,000</td>
<td>4,795,000</td>
</tr>
<tr>
<td>10</td>
<td>VIETNAM</td>
<td>Hai Phong</td>
<td>794,000</td>
<td>4,711,000</td>
</tr>
<tr>
<td>11</td>
<td>EGYPT</td>
<td>Alexandria</td>
<td>1,330,000</td>
<td>4,375,000</td>
</tr>
<tr>
<td>12</td>
<td>CHINA</td>
<td>Tianjin</td>
<td>956,000</td>
<td>3,790,000</td>
</tr>
<tr>
<td>13</td>
<td>BANGLADESH</td>
<td>Khula</td>
<td>441,000</td>
<td>3,641,000</td>
</tr>
<tr>
<td>14</td>
<td>CHINA</td>
<td>Ningbo</td>
<td>299,000</td>
<td>3,305,000</td>
</tr>
<tr>
<td>15</td>
<td>NIGERIA</td>
<td>Lagos</td>
<td>357,000</td>
<td>3,229,000</td>
</tr>
<tr>
<td>16</td>
<td>COTE D’IVOIRE</td>
<td>Abidjan</td>
<td>519,000</td>
<td>3,110,000</td>
</tr>
<tr>
<td>17</td>
<td>USA</td>
<td>New York-Newark</td>
<td>1,540,000</td>
<td>2,931,000</td>
</tr>
<tr>
<td>18</td>
<td>BANGLADESH</td>
<td>Chittagong</td>
<td>255,000</td>
<td>2,866,000</td>
</tr>
<tr>
<td>19</td>
<td>JAPAN</td>
<td>Tokyo</td>
<td>1,110,000</td>
<td>2,521,000</td>
</tr>
<tr>
<td>20</td>
<td>INDONESIA</td>
<td>Jakarta</td>
<td>513,000</td>
<td>2,248,000</td>
</tr>
</tbody>
</table>

Figure 8: Top 20 cities in the world ranked based on size of population that will be exposed to coastal flooding in the 2070s with current exposed population comparison as well (Nicholls, R.J., et al.)
Figure 9: Top 20 cities in the world ranked based on amount of assets potentially exposed to coastal flooding in the 2070s with current exposed assets comparisons as well (Nicholls, R.J., et al.)

Conclusion

Coastal erosion is a threat that needs to be acknowledged and addressed as soon as possible. While it is easier oftentimes to ignore or overlook issues that happen at a more gradual, subtle pace, it is important that governments should start thinking of ways to either prevent the rapid worsening of rising sea levels and coastal erosion, or to mitigate the effects. To help prevent the exponential sea level rise, we need to recognize that sea level rise and resulting coastal erosion are due to the ocean warming, melting of polar ice caps and glaciers, and pollutants and carbon molecules trapped in the atmosphere.
If future generations are to visit Shanghai, New York City, Amsterdam, Miami, and New Orleans, to name a few major metropolitan areas, and to see the historic sites and attractions and experience the culture and excitement of these locations, actions must be taken now to ensure that these regions do not go underwater as they are on track to do so in the coming decades or centuries.
References


Potential Losses and Data Value for Cybersecurity Breaches

Courtland Climer
Research Mentor: David Waguespack

Courtland Climer is a SURE Fellow in the Department of Computer Science at the University of Maryland, College Park, MD. Their email address is: courtlandclimer@gmail.com.

Abstract

The goal of this research is to measure potential losses that a company could expect as an impact of a security breach based on an analysis of previous cases of changes in stock prices of breached firms. A secondary goal is to propose reasons as to why this figure also serves as a means of measuring the financial value of data lost to corporations. From these two goals, the following questions can be derived: What is the average percent impact of cybersecurity breaches on the stock prices of firms, and how can that average be applied to determine the value of a firm’s potential losses? Since this paper serves as an extension to the work done by Spanos and Angelis in their 2016 literature review on the economic impact of cybersecurity incidents, the basic design of this research is as
follows: re-review the works in their review; collect the results from these papers and concatenate them; draw new and distinctive connections between theory, real world events, and studies; and explain where additions need to be made.

This paper uses both qualitative and quantitative methodologies to provide a brief analysis based on a literature investigation of this field. Ultimately, it was found that the weighted average for the results across the studies was -1.74%. It was also found that for large US-based publicly-traded companies breaches are still impactful at low probabilities. Finally, this newfound figure could be used to better approximate values for insurance companies, financial statements, and the Gordon-Loeb model. This research also demonstrates a need for standardization in methodology in this field when performing the calculations associated with event studies. While not a perfect metric, this research serves as another tool that could be used in determining appropriate cybersecurity investments and approximating related economic information.

Introduction

The Information Age has had massive impacts on all facets of life. The significance of computers to modern living exists at all levels of
society from personal to global, from every phone screen owned by billions to the massive server farms supporting even more websites and services.

One specific instance of the impact of the digital era is the mass integration of technology into the corporate world. Virtually every major company on the globe now relies on digital systems for day-to-day functionality. The storage, trading, and even protection of the data on these digital systems has immense value to most companies.

Malicious actors around the world have realized this value and as a result, countless cyber attacks on major corporations now occur every year. Worse is that these attacks are increasing every year. Each new integration of technology for a company becomes a new threat vector for a hacker. These relentless attacks cost major companies millions to prevent, but what about the cost when an attack gets through a company’s cyber defenses and a cybersecurity breach occurs? If you are one of these massive corporations, how can you measure the substantial implicit costs incurred by a breach and the value of the compromised data?

The goal of this research is to measure potential losses that a company could expect as an impact of a security breach based on an analysis of previous cases of changes in stock prices of breached firms. From this exists the questions: What is the average percent impact of cybersecurity breaches on the stock prices of firms? How can that average
be applied to determine the value of a firm’s potential losses? A secondary goal of this research is to propose a means of measuring the financial value of data lost to corporations as the potential losses typically contain the implicit cost of the data lost. This prompts the question: How does potential loss represent the value of the data compromised? The inspiration for this research came from an interest in collecting the numerical results of previous studies and analyzing the methodologies of these studies in an effort to combine their findings in a meaningful way.

Theoretical Framework and Literature Review

The theoretical framework for this work is well-grounded in many recognized theories and principles of accounting and finance. The study of some given occurrence’s impact on the stock price of a company is commonly referred to by academics as event studies. This is usually done by applying mathematical models such as the Capital Asset Pricing Model (CAPM) or the Fama-French three-factor model to determine abnormal returns after economic and potentially significant events. The methodology of event studies allows one to develop market returns and test an event to see if there is any kind of noticeable change in the regression model. The Capital Asset Pricing Model says that the return one gets, in an efficient
market, is equal to risk-free rate and the product of the sensitivity of the excess asset returns compared against the market’s excess asset returns and the market premium (Gordon). This approach of event studies through CAPM relies heavily on the efficient market hypothesis. This is the notion that shareholders make fully-informed decisions and that information concerning the events is widely and publicly available/known (Fama). This suggests that stock prices are mostly accurate and serve as knowledgeable indicators of the value of a company. More specifically, this notion proposes that the price of stock for a company reflects the collective wisdom of the actionable members of the market in an effort to reflect the present value of future cash flows.

The recent changes in EU regulations through the GDPR require companies to quickly report any news of breaches to the public. This is more likely to produce a profound effect on the current financial evaluation of the firm rather than if the event were disclosed weeks or months after the damages occurred. A company being able to potentially delay the release of this sort of information to time it with certain business conditions, both market-wide and internal, could be considered a “big bath” write-off, an accounting term for making bad results look even worse than they actually were, in order to synthetically make future results look stronger. This time constraint provided by new regulations will provide a more time-efficient
market response to the cybersecurity breach as the information is provided as quickly as possible.

The event studies methodology was first applied to the field of cybersecurity and the analysis of the impact of cybersecurity breaches back in 2003 in a paper by Campbell, Gordon, et al in the *Journal of Computer Security*. This was the first study to find strong negative significance for the impact of cybersecurity breaches on the stock price of firms, albeit only in the case of loss of confidential data. Since then, there have been dozens of papers published on the topic. The core question of many of these papers is the same, but as the papers get more recent, there seems to be a focus on the impact of specific technical types of cybersecurity breaches. These studies also slowly tend towards using the Fama-French model or some four-factor model variant for their analysis. The “events” in these studies were not the cybersecurity breach itself, but rather the announcement or report date of the cybersecurity breach. This paper will refer to the event as the cybersecurity breach for simplicity’s sake. This is an important distinction, as it has sometimes been the case that there is some significant time between the actual occurrence of the breach and when the news of the breach is made public. Hopefully, with the implementation of the GDPR as aforementioned, this timing diminishes in all cases.

An excellent literature review of this field was conducted by Spanos
and Angelis in their paper titled “The Impact of Information Security Events to the Stock Market: A Systematic Literature Review”. The purpose of the paper was to organize the current research on the “economic consequences of [cyber]security incidents” and gain insight into what general conclusions or trends existed among the studies found (Spanos and Angelis). A key finding of their work is that a vast majority of the collected studies concluded that cybersecurity breaches have a significant impact on the stock prices of firms.

While a useful paper for understanding the current state of affairs in this realm of research, there is a significant unanswered question brought about by the Spanos and Angelis paper. No substantive work has been done to take the average of the results of events studies focusing on the impact of cybersecurity breaches on the stock prices of firms and apply probabilities of a breach to this figure. This research serves as an extension of the work done in the literature review by Spanos and Angelis, since that paper’s focus was to understand the general trends of the field and not arrive at any conclusions based on the aggregation of the statistically-significant results of the studies they examined. This research used their data and papers to streamline paper collection and data gathering. The value of data lost in a cybersecurity breach was also not in the scope of the Spanos and Angelis paper, so many comments are added in this paper as an extension of their
work. A more in-depth explanation of the process used in this paper follows.

**Methodology**

This research analyzed fourteen academic papers, all referenced and used in the Spanos and Angelis article, which looked at a total of 1,357 security breaches during an eighteen-year period between 1994 to 2012. This is longer than any analysis period used by any one study in this analysis, which went from 2003 to 2014. They were published in an array of academic journals, ranging from financial and economic to technical and security-focused publications. The standard process for calculating averages and weighted averages was used. This research is an experimentation of combining results from an array of studies that utilized varying methods in their analysis. Probabilities of a breach are used to calculate potential losses to provide a better system for scaling the value of the loss based on unique conditions of each corporation, such as previous cybersecurity investments and type of industry.

There were several limitations to this research, specifically in terms of the methodology and approach of the work. Knowing a true average across the works observed is difficult, as many of them utilized varying methodologies amongst themselves. This lack of cohesion beyond general
practice of using the CAPM lends itself to a few issues when considering the combination of their results. The first issue is that many studies used different evaluation windows for CAPM. Some used a few days of event analysis while others used a single or double-day analysis. This could change the results of this research by artificially weighing the values of smaller windows more highly. Along with this, some studies chose to look at the range of the day before the breach to a day after the breach, while others looked at just the day of the breach or from the day of the breach to two days after the breach. While unlikely to have any significant impact on the research, it is important to mention that this detail is not standardized across these studies. Some studies also made the decision to use an adjusted CAPM that included factors seen in more advanced models. For the purposes of this paper, these adjustments were ignored and treated as a regular CAPM analysis, as there was little observable impact of these adjustments.

Another significant limitation is that there likely exists overlapping data from across the fourteen studies, so the average, in reality, might be smaller than appears in this paper. There is no way of being certain, as not all of the studies included the specific events they observed, and some counted multiple events as a single event. Given more time, a standardization and re-analysis of these studies and their events would be
performed that focus on resolving these limitations. Unfortunately, this would be an incredibly time-consuming task, which may or may not yield significantly different results.

In addition, it is well documented in this field of research, as displayed in the works of Gordon et al, that the Fama-French model is typically more accurate than its CAPM counterpart. This paper analyzes studies using the CAPM over the Fama-French model, because these papers are more numerous, and they ultimately provided more events. The results of this paper could likely be different if the Fama-French model had been used for the event studies. A key example of this is illustrated in the Goel et al paper, where the CAPM produced a result of -8% and the Fama-French model produced a result of -1%. The results of this paper can only be applied to publicly-traded companies that are on US stock markets. The corporations that experienced security breaches in this analysis were also quite large on average, so these results should not necessarily be applied to smaller companies.

The explicit costs of a cybersecurity breach on a firm are readily identifiable. These are typically the costs associated with identifying, correcting, and monitoring a cybersecurity breach. This means that these are usually actionable costs, where someone can clearly see how much of the finances are being allocated to some given incident-handling activity.
While easily calculable, these costs do not tell the entire economic story of a cybersecurity breach. There also exist implicit costs, which are often far more severe, but are always harder to directly observe. These costs are necessary to understand the entire financial picture of a security breach, but there is no true metric for measuring them. Potential means of measurement for implicit costs as a result of a cybersecurity breach could be revenue difference between an average revenue and a year-of-breach revenue or changes in a company’s Brand Index Rating. A comprehensive study combining these metrics and the metric of this paper would be an interesting addition to the field as it stands.

The explicit costs are also somewhat factored into the implicit costs found in this case, as the available information under the efficient market hypothesis includes the knowledge of how much a breach could explicitly cost a company, which usually impacts a shareholder’s decisions about the stock they hold in that company. This study will instead focus on measuring potential losses using changes in the stock price of breached firms. This study chooses this method as there is far more backing in the literature for this particular methodology, and the data and practices around this methodology are richer and more interesting.

To practically combine the results of these studies, certain common conditions had to be created to best choose which of the dozens of previous
studies to include in this study for analysis. The first condition was that the studies had to solely focus on breaches related to US-based companies that operated in US stock markets. This qualification was included in order to remove the question of cultural or political differences toward cybersecurity breaches that might exist between American and non-American markets. It was also included because evidence suggests that US breaches are more impactful than non-US cybersecurity breaches (McCarthy).

Another condition was that the study had to use a one-factor CAPM or adjusted CAPM for their event study. While some of the studies used the Fama-French model, more of the studies used the CAPM or some slight adjustment of CAPM. Due to this, this study chose to look at papers using the CAPM. The next requirement for the papers analyzed in this study was that they had to be categorized as works focusing on the impact of security breaches to breached firms. Many other papers in this area of study instead look at the impact of security breaches to the general stock market or competitors of a breached firm. These were outside of the scope of this research because they did not focus on the impact of a breach to a corporation.

An important requirement for these papers was that they had to have found significance in the abnormal returns of breaches, meaning they had to have had some results which suggested that cybersecurity breaches had any
impact at all on the stock prices of firms. In discussion of significance, it is important to note that there is no guarantee that there will be losses as a result of a cybersecurity breach. Some studies not included in this paper found no significance for abnormal returns as a result of a cybersecurity breach. This would mean that breaches could have produced abnormal returns, but these abnormalities would be so minimal that they were essentially zero or that the stocks behaved completely as expected regardless of the cybersecurity breach.

One study that found a return of essentially zero was included in this work because the work was included in the “Significant” category in the Spanos and Angelis paper. Other studies, like the paper by Gordon, Loeb, et al, found that there was only significance in the case of breaches in the confidential categorization (Gordon). Clearly, the results in academia are mixed. There are many different reasons for the variances in results, but a discussion of those differences is not within the scope of this paper. For the purposes of this paper, the paper agrees that breaches are, in general, significant to the changes of stock prices of firms. The final requirement for these papers was that they had to be relatively recent studies, with no publication prior to 2000. This means that there are many breaches in this study from before 2000, but the studies themselves were strictly not, and they had a few decades to standardize some important aspects of their
methodologies.

Results

After carefully analyzing these other studies, several interesting findings came to light. First, the highest abnormal return found across studies was 8% and lowest was approximately 0%. Most of these results hovered around 0.5% and 2%, suggesting that the minimum and maximum abnormal returns are outliers. With a data set of 14 studies, the decision was made to keep both the highest and lowest abnormal returns for further calculations. Second, the range for the analysis window for CAPM models was 1-3 days. The average model of these studies used 2.36 days for their post-window. This would functionally translate into a two-day window of the day of the breach announcement and the day following it. Then, the non-weighted average of the studies was similar to the weighted average at approximately 1.89%. While not a particularly significant figure, this suggests a fairly even distribution of number of events analyzed across the studies.

Another interesting finding was that the relation of abnormal returns and year of publication, abnormal returns and number of events, abnormal returns and sample time, and abnormal returns and length in days of the
post-window used to produce abnormal returns using CAPM was insignificant (p-value of 0.816), insignificant (p-value of 0.754), insignificant (p-value of 0.716), and insignificant (p-value of 0.0796), respectively. This shows that, as expected, qualities like when the study was performed or what post-window the study used did not impact their results significantly. This means that the impact of the difference in post-window limitation is likely very loose and strengthens the legitimacy of the methodology of this work. Next, the largest number of events studied in any one paper was 306 and the smallest was 10. While this is an exceptionally large range, this range shows that none of these studies had enough studies on their own to support large or definitive claims about all major companies. By combining the results of their work, this paper brings that figure up quite considerably. Finally, and most important, the weighted average of the abnormal returns across the 1,357 events was -1.74%. The concatenation of these studies would then suggest that, on average, a US publicly-traded company could expect potential losses based on its total market value, the above weighted average, and the probability of a cybersecurity breach occurring. Formally, the following model can be used as a sum of the results of this paper:

\[ L_p \approx 0.0174 \times MV \times P \]
Here, \( L_p \) represents the potential losses, \( MV \) represents the total market value, and \( P \) represents the expected probability. This is also a model for a rough estimate of the value of the data compromised.

Figure I: Data Collected from Studies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a2</td>
<td>2003</td>
<td>22</td>
<td>1996-2002</td>
<td>-4.5</td>
<td>3</td>
</tr>
<tr>
<td>a3</td>
<td>2003</td>
<td>23</td>
<td>1998-2002</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>a5</td>
<td>2004</td>
<td>66</td>
<td>1996-2001</td>
<td>-1.36</td>
<td>2</td>
</tr>
<tr>
<td>a8</td>
<td>2006</td>
<td>79</td>
<td>2000-2006</td>
<td>-0.6</td>
<td>2</td>
</tr>
<tr>
<td>a14</td>
<td>2009</td>
<td>168</td>
<td>2004-2008</td>
<td>-8</td>
<td>3</td>
</tr>
<tr>
<td>a15</td>
<td>2009</td>
<td>97</td>
<td>2001-2007</td>
<td>-0.31</td>
<td>1</td>
</tr>
<tr>
<td>a16</td>
<td>2009</td>
<td>151</td>
<td>2005-2008</td>
<td>-0.72</td>
<td>3</td>
</tr>
<tr>
<td>a17</td>
<td>2010</td>
<td>41</td>
<td>1997-2003</td>
<td>-3.18</td>
<td>3</td>
</tr>
<tr>
<td>a18</td>
<td>2010</td>
<td>77</td>
<td>2004-2006</td>
<td>-0.84</td>
<td>1</td>
</tr>
<tr>
<td>a23</td>
<td>2011</td>
<td>123</td>
<td>1994-2006</td>
<td>-0.9</td>
<td>3</td>
</tr>
<tr>
<td>a24</td>
<td>2011</td>
<td>306</td>
<td>2000-2010</td>
<td>-0.3</td>
<td>2</td>
</tr>
<tr>
<td>a26</td>
<td>2011</td>
<td>10</td>
<td>2000-2005</td>
<td>-3.45</td>
<td>3</td>
</tr>
<tr>
<td>a31</td>
<td>2013</td>
<td>89</td>
<td>1997-2008</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>a35</td>
<td>2014</td>
<td>105</td>
<td>2008-2012</td>
<td>-0.33</td>
<td>1</td>
</tr>
</tbody>
</table>

The \( \# \) under publications in Figure I corresponds to the number of the article as described in the Spanos and Angelis article.

**Examples of Applying Probabilities**

Now that the average is known, the rest is relatively simple. As mentioned above, one would need to multiply this figure by the total market value and some probability in order to determine potential losses.
The reason for multiplying by the total market value is that the average percent impact is for every share, and the total market value is the sum of the value of the shares. The use of a probability adds a layer of dynamism to the model so it can be more accurately adjusted for the conditions and requirements of some specific company. To put these findings into a clear perspective, it is useful to provide some theoretical examples of what this average could mean to a large publicly-traded US-based corporation.

Google, with a total market value of roughly 816 billion dollars as of early April 2019, could potentially lose 1.42 billion dollars from a security breach under a ten percent chance of the breach occurring. Netflix, with a total market value of about 156 billion dollars as of early April 2019, could potentially lose 135 million dollars from a cybersecurity breach under a five percent of the breach occurring. Facebook, with a total market value of approximately 476 billion dollars as of early April 2019, could suffer potential losses of over 2 billion dollars from a security breach under a thirty percent chance of the breach occurring. This is similar to the -3% market value drop seen in the first day of Facebook’s breach on September 28, 2018, given a very high probability.

To further this point, Figure II, below, illustrates the percent chance of a breach occurring against the cost of potential losses for the product of the found average. The total market value of ten large US-based tech-
focused companies is provided below. The probabilities assigned here are chosen at random. Production of a meaningful probability is another factor that is incredibly difficult to properly determine and lies outside the scope of this research.

In general, the probabilities applied to achieve potential losses can be adjusted according to the industry of the afflicted company. According to the Verizon 2018 Data Breach Investigations Report, the industries most at risk are finance, healthcare, public sector, retail, and accommodation (Ekran Systems). This means that if a company in one of these five industries were interested in calculating its potential losses as a result of a cybersecurity breach, they would want to use a higher value for the probability that they apply to the product of the total market value and average abnormal return than other companies. While some of the studies analyzed in this paper suggest that cybersecurity breaches impacted tech-focused companies harder, they did not suggest that the probability of a breach occurring to tech-focused companies was higher than the probability of a breach occurring to non-tech-focused companies. The level of investment in cybersecurity systems would be a clear example of a way to decrease the probability of a breach occurring.
This paper argues that the potential loss a company sees through its abnormal returns in stock price as a result of a cybersecurity breach is a useful figure in understanding the economic value of the data compromised in a breach. The potential loss figure of the paper is a valid reflection of a very rough value of the data compromised in a cybersecurity breach, because the data being exfiltrated or removed is most often the economic impetus for hackers or malicious users. Of course, some cybersecurity breaches are for political instability purposes or recognition, but, by and large, cybersecurity breaches result in a removal or transfer of data. Some
breaches, such as those classified as Distributed Denial of Service (DDoS) attacks, do not engage directly with the exfiltration of data, but rather deny the exchange of data between a company and its consumers. This inability to exchange data is similar to the cases where actual data is removed because there likely exists some measurable potential data the company could have gained during unaffected operational hours that has the same value as the regular data removed, but this is outside of the scope of this research.

One of the applicable issues of the value of the data is how one ultimately records cybersecurity breaches on financial statements for both internal and external uses. Commonly, extraordinary events --which could include natural disasters, such as floods or tornados -- are included separately on financial statements. However, as cybersecurity breaches become a more universal and persistent threat, they could be required to be constantly reported, monitored, and evaluated as a potential risk to the company in the near future. Having some grounded quantitative estimate as to the value of the data that could be compromised could more accurately inform investors of the significant costs associated with the financial risks of cybersecurity breaches. This would be a substantial improvement over cookie-cutter disclosures on financial statements with regard to cybersecurity threats of the status quo.
Conclusion

The average data breach could cost a publicly-traded US company a 1.74% hit to their stock value over a multi-day evaluation window. For large companies, findings suggest that breaches are still impactful at low probabilities. These findings also offer arguments that explain how there is value in the digital information lost due to a breach. It is difficult to say how accurate this is as a measurement, methodologies for this analysis were quite different across studies. While not a perfect metric, this research serves as another tool that could be used to determine appropriate cybersecurity investments.

With some fine-tuning, these results could easily be applied to the Gordon-Loeb Model. The Gordon-Loeb Model is an economic model for better approximating how much to invest in cybersecurity, given data value, data risk, and the vulnerability of an attack. The results of this paper could be applied as a base case approximation for the model for given data value and vulnerability of an attack. This is useful because approximating the factors of the Gordon-Loeb model is incredibly difficult -- in fact it is likely impossible to do it perfectly. By contributing some stronger, but still rough, estimates for these figures, one can tighten the bounds of variability in their approximations as a whole.
Although many of these studies cannot provide a definitive answer, the collection of these findings generates a step in the right direction for the insurance underwriting industry. Insurance plans are dependent upon both the magnitude of damages associated with an extraordinary event and the probability of said event occurring during a specific time period. Both of these qualities are required of insurance companies to provide reasonable and useful products to their customers. These findings enable insurance companies to find some justifiable rates for their services with regard to cybersecurity breaches. This approach ultimately bolsters the legitimacy of cybersecurity insurance companies, who are already fighting an uphill battle for defining the terms of their service, as their requirement of attribution of a cyber incident is often difficult for companies to provide.

There are several straightforward examples of future work stemming from this research. One useful extension would be to perform further analysis using more recent information from a database like VERIS, a large dataset of collected and visualized publicly-known data breaches. An interesting work focus in this case could be measuring how the impact of cybersecurity breaches has changed over time and if there is a correlation between the rise in technological use in corporate settings and the frequency of cybersecurity breaches in a given year. Another possibility would be to analyze studies that used the Fama-French model and compare
their results to the CAPM results. As aforementioned, it is well-documented that the three-factor model of Fama-French produces much stronger and tighter results. Research into this could clear a more convincing path for event studies researchers to prefer the use of the Fama-French model. This could also lead to a heavier consideration for some of the available four-factor models in future works.

A final extension of the work could be to factor in window lengths of the CAPM and study the impact of using different windows. Perhaps the most substantial limitation of this research is the inability to know the impact of the different window lengths for the CAPM model. While this study performs a brief test that suggests no significant correlation between the results of the studies and the post-windows they used, that p-value was by far the closest to being significant and so a deeper and more mathematically-involved analysis is likely needed. If other research were to delve into this question, that work could greatly empower the value of this work or demonstrate a further need for standardization in this area of research.

Another way to approach the topic of measuring the implicit costs through an analysis of the economic impact of cybersecurity breaches would be to look at the changes in the Brand Index Rating of a company before and after a cybersecurity breach. There have been many articles
written on this impact in the case of one or two companies, but a more comprehensive analysis with a larger data set would be more statistically significant and could help further display the implicit costs of a cybersecurity breach to corporations. Particularly, research in this area would more fully capture the direct consumer impact that is assumed to be known and utilized when looking at the changes in the stock prices.

Of course, this paper has ignored an important consideration thus far. The CAPM is limited to an analysis of the abnormal adjustments in stock price over a short and limited window after the event. But how is the stock price affected in the long term? The results across academia are mixed. Traditionally, the answer in academia has been that there is no significant long-term impact through the use of the CAPM and Fama-French model.

On the other hand, a new study titled “How data breaches affect stock market share prices,” by the CompariTech team, found that the negative impact of a cybersecurity breach can still be seen up to three years after a breach (Bischoff). Instead of the exact traditional methods of event studies, this study focused on tracking the market performance, in various components, of affected companies over the course of three years after a breach and comparing it against the general performance of the market. While this study had a relatively low sample size of about thirty major
companies, it at least serves as a counterargument against the conventional view and provides another method, albeit a very similar one to CAPM, for analyzing the impact of cybersecurity breaches on stock prices.

To further the discussion of the longevity and severity of cybersecurity breach impacts, it is useful to mention that, unlike the financial scandals of the late 1990’s and early 2000’s, such as the easy example of Enron, no notable company has recently gone under as a result of a cybersecurity breach. This suggests that, while cybersecurity breaches can be economically impactful, they are not permanently damaging, and if long-term impacts exist on the stock prices of companies, they are relatively minimal. Figures III, IV, and V provide a clear look at how minimal the massive breaches were to the long-term market performance of Equifax, Target, and Home Depot.

If it is, in fact, the case that the impacts of these cybersecurity breaches are short-term, then this research would suggest that an optimal time to invest in a company would be immediately following a data breach, based on the philosophy of “buying low and selling high,” as it has been shown that, on average, stock prices of corporations are impacted temporarily. Furthermore, work in this field has been done on whether or not announcements of major cybersecurity investments following a cybersecurity breach impact the stock price of a firm positively with mixed
results. The positive results from these studies would suggest that the market value of a corporation can increase in the long term if handled well.

Figure III: Equifax 1-Year Following (Yahoo Finance)

Figure IV: Target 1-Year Following (Yahoo Finance)

Figure V: Home Depot 1-Year Following (Yahoo Finance)
The findings of this research are meant to illustrate potential costs. That means that the results of this paper do not constitute the whole costs of a cybersecurity breach and that most breaches will cost more than the amount suggested in this research. There is no “right” or “perfect” way to measure the implicit costs of any cybersecurity event. In fact, there is no “right” approach to answering most of the broad-stroke questions related to the economic impact of cybersecurity breaches or the value of the information lost in a data breach. This research is just meant to be an addition to the metrics for determining the economic impact of a cybersecurity breach and the value of the data associated with a cybersecurity breach. As technology continues to become a more integral part of our daily lives, the need for strong cybersecurity and the ability to analyze the economic and financial aspects of information security will become increasingly important.
Acknowledgements

I would like to thank the Ed Snider Center faculty for sponsoring this research. I would also like to thank my research mentor, Dr. David Waguespack, who greatly assisted this work by providing his research insights and by dedicating time to meet with me. Finally, I would like to thank Dr. Lawrence Gordon. His expertise in the accounting and economic aspects of cybersecurity was immensely useful in the directing of this research.
References


The Effects of Decriminalization of Drugs: Lessons Learned from Portugal’s Public Health Approach

Nadia Doherty

Foreword by Research Mentor, Dr. Katie Zafft

The United States is in the midst of an opioid epidemic with deadly consequences across the country. Nationally, policies to handle illicit drug use and trafficking reflect a punitive response to this epidemic and use of other controlled substances with little impact on overdose deaths. As states around the country begin to grapple with how to stem the tide of overdose deaths, we can learn from countries around the world that have made progress using a public health approach to drug policy.

SURE Fellow Nadia Doherty provides a cross-country comparison of Portugal and the United States in an in-depth look at how the United States might learn from Portugal’s implementation of a public health approach to their opioid epidemic of the 1990s. Nadia’s project grew out of curiosity about the effects of changing drug policies across the United States, and how decriminalization of drugs may impact use rates and whether this effort is a step in the right direction to adequately respond to drug overdoses. Nadia uses sources detailing historical drug law changes and trends in health statistics for both countries to explore differences and potential policy
effects. Her synthesis of drug policy in the United States and Portugal offers insight into how decriminalization and a shift to public health may work in the United States to reduce the damaging effects of our opioid epidemic.

Nadia Doherty is a SURE Fellow in the Government & Politics Department and Criminology & Criminal Justice Department at the University of Maryland, College Park, MD. Their email address is: nadiaadoherty@gmail.com.

Abstract

In the 1990s, Portugal experienced a severe drug crisis. The nation was rampant with drug use, addiction, and had one of the highest rates of HIV/AIDS related to drug injections. After recognizing the scope of its drug problem, Portugal decriminalized the use and possession of all substances in 2001. Decriminalization was supplemented with several programs related to harm reduction and providing treatment to users who needed it. By creating greater access to treatment and promoting safer use of drugs, Portugal exemplified the importance of a public health approach in dealing with a drug crisis. By focusing efforts on the user and treating drug abuse as an illness rather than a crime, Portugal experienced significant positive impacts. Drug use was contained and remained relatively low compared to other European countries. HIV/AIDS infection rates related to drug injections dropped, and overdose deaths were minimized. The current “War on Drugs” in the United States is a public safety attempt to remedy the nation’s drug crisis and control the overall
drug supply. Unfortunately, it has been largely unsuccessful. Drug use and abuse in the United States continue to increase, as millions of Americans abuse substances like opioids, and thousands die from overdoses annually. The United States has treated users and addicts as criminals, reprimanding them with punitive measures rather than considering them as patients who need treatment. To mitigate its drug problem, the United States must recognize drug use as the public health crisis it is. Portugal has shown that a public health approach which prioritizes the safety, treatment, and rehabilitation of the user produces significant results. If the United States does not change its current policy approach, millions of Americans will continue to suffer from drug abuse, and more lives will be lost.

Keywords: decriminalization, drugs, use, abuse, addiction, public health, public safety, opioids, treatment, rehabilitation, Portugal, United States

Introduction

Drug use and addiction are problems that countries around the world face. In 2015, approximately 5% of the global population reported using drugs at least once and .0.6% of the population-- 44 million people at the time -- suffered from drug addiction. An estimated 28 million healthy years of life, adjusted for disability, were lost as a result of drug use. 17 million
healthy years of life were lost because of drug use disorders (UN Office on Drugs and Crime, 2017).

Portugal has shown the world that with the right approach and infrastructure, these deaths can be minimized and problem drug use can be contained. Following rampant drug use by its citizens in the 1990s, Portugal restructured its approach and developed laws and programs that targeted treatment of drug users. With the decriminalization of drugs in 2001, Portugal marked drug use and addiction as a public health problem rather than a matter of criminal justice. Portugal has since seen significant progress in drug addiction and disorders, overdoses, and other drug-related issues.

The United States, meanwhile, is dealing with one of the worst drug problems in its history. The United States accounts for one-quarter of the estimated drug related deaths worldwide, especially overdose deaths. Drug use continues to rise. Something must change if we want to make a difference. Portugal’s success serves as an example of what can happen when drug use is treated as a public health issue, rather than a public safety issue. This paper will analyze Portugal’s drug policy and outcomes of its decriminalization, treatment, and harm-reduction policies. Through this analysis, this paper will highlight the importance of a public health approach and lessons the United States can learn from Portugal.
Literature Review

A country can take a number of approaches to address drug use and addiction. The 2016 United Nations General Assembly Special Session on drugs identified drug addiction as a “health disorder characterized by chronic and relapsing nature,” (Volkow et al., 2017, para. 1). This identification also recognized that addiction is preventable and treatable and is an illness rather than criminal behavior or moral failure. Addiction experts from the Special Session also stated that drug disorders are brain disorders that people can recover from with evidence-based treatment and social support, rather than punitive measures and criminal sanctions.

A public health-based approach emphasizes the social protection of users and access to prevention, treatment, and recovery measures for substance abuse (Volkow et al., 2017). An important method in a public health approach is harm reduction. According to the American Public Health Association (2013), harm reduction policies focus on reducing the adverse effects of drug use and ineffective drug policies. These policies accept that drug use is present and using can be harmful, and therefore do not aim to unrealistically eliminate use. In Portugal, harm reduction policies include needle and syringe exchange programs, low-threshold substitution programs, drop-in shelters, contact units, and outreach teams.
Drug use and addiction are also addressed as a matter of public safety. According to Sacco (2014), a public safety approach focuses on supply reduction through strict domestic drug enforcement. The idea is that reducing the supply will restrict and reduce the availability of illicit drugs, therefore driving down overall drug use down. This strict enforcement is demonstrated by the thousands of drug-related arrests, seizures, and prosecutions that occur in the United States every year (Sacco, 2014). In an issue brief on heroin abuse from Pew Charitable Trusts (2015), researchers suggest that a combination of law enforcement to reduce trafficking and limit the emergence of new markets is a viable public safety solution for mitigating growth of heroin abuse. Chandler, Fletcher, and Volkow (2009) identify mass incarceration and criminal supervision as results of stricter drug laws and enforcement, influenced by a public safety approach. Many policymakers and researchers suggest public safety initiatives as part of a larger, comprehensive approach to drug use blended with other solutions.

The difference between a public health and public safety approach is how the user is treated. In a public health approach, such as Portugal’s, an individual who abuses or is addicted to drugs is treated as someone who is suffering from an illness, rather than a criminal. Substance use disorder is considered a serious illness and health crisis and is addressed by providing the resources to treat and rehabilitate the user. A public safety approach,
such as the one used by the United States, seeks to cut off the supply of
drugs from users and bring punitive measures against people who commit
drug offenses. Substance use disorders are seen as a criminal matter that
must be met with strict enforcement and punishment to deter future use.

Methodology

Portugal was the first European country to decriminalize the
personal use and possession of all illicit drugs and abolish criminal
penalties. Portugal’s progressive approach has subdued its drug crisis and
transformed how drug use and addiction are treated. The United States and
Portugal share similar forms of government and, at one time, faced similar
issues related to drugs. This paper focuses on a comparison of Portugal
and the United States. This comparison will be developed with qualitative
inquiry, which allows for simplification of data within the context of each
country. Qualitative inquiry generates new ways of understanding existing
data, which will be applied by exploring drug-related changes and their
relation to public health and public safety (Ochieng, 2009).

The context of each country’s drug problem and respective drug
policies will be examined. Various outcomes and data related to Portugal’s
decriminalization will be observed and analyzed. Data and statistics from
the United States also will be analyzed. The data from both countries will
explore several issues related to drugs, including use, overdose deaths, HIV/AIDS infections, incarceration, and treatment. Throughout the data analysis, this paper will seek to identify and compare how each policy approach produces similar or different results in the selected topics related to drugs. Following the data analysis and comparisons, this paper will propose recommendations for the United States to address its drug problem and lessons that can be learned from Portugal’s approach.

**Portugal’s Drug Problem**

Following the political revolution of 1974 and booming economic growth, Portugal experienced several changes. This economic growth brought greater population concentration to cities, especially the capital, Lisbon, and popular port city Porto. It also brought challenges, including drug use. While data on the extent of drug use during this time period is scarce, Portugal was experiencing serious problems related to drug use and abuse and other drug-related issues (Van Het Loo, Van Beusekom, & Kahan, 2002). Of Portugal’s population of 10 million, approximately 100,000 people, or 1% were addicted to drugs. Many addiction cases involved heroin, the highly addictive opioid that is often injected.
The drug problem had many negative societal impacts. In 1999, Portugal had the highest drug-related HIV infection rate in the European Union, and second highest prevalence of HIV among injecting users. The number of recorded incidents of drug use requiring treatment increased over 400%, from 56,438 cases in 1990 to 288,038 cases in 1999 (Laqueur, 2015). In 1991, 4,667 people were arrested for drug-related offenses. By 1998, that number increased to 11,395. Also in 1998, 61% of arrests were for drug use or possession, 45% of them heroin-related. Heroin and cocaine seizures had the largest increase from 1990 to 1998, with cocaine seizures increasing by approximately 300% and heroin seizures by approximately 200% (Van Het Loo et al., 2002). It was clear that drug use was on the rise, and a new strategy was needed to slow its progression.

**Portugal’s Current Drug Policy**

Throughout the 1990s, Portugal struggled to find ways to deal with the rising drug use and addiction. According to a legal report from the Soares (2016), the policy transformation involved a number of bureaucratic and legislative changes. On February 16, 1998, Portugal took measures to change its approach by creating the Commission for National Strategy for Drug Control. Tasked with proposing a new strategy to combat drugs and
drug addiction, the commission prepared a report on its findings and proposed solutions. Arguably the most important recommendation was the decriminalization of drug use. Decriminalization, the report said, had the potential to reduce harm and achieve better control over the current drug crisis while discontinuing useless punishment. This recommendation called for the use, possession, and purchase of all drugs to be reclassified as administrative offenses. It noted that the cultivation of drugs was still subject to criminal sanctions, as it was closely related to drug trafficking. In response to the report, on April 22, 1999, the Portuguese government created the National Strategy for the Fight Against Drugs and Drug Addiction.

The newly proposed National Strategy was implemented through three major pieces of legislation: Law No. 30, Decree-Law No. 130-A, and Decree-Law No. 183. Law 30, approved in November of 2000 and implemented on July 1, 2000, officially decriminalized the use, possession, and purchase of all drugs. As administrative offenses, a new system was formed to address the necessary ramifications. Under Law 30, the Commission of Dissuasions was formed and managed through the Ministry of Health, rather than the Ministry of Justice (Moreira, Hughes, Costa Storti, & Zobel, 2011). An individual is referred to the Commission of Dissuasions when they are found using or in possession of no more than a
10-day supply of drugs, which are subsequently seized. Any individual found in possession of greater than a 10-day supply will be prosecuted in court for a criminal offense.

An administrative drug offense case is transferred to one of Portugal’s 18 district commissions, where the individual will meet with three appointed officials, who include a legal worker appointed by the Ministry of Justice and medical professional and social worker appointed by the Ministry of Health. The offender initially meets with a small team of practitioners with similar roles as the officials to evaluate the present case and provide an assessment. This evaluation identifies the extent and circumstances of the use, substances used, economic status of the user, and whether the user is an addict. The case is then sent to the officials where they seek to understand the drug use and recommend treatment and rehabilitation. The Commission also has several options for mitigating offenses and determining appropriate consequences, including warnings, banning from public places or interaction with others, removal of a professional license, required periodic visits to a defined place, or impose fines. These sanctions can be imposed for a minimum of one month to a maximum of three years (Moreira et al., 2011).

On April 23, 2001, Decree-Law No.130-A was enacted to establish the organization, process, and operating system of the Commission of
Dissuasions, which would become active in July. The law discusses the territorial jurisdiction of the commission based on district lines and sets forth reporting requirements and preliminary measures for the police officer aware of the offense. Cases are reported to the commission within 36 hours of occurrence to ensure efficiency. Offenders are required to meet with the commission within 70 hours from the occurrence (Soares, 2016).

The final piece of legislation implementing the National Strategy was Decree-Law No. 183, enacted on June 21, 2001. This law focused on creating programs and public health structures to address harm and risk reduction and prevention strategies. It discusses the duty of the state to provide and promote access to these programs for all drug users to protect public health and the health of the users themselves. These programs and structures include supporting offices for drug addicts, reception and shelter centers, points of contact and information, mobile spaces for prevention of infectious disease, replacement programs, needle exchange programs, street teams, and programs for supervised use (Soares, 2016). Many of these programs attempt to minimize harm and risk, promote safer drug use and treatment, reduce drug use, and develop outlets of support.

**United States Federal Drug Policy**
In the United States, overdose deaths, mostly involving opioids, have tripled since 1999. Far more people die from an overdose than in car accidents or violence (UN Office on Drugs and Crime, 2017). Drug use continues to rise and more people continue to abuse drugs, like opioids each day. Over the past 30 years, United States drug policy has remained largely unchanged, with exceptions of changes related to marijuana use at the state level. The Federal Comprehensive Drug Abuse Prevention and Control Act of 1970 was signed into law by President Richard Nixon and became effective May 1, 1971. This act placed all substances that were under existing federal regulation into one of five drug classification schedules. The goal of the Controlled Substances Act was to “improve the manufacturing, importation and exportation, distribution, and dispensing of controlled substances” through three different pillars or titles (Gabay, 2013, para. 2). Title I addresses development of rehabilitation programs for drug abusers; Title II, registration and distribution of substances; and Title III, import and exportation of controlled substances and related issues. The Controlled Substances Act, Title II of this law, includes the foundation of the United States’ fight against drug and substance abuse. As part of this law, manufacturers, distributors, and dispensers of controlled substances are required to register with the Drug Enforcement Agency. This registration establishes a “closed system” of disclosure and traceability
from manufacturers to dispensers of these substance, and overall accountability in the pharmaceutical industry (Gabay, 2013).

Under the Controlled Substances Act, these substances are categorized under five schedules. These substances are identified as medications or drugs that are easily abusable and placed into their respective schedules for a variety of reasons about the drug’s acceptable medical use or abuse and dependence potential. The act outlines eight reasons for a substance to be controlled or removed from control: (1) its actual or relative potential for abuse; (2) scientific evidence of its pharmacological effect, if known; (3) state of current scientific knowledge regarding the substance; (4) its history and current pattern of abuse; (5) the scope, duration, and significance of potential abuse; (6) consideration of any risks to public health; (7) its psychic or physiological dependence liability; and (8) whether the substance is an immediate precursor of a substance already controlled under this subchapter (United States Drug Enforcement Agency, n.d). Heroin and marijuana are among other drugs that fall under Schedule I substances, the most dangerous category, defined as having no current acceptable medical use and a high potential for abuse (United States Drug Enforcement Agency, n.d).

Violation of the Controlled Substances Act and other drug-related laws results in a variety of consequences. A federal drug offense warrants
the imposition of fines or imprisonment, which varies by the type of violation, substance, and amount of the substance that was seized. The first major offense is trafficking, the “unlawful distribution or possession with intent to distribute, manufacture, importation, or exportation of a substance,” (Yeh, 2016, p. 4). For example, an individual charged with a federal trafficking offense of heroin is subject to fines up to $10 million and imprisonment between ten years to life. An individual found in possession of marijuana under 50 kilograms or 49 plants is subjected to fines up to $250,000 and a maximum of 5 years in prison. If a trafficking offense involves the death or serious bodily injury of another person, all Schedule I and II substances are subject to 20 years to life imprisonment.

The second major offense category is simple possession, where the fines and terms of imprisonment increase as the number of subsequent offenses increases. The third major offense category is controlled substance violations and other related crimes, which includes a variety of offenses, from attempt and conspiracy and establishing manufacturing operations, to robberies involving controlled substances. The fourth major offense category is racketeering, defined as commission of a pattern of activity involving one or more dangerous drug felonies to invest in, acquire, operate, or participate in the affairs of an interstate enterprise. The other offense categories include smuggling, laundering money from controlled
substances, and tax offenses involving income from controlled substance violations (Yeh, 2016).

Following the Federal Comprehensive Drug Abuse Prevention and Control Act, another major piece of legislation related to drug policy and enforcement was passed. The Anti Drug-Abuse Act, passed in 1986 and amended in 1988 following a moral panic surrounding rampant crack cocaine use revised federal drug laws to increase and improve the enforcement and sentencing of drug offender, as well as authorize more funding for treatment and rehabilitation. Substantially fewer funds than appropriated were actually used for treatment purposes. But the act’s biggest impact was the implementation of minimum sentencing laws. Mandatory minimum sentencing laws require a judge to impose a prison sentence based on the nature of drug charges of which a defendant is convicted. The Anti Drug-Abuse Act imposed minimum sentences of 5 years in prison for specified amounts of drugs such as 5 grams of crack, 500 grams of cocaine, or 1 kilogram of heroin. It also imposed minimum sentences of 10 years for specified amounts, such as 50 grams of crack or 5 kilograms of cocaine. These mandatory minimum sentence laws have since contributed to the substantial growth of the federal prison system and have been applied disproportionately to lower level offenders as well (Criminal Justice Policy Foundation, n.d).
While the United States federal drug policy has not experienced much change or reevaluation, many states have taken their own actions toward a progressive approach. In 2000, California passed a ballot initiative, Proposition 36, which mandated that all adults convicted of possession or use of illegal drugs were to be sent to treatment in lieu of incarceration. The law also required California to designate funds to support this plan (Klein, Miller, Noble, & Speiglman, 2004). The state recognized that rather than considering penal consequences, an individual’s need for treatment must take priority in determining how to intervene. Similarly, many states have adopted more liberal drug policies in relation to marijuana. Fourteen states have decriminalized the recreational use of marijuana, and ten states have legalized its use. Thirty-three states have legalized medical marijuana use (The National Organization for the Reform of Marijuana Laws, n.d).

Many states also have implemented harm reduction policies to promote safer drug use and minimize associated risks. Several states have syringe exchange programs, where users have free access to sterile needles and syringes and can exchange or dispose of used ones. These programs help HIV prevention and reduce the risk of hepatitis-C infection. As of 2016, states and local communities now can access federal funding to support certain aspects of these exchange programs (Center for Disease
Control and Prevention, 2018). Another major harm reduction program is the supervised consumption facility, or safe injection site, which provide a safe and hygienic environment for users to inject drugs under medical supervision. While its implementation is fairly new, cities such as New York and Philadelphia predict great success in preventing overdoses, saving money in healthcare costs, and reducing HIV cases related to injection (Larson, Padron, Mason, & Bogacyzk, 2017).

**Data Analysis**

*Substance Use and Disorders*

Portugal experienced many changes in drug use after 2001. The National Population Survey on Psychoactive Substances in the Portuguese Population was administered in 2001, in 2007 and 2012. In 2001, the study recorded that 7.8% of respondents aged 15-64 reported using illicit drugs at least once in their life. The number increased to 12.0% of respondents in 2007, then decreased to 9.5% in 2012. The largest increase during this period for lifetime prevalence use was for marijuana usage, which increased from 7.6% of respondents in 2001 to 11.7% in 2007, then decreased to 9.4% in 2012. Heroin usage for this measure increased slightly from 0.7% of respondents in 2001 to 1.1% in 2007, then decreased to 0.6% in 2012.
Other substances did not experience significant changes or maintained the same percentage (Santos, Duarte, & Maia, 2013).

There were similar trends when measuring responses for drug use in the previous year. Illicit drug use in the past year among respondents increased from 3.4% in 2001 to 3.7% in 2007 and decreased to 2.7% in 2012. Similarly, for marijuana usage, 3.3% of respondents said they had used in the past year in 2001, 3.6% in 2007 and 2.7% in 2012. Heroin usage in the previous year remained mostly unchanged, with fluctuation between 0.1 to 0.3% between 2001 and 2012 (Santos et al., 2013). Overall, drug use in people ages 15-64 mostly increased following decriminalization in 2007 and subsequently decreased in 2012.

Many Portuguese policymakers have emphasized the importance of fighting drug use and addiction among young adults, aged 15 to 34. The belief is that young adults can be educated about the prevention of drug use and abuse and potentially be treated for addiction more effectively than older adults. Marijuana is the most commonly used drug among young adults and the Portuguese population in general. Following marijuana, young adults also reported using cocaine and MDMA/ecstasy. For marijuana use, approximately 5.1% of young adults said they used in the past year in 2012 and 8.0% of young adults in 2016. For all other substances, young adult usage has steadily declined since 2007. Cocaine
use was reported by 0.3% of young adults in the past year, while MDMA usage was 0.2%. Comparatively, drug use among both young adults and the general adult population in Portugal, with the exception of high-risk heroin usage, is much lower than other European countries. High-risk opioid usage, which is characterized as recurrent, harmful use, is found in approximately 5.2 per 1000 people of the Portuguese adult population. This is slightly above average compared to other European countries (EMCDDA, 2018).

Illicit drug use is evidently prominent in the United States. The National Survey on Drug Use and Health (NSDUH) collects data and information on ten categories of drugs in the United States: marijuana, cocaine, heroin, hallucinogens, inhalants, methamphetamine, and the misuse of prescription pain relievers, tranquilizers, stimulants, and sedatives. In 2017, approximately 30.5 million Americans aged 12 and older -- 1 in 9 -- reported currently using illicit drugs, in the past month. This number is up from 28.6 million using in 2016, and 22.6 million in 2010. Marijuana was the most commonly used substance, which 26.0 million people reported using. Approximately 3.2 million people reported misusing prescription pain relievers, the second highest use of illicit drugs. Roughly 241.6 million respondents (89%) reported not using illicit drugs in the past month at all. When broken down by age group, in 2017
approximately 2 million adolescents aged 12-17 reported using in the past month. This represents about 7.9% of total adolescents. Young adults aged 18 to 25, approximately 24.2%, about 8.3 million, reported using. Approximately 9.5% of adults age 26 or older, 20.2 million people reported using illicit drugs in the past month (Substance Abuse and Mental Health Services Administration, 2018).

Statistics about the misuse of opioids, including heroin, are extremely important especially considering the perceived epidemic of opioid use in the United States. Opioid misuse includes the use of prescription pain relievers, such as oxycodone, heroin and fentanyl. In 2017, there were approximately 11.4 million opioid misusers aged 12 or older in the United States in the past year. That is approximately 4.2% of the adult population. The vast majority of these individuals misused prescription pain relievers. When analyzed by age, approximately 3.1% of adolescents, about 769,000, reported misusing opioids. About 7.3% of young adult respondents, or 2.5 million young adults, reported misusing opioids. For adults, approximately 8.1 million, 3.8% of total adult respondents, reported misusing opioids in the past year (SAMHSA, 2018).

The NSDUH also collected data on people with drug disorders in the United States. Drug disorders are defined in the survey as the dependence on or abuse of one or more illicit drugs. A person is defined as having a
drug disorder for a particular drug if they meet at least three out of the seven criteria points. The criteria include characteristics of dependence relating to continued use, tolerance, and withdrawal symptoms. Users not meeting the dependency criteria were then filtered through criteria for drug abuse. From this application, the survey found that in 2017, approximately 7.5 million people, or 2.8% of people aged 12 or older, had a drug disorder. When broken down by age, approximately 3% of adolescents, or 741,000 adolescents, had a drug use disorder. Approximately 2.5 million young adults, or 7.3%, and approximately 4.3 million adults, or 2.4%, had a drug use disorder. (SAMHSA, 2018).

**Overdose Deaths**

The number of overdose deaths in Portugal has undergone significant change in the past decade. Portugal has used different indicators for deaths attributed to drug intoxication. The primary indicator up until 2010 was the National Institute of Forensic Medicine indicator of drug-related deaths, which measured the “number of deaths that involve a positive post-mortem toxicological test for the presence of illicit substances.” Unfortunately, this indicator came with limitations, as it was highly dependent upon changes in recording practices, such as the overall number of autopsies performed. It was also an indirect indicator, as one can
have traces of drugs in the body, but that does not necessarily mean the
drugs were the cause of death. Due to these limitations, overdose deaths
have not been measured before or directly after the reform (Hughes &
Stevens, 2012). Data on overdose deaths became widely available in 2010
after a call by the European Monitoring Centre for Drugs and Drug
Addiction (EMCDDA) to improve this indicator. From the newly available
data collected by the EMCDDA, Portugal recorded 94 overdose deaths in
2008. The number of overdose deaths dropped to its lowest in 2011, with
19 deaths, and has risen to 27 as of 2016. Over the eight years that were
recorded, the number of overdose deaths has decreased or fluctuated within
reason and maintained a lower level than prior to decriminalization
(European Monitoring Centre for Drugs and Drug Addiction, 2018).

Another indicator to measure drug-attributed deaths, defined as
people determined by doctors to have died from drugs, was created by the
National Statistics Institute. This measure coincides with the EMCDDA’s
overdose measure and is able to backtrack to the beginning of the
decriminalization era in 2001. The number of deaths attributable to drugs
was approximately 375 in 2001, plunging to its lowest at approximately 10
drug attributed deaths in 2005. From 2005 to 2008 there were slight
increases in the number of deaths, but the number remained much lower
than at the time of decriminalization (Hughes & Stevens, 2012).
The United States has been experiencing a severe number of drug-related overdose deaths, especially with the rise of opioid addiction and the prevalence of fentanyl. Every day in the United States, more than 130 people die from an opioid-related overdose (National Institute on Drug Abuse, 2019). According to the CDC, the rises in overdose deaths came in three waves. The first wave started in the 1990s, with the increased prescription of opioids which led to the rise of related deaths beginning in 1999. The second wave began in 2010, with a rapid increase in overdose deaths involving heroin. The third and most severe wave began in 2013, with synthetic opioids, such as illicitly-manufactured fentanyl, in overdose deaths (Center for Disease Control and Prevention, 2018). In 1999, the United States had 16,849 drug overdose deaths. The numbers rose to 36,010 in 2007, and to 70,237 in 2017. The number of overdose deaths from opioids significantly attributed to this large increase over the 18-year period. In 1999, the number of overdose deaths involving opioids was 8,048. 18,515 in 2007. 47,600 in 2017. Overdose deaths involving opioids made up approximately 68% of the total overdose deaths in 2017 (National Institute on Drug Abuse, 2019).

**HIV/AIDS Infections**
Portugal’s approach and strategy transformation were largely prompted by their HIV/AIDS infection rates from drug-related injections. According to data from Portugal’s Institute on Drugs and Drug Addiction, new cases of HIV among drug users in treatment dropped from 1,482 new cases in 2000 to 116 in 2010. In relation to AIDS diagnoses, the number of new cases among drug users in treatment in 1999 was 675 and declined to 88 in 2010 (Laqueur, 2015). For new cases attributed to drug injections among the general population, there were 493 cases in 2006. This number dramatically decreased over subsequent years to just 30 new cases in 2016 (EMCDDA, 2018). It is important to note that while Portugal has experienced a decline of new infections from injecting drug users, they still have one of the five highest rates in the EU.

The United States deals with similar risks and concerns related to HIV/AIDS infections from drug-use infection. From 2008 to 2014, the United States experienced an overall decrease in annual HIV diagnoses among injecting drug users. In 2015, 6% (2,392) of the approximate 39,513 new HIV diagnoses were attributed to injection use. The following year, in 2016, people who inject drugs accounted for 9% (3,425) of the 39,782 new HIV diagnoses in the United States (Center for Disease Control and Prevention). Overall, the percentage of HIV diagnoses attributed to
injecting drug users in the United States has remained at a moderately low amount in the past decade with reasonable fluctuation from year to year.

**Incarceration**

Incarceration is used throughout the criminal justice system as punishment for committing a crime and deterrence for others not to commit crimes. Throughout the 1990s and by 2000, Portugal had one of the highest rates of incarceration in Western Europe and the highest proportion of drug offense prisoners. Portugal had 14,276 drug-related arrests and 3,829 people in prison for drug offenses in 2000. Of the those incarcerated for drug offenses, only 25 were convicted for use (Laqueur, 2015). This shows that even prior to its policy change, Portugal did not incarcerate many individuals for use, but rather focused on trafficking-related charges. These 3,829 people in prison for drug offenses made up 43% of the prison population. Four years after decriminalization, in 2005, drug offenders decreased to 28% of the total prison population. In 2010 only 1,950 people were in prison for drug-related offenses, which was only 21% of the total prison population. In both 2000 and 2010, the vast majority, approximately 90%, of drug-related offenders were sentenced for trafficking offenses, while the rest were mostly minor trafficking and traffic-use offenses (Laqueur, 2015).
The United States has very different statistics on incarceration and prison population related to drug offenses. Unlike Portugal, the number of people in federal prison for drug offenses has followed an overall increasing trend over time. In 1980, only 4,700 people were in federal prison for drug-related offenses (Sentencing Project). According to the Federal Bureau of Prisons, as of 2019, 167,763 people are in federal prison, 76,191 are for drug offenses. This means that 45.4% of the current federal prison population are drug offenders. (Federal Bureau of Prisons, 2019). In prisons under state jurisdiction, 1,306,305 individuals were in state prisons at the end of 2017, accounting for approximately 88% of the total prison population including both federal and state prisoners. About 190,100 of prisoners in state facilities or 15% of the total state prison population, had been convicted and serving time for a drug offense, with approximately 24% serving sentences for drug possession (Bronson & Carson, 2019).

Treatment

Portugal’s statistics regarding treatment services can be largely attributed to its efforts to increase access and promote quality treatment for drug users. The first year of data collection on treatment centers was 1998, when it was reported that approximately 23,654 drug users received some
form of drug treatment. Following decriminalization, in 2008, the total number of users in treatment was 38,532. There was also an increase in the number of outpatient facilities, from 53 in 1998 to 79 in 2010 (Laqueur, 2015). In 2016, 30,653 people were receiving treatment, the majority of whom were in specialized drug treatment centers as part of outpatient care. Approximately 2,729 patients received inpatient care through therapeutic communities and residential drug treatment. Approximately 42% of people entering treatment were there for heroin abuse, while 39% were for marijuana abuse. Despite the majority of people entering treatment for heroin, the number of first-time entrants for heroin has been decreasing since 2009, while the number of first-time entrants for marijuana have been increasing (EMCDDA, 2018). This pattern in first-time entrants is in line with current trends in drug use, as marijuana use has become more popular and has increased in Portugal.

According to the 2017 National Survey of Substance Abuse Treatment Services conducted by Substance Abuse and Mental Health Services Administration (2018), there were approximately 14,000 substance abuse facilities in the United States. The actual number is slightly higher, as about 11% of facilities did not respond to the survey. On March 31, the survey reference date, there 1,356,015 people were enrolled in substance abuse treatment. The proportion in treatment for drug abuse
increased from 36% in 2007 to 47% in 2017. Outpatient treatment was offered by 82% of facilities and received by 91% of individuals in treatment. Residential treatment was offered by 23% of all facilities and received by 7% of individuals in treatment. Medication-assisted opioid therapy, which includes receiving substitute substances such as methadone, was available in 10% of all facilities.

In 2017, the National Survey on Drug Use and Health (NSDUH) reported that approximately 7.6% or 20.7 million people aged 12 or older needed substance use treatment. But only 4 million people, or 1.5%, actually received substance use treatment in 2017. About 94.3% of people aged 12 or older who needed substance use treatment, but did not receive it, believed they did not need treatment in the past year for their substance abuse (SAMHSA, 2018).

**Discussion**

Portugal has experienced significant changes following the implementation of its new strategy in 2001. These changes are not necessarily a direct effect of decriminalization, but rather an overarching shift of addressing drug use and addiction as a public health crisis. After evaluating the data and statistics of several areas related to drug use,
addiction, and law enforcement, we can observe significant differences between Portugal and the United States.

Drug use is a complex issue to understand in terms of how it was affected by Portugal’s policy. Many believed that decriminalization would promote greater drug use. Others believed that, like other factors in Portugal, it would dramatically decrease. But, neither of these changes occurred. Drug use fluctuated and changed depending on the type of drug and age group. Marijuana use increased overall, following trends throughout Europe and across the world. The increased use of marijuana was occurring even before decriminalization and continued to prevail post-decriminalization. As marijuana use increased, heroin use decreased or remained comparatively lower than pre-decriminalization levels. One can argue that rising marijuana use levels are less severe than potentially rising use levels of heroin. An explanation for lower rates of heroin use and other illicit drugs may be the strength of the economy. Portugal’s economy has grown and experienced great success, which creates a better social fabric and quality of life and thus can lead to less self-medication, a common coping practice with stress or financial burden. Other factors that contribute to the prevalence of drug use include availability and price, and broader social or cultural trends, as is seen in marijuana use.
Portugal’s public health initiative largely affected the decline in HIV/AIDS diagnoses and overdose deaths. Despite some fluctuation and different measures of overdose deaths, Portugal has experienced significantly lower amounts of drug-induced deaths since decriminalization. The drop in new cases related to drug injections and deaths can be attributed to greater treatment access and harm reduction policies. Portugal distributes hundreds of thousands of clean syringes through their exchange programs, which helps prevent many users from injecting drugs with used or dirty needles, a primary cause of HIV/AIDS infections. Additionally, users have greater access to treatment in outpatient and inpatient centers, as well as substitution programs and drop-in shelters. As part of the substitution programs, methadone is provided for free to people addicted to opioids such as heroin. Methadone is ingested orally and prevents withdrawal and sedative effects. Methadone is highly effective, requiring only a single daily dose compared to multiple doses of heroin a day (Jacinto, Madeira, Santos, Goulão, & Pádua, 2015). Methadone treatment also reduces overdoses and HIV/AIDS infections related to injections, as well as contributing to lower rates of heroin use.

By decriminalizing drug use and prioritizing the access and resources available to drug users for treatment, more people have used drugs more safely or pursued treatment. Once criminal sanctions were
removed, public attitudes and opinions were ultimately changed. Stigma surrounding drug use and addiction disappeared, and users were able to seek treatment without feeling shame or attacks on their dignity. Prior to decriminalization, a substantial barrier for addicts seeking treatment was the hesitation to engage with or discuss their use with government officials in fear of arrest and punishment. Now users can feel comfortable seeking treatment and ways to use drugs safely to reduce potential harm.

Prioritizing the health concerns surrounding use and addiction and directing punitive efforts towards drug trafficking has led to safer drug use overall.

The United States’ approach to drugs has not produced significant progress in treating drug use and addiction. The United States has focused most of its efforts on controlling the supply of drugs, rather than treating the demand and adverse effects of use. While the recent government administration has declared an opioid crisis, the United States continues to address this issue through public safety measures (The White House, 2018). It has perceived the drug problem as being fueled by drugs supplied from south of the border, becoming part of a broader agenda on immigration. The concentration of efforts made by the United States is continually presented with a punitive approach that punishes substance abuse disorders and supplemented by a lack of efforts to improve policy at the federal level. Attempts to restrict supply has failed to curb drug use or rising addiction
and overdose deaths. Rather, the United States has punished thousands of nonviolent drug addicts for use or small-scale possession. Clearly, this has not reduced the supply of drugs or deterred future use, because the drug problem continues to persist. Supply control has led to an increase in illegal drug markets and sales, which in turn has increased illicit drug use.

Significant federal funding for treatment and recent development of harm reduction policies still falls short of adequately addressing the drug problem in the United States. Facilities are not well-equipped with proper treatment methods or medications, and there is not sufficient access to treatment and rehabilitation services. Even if there were sufficient access, the stigma surrounding addiction and fear of prosecution can make users afraid to seek help. Many states implement court-ordered participation in treatment for drug offenses. However, the treatment is for short periods of time and has found to be ineffective. Studies have shown that mandated treatment has little effect on use and recidivism (Werb et al., 2016). Additionally, a study by the Massachusetts Department of Public Health found that many users who were involuntarily committed for opioid abuse treatment were twice as likely to die from a subsequent opioid overdose than those who chose to enter treatment (Massachusetts Department of Health, 2016). By forcing users to attend treatment, this ignores whether the user is actually committed to getting better and the extent of treatment and
resources the user may need beyond the mandated treatment. Similarly, a study of prisoners in the state of Washington found that prisoners are almost 13 times as likely to die from an overdose within their first two weeks of release than the general population (Binswanger et al., 2007). This public safety approach in the United States is forcing people into either prison or ineffective treatment, neither of which are viable solutions. This drug crisis is not receiving federal attention or recognition as the public health crisis that it is. Spending billions of dollars on a public safety approach and incarcerating thousands of people each year is clearly not working, as people continue to die each day from drug use and overdoses.

There were several limitations related to the scope of the data, specifically in Portugal. Portugal did not collect sufficient data on drug use prior to decriminalization, so comparisons before and after 2001 cannot be made. Many studies in Portugal, like its survey report on drug use, were not conducted regularly or did not collect extensive data. The primary national study reported data in 2001, 2007, and 2012. This study has not released recent data, rather different studies have released similar data in 2015. The lack of continual data collection created difficulties in identifying trends or changes in use during the past twenty years due to gaps in the data. Similarly, with overdose deaths, different indicators created a difference in drug-related deaths versus drug-induced deaths. Therefore, any early data
before or around the period of decriminalization related to drug-related
deaths were not used, as it was an inaccurate indication of deaths caused by
drugs. These outcomes, both positive and negative, are not solely
attributable to decriminalization. While this paper intends to show that
decriminalization combined with public health initiatives, such as the
expansion of treatment and harm reduction services, influenced the results,
there are also a variety of economic, social, and cultural factors that can
influence the observed data.

**Conclusion**

Portugal and the United States have shared similar challenges and
experiences related to drug use and addiction in the past twenty-five years.
Drug use and addiction have worsened over time and led to the deaths of
thousands of individuals each year. But, unlike the United States, Portugal
has addressed this issue of drugs from a different perspective. Portugal
decriminalized the use and possession of all drugs and began focusing on
the treatment and rehabilitation of the user. By addressing drug-related
issues as a matter of public health and treating users as humans with an
illness rather than criminals, Portugal has seen tremendous success over
time. Portugal’s transformation and overall success in mitigating its drug
crisis serve as a lesson to the United States of how to tackle its current drug
problem. The United States must reduce the stigma and public attitude surrounding drug users and addicts. Reducing stigma and negative attitudes reduce any fear individuals may have to seek treatment. By recognizing the importance of all lives and working to treat users, the United States can save thousands of lives each year.

If the United States wants to achieve any significant change in its current situation, it must consider Portugal’s methods. While a complete transformation of policy and drug enforcement infrastructure at the federal level is unlikely at the moment, one recommendation may be to start at the state level. Harm reduction policies, like syringe exchanges and substitution treatment, should become more widely adopted across all states to promote safe usage and minimize the risk of an overdose. Additionally, the prosecution of individuals using or possessing small amounts of drugs must not be prioritized. The United States cannot continue to fill prisons with drug users and addicts when they pose little threat to society. They need to be helped, not punished. The United States must designate more federal attention and funding towards the development and improvement of treatment programs and facilities. It must ensure that users who need help have access to quality treatment.

Further research may be conducted to determine how the United States would adopt the Portuguese model of decriminalization and public
health prioritization. If adopted federally, research may explore how the infrastructure of the criminal justice system, enforcement agencies, and funding would change to accommodate a new public health approach. Because the changes in Portugal cannot be claimed as a direct effect of decriminalization and new policies, further research may consider whether the United States would experience similar progress towards substance use and abuse, and any obstacles they may encounter. This paper solely focused on the lessons that the United States can learn from Portugal. Other research may be conducted to compare Portugal and apply any relevant lessons to different countries around the world that may be experiencing problems similar to the United States.

The United States must understand that a drug-free society is unattainable, but a society with limited drug-related deaths and open access to treatment for users and addicts is. Decriminalization does not legalize or justify the use of drugs, rather it allows for drug use and addiction to be mitigated through health and rehabilitation programs, rather than punitive measures. Portugal has shown the United States that in order to combat a drug crisis, it must address use and addiction as health challenges rather than criminal matters. If the United States wants to alleviate its current drug crisis, it must focus efforts on treating drug users, preventing drug-related harms, and addressing their illnesses. Without the recognition and
implementation of public health initiatives, drug use and addiction will continue to take the lives of thousands of Americans each year.

Acknowledgements

I would like to thank Dr. Katie Zafft, of the Department of Criminology and Criminal Justice, and the Ed Snider Center faculty for supporting this research project.
References


Syringe Service Programs. Retrieved from


Understanding the Epidemic. Retrieved 2019, from

https://www.cdc.gov/drugoverdose/epidemic/index.html


Improving Public Health and Safety. *JAMA, 301*(10), 183-190.

doi:10.1001/jama.301.10.1024-d


https://www.cjpf.org/mandatory-minimums


https://www.bop.gov/about/statistics/statistics_inmate_offenses.jsp

doi:10.1310/hpj4806-473


doi:10.17265/2328-7136/2015.02.005


Incorporating a public health approach in drug law: lessons from local expansion of treatment capacity and access under California's


doi:10.1111/lsi.12104


European Monitoring Centre for Drugs and Drug Addiction website:


*Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health* (HHS Publication No. SMA 18-5068, NSDUH Series H-53).

Retrieved from https://www.samhsa.gov/data/


Differences in Media Representation in Low Casualty Terror Attacks

Elizabeth Geary

Foreword by Research Mentor, Dr. Amy Pate

The impact of terrorist attacks and how society responds to them is mediated in part by how attacks and their perpetrators are represented in the media, both traditional and new. Additionally, reporting in news outlets of terrorist attacks is almost inevitably mediated by societal biases, potentially amplifying distorted understandings of the phenomenon.

In her project, SURE Fellow Elizabeth Geary analyzes how political leanings of U.S.-based news outlets relates to the coverage and representation of terror attacks in the United States, comparing attacks perpetrated by Muslims with attacks by other perpetrators. Notably, she finds that even though Muslim perpetrators were responsible for a minority of attacks, they received more coverage than non-Muslim perpetrators.

Elizabeth was a student in my course on “Terrorist Motivations and Behaviors” when she approached me about mentoring her for her SURE project. It has been rewarding to support her as she goes through the research process, including building an 152 original dataset for analysis. I’m pleased that she hopes to continue data collection and analysis on the project during
the summer on her own, to deepen her understanding of how characteristics of terrorist attacks influence their portrayal in a variety of media sources. Elizabeth Geary is a SURE Fellow in the Criminology & Criminal Justice Department and School of Public Policy at the University of Maryland, College Park, MD. Their email address is: egeary1@terpmail.umd.edu.

Abstract

The way the public views terrorism is impacted by the way the media represents terror attacks. Since people are creatures of habit and typically stick to their media outlet of choice, does the leaning of the media outlet, left, center/unbiased, or right leaning, change the context of terror attacks and therefore inform the public about terror attacks differently? In this research, 776 articles across the political spectrum were read to understand the different context used for Muslim and non-Muslim perpetrators and if there was a difference in the context utilized by the different leanings of the media outlets. Overall, by perpetrator there was very little variation in the context utilized by the media outlets, except that right leaning outlets were more likely to reference ISIS or al Qaeda ties or links and were more likely to use the phrase “radical Islamic terrorism” than their counterparts. Non-Muslim also was found to have very little variation in the context utilized, unless the perpetrator had a Muslim sounding name. Even if the attack was not jihadi motivated, non-Muslim perpetrators with Muslim sounding names were more likely to be
contextualized in the media the same way as Muslim perpetrators. Muslim perpetrators were also more likely to receive higher media sensationalism than non-Muslim perpetrators.

**Introduction**

Terror attacks are a real and known threat around the world. The way we learn about terror attacks is through media coverage. However, not all media outlets are the same, and most people stick with their media outlet of choice because of the content they choose to share. This begs the question: Do different media outlets report on terror attacks differently?

For this research, the definition of terrorism, as defined by the Study of Terrorism and Responses to Terrorism (START), is “the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation” (*Codebook: Inclusion Criteria and Variables*, 2018). To be included in this research, the attack must meet all three of the criteria as defined by START: the act must be aimed at attaining a political, economic, religious, or social goal; there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims; and it must be outside the context of legitimate warfare activities. All of the terror attacks chosen for
this research were found using START’s Global Terrorism Database, available through the START website.

Media attention to these attacks matters. It signifies to the public that this is something they need to be informed about and proceeds to inform them about what they think the public needs to know. If the context of this information or the level of reporting differs among news outlets, depending on their bias leaning, that informs the public differently on types of terror attacks, depending on their news outlet of choice. The term ‘bias’ is not used negatively, since all of the news sources chosen for this research have received very high factual reporting ratings. Instead, bias here refers to the likelihood of the news source reporting on events that align with the news source’s viewpoint and reporting in such ways that follow their agenda. Therefore, I hypothesized that the different leanings of the news outlets will affect their level of media attention and context in which they give that attention. This research aims to determine if the quality and quantity of media attention differs based on the perpetrator and the sensationalism of the attack.

**What receives more media attention?**

Previous research has attempted to determine what receives more media attention and why. Most recently, Kearns (2018) has looked at which
factors caused certain terror attacks to receive more news coverage than others. Overall, their findings reveal that the most important factor in the terror attack receiving media attention was if the perpetrator was Muslim. Controlling for other factors, they found that Muslim perpetrators received between 366% more media coverage when controlling for outliers. Based on this finding, it would be a reasonable assumption to think that Muslim jihadis are committing the majority of terror attack inside the United States. However, a 2017 report by the United State Government Accountability Office found that following the 9/11 attacks and through 2016, far-right extremists accounted for 62 terror attacks while Islamic extremists accounted for 23 attacks (Countering Violent Extremism: Actions Needed to Define Strategy and Assess Progress of Federal Efforts, 2017).

When looking at which types of crimes get more coverage in newspapers, Katz (1987) found that crimes that are less likely to happen, such as murder, are more likely to get reported. There were also themes of “moralized political conflict,” where the reader was able to decide moral judgement of the perpetrator for the more horrific crimes. Both of these findings translate into the public’s consumption of terrorism news, since the likelihood of being a victim of a terror attack is incredibly low, and it is easy to declare the actions of a terrorist as absolutely morally wrong. Because terrorism, a moralized political conflict, tells more of a story than
just a ‘typical’ murder, Katz says these crimes are more likely to be reported.

Similarly, in a 1973 study of British daily newspapers, Roshier found that “whimsical circumstances” played the largest role in a crime being covered in the newspaper. While not entirely “whimsical,” terror attacks offer a form of drama to the news that is not typical in everyday life or in the news. In previous studies on some of the factors that influence the newsworthiness of a story, Galtung and Ruge (1965) found that certain factors, including meaningfulness, unexpectedness, and reference to something negative, increase the likelihood of the story receiving a sensational amount of media coverage. In a follow-up study to Galtung and Ruge, Harcup and O’Neill (2001) found that similar factors, such as surprise, bad news, magnitude, and relevance all play an increased role in the story receiving media attention. These factors are all present in terror attacks and shape the way the media reports these attacks. Another key factors that Harcup and O’Neill (2001) found was the newspaper’s agenda. If a terror attack is used as a way for the media outlet to push their own agenda, we will see an increase in reporting of that attack by that media outlet.
What are the different frames?

Since September 11, 2001, the United States and the media have framed terror attacks in a different light. Many studies have been devoted to exploring this shift and understanding the way that terror attacks are now portrayed in the media. Commonly, terror attacks with Muslim perpetrators have led to news outlets attributing the attack to international ties and framing the attack as international, even if the perpetrator is an American citizen (Powell, 2011; Powell, 2018; Mitnik, 2017). This frame pushes the burden of blame off American shoulders and onto the country the media finds the perpetrator to have traveled to, most commonly Middle Eastern countries such as Yemen. Powell (2018) even found that some news outlets went as far as to force international angles on Muslim perpetrators, even if there was no international travel or ties. Mitnik (2017) also found that non-Muslim perpetrators did not have the same international angle forced upon them. They were more commonly referred to simply as “domestic terrorists.”

There is debate among researchers about whether post 9/11 terror attacks have episodic or thematic framing. Nacos and Torres-Reyna (2003) define episodic frames as news articles that focus on a singular event while thematic frames focus more on context and continuity of the story over the single event occurrence. Ruigrok and Atteveldt (2007) found that after
9/11, news outlets shifted away from episodic framing and focused more heavily on thematic framing. 9/11 brought more light on terror attacks and, overall, increased their newsworthiness, shifting reporting from focusing on the single event to broader coverage of not only that attack, but also related attacks and the war on terror. Papacharissi and Oliveira (2008), however, claim that terror attack news coverage is still focused on episodic framing, based on their research of The New York Times and The Washington Post reporting. Because Ruigrok and Atteveldt (2007) and Papacharissi and Oliveira (2008) use different sources, with Papacharissi and Oliveira (2008) solely studying articles written by The New York Times and The Washington Post.

Do these frames influence readers?

Even with these researched frames made by media outlets, frames mean nothing if they have no effect on the audience. Gilliam and Iyengar (2000) looked at the different ways that media can influence people’s opinion on criminal justice attitudes. When people were exposed to a more “radical” and harsh script by journalists, they were more likely to be supportive of more severe criminal justice practices, such as capital punishment and mandatory minimums. Gilliam and Iyengar also found that
media make use of stereotypes and uses “racial imagery” they studied to see if this type of reporting affects how readers view the perpetrators. They found that those exposed to the harsher script were more likely to have negative views of minorities. They also found that the media is more likely to present crime news on minority perpetrators, even though they did not make up the majority of perpetrators. By describing the violent nature of the crime and giving racial characteristics to the perpetrator, the readers drew conclusions about who the dangerous people in society are and how they should be handled within the criminal justice system. Gilliam and Iyengar concluded that the media’s reporting did have an impact on readers’ understanding and future ideas. This also applies to the way news outlets report on terror attacks, especially if there are discrepancies in how they report on Muslim perpetrators and non-Muslim perpetrators.

In studies on people’s perceptions of terroristic behaviors, researchers have found that frames do influence perceptions of who is a terrorist (West & Lloyd, 2017; Kearns et al., 2019; Haider-Market et al., 2006) but in different ways. West and Lloyd (2017) found that perpetrators with Muslim sounding names were more likely to have their actions labeled as terrorism than perpetrators with white names. Kearns (2019) found that personal factors, such as prior Islamophobic feelings and trust in the news outlet, impact the way individuals view a frame of terrorist reporting, as
well as the frame the news outlet gives the attack. Finally, Haider-Markel (2006) found that the framing of news stories influenced readers’ perceptions of future terror attacks but did not influence how they voted on future policy initiatives. Overall, Kearns and Haider-Markel found that the way the media frames terror attacks does have an influence on the way readers perceive terror attacks. These findings continue with West and Lloyd’s research and build off of Gilliam and Iyengar’s research, which shows how media frames cause readers to perceive Muslims as the main perpetrators of terror attacks, which creates a negative and stereotypical viewpoint of Muslims.

**Methodology**

Using the criteria set about by the Global Terrorism Database from START and setting the date of attack timeline between January 1, 2008 and December 31, 2017 and the casualty count between 1 and 10, 70 terror attacks occurred within the United States. LexisNexis was then used to find reputable nationwide media coverage of the event and the thirty days following the attack. Using a media bias fact check, I organized the news sources into three categories of bias: left leaning, right leaning, and center/unbiased. Each article had to call these incidents “terrorism” to be
included in the research. Articles that referred to the attacks only as hate crimes or crimes of violence were not included.

For this research, I hypothesized that the context of reporting would differ among these three categories of news sources depending on whether the perpetrator was Muslim or not and if there was low or high sensationalism. In this context, sensationalism describes terror attacks that received at least twenty articles from any of the news outlets. This led to there being 42 terror attacks that resulted in no articles; 9 terror attacks that resulted in articles when the assailant was Muslim, 5 of which were highly sensationalized; and 20 terror attacks that resulted in articles when the assailant was non-Muslim, 2 of which were highly sensationalized.

To gain an understanding of the context of the media reporting, the articles were skimmed to learn the main points the media tried to get across. The articles were then reread and the context for each article recorded. For Muslim perpetrators, the main context points were mentioning of; background on the attack; ISIS or al Qaeda; international travel, especially travel to the Middle East; policy issues or changes relating to the attack; the phrase “radical Islamic terrorism” and questioning if the attack is terrorism.
“Background” refers to a significant amount of information reported in regards to the terror attacks. Background can include information about the attack, the perpetrator, or the victims. “ISIS or al Qaeda” refers to the media outlet either confirming a relationship between the perpetrator and ISIS or al Qaeda or questioning if there is a relationship. This can include both ISIS or al Qaeda directed attacks or inspired attacks. “International travel” refers to either the media outlet reporting on the perpetrator’s international travel or questioning if they have travelled internationally.

“Policy issues or changes” refers to the media outlet using the terror attack to discuss policy issues or changes resulting from the terror attack. For example, some terror attacks sparked debates about terrorism watch lists, gun laws, or airport security. The terminology for “radical Islamic terrorism” refers to any phrase that positions the Muslim faith with terrorism, such as radical Islamic terrorism, Islamic extremists, Muslim extremists, etc. Phrases such as radical extremism were not included since they did not link the Islamic faith with the terror attack.

The articles were read blindly; I did not know which news outlet I was reading before or while I was skimming the article to avoid any accidental bias. For the left leaning news sources, I used MSNBC and CNN. For the right leaning news source, I used Fox News. For the center/unbiased news sources, I included sources such as PBS, NBC, ABC,
CBS, USA Today, New York Times, The New York Post, and Washington Post. I chose these sources because they are nationwide, widely digested, and highly reputable. These attacks were then separated into two categories: if the Global Terrorism Database defined the attack as jihadi-inspired or not. They were then split further based on whether attack generated any articles about it. These 29 attacks, of 9 Muslim perpetrators and 20 non-Muslim perpetrators, were read for context and sensationalism to determine what, if any, difference there was in reporting based on the media outlet’s leaning.

I then completed Two Sample T Tests for the data to determine the differences. For each contextual category, I found the average number of times the context was used per attack. For example, in the Detroit attempted terror attack of an airplane, center outlets linked the perpetrator to al Qaeda in 46 of their 75 or 61.33%, of their articles. These averages were used to find the mean and standard deviation for the outlet overall per category. Any attack that did not generate articles by one of the outlets was omitted instead of counting it as a zero, so that only reported on context would be included in this research.

**Results**

1. **Muslim and Non-Muslim Perpetrators**
The first analysis was on the difference in context in reporting where the perpetrator was Muslim. These accounted for 9 of the attacks that were used in this research. After reading all of the articles used in this study, I noticed the six main themes of mentioning background, ISIS or al Qaeda, international travel, policy issues or changes, the phrase “radical Islamic terrorism,” and questioning if the attack was in fact an act of terror. The use of the media outlets discussing background, international travel, policy issues or changes, and questioning if the attack was terrorism was not statistically significantly different among all three types of media outlets. In looking at the outlets’ mentioning of ISIS or al Qaeda, left leaning and center had no statistically significant difference in their reporting, but there was a difference in right leaning reporting for both left leaning and center outlets. At the 95% confidence interval, there was a statistical significance that right leaning articles discuss ISIS and al Qaeda ties more than left leaning articles. Similarly, right leaning outlets reported more often than center outlets at the 90% confidence level. There also was a statistically significant difference in the use of the phrase “radical Islamic terrorism.” When comparing left leaning and center articles, there was no statistically significant difference. However, when comparing right leaning articles to both left leaning and center articles, there was a statistically significant
difference that right leaning articles use the phrase “radical Islamic terrorism” more often than the counterparts at the 98% confidence interval.

Table 1 - Muslim Perpetrators

<table>
<thead>
<tr>
<th></th>
<th>Background</th>
<th>ISIS/al Qaeda</th>
<th>Travel</th>
<th>“RIT”</th>
<th>Policy</th>
<th>Terrorism?</th>
<th>Article Count Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean %</strong></td>
<td>85.6, 85.6, 87.21</td>
<td>44, 48.37, 73.3</td>
<td>24.2, 30.38, 26.35</td>
<td>6.1, 4.568, 29.48</td>
<td>34.44, 46.27, 48.46</td>
<td>10.97, 2.44, 5.47</td>
<td>40.78, 16.56, 9.44; 66.78 (not %)</td>
</tr>
<tr>
<td><strong>Std dev</strong></td>
<td>0.1763, 0.1843, 0.2011</td>
<td>0.291, 0.2873, 0.33</td>
<td>0.2373, 0.3418, 0.273</td>
<td>0.08766, 0.06935, 0.2451</td>
<td>0.3065, 0.412, 0.3416</td>
<td>0.1683, 0.06378, 0.1169</td>
<td>56.588, 24.54, 9; 88.518</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>9, 8, 8</td>
<td>9, 8, 8</td>
<td>9, 8, 8</td>
<td>9, 8, 8</td>
<td>9, 8, 8</td>
<td>9, 8, 8</td>
<td>367, 149, 85; 601</td>
</tr>
</tbody>
</table>

(Left leaning/center/right leaning/total numbers; total only applies for article count total)

The second analysis was on the difference in context in reporting where there was a non-Muslim perpetrator. 20 attacks were studied in this section. There were some of the same themes found in both Muslim and non-Muslim, such as both kinds of attacks brought about discussion over background, mentioning policy issues or changes, and questioning if the attack is terrorism. Both mentioning policy issues or changes and questioning if the attack is terrorism had no difference among the three types of media outlets reporting. When measuring the use of background
information, left leaning and center, as well as right leaning and center, were not statistically significantly different, while right leaning articles were statistically significantly larger at the 98% confidence level.

Table 2 - Non-Muslim Perpetrators

<table>
<thead>
<tr>
<th></th>
<th>Background</th>
<th>Terrorism?</th>
<th>Policy</th>
<th>Article Count Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean %</strong></td>
<td>79.33, 83.75, 98.21</td>
<td>47.69, 38.75, 22.5</td>
<td>26.24, 27.08, 23.48</td>
<td>6.15, 1.5, 1.1; 8.75 (not %)</td>
</tr>
<tr>
<td><strong>Stan dev</strong></td>
<td>0.3331, 0.3538, 0.05051</td>
<td>0.4374, 0.4657, 0.42</td>
<td>0.3607, 0.407, 0.3474</td>
<td>12.06, 2.259, 1.971; 15.4949</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>19, 10, 8</td>
<td>19, 10, 8</td>
<td>19, 10, 8</td>
<td>123, 30, 22; 175</td>
</tr>
</tbody>
</table>

(left leaning/center/right leaning/total numbers; total only applies for article count total)

Comparing the average amount of time policy issues or changes were discussed by the media outlets found no statistically significant difference if the perpetrator was Muslim or non-Muslim. For non-Muslim perpetrators, across all media outlets, policy issues or changes were brought up approximately 25% of the times that the terror attack was discussed. For Muslim perpetrators, this average across all media outlets was just under half of the articles, with left leaning outlets bringing down the average. These were not found to be statistically significantly different, however, because all of these variables have very high standard deviations,
demonstrating a lot of variability in media’s use of mentioning policy issues or changes, regardless of the perpetrator.

In comparing the differences in coverage of media outlets questioning if an attack was terrorism for Muslim and non-Muslim perpetrators, there were statistically significant increases in questioning if the attack was terrorism for both left leaning and center media outlets at the 99% and 95% confidence interval, respectively. Right leaning media outlets, however, were found to have no statistically significant difference in questioning if the attack was terrorism regardless of whether the perpetrator was Muslim. This can be explained, however, since right leaning outlets were less likely overall to question if an attack was terrorism, regardless of whether the perpetrator was Muslim. In reading the context of these articles, right leaning articles were more definitive in their word choices in describing these attacks. Very rarely did the reporter waver on calling an attack terrorism, while left leaning and center news outlets were more likely to be hesitant in handing out a “terrorism” label to an attack, especially for terror attacks perpetrated by non-Muslims.

Another finding in the non-Muslim, non-jihadi inspired attacks was that the news outlets, regardless of leaning, gave the same contextual coverage to attacks where the perpetrators had Arabic sounding names as they did to attacks by Muslim perpetrators. The attacks led by Ahmed
Aminamin El-Mofty, Kori Ali Muhammad, and Amor Ftouhi had all media outlets report on possible ISIS or al Qaeda ties, international travel, and the use of Arabic words being spoken during the attack. None of the other 17 non-Muslim, non-jihadi attacks include any form of international travel or international terrorist ties that are commonplace contextual frames in all outlets for Muslim perpetrators. Media outlets, regardless of their leaning, were more likely to try to associate Muslim perpetrator characteristics on attacks that were not jihadi motivated.

Of all the 20 terror attacks by non-Muslim perpetrators that occurred in this ten-year timeframe, these three were the only attacks with names of Middle Eastern descent. The way the media tried to frame these non-jihadi attacks under the same lens as jihadi attacks shows that it is less about the motive of the attack and more about the public’s perception of the attack. If the public, told by the media, thinks the attack is perpetrated by someone of Middle Eastern descent and is Muslim, demonstrated by their name, they are more likely to label the attack as terrorism. For example, Amor Ftouhi, attacked and killed a man at the airport in Flint, MI, with an anti-government motivation. This non-jihadi attack received the second most media attention of all the non-jihadi attacks, topped only by Dylan Roof’s attack on an African-American church, which killed nine people. A similar attack happened at a Louisiana airport, where an assailant attacked a TSA
agent. This attack received no attention from any media outlet designating the attack as “terrorism.” One difference about the attackers, however, was that the Flint airport attacker was named Amor Ftouhi, while the Louisiana airport attacker was named Richard White.

2. Sensationalized Coverage

In terms of coverage of terror attacks, left leaning news outlets are substantially more likely to cover the attack and have more articles about the attack. In low sensationalized attacks, attacks with less than twenty articles covering the attack, left leaning outlets covered 21 of the 22 attacks. Center outlets covered 12 of the 22, and right leaning outlets covered only 8 of the 22 terror attacks. Left leaning outlets composed almost three times the number of articles as center outlets and more than four times as many articles as right leaning outlets. This shows a considerable difference in where the coverage of these low sensationalized terror attacks is coming from. Left leaning outlets were found to cover low sensationalized terror attacks statistically significantly more than center outlets and right outlets, both at a greater than 99.99% confidence level. Center outlets and right leaning outlets, however, were not found to report on low sensationalized terror attacks at a statistically significant difference.

Table 3 - Low Sensationalized Attacks
<table>
<thead>
<tr>
<th></th>
<th># of Left Leaning</th>
<th># of Center</th>
<th># of Right Leaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>2.9545</td>
<td>1.04545</td>
<td>0.72727</td>
</tr>
<tr>
<td><strong>Stan Dev</strong></td>
<td>2.3397</td>
<td>1.3266</td>
<td>1.2792</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>65</td>
<td>23</td>
<td>16</td>
</tr>
</tbody>
</table>

Of the 29 terror attacks, 22 were considered low sensationalism, meaning these attacks generated less than twenty articles combined among all three types of media outlets. Of these 22 terror attacks, 4 were committed by Muslims with jihadi motivations, making these attacks 18.2% of the low sensationalized attacks. However, media outlets overall reported more articles on these attacks than on the non-Muslim perpetrators, even within the low sensationalism group. Muslim attackers accounted for 24.6% of left leaning outlets’ articles, 26.1% of center outlets’ articles, and 37.5% of right leaning outlets’ articles. Even as low sensationalized attacks, terror attacks committed by Muslims with jihadi motivations generated more articles.

For high sensationalized terror attacks, where the attack received more than twenty articles of coverage from all three types of outlets combined, left leaning outlets once again triumphed in leading over center and right leaning outlets. Left leaning outlets reported again almost three
times as many articles as center outlets and almost five times as many articles as right leaning outlets. Statistically speaking, left leaning outlets were found to have significantly more coverage than both center and right leaning outlets at a 99.99% confidence level. Center outlets, as well, reported more frequently than right leaning outlets at a 99.99% confidence level, with center outlets reporting almost two times as many articles as right leaning outlets.

However, three of the high sensationalized terror attacks only reached their 20-article coverage through the heavy use of left leaning outlets supplying the articles. I then looked at terror attacks that resulted in at least ten articles per type of media outlet to account for this discrepancy. This did not account for a large difference in findings, since left leaning outlets still reported over two times as many articles as center outlets and over four times as many articles as right leaning outlets. Center outlets were also found to report almost twice as many articles as right leaning outlets. Again, even accounting for this possible discrepancy, left leaning outlets were found to report more high sensationalized terror attacks than center or right leaning outlets at a 99.99% confidence level. Center outlets, as well, reported more high sensationalized terror attacks than right leaning outlets at a 99.99% confidence level. However, due to the fact that there were only four terror attacks used in this statistical analysis, outliers will heavily
affect the data and could be the cause for the extremely low p-values that were calculated.

Table 4 - All High Sensationalized Attacks

<table>
<thead>
<tr>
<th></th>
<th># of Left Leaning</th>
<th># of Center</th>
<th># of Right Leaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>60.7143</td>
<td>22.4286</td>
<td>12.857</td>
</tr>
<tr>
<td>Stan dev</td>
<td>54.7623</td>
<td>25.317</td>
<td>7.5813</td>
</tr>
<tr>
<td>N</td>
<td>425</td>
<td>155</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 5 - High Sensationalized Attacks with Each Leaning having over 20 Articles

<table>
<thead>
<tr>
<th></th>
<th># of Left Leaning</th>
<th># of Center</th>
<th># of Right Leaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>80.25</td>
<td>35.25</td>
<td>17.75</td>
</tr>
<tr>
<td>Stan dev</td>
<td>67.874</td>
<td>27.6451</td>
<td>6.131</td>
</tr>
<tr>
<td>N</td>
<td>321</td>
<td>141</td>
<td>71</td>
</tr>
</tbody>
</table>

One of the most interesting results in the context differences between Muslim and non-Muslim perpetrator reporting was the different context used by all or some of the media outlets. For example, across all media outlets, terror attacks with Muslim perpetrators led to discussions about links or ties to international terrorist organizations, specifically ISIS.
and al Qaeda. Some instances, such as the Detroit airplane attempted terror attack, were al Qaeda directed attacks. However, most of the terror attacks by Muslim perpetrators were labeled by the media as ISIS or al Qaeda inspired. Attaching this tagline of “inspired” to these terror attacks creates an international link for Muslim perpetrators, even though the perpetrator may be a U.S. citizen. As Powell (2018) discusses in her study of terror attacks, giving Muslim perpetrators international ties or links and not giving these same links to non-Muslim perpetrators creates a fear of the “international Muslim terrorist” who will come into the United States to commit terror attacks. It creates an “us vs. them” ideology, which, in turn, influences the way American citizens view Middle Eastern countries and the people who live there.

3. Attacks that led to no media attention

During this ten-year period from January 1, 2008 until December 31, 2017, there were 70 terror attacks in the United States. However, only 29 of these attacks resulted in media outlets referring to them as “terrorism.” The other 41 attacks either received no media attention or no media outlets calling the incident “terrorism.”

Of the 41 attacks, 26 resulted in zero deaths, 9 resulted in one death, and 6 attacks resulted in two deaths. None of the terror attacks that received
no articles had more than two deaths, even if more injured victims were involved. Of the attacks that did result in media attention, it does not appear that the number of articles is linked to the total number of victims killed, but more on the impact of the attack. The attack that led to the most deaths was the Charleston church shooting by Dylan Roof, with nine victims killed. However, only 69 articles were published in total about this attack, making it far from the attack that generated the most media attention. The attack that did, in fact, receive the most media attention was the failed terror attack on an airplane in Detroit, which led to 277 articles being published by the media. This attack did not even result in a single fatality, only injuring the attacker and a good Samaritan who made sure the fire the attacker attempted to create did not catch the plane on fire. The attack that generated the third most media attention, at 81 articles, was again another attempted terror attack with no fatalities. when a man tried to blow up a bomb attached to himself at Port Authority in New York City. With these two attacks in particular, they had less fatalities than attacks that received no media attention. However, these attacks had the potential for an egregious amount of innocent human lives lost. The Detroit airline had well over 300 passengers. Port Authority has hundreds of people regularly passing through.
The second most reported on attack was a terror attack on a Navy Operational Support Center in Chattanooga that resulted in five fatalities. The fourth most reported attack was the Charleston church shooting. Both of these attacks resulted in a large number of fatalities, so it is intuitive to understand why these attacks generated large amounts of media attention. Taking into account the number of people who could have died makes it easier to understand why these attacks received such high sensationalism. Conversely, an attack against police and civilians in Las Vegas resulted in five fatalities but received only three articles from the media, two of them from left leaning outlets and the third was from a center outlet.

Of these 41 attacks, four were jihadi motivated attacks. Percentage wise, 9.76% of the attacks with no media articles were jihadi motivated terror attacks. Of the 29 attacks that did receive media attention, nine were jihadi motivated. Percentage-wise, 31% of attacks with articles were jihadi motivated terror attacks. Of all the articles that were used for this research, 601 covered the nine jihadi motivated attacks, compared to just 175 articles covering the 20 non-jihadi motivated attacks. Therefore, 77.4% of the articles were concerning the nine jihadi-motivated attacks while only 22.6% of the articles covered the 20 non-jihadi motivated attacks.

For attacks with Muslim perpetrators, left leaning outlets accounted for 61.1% of the articles, center outlets accounted for 24.8% of the articles,
and right leaning outlets made up 14.1% of the articles. This breakdown, of left leaning outlets reporting the most, then center outlets, followed by right leaning outlets, continues with non-Muslim perpetrators; left leaning outlets accounted for 70.2% of the articles, center outlets accounted for 17.1% of the articles, and right leaning outlets concluded the final 12.6% of the articles. This breakdown matters because it shows that left leaning outlets report overall more on terror attacks than their counterparts. However, it also shows that this breakdown is less pronounced in terror attacks committed by Muslim perpetrators, showing that center and right leaning outlets report more often on attacks if the perpetrator is Muslim and less likely to report as often on non-Muslim perpetrators. The result is skewing public perception of terror attacks.

The effect this can have on the public

The way these news outlets contextualize these terror attacks is their frame of the event. The public relies on the media to present the news so they can understand what is important to know about these terror attacks. People, in general, rely on their one, reliable, news source to gather their information. With these similarities and differences in news outlets reporting on terror attacks, the public is not equally aware of all of the same things. With different framing of each terror attack by different media
outlets, depending on the leaning of the outlet, the public is not equally informed and does not perceive terrorism the same way. For example, people who follow more right leaning news outlets may be more likely to perceive the terror attack to be due to “radical Islamic terrorism,” since right leaning outlets use this phrase significantly more than left leaning or center new outlets. Similarly, people who follow more left leaning news outlets may exaggerate the threat of terrorism since left leaning outlets were more likely to report on terror attacks in general than their counterparts.

While research has mixed findings on the way framing can affect the public perceptions, it cannot be argued that these different frames do exist and do depend on the leaning of the media outlet. While much of the context remained constant across the attacks, the differences that were found could have nationwide effects on how terrorism is viewed by the public.

**Conclusion**

When separating the attacks by whether they were committed by a Muslim or non-Muslim, there was not a lot of variation in the context that three different types of media outlets provided. For terror attacks committed by Muslims, there was found to be no statistically significant difference in the use of discussing background, international travel, policy
issues or changes, and questioning if the attack should be considered terrorism. The differences, however, show how the leaning of the media outlet affects the context that they choose to broadcast to the public. Right leaning outlets were found to more often discuss links or ties to ISIS or al Qaeda than left leaning or center outlets. Right leaning outlets were significantly more likely to use the phrase “radical Islamic terrorism” for Muslim perpetrators than either left leaning or center leaning media outlets. These differences show that those who watch and read right leaning outlets are more likely to understand terror attacks perpetrated by jihadists as being linked to Islamic faith and links them to international terror groups, even if they are United States citizens.

Within non-Muslim perpetrators, there was very little difference in the context presented in the three different leanings of media outlets. Right leaning outlets were found to report more background information than left leaning outlets, but, overall, there was no major difference in the reporting context.

While again there was very little variation for non-Muslim perpetrators in the context shared by all three types of media outlets, the more interesting findings were in the similarities and differences in the context for non-Muslim perpetrators and their Muslim counterparts. Some of the context between the two perpetrators had overlap: attacks by both
perpetrators generated discussion on background, policy issues or changes, and questioning if the attack is terrorism. When studying policy issues and changes between Muslim and non-Muslim perpetrators, there was no statistically significant difference in the amount of coverage spent on policy issues or changes, regardless of the perpetrator type. There were, however, differences in the amount of coverage from media outlets questioning if the attack was terrorism or not in both left leaning and center outlets. Right leaning outlets did not have a significant difference, but for both types of attacks there was very little debate about whether the attack was terrorism.

Knowing how frames can impact the public’s perception of terrorism, it is important that the media give a fair and representative analysis of the terror attacks that occur in this country. With media outlets reporting differently on terror attacks depending both on whether or not the perpetrator is Muslim and the leaning of the media outlet, the public is not receiving fair and equal understanding. Based on this research and Kearns (2018) previous research, it is evident that Muslim perpetrators receive more media attention from all leanings of media outlets. This fact alone skews public perception of terror attacks to thinking that Muslims commit the majority of these terror attacks, when in reality, they commit the minority of attacks. The number of fatalities also is not the deciding factor in the amount of media coverage an attack receives, it matters on the
potential for harm as much as the actual harm done. Both extremes of leanings of media outlets have created framing issues in the public’s mind. Consumer of right leaning media are more likely to call jihadi-motivated terror attacks “radical Islamic terrorism” and to look for ties to ISIS or al Qaeda. Consumers of left leaning media are more likely to question if an attack committed by a non-Muslim is, in fact, an act of terror.

Areas for future research

This research was limited to attacks where the casualty numbers were between one and ten. It would be interesting to see further research done to compare these statistics with results from terror attacks with more than ten casualties. At the time of this research, only terror attacks through 2017 were available through the Global Terrorism Database. It would be interesting to see if these results remain true with terror attacks that were committed in 2018 and years to come.

This research also did not study the way these frames directly impact the public. It would be interesting to survey a representative sample of the population who watch and read these media outlets to see if the frames employed by the media outlets impact their views and how they perceive terrorism.
This research was done as a building block on Kearns (2018) research on which terror attacks receive more media attention. This research delved into the areas that Kearns mentioned in her research exploring the difference in context and being able to use broadcasted television news, both of which were done in this research. Hopefully more research can be done to explore other areas in the media representation of terror attacks and the way these frames impact public opinion.
References

*Codebook: Inclusion Criteria and Variables*, Global Terrorism Database, July 2018. Available


Issues with Public Disclosure of Non-GAAP Financial Metrics

Michael Gordon

Foreword by Research Mentor, Dr. Progyan Basu

I have enjoyed working with Michael Gordon in his research project sponsored by Snider Center’s SURE program. Michael’s research was on a highly relevant Accounting topic – disclosure of non-GAAP financial metrics. Publicly held companies often report their earnings based on accounting methods not endorsed by the Financial Accounting Standards Board. These non-GAAP measures can often lead to ambiguity and confusion among investors and analysts. A number of companies have been a number of situations, where companies have taken advantage of this loophole to manipulate their reported performance measures. Michael has done a thorough job of studying the financial disclosures of companies who have used non-GAAP metrics, and the impact of such disclosures on the investment community. He has also provided possible solutions to this topic, and opens avenues for further empirical research in this area.

I am very impressed by Michael’s passion for Accounting, and his deep and genuine thirst for knowledge. More importantly, he is a self-motivated individual, and performed with distinction with minimal supervision. His paper is worthy of submission to an academic journal, and we are planning to submit his paper to the 2019 Annual Meeting of the American Accounting
Association under the category of “Faculty-Student Collaborations in Accounting Research.”

Michael Gordon is a SURE Fellow in the R.H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: mgordon@umd.edu.

Abstract

The use of non-Generally Accepted Accounting Principles (non-GAAP) metrics by firms has increased dramatically in recent years. However, the regulatory structure for ensuring these metrics do not cross the line from appropriate discretion by managers to misleading investors has not kept pace. The last major pronouncement from the SEC addressing the disclosure of non-GAAP metrics was several years ago and has only been supplemented with unofficial clarifications to deal with highly technical aspects of securities laws. This paper examines the widespread usage of non-GAAP metrics and why the current lack of regulation with regards to disclosure of these data in unofficial settings, such as social media, is a problem that regulators need to address for protection of the investing public.
Part 1. Introduction

The history of corporate earnings publication in the US is long and complex. Regulators must balance the need for companies to communicate with investors in ways deemed appropriate for both management and their investors, with the necessity for reliable and standardized information across companies and industries. Over the past several decades, researchers have conducted many studies to determine the accuracy and impact of earnings measurements. With the growth in the use of unstandardized metrics accelerating in recent years, as will be expounded upon, accounting regulatory bodies must effectively respond to this trend to protect investors from inaccurate, incomplete, and false information. In recent years, various accounting regulators have spoken out against the continued use and promotion of adjusted earnings metrics, resulting in some action to curb the use of these measures in official documents, as will be discussed later. However, little has been done to combat the prevalence of adjusted measures disclosed in unofficial settings.

To explore this gap in policy, it is important to understand several things about the nature of disclosure rules. This paper will explain the relevant history of accounting laws and regulations to provide context for the current issues. With this foundation, the paper will review the motivations for, and the problems with, the use of non-GAAP earnings
disclosures, followed by the effect non-GAAP reports have on the investing public. We will then explain why this is important in the accounting world, and the motivations underlying each side. Lastly, this paper examines relevant data on this subject and propose potential solutions for this problem.

**Part 2. History of Disclosure Regulations in the United States**

The origin of modern financial regulations dates back to 1934, when Congress enacted the Securities Exchange Act (1934 Act), which brought the Securities and Exchange Commission (SEC) into existence. The 1934 Act was created in response to a string of fraudulent companies that fell apart during the Great Depression in the 1920s and 1930s. Following the birth of the SEC, disclosure regulations and principles were enacted primarily by the Financial Accounting Standards Board (FASB) and the American Institute of Certified Public Accountants (AICPA), with the input and consent of the SEC. Over the course of several decades, the trio of entities codified Generally Accepted Accounting Principles (GAAP) as the official guideline of public company financial disclosure.

Beginning in the 1980s, companies began presenting adjusted metrics in lieu of, or alongside, the required GAAP metrics. To address the concern of many investors regarding the possibly misleading nature of the
non-GAAP metrics, the SEC, in December 2001, issued a public warning to publicly-traded companies that adjusted metrics-- called “pro forma”-- have the capability of misleading investors and should be presented cautiously.\(^1\) Despite the increased scrutiny of financial metrics, many companies still failed to shield investors from misleading financial performance metrics.

In early 2002, the SEC issued its first Enforcement Action on the basis of pro-forma earnings. The SEC accused Trump Hotels & Resorts Inc. of “knowingly or recklessly issu[ing] false and materially misleading statements” in its financial statements (Securities and Exchange Commission, 2002). The adjusted earnings presented by the company had excluded a one-time charge, allowing it to beat Wall Street analyst expectations (Securities and Exchange Commission, 2002). Shortly after the Enforcement Action, the International Organization of Securities Commissions (IOSCO) released a report called “Cautionary Statement

\(^1\) It is important to distinguish between financial and nonfinancial non-GAAP metrics. This paper intends to address only financial metrics. Nonfinancial non-GAAP metrics, such as “same store sales” and “revenue per square foot” are beyond the scope of this paper. Additionally, other institutionalized metrics for specific sectors (e.g. Funds from Operation (FFO) in the Real Estate (REIT) and Master Limited Partnership (MLP) sectors are not intended to be addressed in this paper.
Regarding Non-GAAP Results Measures.” In the report, the IOSCO stated that many public companies had started publishing company-specific measurements of earnings that did not align with GAAP. Since individualized metrics are, by definition, not standardized across companies, the discretion exercised by management teams lacks the fundamental quality of comparability and consistency (Black D. E., Christensen, Ciesielski, & Whipple, 2018). The lack of comparability is due, in large part, to the prevalence of various exclusions across industries, as well as the proportion of firms within each industry that exclude various items.

The IOSCO report also addressed public companies’ disclosures in the form of press releases, letters to shareholders, and official documents filed with the SEC. Filings with the SEC rarely include non-GAAP metrics, due to the legal nature of official filings. When non-GAAP measures are shared, in documents filed with the SEC or disseminated to

2 Comparability and consistency are fundamental characteristics of the FASB’s conceptual framework that serve as the foundation of GAAP guidelines.

3 For example, 10Ks (annual earnings reports), 10Qs (quarterly earnings reports), 8Ks (disclosures of material events), and other documents.
shareholders, companies are required to reconcile the adjusted figure to the most directly related GAAP metric (Sherman & Young, 2017). These reconciliations can become exceedingly confusing to investors, to the point that they become nearly meaningless and worthless to investors. For example, in PepsiCo Inc.’s 2015 Annual Report to Shareholders, the company spent three full pages reconciling various non-GAAP metrics to the related GAAP metrics (PepsiCo Inc., 2015).

Many agree that some non-GAAP metrics can be useful when employed in moderation (e.g. to supplement GAAP information or explain a nonrecurring event that caused a material fluctuation in earnings), but these measures can become problematic and misleading when used inconsistently, excessively, and without being adequately defined. Many companies in the years leading up to the Dot-com stock market crash took advantage of loose regulations relating to non-GAAP metrics to paint a positive and unrealistic image of firms’ health and operational performance. The unearthing of massive frauds, such as Enron, WorldCom, and Global

__________________________

4 This legal requirement will be further discussed below.
Crossing, prompted Congress to address a serious lack of financial
disclosure rules.

**Part 2a. The Sarbanes-Oxley Act of 2002** and Subsequent Regulations

To address the lack of regulation of public companies’ disclosures to
investors, Congress enacted the Sarbanes-Oxley Act (SOX) to “fix auditing
of US public companies” (Coates, 2007). In §401(b) of SOX, Congress
empowered the SEC to issue rulings regarding pro-forma earnings data,
since having each company reporting customized metrics is too harmful to
investors to be tolerated. Many recent studies have shown the prevalence of
non-GAAP financial reporting in the investing world. A 2015 study found
that 94% of companies in the S&P500 index reported at least one non-
standardized financial metric (Shumsky & Francis, 2016). In 2013, a
McKinsey study found that all 25 of the largest US-based nonfinancial
companies reported some form of non-GAAP earnings (Sherman & Young,

5 The law is officially known as the Public Company Accounting Reform and Investor Protection
A PwC study found that 95 of the companies forming the FTSE100 index used a form of adjusted earnings (Sherman & Young, 2017). In compliance with the SOX requirement to issue guidelines, in 2002, the SEC enacted SEC Release 33-8145, which instituted Regulation G, as well as amended Regulation SK item 10(e). The enactment and revision, respectively, of these regulations addressed the method of disclosing earnings releases. Specifically, the SEC now required companies to file quarterly and annual earnings reports on a Form 8K. The SEC defined non-GAAP measures as “a numerical representation of a [company’s] historical or future financial performance, financial position, or cash flows that excludes [amounts subject to GAAP]” (Securities and Exchange Commission, 2002). Reg. G applies whenever a registrant.

While British companies do not follow US GAAP (the FTSE 100 is composed of UK-based firms), the adjusted metrics they report stray outside International Financial Reporting Standards (IFRS), which is the international equivalent of US GAAP, codified by the International Accounting Standards Board (IASB).

These will be referred to as “Reg. G” and “Reg. SK.”

A registrant is a public company that has a class of securities registered under §12 of the 1934 Act or is required to file reports under §15(d) of the 1934 Act.
discloses material information that includes non-GAAP data points to the public, whereas Reg. SK 10(e) addresses the presence of non-GAAP data contained in filings with the SEC.

With the institution of Reg. G, the SEC established two overarching requirements for public companies to abide by in financial disclosures (International Financial Law Review, 2017):

1. **Reconciliation**: When public companies disclose material information\(^9\) that includes non-GAAP metrics\(^{10}\), the registrant must include a presentation of the most directly comparable GAAP financial measure and a quantitative reconciliation of the difference(s) in calculation.

2. **General Disclosure requirement**: Companies cannot publicize non-GAAP measures that, along with accompanying data and disclosures, have material misstatements or are lacking information relevant to investors.

\(^9\) In any format, such as SEC filings, earnings calls, investor presentations, etc.

\(^{10}\) This only applies to historical non-GAAP metrics.
Reg. SK 10(e) supplements Reg. G by providing guidance on what companies must include in financial statements as adequate notice to investors about the use of non-GAAP data. Specifically, Reg. SK requires that:

1. Companies provide equal or greater prominence of GAAP measures compared to non-GAAP measures.

2. Companies must provide a quantitative reconciliation from non-GAAP to GAAP and show the most directly comparable GAAP measure for each non-GAAP measure.

3. Management must include a statement explaining why the non-GAAP measure is believed to be a better way to evaluate the company’s performance.

4. Management must include a statement explaining how non-GAAP measures are used.

To further reinforce restrictions on non-GAAP reporting, out of concern of misleading investors, the SEC formed a special taskforce in 2015. Showing how seriously they took the issue, the taskforce labeled non-GAAP metrics a “fraud risk factor” (Black E. L., Christensen, Kiosse, & Steffen, 2015). In May 2016, the SEC issued new guidance on non-GAAP disclosure methods, including disallowing companies to use non-GAAP metrics as the only financial measure in public communications (Dicker,
2016) and adding new rules about how the measures can be presented in press releases.

Similarly, in 2014 the FASB began researching whether redesigning the traditional income statement would alleviate any of the issues with reliance on non-GAAP measures. The project, called “Financial Performance Reporting,” is still underway, and no relevant findings have been published.

**Part 3. Why Non-GAAP Metrics are Problematic**

Many individuals in the financial world have voiced concern over the veracity of non-GAAP metrics. In his 2014 annual letter to shareholders, Berkshire Hathaway Chairman Warren Buffet remarked “When CEOs tout EBITDA\(^\text{11}\) as a valuation guide, wire them up for a polygraph test” (Buffet, 2014). Former SEC Chairwoman Mary Jo White said that the use of non-GAAP metrics “deserves close attention,” because they can be a “source of confusion” to investors (White, 2015). A prime example of how the metrics can confuse investors is Valeant

\(^{11}\) Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITDA) is a popular non-GAAP metric.
Pharmaceuticals Inc. In 2015, Valeant reported an adjusted profit of $2.84 billion, significantly divergent from their GAAP loss of nearly $300 million (Rapoport, 2016). Nonprofessional investors likely did not pick up on the massive difference between what the company said they earned and what the company actually earned.

Prior research found that about 11% of companies report non-GAAP EPS in both their earnings announcement and their proxy statement.¹² Within that group of companies, only at about 31% of companies do the board and managers use the same metrics (Black, Black, Christensen, & Gee, 2018).

In 2016, former SEC Chief Accountant James Schnurr stated that non-GAAP metrics should “supplement the information in financial statements and not supplant the information” (Schnurr, 2016). Nevertheless, many companies did not abide by this statement, forcing the SEC to take definitive action. In the second quarter of 2016, about 40% of firms in the S&P500 index displayed a non-GAAP metric before the

¹² Study included non-GAAP EPS metrics from proxy statements and earnings announcements of companies in ExecuComp, between 2009 and 2015.
comparable GAAP metric in the headline text of earnings releases. The SEC responded to this trend by issuing tougher rules on placement of non-GAAP numbers in earnings releases, in accordance with Reg. SK 10(e). Following the institution of the new rules, only 6% of companies continued placing non-GAAP metrics before the comparable GAAP metric (Davine, Mitrovich, & Mazor, 2016).

In a message similar to that of Schnurr’s, International Accounting Standards Board (IASB) Chairman Hans Hoogervorst said “It is important to remember that non-GAAP measures represent a selective presentation of an entity's financial performance. Often, that selection is not free from bias” (Hoogervorst, 2015). In the same speech, Hoogervorst said that more than 88% of S&P500 firms include non-GAAP metrics in earnings releases, and, of those firms, 82% reported non-GAAP figures that were greater than the GAAP EPS figure (Sherman & Young, 2017).

Prior research has shed light on the varying usage of non-GAAP metrics across industries. Overall, the percentage of companies reporting non-GAAP earnings increased from just over 50% in 2009 to more than

13 Some did not even mention a GAAP figure until much later in the release.
70% in 2014 (Black D. E., Christensen, Ciesielski, & Whipple, 2018). By 2016, 96% of S&P500 companies filing quarterly reports disclosed at least one non-GAAP metric. The average number of non-GAAP metrics per filing grew from 2.35 in 1996 to 7.45 in 2016 (McKeon, 2018).

![Percentage of Companies Reporting Non-GAAP Metrics, by Industry](image)

**Figure 1: Percentage of companies in each S&P500 sector that report non-GAAP metrics**

Four S&P500 sectors in particular tend to report non-GAAP earnings disproportionately (see Figure 1):  

- 86% of companies in the healthcare sector  
- 78% of companies in the information technology sector  
- 75% of companies in the materials sector  
- 72% of companies in the utilities sector  

One possible explanation for why these sectors tend to report non-GAAP earnings more than others may be due to the higher regulatory and
other fixed costs absorbed by companies in some of those industries. High costs create higher barriers to entry, so regulators place more restrictions to prevent a monopoly or oligopoly from forming. In the information technology industry, the vast majority of newer companies are unprofitable when measured by traditional GAAP metrics. Between 2001 and 2017, only 36% of the 601 IPOs in the sector were profitable (Ritter, 2018). This lack of profitability could be a factor that motivates management teams to obscure the lack of profits by promulgating company-specific metrics excluding various expenses.

**Part 3a. The Effect on the Investing Public**

The effect that disclosure of non-GAAP metrics has on nonprofessional investors is exacerbated by the fact that existing regulations do not extend to social media posts. Authoritative SEC guidance applies only to documents filed with or furnished to the SEC, except for cases of outright securities fraud (Brown, Elliott, & Grant, 2018). According to the SEC, the only requirement companies must meet when using a social media platform to convey information to investors is that the company must inform investors which platform will be used, in accordance with Regulation Fair Disclosure (Reg. FD) (Blankespoor, Miller, & White, 2014). In terms of regulation the SEC treats social media
posts similar to companies’ websites, (Securities and Exchange Commission, 2013). Some securities experts suggest that, to prevent investors from being misled by financial information in terse social media posts, the SEC should require companies to include a hyperlink to an officially-filed document that contains a reconciliation required by Reg. SK 10(e). A 2014 study found that 75% of earnings-related tweets include a hyperlink to the full press release (Brown, Elliott, & Grant, 2018).

However, for two reasons, this may not alleviate the underlying issue of investors being misled. First, investors may not be interested in clicking on the link. Second, since a significant portion of average daily trading volume on US stock exchanges are algorithm-driven, the inclusion hyperlinks to external sources would not make a difference to those computer programs.

This regulatory guidance resulted from an SEC Division of Enforcement investigation following a controversy involving Netflix Inc. CEO Reed Hastings. Hastings had posted on his Facebook page that Netflix had reached one billion viewing hours for the first time, without filing an accompanying Form 8K or press release. Since neither he nor the company had ever used Facebook as a means of disclosing information, the SEC looked into the ramifications of the lack of rules relating to social media posts. The SEC then issued this regulatory guidance.

Data from Tabb Group estimates that computer-driven models account for about 29% of daily trading volume. JPMorgan estimates that, including other types of passive investment flows, the
A study tested whether disseminating images containing prominently presented non-GAAP financial measures, coupled with a hyperlink, has any distinct effect on investors. Prior research had already established that prominently-displayed pieces of information are likely to attract attention and can lead to overreliance on the information. Complementary research had similarly found that people tend to remember information displayed in image form more than information in text form. Based on these prior discoveries, the authors studied how human information-processing deals with financial information (Brown, Elliott, & Grant, 2018). The results showed that, as the authors predicted, “Earnings release prominence and image-containing] tweets have an interactive effect on investors’ valuation judgements,” meaning that, all else being equal, the formatting of content within a press release can distort investors’ perception of the information. This reaction is not limited to making valuation decisions on the basis of percentage of volume driven by algorithms is as high as 85%. These algorithms are programmed to search for keywords in news headlines and summaries and do not distinguish between GAAP and non-GAAP metrics (Zuckerman, Levy, Timiraos, & Banerji, 2018).

A possible reason for this, according to the study is, because images stimulate both verbal and visual portions of the brain-- as opposed to text, which does not-- the dual-processing can lead to longer-term retention of the information.

16 A possible reason for this, according to the study is, because images stimulate both verbal and visual portions of the brain-- as opposed to text, which does not-- the dual-processing can lead to longer-term retention of the information.
only non-GAAP metrics; the authors found that when GAAP metrics are
presented more prominently than non-GAAP metrics, investors arrive at a
lower valuation judgement. Nevertheless, if an image-containing tweet
contains both a GAAP and non-GAAP figure, the resulting valuation
judgement is higher, regardless of which metric was disclosed with greater
prominence.

![Image of Alcoa Inc. tweet]

*Figure 2: Image-containing Tweet Containing Non-GAAP Information (Brown, Elliott, & Grant, 2018)*

![Image of Coca-Cola Co. tweet]

*Figure 3: Image-containing Tweet Containing Non-GAAP Information (Brown, Elliott, & Grant, 2018)*
Figures 2 through 5 are examples of image-containing tweets that seem to violate the reconciliation requirement of Reg. SK. They contain only non-GAAP results for the non-revenues figures, so the results may be misleading. For instance, Coca-Cola reported a “+6% comparable currency
neutral income before taxes” (see Figure 3), but the Coca-Cola 2015 Form 10K GAAP pre-tax income was up only 3% (The Coca-Cola Company, 2015).

Overall, the authors found that there is a cognitive effect on investors’ perceptions of earnings stemming from earnings information being released in image form on social media. The effect of these images, coupled with the lack of regulation covering this aspect of securities disclosures, is a gap that should be addressed by the SEC.

A different study corroborated the above findings by showing how investors pay more attention to non-GAAP metrics when assessing the performance of a company. This could be due, in a large part, to the overwhelming preference of companies to announce selectively customized metrics to paint an unrealistic picture of the accounting period.

17 While Wall Street analysts also base their estimates on non-GAAP metrics, a discussion of the implications of this is beyond the scope of this paper.
In 2014, companies excluded an average of 3.6 line-item expenses from their non-GAAP metrics (see Figure 6 above). In monetary terms, these expenses exclusions amounted to adding an extra $1.03 to earnings per share per company (Black D. E., Christensen, Cieseilski, & Whipple, 2017).

These aforementioned studies demonstrate the effects, whether intended or unintended, that non-GAAP metrics have on investors when released via social media. Under §13(a) and §15(d) of the 1934 Act, companies are required to disclose material events to keep investors informed of the company’s status. The SEC, on this basis, mandated that companies must file a Form 8K to disclose material events. Information disclosed on an 8K is thus subject to all relevant securities laws and
regulations. Since annual and quarterly reports are already filed on official forms, it would seem logical to require the same level of disclosure regulation to apply to information released to the public in a medium like Twitter or Facebook.

**Part 3b. How the Differences Are Manifested**

![Exclusions from Non-GAAP Earnings by Type and Frequency](image)

*Figure 7: Exclusion of line items from non-GAAP earnings by frequency* (Black D. E., Christensen, Ciesielski, & Whipple, 2018)

---

18 The legal source for the Forms 10K and 10Q is the same as the source for the Form 8K.
In recent years, there have been growing divergences between intracompany GAAP and non-GAAP EPS estimates by analysts. Much of this split can be attributed to the exclusion of expenses deemed to be “non-recurring” or “non-cash” in nature (Bradshaw, Moberg, & Sloan, 2000). Within the overarching categories of “non-recurring” or “non-cash” expenses, the increasing frequency and magnitude of certain charges, in particular, are especially astounding. Some of the most common exclusions include merger and acquisition costs, stock-based compensation expenses,

---

19 Some businesses, such as Danaher Inc., are known for routinely buying existing firms. In this case M&A expenses could be considered a recurring expense.
amortization expense\textsuperscript{20}, restructuring costs,\textsuperscript{21} and asset impairments. \textsuperscript{22} See Figure 7, above, for the frequency of various exclusions used by companies. The figure shows the frequency of various exclusions of expenses from non-GAAP metrics. For example, about 47\% of companies excluded restructuring charges, and about 27\% excluded amortization expenses (Black D. E., Christensen, Ciesielski, & Whipple, 2018).

\textsuperscript{20} Amortization is the allocation of the cost of an intangible asset (e.g. patents, copyrights) to future periods in which the company expects to benefit from the asset. This is recorded as an expense on the income statement and a reduction in the value of the intangible asset on the balance sheet.

\textsuperscript{21} While the term “restructuring charge” is not explicitly officially defined, the SEC has stated in the past that restructuring expenses can arise from many scenarios including “from the consolidation and/or relocation of operations, or the disposition or abandonment of operations or productive assets…a business combination, a change in an enterprise’s strategic plan, or a managerial response to declines in demand, increasing costs, or other environmental factors” (Securities and Exchange Commission, 1999).

\textsuperscript{22} Asset impairments occur when the expected future benefit of an asset (measured in terms of the expected future net cash flows) decreases to below the current book value of the asset on the firm’s balance sheet. When this occurs, the asset is reduced to its fair market value and the company records an impairment loss.
Definitionally, these types of expenses should not be considered a common aspect of doing business. The SEC has prohibited the classification of line items as non-recurring, infrequent, or unusual if “the nature of the charge or gain is such that it is reasonably likely to recur within two years, or there was a similar charge or gain within the prior two years” (Securities and Exchange Commission, 2018). Some companies have abused this standard. One of the most prominent examples is Whirlpool Inc., which reported a restructuring charge every quarter from 1990 through the present, prompting a letter from the SEC in 2016 asking for an explanation of how these charges cannot be considered routine (Schilit, Perler, & Engelhart, 2018, p. 76).

Another prominent example of a company that was publicly admonished by the SEC for excluding too many items from their non-GAAP earnings was Groupon Inc. Prior to its 2011 Initial Public Offering (IPO), Groupon filed its financial statements. Included in the filing was a

**Adjusted CSOI**

The following is a reconciliation of Adjusted CSOI to the most comparable GAAP measure, ",(Loss) income from operations," for the years ended December 31, 2008, 2009 and 2010 and the first quarter of 2010 and 2011:

<table>
<thead>
<tr>
<th></th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Loss) income from operations</td>
<td>$(1,632)</td>
<td>$(1,077)</td>
<td>$(8,571)</td>
<td>$(117,148)</td>
</tr>
<tr>
<td>Adjustments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online marketing</td>
<td>162</td>
<td>4,446</td>
<td>241,546</td>
<td>9,064</td>
</tr>
<tr>
<td>Stock-based compensation</td>
<td>24</td>
<td>115</td>
<td>36,168</td>
<td>116</td>
</tr>
<tr>
<td>Acquisition-related</td>
<td></td>
<td></td>
<td>203,183</td>
<td></td>
</tr>
<tr>
<td>Total adjustments</td>
<td>186</td>
<td>4,561</td>
<td>480,897</td>
<td>4,020</td>
</tr>
<tr>
<td>Adjusted CSOI</td>
<td>$(1,446)</td>
<td>$3,084</td>
<td>$60,553</td>
<td>$12,901</td>
</tr>
<tr>
<td>Adjusted Segment Operating Income:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>$(1,446)</td>
<td>3,084</td>
<td>88,036</td>
<td>12,591</td>
</tr>
<tr>
<td>International</td>
<td>$(1,446)</td>
<td>3,084</td>
<td>27,683</td>
<td>62,619</td>
</tr>
<tr>
<td>Adjusted CSOI</td>
<td>$(1,446)</td>
<td>3,084</td>
<td>60,553</td>
<td>$12,901</td>
</tr>
</tbody>
</table>

*Figure 8: Groupon Inc.’s comparative income statement from 2008 through 2010 (source: the company’s pre-IPO Form S-1 filing with the SEC).*
unique non-GAAP metric called “Adjusted Consolidated Segment Operating Income” (CSOI), which the company touted as a sign of enormous success. Groupon’s 2008 loss from operations totaled $1.6 million. The operating loss grew to over $420 million by 2010. At the same time, the CSOI went from -$1.4 million to about $60.5 million, respectively. The line items that Groupon decided to omit from its CSOI included online marketing and stock compensation, both of which are routine expenses for most companies (See Figure 8 above).

In response to the disclosure of this controversial metric, the SEC sent the company a letter on June 29, 2011, asking for justification for excluding these recurring expenses from CSOI. While Groupon complied with the Reg. G requirement to reconcile all non-GAAP metrics to the closest GAAP metric, the company had violated the Reg. SK 10(e) prohibition against classifying recurring expenses as non-recurring if the charge has occurred, or will occur, in a two-year window of time.

Part 4. Why Non-GAAP vs. GAAP Matters

The use of non-GAAP metrics has more ramifications and uses beyond management teams trying to show investors a rosier picture of the business. Many companies’ boards use non-GAAP metrics as benchmarks for executive compensation. This can create negative incentives to adjust
the GAAP metrics as much as possible to maximize the size of management’s salaries. For example, at Valeant Pharmaceuticals Inc., the board pegged executive compensation to a plan called “Total Shareholder Return” (TSR), which essentially mirrored the stock price performance (Schilit, Perler, & Engelhart, 2018, p. 22). This payment structure motivated CEO Michael Pearson to drive up the stock price through any means possible. Since non-GAAP metrics can have the effect of smoothing out the earnings quarter to quarter, Pearson recklessly used adjusted earnings to artificially inflate the stock price. In turn, he received generous salaries that made him worth $3 billion at his peak. This risky compensation plan is not limited to Valeant, or even just a few companies. At Allergan Inc., between 2014 and 2015, the CEO pay, which totaled $58 million, was based 68% on stock performance and 28% on non-GAAP earnings targets. In 2009, boards at 249 public companies based the payment plan on non-GAAP metrics. By 2013, the number grew to 542 companies, which represented 28% of large-accelerated filers (Usvyatsky, 2014).23 The marginal salary to be gained from reporting excessively high

23 The term “large-accelerated filer” is defined under 17 CFR §240.12b-2(2) as a company meeting the following conditions: (1) The issuer had an aggregate worldwide market value of the voting and non-voting common equity held by its non-affiliates of $700 million or more, as of the
metrics is a clear motivator. A prior study found that, all else being equal, CEOs of firms that make large adjustments to arrive at non-GAAP metrics make an average of $1.9 million, or 16%, more than should be expected under the study’s model (Guest, Kothari, & Pozen, 2018). The authors of the study found that non-GAAP earnings are better predictors of executive compensation than either GAAP EPS or GAAP Income from Operations.

There also tend to be variances in how stocks react based on whether the company reports only GAAP metrics or both GAAP and non-GAAP metrics. In conducting original research, the data showed that there is a statistically significant difference between how GAAP and Both companies.

Part 5. Quantitative Analysis

last business day of the issuer’s most recently completed second fiscal quarter, (2) the issuer has been subject to the requirements of §13(a) or §15(d) of the 1934 Act for a period of at least twelve calendar months, and (3) the issuer has filed at least one annual report pursuant to §13(a) or §15(d) of the 1934 Act.

24 For simplicity, companies that report only GAAP metrics will be referred to as “GAAP companies,” and companies that report GAAP and non-GAAP earnings will be referred to as “Both companies.”
To understand the real impact that the use of non-GAAP metrics has on the stock market, stock price and earnings data spanning multiple quarters were gathered. The analysis was based on the hypothesis that companies reporting only GAAP metrics in their 10Qs and 10Ks exhibit greater volatility in response to their earnings report, due to the lack of earnings management at companies reporting only GAAP metrics.

Methodologies

To collect a sample of companies, all S&P500 companies, excluding multiple classes of shares of the same company (e.g., Alphabet Inc.) were downloaded. The companies were organized by sector and type of reporting (GAAP-only [herein defined as “GAAP”] or both GAAP and Non-GAAP [defined herein as, “Both”]). To test the hypothesis that GAAP and Both companies react differently to earnings reports, five S&P500 sectors were

25 Earnings management is the practice of adjusting earnings for the sake of beating Wall St. analysts’ expectations. Since GAAP-only companies do not adjust their earnings, the quarterly earnings reports should tend to be more volatile, reflective of a constantly-changing business environment.
chosen at random. Within the five sectors, one GAAP and one Both were selected (see Table 1, below). The companies analyzed were:

Table 1: Companies analyzed for quantitative research

<table>
<thead>
<tr>
<th>Industry</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer Discretionary</td>
<td>Nike Inc. (NKE)</td>
</tr>
<tr>
<td></td>
<td>The Walt Disney Co. (DIS)</td>
</tr>
<tr>
<td>Financials</td>
<td>The Goldman Sachs Group Inc. (GS)</td>
</tr>
<tr>
<td></td>
<td>Bank of America Corp. (BAC)</td>
</tr>
<tr>
<td>Consumer Staples</td>
<td>PepsiCo Inc. (PEP)</td>
</tr>
<tr>
<td></td>
<td>Costco Wholesale Corp. (COST)</td>
</tr>
<tr>
<td>Telecommunication Services</td>
<td>AT&amp;T Inc. (T)</td>
</tr>
<tr>
<td></td>
<td>T-Mobile US Inc. (TMUS)</td>
</tr>
<tr>
<td>Energy</td>
<td>Valero Energy Corp. (VLO)</td>
</tr>
<tr>
<td></td>
<td>Exxon Mobil Corp. (XOM)</td>
</tr>
</tbody>
</table>

For each of the 10 companies, historical stock price history from January 4, 2016 through February 19, 2019 was downloaded from Yahoo.

For some of the industries, there were no companies that reported only GAAP earnings, so new sampling to replace the companies was done.
Finance. The earnings data for each company were obtained through Fidelity.com and matched to the corresponding daily stock price. Earnings reported before the market open were recorded as having occurred on that day. To reflect the true market reaction, earnings reported after the market close were recorded as having occurred on the next day. This data collection resulted in 120 unique records with which to analyze the difference in reactions post earnings release.

Figure 9 was created from the average daily return\(^{27}\) for each company, by GAAP (blue bars) and Both (red bars), on their quarterly

\[^{27}\text{Calculated as the natural log of } (t_0 \div t_1)\]
earning day (day 0), each of the 14 days prior to the report date, and each of the 14 days following the report date. A four-period moving-average standard deviation was added for each line to show the average stock volatility.

Results

To test whether the response to the earnings announcements differed between the GAAP and Both companies, an analysis of variance was run. There were four response categories -- days prior to earnings day (instead of 14 days, 5 days was chosen to represent the week before earnings day), earnings day, 5 days after earnings day, and all other days.

Table 2 shows a statistically significant difference in the stock price reaction of GAAP and Both companies on the earnings day. The average stock return of GAAP companies on earnings day is 1.03%, while Both companies have an average stock return of -0.98%. In the five days preceding earnings day, there is a small difference in the average stock price movements for GAAP and Both companies (-0.06% and 0.01%, respectively). The price gap widens slightly in the five days following earnings day, to a loss of -0.14% for GAAP companies and a gain of 0.07%
for Both companies. Other than the earnings day, none of the differences in the average stock prices is statistically significant.

*Table 2: Difference Between GAAP and Both Companies Average Daily Returns*

<table>
<thead>
<tr>
<th>Reporting Type</th>
<th>Earnings Day</th>
<th>5 Days Prior</th>
<th>5 Days After</th>
<th>All Other Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both</td>
<td>-0.98%</td>
<td>0.01%</td>
<td>0.07%</td>
<td>0.03%</td>
</tr>
<tr>
<td>GAAP</td>
<td>1.03%</td>
<td>-0.06%</td>
<td>-0.14%</td>
<td>0.06%</td>
</tr>
</tbody>
</table>

F-Test = 2.2, p < .05

Table 3 answers the hypothesis that GAAP companies will be more volatile than Both companies. The average standard deviation of GAAP companies on earnings day is 2.36%, while Both companies have a standard deviation about half of that, at 1.23%. This difference was statistically significant. Similarly, in the five days preceding earnings day, the standard deviation of GAAP companies was twice that of Both, 2.19% and 1.12%, respectively. The difference in the standard deviation for the five days after the earnings and all other days showed no significant difference.
Table 3: Difference Between GAAP and Both Companies Standard Deviations

<table>
<thead>
<tr>
<th>Reporting Type</th>
<th>Earnings Day</th>
<th>5 Days Prior</th>
<th>5 Days After</th>
<th>All Other Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both</td>
<td>1.23%</td>
<td>1.12%</td>
<td>0.95%</td>
<td>1.02%</td>
</tr>
<tr>
<td>GAAP</td>
<td>2.36%</td>
<td>2.19%</td>
<td>0.93%</td>
<td>1.08%</td>
</tr>
</tbody>
</table>

F-Test = 35.0, p < .01

The results clearly show a difference in how the stock market responds to GAAP companies compared to Both companies in the run up to and the day of their earnings release. GAAP companies have higher average returns on earnings day, but the average volatility is far higher, both on earnings days and the five days preceding.

Part 5. Why Managers Like Using Non-GAAP Metrics

To understand why non-GAAP metrics make a difference, it is important to understand why management teams feel compelled to use them in the first place and what they hope to accomplish by using these metrics. While the line items excluded from non-GAAP metrics can seem arbitrary, the underlying motivations for excluding items are relatively
consistent across companies and industries. A survey of CFOs was conducted to ascertain why managers use non-GAAP metrics. When asked what the concept of “earnings quality” means to top-level managers, only 8% of respondents said, “accurate application of GAAP rules,” and only 6% said “consistent application of GAAP.”

When asked what the strongest motivations are for companies to misrepresent true economic performance when disclosing earnings, about 94% said “to influence stock prices;” about 93% said “to hit earnings benchmarks;” and about 88% said “to influence executive compensation.” When asked to share examples of red flags that can mean earnings misrepresentation, about 8% said “using non-GAAP and/or changing metrics.” The results of this survey show how, at the vast majority of firms, the use and disclosure of non-GAAP metrics is not a concern of management (Dichev, Graham, Harvey, & Rajgopal, 2015). The prospect of having a high stock price, which leads to high pay for many companies, is a strong incentive that encourages the use of non-GAAP metrics. The current regulatory structure does not place enough

28 Interestingly, these percentage are very similar to the percentage of companies that do not disclose non-GAAP metrics in their quarterly and annual reports. In 2015, 94% of companies disclosed a form of non-GAAP metrics in their earnings report (see Part 2a, above).
restrictions on the usage of these metrics, a gap that can result in harm to investors.

Managers of GAAP-loss firms, or companies that have consistent net losses on a GAAP basis, are particularly likely to report non-GAAP metrics. This probably stems from the desire to cover up the losses on the income statement as much as possible. The number of GAAP-loss firms reporting non-GAAP metrics has steadily grown over time. In 2006, about 30% of GAAP-loss firms used non-GAAP metrics. By 2014, the number of loss-firms reporting non-GAAP metrics grew to 46% (Leung & Veenman, 2018). More recent data would likely show a further rise in this percentage. The authors of the GAAP-loss firm study discuss why companies exclude certain items from non-GAAP metrics. One explanation is that certain items required to be disclosed under GAAP are perceived to be less predictive and relevant than other expenses. The most popular exclusions that fit that description include stock-based compensation and amortization of acquired intangible assets. Some would counter that those are relatively routine aspects of running a public company. Stock-based compensation is a popular and commonly used method of rewarding management in a non-monetary manner. Amortization of any form, including intangible assets, is, by nature, intended to be carried out over a predetermined time and a predetermined amount. Nevertheless, these expenses are frequently
excluded from non-GAAP metrics. Additionally, the authors found that 40% of firm-quarters excluded recurring expenses from non-GAAP metrics, a sign that managers at some companies arbitrarily determine which types of expenses they think fail to paint a faithful representation of the firm’s performance.

Part 6. Possible Solutions

Non-GAAP reporting is clearly a problem that needs to be addressed. The legal and regulatory requirements are not clearly defined in all cases and have some serious gaps that allow companies to disseminate incomplete or inaccurate information. Public disclosure via social media, in particular, is an area seriously lacking in protections for nonprofessional and professional investors alike. Short of outright banning the disclosure of non-GAAP earnings, which would have intense backlash from companies, the SEC and FASB need to tighten the rules on the use and publication of non-GAAP metrics. One possible solution to social media disclosure could be to require companies to file all financial information-containing social media posts on a Form 8K, thereby subjecting the information to Reg. G, Reg SK, and other regulations. Forcing all information disclosed to the public to comply with securities laws and regulations could limit the influence of non-GAAP metrics on investors. This proposal is similar to the
implementations of SEC Release 33-8145 (see Part 2a above), which required official earnings releases to be disclosed on a Form 8K. Extending the requirement to include all forms of financial disclosure in the public sphere-- both official and unofficial settings-- is a logical extension of the regulation.

The study by Brown, Elliott, and Grant, mentioned earlier, discussed this issue of regulations not currently covering financial releases in unofficial forums, such as Twitter and Facebook. Furthermore, the SEC has previously stated that omitting comparable GAAP measures from an earnings release headline or caption containing a non-GAAP figure, or describing a non-GAAP measure as “record performance,” “exceptional,” or any other praising descriptor without an at-least-equally-as-prominent GAAP figure characterization could be considered a violation of Reg. SK 10(e)(1)(i)(A) (Securities and Exchange Commission, 2018). However, many companies fail to comply with this guidance. Of 100 randomly selected S&P500 companies’ tweets, 36 contained a non-GAAP-metric-containing image. Many of them did not include a related GAAP figure.

29 This guidance was given in the form of an answer by SEC staff to companies through a Compliance and Disclosure Interpretation on April 4, 2018.
Official rules from the SEC, as opposed to simple suggestions, could curb the current regulatory gap.

**Part 7. Conclusion**

The history of non-GAAP metrics is long and complex. SEC regulations must balance the allowance for companies to communicate with investors in ways deemed appropriate to management with the need for accurate and standardized information across companies and industries. From before the first SEC Enforcement action in 2002 until the present, immeasurable research has been conducted to determine the accuracy and impact of non-GAAP measures. In recent years, SEC officials and other accounting regulators have spoken out against the continued use and promotion of non-GAAP metrics, resulting in some action to curb the use of these measures in official filings. However, little has been done to combat the prevalence of adjusted measures disclosed via social media. Aside from the requirement for companies to notify investors precisely as to which platform will be used for information-sharing, little regulation applies to unofficial disclosures. It has been well-documented that there is a measurable and significant difference in how stock prices react to earnings releases based on whether the company releases non-GAAP metrics or only
GAAP. The ability of managers to control the direction of stock price movements by promoting adjusted figures in public a tactic employed by as many as 96% of S&P500 companies, is a serious problem.

Future research could, to the extent possible, explore the potential for tighter regulations on unofficial disclosure and the resulting impact on investor perception of earnings and the trustworthiness of management. While an outright ban of all non-GAAP metrics could increase the quality of public financial information, such an action would likely fail to be implemented due to unpopularity. Less drastic, though still pronounced, actions, such as increasing regulatory scrutiny of disclosure in unofficial formats, could improve the reliability of earnings data. Exploring various routes for implementation of this is a question for future research to answer.

Acknowledgements

I would like to thank my research mentor, Dr. Progyan Basu, of the Department of Accounting and Information Assurance at the Robert H. Smith School of Business, for guiding and supporting me throughout the research process. I would also like to thank the staff and student directors at the Ed Snider Center for Enterprise and Markets for organizing and managing the Snider Undergraduate Research Experience program.
Appendix of Figures

Figure 1: Percentage of companies in each S&P500 sector that report non-GAAP metrics (Black D. E., Christensen, Ciesielski, & Whipple, 2018) 173

Figure 2: Image-containing Tweet Containing Non-GAAP Information
(Brown, Elliott, & Grant, 2018) 177

Figure 3: Image-containing Tweet Containing Non-GAAP Information
(Brown, Elliott, & Grant, 2018) 177

Figure 4: Image-containing Tweet Containing Non-GAAP Information
(Brown, Elliott, & Grant, 2018) 178

Figure 5: Image-containing Tweet Containing Non-GAAP Information
(Brown, Elliott, & Grant, 2018) 178

Figure 6: Exclusions from Non-GAAP earnings per share (Black D. E., Christensen, Ciesielski, & Whipple, 2017) 180

Figure 7: Exclusion of line items from non-GAAP earnings by frequency
(Black D. E., Christensen, Ciesielski, & Whipple, 2018) 181
Figure 8: Groupon Inc.’s comparative income statement from 2008 through 2010 (source: the company's pre-IPO Form S-1 filing with the SEC). 184

Figure 9: Average Daily Change and Volatility for GAAP and Both companies. 190

Appendix of Tables

Table 1: Companies analyzed for quantitative research 189
Table 2: Difference Between GAAP and Both Companies Average Daily Returns 192
Table 3: Difference Between GAAP and Both Companies Standard Deviations 193
References


Evaluating Corporate Self-Governance in Eradicating Forced Labor

Kunal Harmalkar

Foreword by Research Mentor, Dr. Virginia Haufler

Why do people still suffer from the horrors of slavery today in the 21st century? “Modern slavery” remains a scourge across the world, touching every country from the most to the least developed. Human trafficking is the second most lucrative illicit business in the world (after drugs). The people who make our chocolate, coffee and tea, or our t-shirts and jeans, are often coerced, unpaid and abused. What can we do to end this?

A wide array of actors are engaged in efforts to eliminate human trafficking worldwide. One new approach gaining increasing attention is government regulation of global supply chains. They seek to leverage the power of lead companies in global trading networks over their suppliers. A prominent example of this is the UK Modern Slavery Act of 2015. This regulation requires companies based in the UK to report publicly on the steps they have taken to combat modern slavery within their global supply chains.

Kunal Harmalkar came to me the summer before he started at the University, asking for a research project to pursue. I suggested he examine the reports submitted by UK firms to comply with the law, and directed him to SURE.
Kunal brought both enthusiasm and technical skills to the job of analyzing hundreds of reports. Kunal collected the reports, cleaned the data, and analyzed the text—with some advice from Andrew Lugg, a graduate student in GVPT. he sought to identify the factors that are associated with lower levels of slavery, correlating the topics found in the reports with different characteristics of the firm. His work provides interesting insight into the utility of public reporting and supply chain regulation to address the global problem of human trafficking.

Kunal Harmalkar is a SURE Fellow in the Government & Politics Department at the University of Maryland, College Park, MD. Their email address is: kunalh@terpmail.umd.edu.

Abstract

Although slavery is often seen as an outdated practice, forced labor is prevalent in goods that Americans purchase daily. After the passage of the UK Modern Slavery Act in 2015, large corporations listed on the London Stock Exchange were required to publish a disclosure outlining steps they have taken to eradicate slavery from their supply chains. This research seeks to evaluate voluntary disclosure transparency legislation in modifying corporate behavior. To test the effectiveness of the UK Modern Slavery Act, text analysis was used to compare topics within mandated corporate disclosures and transparency scores. A lack of significant correlation implies that the contents of corporate disclosures are not actually reflective of real-world behavior and that alternatives to voluntary
disclosure transparency legislation ought to be pursued while creating global governance initiatives.

*Keywords: corporate governance, modern slavery, forced labor, text analysis, text mining, structural topic model, R, global governance*

**Acknowledgements**

This research was conducted under the guidance of Dr. Virginia Haufler from the Department of Government and Politics at the University of Maryland, College Park. Thank you to Dr. Haufler, Andrew Lugg, Dr. Elson, and the Ed Snider Center for Enterprise and Markets at the Robert H. Smith School of business for supporting and guiding this investigation.

**Introduction**

To understand how modern slavery can be eradicated from corporate supply chains, the term has to be contextualized to the form it takes within globalized supply chains. Modern slavery is broadly defined as “any form of forced human exploitation for labor or service, such as human trafficking and forced labor” (SourceIntelligence), but only the forced labor aspect of
modern slavery is pertinent to this investigation. Although modern slavery is often viewed as an antiquated practice, it is incredibly prevalent in certain sectors of developing countries.

Slavery in corporate supply chains often takes subtle forms. Instead of physical bondage, Verité, a transnational non-governmental organization, found that “The charging of recruitment fees and expenses to migrant workers is the most significant contributor to the shameful ongoing presence of debt bondage, human trafficking, forced labor, or modern slavery in global supply chains” (Verité). When local recruiters hire workers for factories which subcontract from transnational corporations, the recruiters sometimes charge “recruitment fees.” These fees carry a high interest rate, which means that workers can never fully pay off the balance.

This forces the workers to continue working despite terrible conditions and low pay. In addition, if recruiters move the workers across an international boundary, they may take away a worker’s passport until the recruitment fee has been paid off. The workers are then kept against their will in a foreign country, forced to work to pay off an impossible debt.

Unfortunately, corporations cannot easily verify that the contracts they give to factories are being fulfilled without unsafe and unethical practices. In one instance, local regulatory bodies certified a garment
factory in Karachi, Pakistan as safe just a few weeks before the factory burned down, killing more than 300 workers (Al Jazeera). Upon investigation, local authorities found evidence of child labor and “prison-style” living arrangements for workers. This large-scale disaster was not an isolated incident. A Bangladeshi factory collapse in 2013 killed 400 workers who were producing clothes for JCPenney, Dress Barn, and Wal-Mart (CNN). The resulting customer backlash against these corporations made it clear that removing slavery from supply chains is more than just a priority for human rights groups. It should be a priority for corporations who are held accountable by conscientious consumers.

To reduce the prevalence of modern slavery, the United Kingdom passed the Modern Slavery Act of 2015, which requires corporations listed on the UK Stock Exchange publish the steps they take to eradicate modern slavery practices from their supply chains. Although many corporations publish statements, it is still unclear whether this law has actually led to supply chain transparency. Similar bills have been passed in Australia and California, but their effectiveness is unclear (International Corporate Accountability Roundtable).

Rather than analyzing where slavery exists, this investigation will analyze the effectiveness of counter-slavery legislation. This paper will
help stakeholders evaluate the success of international governance initiatives that have been put into force.

**Literature Review**

Modern slavery is a pervasive facet of corporate supply chains, fueled by consumer preferences for cheap goods and by long, opaque supply chains. Attempts to combat slavery practices have found major roadblocks. Corporations which provide contracts and factories that fulfil them are almost always headquartered in different countries. In addition, developing countries have little incentive to increase labor laws out of fear of harming their manufacturing economy. This literature review reviews why slavery is prevalent in certain industries and examines different forms of counter-slavery initiatives.

Crane (2012) found that “labor density, value distribution, elasticity of demand, legitimacy, and regional clustering” are major factors in determining the attractiveness of an industry to slavery practices. Crane’s analysis was one of the first to investigate slavery as a management practice. This paradigm is incredibly important, because although it is useful primarily in analyzing the forced labor aspect of slavery, it can reveal steps corporations and countries can take to make slavery as a
management practice less attractive. Crane concluded that a combination of these factors allowed industries to “deflect institutions which render slavery illegitimate.” Essentially, these factors allow individual managers to maintain their practices despite anti-slavery efforts.

One of the most interesting factors Crane identified was regional clustering. Snyder, Priem, and Levitas (2009) analyzed how illegal practices diffuse among management elites. Although they mostly examined corporate practices, they also concluded that “geographical social networks” act as a channel to spread illicit practices. When slavery is used as a management practice, other managers nearby, within the same social network, observe the success or failure of the choice to employ slavery practices. This leads to regional clustering: societies which can successfully utilize slavery will do so on a large scale, while those which cannot will see almost none. However, this clustering effect applies to slavery only as a management practice, not sexual slavery, trafficking, or recruitment of child soldiers.

Because developing countries lack incentive to protect their own laborers International governance initiatives are needed. Known as the “race to the bottom” (Mosely), countries vie for limited foreign investment by slashing worker protections. In turn, lax worker protection laws translate to higher profit margin for transnational corporations. This leaves the onus
on developed countries and consumers of finished products to modify the behavior of corporations to promote human rights around the world.

One such solution is the UK Modern Slavery Act, which demands that all large corporations listed on the UK Stock Exchange publish a disclosure of all steps taken to eradicate slavery from their supply chains. However, the effectiveness of mandated disclosures on promoting human rights and addressing forced labor is still unknown.

LeBaron and Rühmkorf (2017) compared the effectiveness of the UK Bribery and Modern Slavery Acts of 2010 and 2015 respectively. These two Acts are examples of extraterritorial legislation, laws which threaten to punish corporations operating in the United Kingdom despite potential violations of laws occurring outside of the jurisdiction of the UK. While both Acts seek to promote multinational corporate accountability, the two take different approaches.

The Bribery Act places criminal liability on corporations, which contravene section 7 of the Act, while corporations are free under the Modern Slavery Act to report that “no steps have been taken.” The Modern Slavery Act is a form of “transparency legislation,” while the Bribery Act is a form of “due diligence liabilities.” They observed that anti-bribery measures are far more prevalent in corporate practices than anti-forced
labor measures. In finding that the Modern Slavery Act “fails to establish new public labour standards or enforcement,” LeBaron and Rühmkorf conclude that the trend toward voluntary reporting over corporate liability has “undermined the effectiveness of recent governance initiatives to address forced labour in global supply chains.”

**Research Design**

To evaluate LeBaron and Rühmkorf’s findings that voluntary reporting initiatives do not effectively address forced labor, the primary analysis in this paper will be a quantitative model of transparency, i.e. modeling the content of voluntary corporate disclosures against the transparency of said corporation’s supply chains. To give a corporate supply chain a transparency score, this investigation will use scores assigned by Know The Chain, a benchmark created by a coalition of NGOs. This benchmark score ranges from 0-100 and is created via an additive index of transparency markers.

Unfortunately, the data available from Know The Chain extends to only 38 companies ranked in 2016 and 2018. A low $n$ means that confidence intervals within the analysis will be large, although strong correlations will be statistically significant.
The corporate disclosures were taken directly from corporate websites. In the past, web crawlers have been used to create large data sets. The investigation began with a dataset of $n = 1699$, but a closer look reveals that many of the scraped disclosures were incorrect. Rather than scraping modern slavery disclosures, the web crawlers were instead copying down cookie statements, or disclosures not related to forced labor. Therefore, rather than using a web crawler, the dataset was created by manually copying in corporate disclosures.

Modeling the content of corporate disclosures requires a quantitative text analysis. By using a Structural Topic Model (STM), it is possible to create a model of topic prevalence within individual corporate disclosures. Rather than simply using a frequency analysis to see the most commonly used words within the corporate disclosure, the STM analyzes which words are most commonly used *with each other*. This creates topics which can be modeled via proportion for each document. Modeling the proportion of the disclosure that discusses a certain topic will reveal if the content of voluntary disclosures is significantly related to transparency scores. To create the model, the content of corporate disclosures found on company websites, need to be pre-processed.

Preprocessing the corporate disclosures requires two steps: removing stop words and stemming. By removing stop words, such as “and.” “or,”
“if”, and other common words, the frequency analysis will focus on the topics of discussion itself.

Stemming the statement means removing suffixes from all words. For example, stemming converts both the words “business” and “businesses” to simply “busi.” Although it may be harder to comprehend, stemming words enhances the accuracy topic analysis by showing which topics are actually being discussed, despite different tenses used to discuss them.

One large drawback to the dataset created by combining corporate disclosures with Know The Chain rankings is that corporations not based in developed countries rarely publish any statements. None of the low-scoring Chinese companies contained within the dataset published statements. This will not affect the structural topic model, as only corporations with disclosures to analyze will be included in the output. However, this means that the outputs of the model cannot be extrapolated to include corporations headquartered within China.
Data

Score Distributions:

Interpretation

Corporations analyzed by Know The Chain in 2016 and headquartered in the United Kingdom (red) all scored within 60-70 points on the benchmark, while non-UK countries (white) were normally distributed, although negatively skewed.

Nearly all low-scoring (< 10) corporations were headquartered in China. However, all of the corporations depicted above were required to post a disclosure under the UK Modern Slavery Act of 2015, because they
operate within the United Kingdom. A T-Test found that UK corporations, on average, scored 15.33 points higher than non-UK corporations, with a P-value of .00, indicating statistical significance.

Structural Topic Model
Interpretation

The structural topic output reveals the most prevalent topics within the dataset of 38 corporate disclosures, ranked in order of most to least prevalent. The FREX output for the topic analysis (Appendix B) reveals the words within the topic that are unique between the differently numbered topics.

The most prevalent topic, topic 7, discusses the production of apparel within factories. At a value of .12, it is estimated that 12% of the corpus, or dictionary, is somewhat related to apparel production. The second most prevalent topic, 9, discusses organized labor within the European market, indicating that corporations discuss unionization within their disclosures. Similarly, all topics can be identified via the FREX located in Appendix B.
Estimated Effect Model

The structural topic model predicts the expected topic proportion by transparency score. After using Bayesian statistics to define the ten most prevalent topics across all corporate disclosures within the dataset, the model then determines the proportion of each topic within each disclosure. By passing the Know The Chain data to the model, covariate analysis can be conducted between the proportion and the Know The Chain score.

Interpretation
An increasing proportion of the topic discussing legal compliance is negatively correlated to transparency scores. While this may appear to be counterintuitive, under the UK Modern Slavery Act corporations are allowed to simply publish that they are posting a message in *legal compliance with the act*. Therefore, disclosures which simply address legal compliance with the act, rather than actual steps taken to address forced labor, are more likely to belong to companies with lower transparency scores.

In addition, corporations which largely address corporate social responsibility within their disclosures are more likely to have lower transparency scores. Again, this appears counterintuitive, due to the fact that the STM output is one of proportions.

Topics related to sourcing, apparel, and conflict are all positively correlated with transparency scores. However, the confidence interval for the correlations crosses 0, meaning that the correlation may not be statistically significant at the .05 significance level. As such, this investigation has not found statistical proof that increased discussion of ethical sourcing, clothing factories, or conflict-free mineral sourcing actually leads to a higher transparency score.
Conclusion

Based on the findings above, it is doubtful that statements made within corporate disclosures accurately reflect the truth about transparency. However, it is clear that disclosures which focus only on meeting the legal compliance requirement within the UK Modern Slavery Act of 2015 are significantly linked to lower transparency scores. This means that corporations which are not transparent are still meeting the legal threshold of the anti-slavery goal but not actually addressing forced labor.

In addition, the lack of a statistically significant relationship between the discussion of a topic and transparency score is a relevant finding in itself. A lack of correlation between statements made by a corporation to observed behavior of a corporation is a red flag—an indicator that voluntary statements are not sufficient in addressing forced labor.

Extraterritorial legislation walks a thin line between protecting human rights and violating the sovereignty of other nations. Legislation which harms the industry of other countries can lead to retribution and sanctions. For example, tariffs were placed on Chinese goods in response to the Made in China 2025 initiative (CFR), which simply threatened the United States’ manufacturing industry.
However, failing to act to prevent human rights violations is simply unacceptable. In the 2005 World Summit Outcome Document, all member states of the United Nations endorsed “Responsibility to Protect,” a principal which affirms the view of sovereignty as a responsibility to provide for the welfare of citizens. If a country fails to meet this responsibility, other countries have a responsibility to interfere with the sovereign operations of said country and ensure the welfare of all citizens.

If initiatives which hinge upon voluntary disclosures are truly not effective in modifying transnational corporate behavior, countries interested in promoting human rights should move toward due diligence liabilities. While corporate self-regulation is a seemingly painless solution, the misrepresentation of corporate behavior within disclosures is undermining global governance initiatives. On paper, these initiatives are meeting their targets, while in practice they do not actually increase supply chain transparency.

By combining text analysis of disclosures with empirical observations from Know The Chain, this paper has shown the lack of effectiveness inherent within voluntary disclosure-based governance initiatives. This research validates the findings of LeBaron and Rühmkorf, indicating that effective corporate governance must be enforced rather than voluntary.
References

Al Jazeera. “Pakistan Plant Certified Safe before Disaster.” *News | Al Jazeera*, Al Jazeera, 22 Sept. 2012,


Fox, Emily Jane. “Shoppers Lash out at Stores over Bangladesh.”
*CNNMoney*, Cable News Network, 2 May 2013,
money.cnn.com/2013/05/01/news/companies/bangladesh-factory-shoppers/index.html.

doi:10.18637/jss.v070.b01.


Know The Chain. “Benchmarks.” *KnowTheChain*, 2016,
knowthechain.org/benchmarks/comparison_tool/4/.


## Appendix A: Data from Know The Chain:

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Market Cap (US Billions)</th>
<th>Country</th>
<th>2016</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lululemon Athletica</td>
<td>11</td>
<td>Canada</td>
<td>69</td>
<td>89</td>
</tr>
<tr>
<td>Gildan Activewear</td>
<td>7</td>
<td>Canada</td>
<td>51</td>
<td>47</td>
</tr>
<tr>
<td>BOE Technology Group Co</td>
<td>33</td>
<td>China</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Shenzhou International Group</td>
<td>16</td>
<td>China</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Kering</td>
<td>64</td>
<td>France</td>
<td>27</td>
<td>45</td>
</tr>
<tr>
<td>Adidas</td>
<td>47</td>
<td>Germany</td>
<td>81</td>
<td>92</td>
</tr>
<tr>
<td>Hugo Boss</td>
<td>6</td>
<td>Germany</td>
<td>45</td>
<td>62</td>
</tr>
<tr>
<td>Prada</td>
<td>10</td>
<td>Italy</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Hitachi</td>
<td>38</td>
<td>Japan</td>
<td>34</td>
<td>39</td>
</tr>
<tr>
<td>Murata Manufacturing Co</td>
<td>31</td>
<td>Japan</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>Canon</td>
<td>43</td>
<td>Japan</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Keyence</td>
<td>74</td>
<td>Japan</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Company</td>
<td>Score</td>
<td>Country</td>
<td>Year 1</td>
<td>Year 2</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------</td>
<td>---------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Fast Retailing Co.</td>
<td>45</td>
<td>Japan</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>ASML Hodling</td>
<td>87</td>
<td>Netherlands</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>Samsung Electronics</td>
<td>310</td>
<td>South Korea</td>
<td>54</td>
<td>62</td>
</tr>
<tr>
<td>SK Hynix</td>
<td>29</td>
<td>South Korea</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Industria de Diseño Textil</td>
<td>112</td>
<td>Spain</td>
<td>61</td>
<td>70</td>
</tr>
<tr>
<td>Ericsson</td>
<td>21</td>
<td>Sweden</td>
<td>55</td>
<td>46</td>
</tr>
<tr>
<td>Hennes &amp; Mauritz</td>
<td>8</td>
<td>Sweden</td>
<td>69</td>
<td>65</td>
</tr>
<tr>
<td>Taiwan Semiconductor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing Co</td>
<td>227</td>
<td>Taiwan</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>Foxconn</td>
<td>55</td>
<td>Taiwan</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>Primark</td>
<td>31</td>
<td>United Kingdom</td>
<td>63</td>
<td>72</td>
</tr>
<tr>
<td>Intel Corporation</td>
<td>225</td>
<td>United States</td>
<td>59</td>
<td>75</td>
</tr>
<tr>
<td>HP Inc.</td>
<td>38</td>
<td>United States</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>Apple Inc.</td>
<td>852</td>
<td>United States</td>
<td>62</td>
<td>71</td>
</tr>
<tr>
<td>Company</td>
<td>Market Cap</td>
<td>Country</td>
<td>Market Value</td>
<td>Sales Value</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>----------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Microsoft Corporation</td>
<td>732</td>
<td>United States</td>
<td>57</td>
<td>61</td>
</tr>
<tr>
<td>Cisco Systems Inc</td>
<td>205</td>
<td>United States</td>
<td>58</td>
<td>51</td>
</tr>
<tr>
<td>Qualcomm Incorporated</td>
<td>101</td>
<td>United States</td>
<td>42</td>
<td>49</td>
</tr>
<tr>
<td>Texas Instruments</td>
<td>108</td>
<td>United States</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>Broadcom Inc</td>
<td>102</td>
<td>United States</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>Gap</td>
<td>13</td>
<td>United States</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td>PVH Corp</td>
<td>12</td>
<td>United States</td>
<td>57</td>
<td>69</td>
</tr>
<tr>
<td>VF</td>
<td>23</td>
<td>United States</td>
<td>43</td>
<td>64</td>
</tr>
<tr>
<td>Nike</td>
<td>111</td>
<td>United States</td>
<td>49</td>
<td>63</td>
</tr>
<tr>
<td>Ralph Lauren</td>
<td>9</td>
<td>United States</td>
<td>46</td>
<td>58</td>
</tr>
<tr>
<td>L Brands</td>
<td>14</td>
<td>United States</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>Hanesbrands</td>
<td>29</td>
<td>United States</td>
<td>54</td>
<td>50</td>
</tr>
<tr>
<td>Under Armour</td>
<td>6</td>
<td>United States</td>
<td>38</td>
<td>44</td>
</tr>
</tbody>
</table>
Appendix B: Topic Summary

A topic model with 10 topics, 29 documents and a 1485 word dictionary.

Topic 1 Top Words:
- Highest Prob: supplier, code, conduct, human, trafficking, rba, supplier
- FREX: rba, contractor, depend, illeg, code, nonconform, purchase
- Lift: illeg, industry-wid, thing, depend, nonconform, modul, pursu
- Score: rba, illeg, nonconform, modul, industry-wid, thing, conform

Topic 2 Top Words:
- Highest Prob: supplier, trafficking, human, supply, chain, slaveri, labor
- FREX: combat, violent, effort, team, progress, interview, legal
- Lift: anti-bribery, deep, thereaft, fire, hear, mental, briber
- Score: feder, anti-bribery, alleg, anti-human, miner, combat, robust

Topic 3 Top Words:
- Highest Prob: human, slaveri, trafficking, modern, busi, supplier, code
- FREX: high, modern, servic, area, categori, pleas, sourc
- Lift: complement, advers, north, onboard, york, seminar, “fla”
- Score: low, labour, com, advers, seminar, north, forward-look

Topic 4 Top Words:
- Highest Prob: human, labor, right, supply, corpor, chain, product
- FREX: slave, csr, philosophi, corpor, hotlin, measur, societi
- Lift: utmost, philosophi, enact, observ, csr, slave, transact
- Score: csr, philosophi, slave, utmost, transact, charter, observ

Topic 5 Top Words:
- Highest Prob: chain, supply, vendor, human, slaveri, risk, busi
- FREX: vendor, foreign, migrant, labour, relev, retail, apparel
- Lift: great, feedback, cloth, company-oper, greatest, manual, semi-announc
- Score: vendor, labour, foreign, great, assessor, feedback, apparel

Topic 6 Top Words:
- Highest Prob: supplier, human, conduct, code, audit, polici, labor
- FREX: action, rba’, correct, vap, non-conform, complet, contain
- Lift: life, subset, draft, inclus, teleconfer, vap, academi
- Score: vap, rba’, inclus, academi, saq, rba, conform

Topic 7 Top Words:
- Highest Prob: labor, supplier, factori, includ, risk, brand, forrc
- FREX: factori, brand, leadership, fla, associ, shall, footwear
- Lift: brands’, specialiti, tradit, country-specif, excess, factory’, joint
- Score: factori, brand, athlet, footwear, leadership, factory’, license

Topic 8 Top Words:
- Highest Prob: supplier, labor, gap, facil, work, group, product
- FREX: gap, inc, facil, procedur, brand, charter, back
- Lift: cycl, itali, difficulti, mind, precautionari, reserv, sub-contractor
- Score: gap, itali, vendor, brand, charter, inc, inc’

Topic 9 Top Words:
- Highest Prob: slaveri, human, supplier, busi, labour, modern, group
- FREX: europ, ltd, organis, labour, programm, market, group
- Lift: branch, emea, netherland, portfolio, london, authoris, propro
- Score: labour, emea, ltd, europ, branch, netherland, organis

Topic 10 Top Words:
- Highest Prob: supplier, human, right, work, respons, trafficking, requir
- FREX: devic, grievanc, packag, tier, electron, support, hardwar
- Lift: easili, intervenc, -day, devic, polic, south, video
- Score: devic, packag, hardwar, mine, electron, easili, raw
Appendix C: Market Capitalization and Transparency Score

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-42.948</td>
<td>-12.759</td>
<td>3.995</td>
<td>12.198</td>
<td>38.756</td>
</tr>
</tbody>
</table>

Coefficients:

|                                         | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------------------------------------|----------|------------|---------|---------|
| (Intercept)                             | 41.01856 | 3.92976    | 10.438  | 1.95e-12*** |
| Market.Cap..US.Billions.                | 0.02067  | 0.01918    | 1.059   | 0.182   |

---

Signif. codes:  
*   0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.99 on 36 degrees of freedom
Multiple R-squared: 0.04883,  Adjusted R-squared: 0.0224
F-statistic: 1.848 on 1 and 36 DF,  p-value: 0.1825

Appendix D: RStudio Code

processed <- textProcessor(slaverydata$Disclosure, 
metadata = slaverydata)

out <- prepDocuments(processed$documents, 
processed$vocab, processed$meta)

docs <- out$documents

vocab <- out$vocab

meta <- out$meta
```r
model2 <- stm(documents= out$documents, vocab =
out$vocab, K=10,

    prevalence = ~ score2016, data =
out$meta,

    init.type = "Spectral",seed = 1648)

plot.estimateEffect(prep, covariate = "score2016",
topics =c(7,9,2,1,3),

    model= model2, method =
"continuous",

    main = "Expected Topic
Proportion \n by Transparency Score",

    xlab= "Know The Chain
Transparency Score",

    labeltype = "custom",

    custom.labels = c("Legal
Compliance", "Conflict", "Sourcing", "CSR",

        "Apparel"))
```
Evaluating Patients’ Trust in Physicians Through Yelp Reviews

Beixuan Jia and Siddharaj Vaghela

Beixuan Jia is a SURE Fellow in the Department of Mathematics at the University of Maryland, College Park, MD. Their email address is: jiabx@umd.edu.

Siddharaj Vaghela is a SURE Fellow in the Department of Computer Science at the University of Maryland, College Park, MD. Their email address is: siddhu18.vaghela@gmail.com.

Abstract

Previously, the importance of building patients’ trust in their physicians has been widely recognized. There have been increasing efforts to measure patients’ trust (Hall et al. 2002; Thom et al. 2004). Meanwhile, the rapid growth of digitalization in the healthcare industry has made online reviews a critical supplemental resource to traditional surveys. Studies showed that in the United States, 19% of citizens believed that online doctor ratings were “very important” for them. 59% of citizens said that online doctor ratings were “somewhat important” for them when they searched for a physician” (Hanauer et al. 2014; Hao et al. 2016). The goal of this study was to evaluate patients’ trust in their doctors through online
reviews with text mining and natural language processing techniques. Based on these connections surrounding the trust level prediction, the research summarized possible reasons that might cause patients’ high or low trust toward their doctors. To do so, we first conducted a text mining analysis of Yelp Reviews. From the Yelp website, we extracted 500 online reviews about dentist offices in the New York City area. As shown in Fig. 1 and Fig. 2, each review was assigned a “high” or “low” trust level by volunteers recruited from Mechanical Turk, a crowdsourcing platform. We then generated two word clouds (Fig. 3 and Fig. 4) to visualize online review data and obtained an overview of high frequency words for high and low trust level reviews. Later, we came up with 11 themes which measured both services provided by the physicians and other services in the clinics. In particular, the physicians’ skills consisted of both clinical competencies and service skills.

From the data analysis, we concluded that patients with both high and low trust levels focused on three themes: bedside manner, expertise/competence, and communications. People with high trust levels paid more attentions to overall office experience, and people with low trust levels tended to complain more about waiting time/scheduling and billing/cost. This study also showed that there were actually more online reviews regarding the quality of services from other staff in the office. This
article provided a new perspective for physicians, that to improve patients’
trust would require efforts from both doctors and other staff.

*Keywords: patients’ trust, online reviews, text mining, natural language processing*

**Acknowledgements**

This research for Evaluating Patients’ Trust in Physicians Through Yelp
Reviews was prepared under the guidance of Dr. Gordon Gao, Dr. Michelle
Dugas, and Mikhail Lysyakov and the Ed Snider Center for Enterprise and
Markets at the Robert H. Smith School of Business.

**Introduction**

While investigating the relationship between patients and their
physicians, patients’ trust in their physicians plays a key role in revealing
their acknowledgments and satisfaction to the doctors’ clinics and service
competencies. According to Mark A Hall, Fabian Camacho, Elizabeth
Dugan, and Rajesh Balkrishnan’s study, trust is a crucial factor to “show
patients’ willingness to see care, reveal sensitive information, submit to
treatment, and follow physicians’ recommendations” (Hall, Dugan et al.
2001). Understanding patients’ trust could also help the doctors and
policymakers better recognize patients’ needs and expectations and provide better services to the patients. Overall, evaluating patients’ trust could benefit all participants of the healthcare system, including patients, doctors, and policy makers. From the patients’ perspective, ignoring the trust issue could lead to life-changing consequences, such as not finding the best care and deteriorations of health conditions. From the policy-makers’ standpoint, failing to understand and respond to the issue might ultimately lose the patients’ trust in the whole healthcare system (Thom, Hall, and Pawlson 2004).

Establishing an evaluation system for patients’ trust would improve the quality of healthcare system in the long run. As David Thom, Mark Hall, and L. Gregory Pawlson pointed out, “Measuring trust would help to inform public policy deliberations and balance market forces that threaten the doctor-patient relationship” (Thom, Hall et al. 2004). Furthermore, studies have shown that evaluating patients’ trust in their doctors is important and necessary, because it not only gives the medical relationships intrinsic value, but also has significant meanings in an instrumental fashion (Hall, Camacho et al. 2002).

Meanwhile, the exponential growth of social media has revolutionized people’s lifestyles and values. In particular, people have been using online reviews to help them make decisions before purchasing
products or services. Significant impacts of online customer reviews have been observed in a number of industries. According to Andrea López, Alissa Detz, Neda Ratanawongsa, and Urmimala Sarkar, the Internet has been “increasingly used by patients to find medical facilities or doctors, to research specific medical conditions, and to form support networks specific to health” (López et al. 2012). The Pew Internet & American Life Project 2008’s Tracking Survey showed an increase in the number of Americans who look online for health information from 25% in 2000 to 61% in 2008 (López, Detz, Ratanawongsa, and Sarkar). Therefore, the online reviews had become a key source for more and more people to look up health information. It was reasonable to examine the information behind online reviews beyond the traditional approaches, which had been conducted by previous studies.

Although researchers had been making progress in both patients’ trust in physicians and online review analytics in the healthcare area, much less effort has been made in measuring trust from the perspective of online reviews. In our research project, we hoped to bridge such a gap by examining online reviews and investigating possible factors that might impact patients’ trust levels in their doctors. The article began by presenting the current work in measuring trust issues and online reviews, respectively. It then explained how online reviews have been collected, categorized, and
analyzed. Finally, we interpreted the results from text mining and compared our conclusions with other work.

**Literature Review**

In recent years, much research has been dedicated to evaluating patients’ trust by constructing several domains of trust. For instance, Hall, Camacho, Dugan, and Balkrishnan assumed that “fidelity, competence, honesty, confidentiality, and global trust” were the five domains that contributed to the patients’ trust levels to their physicians (Hall, Camacho et al. 2002). Later, Thom, Hall, and Pawlson’s study refined the “five domains” after examining results from more recent research, and came up a system which measured patients’ trust through “technical competency, interpersonal competency, and agency” (Thom, Hall, and Pawlson 2004). This research supports the idea that general trust is unidimensional. This result indicated that “people do not appear to distinguish trust in the medical profession among the dimensions of fidelity, competence, and honesty” (Hall et al. 2002). They also point out in their research that most participants were not concerned about how doctors in general protect confidentiality, and the variation in concern “is poorly correlated with responses to other trust questions.” However, this result may not apply to minority groups or specialized populations (Hall et al. 2002), which has
been supported by Thom, Hall, and Pawlson’s conclusion that “confidentiality is rarely mentioned by patients and is more weakly associated with the concept of trust as defined by the other domains” (Thom et al. 2004).

Past studies also checked the association between trust and other relevant concepts. In Elena A. Platonova, Karen Norman Kennedy and Richard M. Shewchuk’s work, they investigated the relationship among patient trust, patient satisfaction, and patient loyalty. They concluded that there existed a strong linkage of patient trust to both satisfaction and loyalty. In their research, they treated patient trust as the central construct of their model, “as indicated by the significance and magnitude of its direct and indirect effects on the other examined key constructs” (Platonova et al. 2008). Furthermore, they suggested that Primary care physicians (PCPs) “must attend to the development of trusting relationships with their patients to foster deep loyalty that leads to the positive behaviors of recommendations and continued service” (Platonova et al. 2008). Their research also implied that “switching costs and attractiveness of competing alternatives did not seem to affect our respondents’ level of loyalty to their PCP; thus, switching costs and attractiveness of alternatives may not be barriers in health care,” which was contradictory to their initial hypothesis (Platonova et al. 2008).
Hall et al. (2002) found that, on the one hand, general trust had a strong positive association with factors such as “satisfaction, trust toward one’s personal physician, following doctors’ recommendations.” This statement had also been verified by Thom et al.’s work. They added that trust “has a strong emotional component not present in satisfaction,” as it refers to the relationship between the physician and patient that is based largely on perceptions about the physicians’ motivations, rather than only “refers to the patient’s opinions of the physicians’ actions” (Thom et al. 2004). On the other hand, Hall et al.’s study also pointed out that the general trust has a strong negative association with “prior disputes with physicians, have sought second opinions, and having changed physicians” (Hall et al. 2002). There also was no evidence that there is a relationship between trust and various demographic characteristics, except the participants’ ages. More research should be conducted, since most participants are in good health conditions.

The study of Hall et al. (2002), Thom et al. (2004) concluded that when compared with other factors, patient’s trust is “more strongly associated with adherence and continuity of enrollment,” since only trust showed strong, independent association with these two predictors when trust and satisfaction were placed in the same multivariate model (Thom et al. 2004). From patient focus groups and interviews, patient surveys, and
psychological theory, they also claimed that physician behaviors are related to the increase or decrease of patients’ trust. However, Thom and his team fails to identify what specific behaviors of the physicians might have positive or negative impacts on patients’ trust. Additional research is needed to address this question.

Previous research focused mainly on measuring patient trust using traditional methods, such as in-person surveys and telephone interviews with a large number of randomly selected participants. For instance, in Hall et al.’s study, they developed and tested a multi-item measure for general trust between doctors and patients. They conducted a national telephone survey for a sample of 297 male and female participants. Similarly, Platonova et al. conducted in-person surveys with a total of 554 patients who was waiting for their appointments. Comparing these research teams’ work on the measurement of patients’ trust, few had considered online reviews as a source for the responses.

Although the impacts of online reviews had not been discussed in past research, more recent work has acknowledged the significant values of online reviews platform, such as Yelp. For instance, Ranard et al.’s research stated that “as the most widely used free available commercial website in the U.S. for hospital ratings, Yelp may complement the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)
survey, which is the US standard for evaluating patients’ experiences after hospitalization” (Ranard et al. 2016). Moreover, they endorsed the positive influences of Yelp reviews to the healthcare industry and introduced the advantage of Yelp reviews over HCAHPS. “The large collection of patient- and caregiver-centered experiences found on Yelp can be analyzed with natural language processing methods, identifying for policymakers the measures of hospital quality that matter most to patients and caregivers. The Yelp measures and analysis can also provide actionable feedback for hospitals” (Ranard et al. 2016). The same study also stated that Yelp reviews possessed narrative components which “reflected the features of a hospital experience most important to patients” and could not be checked in structured surveys.

Due to the considerable potential of online reviews, researchers had been exploring the information hiding behind the online reviews and their applications in the healthcare industry. In Ranard et al.’s study, they compared and contrasted the information provided by HCAHPS and Yelp reviews. It showed that the Yelp reviews generated 12 new domains in addition to those established by HCAHPS: cost of hospital visit, insurance and billing, ancillary testing, facilities, amenities, scheduling, compassion of staff, family member care, quality of nursing, quality of staff, quality of technical aspect of care, and specific type of medical care (Ranard et al.
They suggested that Yelp reviews should be used as supplemental information for traditional patient-experience surveys and assist patients, providers and policy makers in their understanding and assessment of hospital quality.

Another study by López et al. measured patients’ satisfactions about their primary care physicians. They concluded that among all 712 reviews, more than 63% were positive, while negative interpersonal reviews “underscore the importance of well-perceived bedside manner for a successful patient–physician interaction” (López et al. 2012). A similar approach had been taken in comparing the healthcare policy among different countries. In Hao et al.’s research, they discussed similarities and differences between the Chinese and American patients’ experience and the impacts of two healthcare systems found in Yelp reviews. They then concluded that Yelp reviews were “useful assets to assist healthcare consumers, providers, and administrators in moving toward a patient-centered care” and “valuable sources for international perspectives on healthcare systems” (Hao et al. 2016).

Although the importance of patients’ trust and various applications of online reviews in the healthcare industry had been recognized, no research had been conducted that assessed patients’ trust through Yelp reviews. In our research project, we hope to bridge such a gap with data
analysis in Yelp reviews by investigating what behaviors of the doctors might lead to the low- or high-level trusts and analyzing the results from people who received different services.

Methodology and Analysis

Through this research process, we worked with the Smith School's Center for Health Information and Decision Systems (CHIDS). We started our research methodology by looking at web-scraped data from patients Yelp reviews of New York City dentists. The data was collected from one of the PhD students who was assisting us with our research project. As the data was collected through web-scraping, each review was given a selective review ID (for identification purposes), and a trust score. As shown in Fig 1 and 2, each review was given a trust score of either “high” or “low,” based on an earlier study given to us by our research mentors. From that earlier study, three independent observers rated each review. Each observer was asked to answer a series of questions regarding each review and to give a score from one to five for each question, one meaning “strongly disagree” and five meaning “strongly agree.” Scores given to these questions were then averaged out. If the average score for a review was between 1 and 2, it was a low trust review. If the average score for a review was between 3 and 5, it was a high trust review.
The data set included 500 total reviews of dentist offices in the New York City area. Of those 500, 250 reviews were comprised of high trust reviews, and the other 250 reviews were comprised of low trust reviews. The data was collected as CSV (Comma separated values) files. This allowed for better organization of the data, as it was easily accessible through spreadsheet applications, such as Microsoft Excel.

Once we had the collected data, we began our process of data analysis. For the first few weeks upon receiving our data, we ran word frequency analysis to isolate the common and most frequent words that existed in our reviews. We ran these frequencies on high and low trust score reviews separately. We used the Python programming language, along with packages such as Pandas, NumPy, collections, Pillow, and nltk. Python was used, because of its ease of use, as well as its ability to work well with data analysis. The word frequency program we used was from a post written by Tirthajyoti Sarkar. Sarkar’s Python script was a simple Python script that extracted the most common stories from a story. We tweaked the code in to match the data we were using. Another issue we had to account for was stop words. Stop words were defined as words related to names or places (such as Dr., Mr.), and terms that could not be used to produce an obvious topic (such as “it.” “for,” “was,” “they”). Essentially, they are words that help form sentences, but are meaningless on their own.
We included a pre-processing step, in which we eliminated stop words from our search results. We also completed an exploratory analysis step to visualize the differences between low and high trust reviews through word clouds. The larger and bolded words indicated a more frequent word.

However, words didn’t paint the whole picture of why patients were writing what they were writing on their reviews. To gain a better perspective of the context behind the most common words, we used n-gram frequencies. N-grams are phrases comprised of n (n being a variable) words. We used an n grams frequency script that was posted on the GitHub site by user benhoyt. We tweaked this program to work with our data. The program allowed us to isolate the number of frequent phrases we wanted, as well as the range of n grams we wanted. For our data analysis, we decided to extract the 100 most frequent phrases for each n gram between 3 words and 5 words. In total, we had 300 phrases collected. We did this for the “high” and “low” trust reviews separately.

This allowed us to see more clearly the themes that were present in both the “high” and “low” trust reviews. After working closely with our mentors, we came up with 11 themes that highlighted a patient’s review. The themes we chose were:

- Billing/Cost:
○ Keywords: costly insurance, payment, price, billing

● Fraud/Upselling
  ○ Keywords: unnecessary, scam, deceived, rip-off, sham

● Waiting Time/Scheduling
  ○ Keywords: appointment, waiting, long, short, openings, cancellations

● Clarity in Communication
  ○ Keywords: explained, clear, informative, options, presented

● Bedside Manner
  ○ Keywords: gentle, kind, considerate, engaged, nice

● Expertise/Competence
  ○ Keywords: informative, knowledgeable, thorough, incompetent, explained

● Office Experience
  ○ Keywords: experience, office, clean, comfortable, enjoyable

● Technology
  ○ Keywords: Netflix, equipment, new, television movies

● Pain
  ○ Painless, painful, bleeding, hurt, ache, irritation

● Privacy/Confidentiality
  ○ Keywords: share, personal, secret, concealed, private
Front Office Experience/Communication

- Keywords: staff, clear, communication, corresponded, reception

Once we had finalized the themes, we went back into our data and added these themes as columns next to our reviews. Next, to understand which themes were the most frequent in the reviews, we individually tagged each review to see which themes were present. We then ran another frequency program in Microsoft Excel to determine the frequency of the themes. We used pie charts, as well as bar graphs, to display the data. This allowed us to accurately view the data in a visual format, helping to navigate the frequency with which certain themes appeared in high and low trust reviews.

**Results**

After conducting data analysis and text mining techniques, we found some interesting observations from both high and low trust reviews. Some data visualization results generated by statistical analysis can be found in the appendices of this paper. Fig 3 and Fig 4 were word clouds generated by Python, where larger words indicate higher frequency, and vice versa. Fig 5 and 6 are pie charts showing the percentage of each theme’s appearance among all high or low trust reviews. Fig 7 is a bar chart which,
showing how many times each theme has been tagged in high and low trust level reviews, respectively. This section divided all reviews into several pairs: high versus low trust level reviews; reviews regarding the doctors’ clinical skills versus service skills; reviews that commented on the doctors versus other staff in the office; and easier tasks versus more complicated tasks. We also checked mixed reviews and potential correlations between certain themes.

❖ Top five themes from high and low trust levels

As shown in Fig. 5, we saw that for the low trust level reviews, the top five themes mentioned most often were: bedside manner (20% of all low-level trust reviews), waiting time/scheduling (16%), front office experience/communication (16%), expertise/competence (14%), and billing/cost (14%). In Fig. 6, we saw that the top five themes mentioned most often in high trust level reviews were: office experience (18% of all high-level trust reviews), bedside manner (16%), expertise/competence (16%), front office experience/communication (13%), and clarity in communication (12%). Both high- and low-level trust reviews focused on bedside manner, expertise/competence, and communications.

Besides the overlaps, we also could see that reviews from people with low trust levels fell mainly into two categories -- waiting
time/scheduling and billing/cost. It showed that patients who had bad experiences with their dentists were usually not satisfied with the long waiting time or the inefficient scheduling system. Additionally, some of them were upset about the expensive cost, upselling, and insurance issues. Meanwhile, patients who were satisfied with their doctors focused more on the overall office experience, such as a welcoming and clean environment.

❖ **Number of topics mentioned by the low and high trust levels**

It was also interesting to note that, compared those who did not trust their doctors, people with high trust levels tended to mention more topics in their reviews. From the bar chart in Fig. 7, we saw that topics from the 11 themes have been mentioned approximately 1,281 times among all 500 reviews.

As our data analysis shows, the average number of topics mentioned by high trust level reviews was 3.52, more than twice the average number of topics mentioned by low trust level reviews, 1.68. This observation matches common sense, since unhappy patients are more likely to focus on a particular aspect they were dissatisfied with, whereas satisfied patients are more likely to mention more good aspects in their reviews.

❖ **Reviews about the physicians’ clinical skills versus the service**
skills in high and low trust levels

Among all the categories, we noticed that there were some themes directly related to the doctor’s expertise or competence, such as “Expertise/Competence” and “Pain,” whereas some themes were more related to the doctor’s communication and services skills, including “Bedside Manner.” and “Clarity in Communication.” 293 reviews commented on doctors’ expertise or competence, 206 with high trust and 87 with low trust. There were 374 reviews regarding doctors’ interpersonal skills and behaviors, 271 with high trust and 103 with low trust. This told us that patients might focus more on the physician’s service rather than competency. For both comments regarding service skills and clinical skills, there were more reviews with high trust levels than low trust levels. It also showed that patients with low trust levels might lose their trust due to reasons other than the services offered by doctors, such as a poor front desk experience, which will be discussed in the next section.

❖ Reviews about the physicians versus the clinic/office in high and low trust levels

The 11 themes could also be divided into two other groups: themes related to the doctors’ skills and the overall experience at the office. Four themes were directly related to the doctors, such as “Clarity in

Sure Journal 2019  |  253
Communication,” “Bedside Manner,” “Expertise/Competence,” and “Pain”.

Six themes were related to the overall experience at the office and the qualities of the services provided by other staff, such as the front desk receptionists and nurses who assisted the doctors. The reviews which focused more on the clinic/office as a whole include “Billing/Cost,” “Fraud/Upselling,” “Waiting Time/Scheduling,” “Office Experience,” “Technology,” and “Front Office Experience/Communication.” The theme “Privacy/Confidentiality” could refer to indicating how either the doctor or other staff properly handled the patient’s sensitive information. Among all 1,281 tags, 754, roughly 58.8%, were related to the overall office service. The remaining 527 tags, only 41.2% of the total reviews, mentioned the doctor’s services. Among all reviews regarding the overall office service, 272 were tagged low trust, and 480 were high trust. For reviews regarding the doctor’s services, 200 were low trust, and 467 were high trust. The total numbers of tags were much more than the total number of reviews, since most reviews were tagged by more than one theme, as we discussed earlier.

Many mixed reviews, which praised the doctor but complained about the other staff in the office, were found among all 500 reviews (250 low and 250 high trust levels). For instance, one of the low trust level reviews mentioned that “the doctors are good. The administration is obnoxious and unfriendly.” This showed that improving patients’ trust
levels would need the efforts of both the physicians and the other clinic staff. Clinics should focus not only on enhancing the doctor’s clinic and service skills, but also improving the other staff’s communication and organizational skills by providing regular trainings.

❖ Mixed Reviews

A review was defined as “mixed review” if the patient made both positive and negative comments on certain themes. Based on our tagging criteria, a mixed review might include some themes tagged as “1” (positive) and the others tagged as “2” (negative). There were many mixed reviews among the 500 reviews. An obvious discrepancy existed in two different trust levels. Among all 250 high trust level reviews, only 7 reviews were mixed reviews, which is 2.8% of the total high trust level reviews. In contrast, more mixed reviews were observed among the low trust level reviews, which include approximately 46 mixed reviews, 18.4% of the total low trust level reviews. This result was expected: in the tagging system, most mixed reviews were tagged as “low trust level,” since a review showed “low trust” as long as the patient complained about at least one theme, even if the patient might have been satisfied with other themes.

❖ Reviews about “easier tasks” versus “more complicated tasks”
Besides the 11 themes, we also tried to label the reviews with “easier” and “more complicated” tasks. A review was tagged “easier task” if the patient received service such as regular cleaning or routine dental check-up, and was tagged “more complicated task” if the patient received more complicated procedures, such as root canal or extraction. Generally speaking, we would expect that the complicated tasks were more expensive and time-consuming than the easier task. We also would like to check if there were different focuses for people who received easier or harder procedures, and if people complained less when they received easier procedures.

From the bar chart in Fig 7, there were 60 among all 500 patients who saw their dentist for routine check-up or cleaning. 14 had low trust levels, and 46 had high trust levels. There were total 86 patients who received more complicated procedures. 21 were tagged low trust levels, and 65 had high trust levels.

Before conducting deeper data analysis, we realized that since our sample size was small, the results would probably be inaccurate and biased. Another limitation for tagging easier or more complicated tasks was that many reviews did not mention what medical service they had received. Therefore, the results could be inaccurate, and we decided not to include this measurement in our conclusion.
Discussion

The results of this study were generally consistent with previous studies of patients’ trust and online reviews. The first finding was that most patients, no matter their trust levels, had expressed their concerns in bedside manner, expertise/competence, and communications. These three aspects reflected the “technical competency”, “interpersonal competency” and “agency” domains of trust in Thom et al.’s study (2004). Studies also showed that these three domains were not independent from each other or from trust more generally (Hall et al. 2002; Thom et al. 2004).

Besides bedside manner, expertise/competence, and communications, patients with low trust levels also complained about waiting time/scheduling and billing/cost. This matched Hao et al.’s research, which claimed that negative reviews from American patients focused more on “the staff, wait time, and insurance” (Hao, Zhang, Wang, Gao 2016). Our study also showed that patients with high trust levels mentioned the overall office experience, such as a welcoming and clean environment, more often than other patients.

Our tagging system measured not only doctors’ clinical skills, but also their interpersonal and service skills. The result showed that there were more comments regarding doctors’ service skills than the clinical skills.
This told us that patients are likely to focus more on a physician’s service than competency. For comments regarding both service and clinical skills, there were more reviews with high trust levels than low trust levels. It also showed that patients with low trust levels might lose their trust for reasons other than the services offered by doctors, such as the poor front desk experience, which would be discussed later.

Surprisingly, more reviews mentioned the overall office experience instead of focusing only on the doctors’ skills. This result helped verify that the quality of doctors’ services was not the only factor which contributed to the patients’ trust levels. Other factors, such as structural/staffing, also played important roles in patients’ trust (Thom et al. 1997; López et al. 2012).

It also showed that there were more patients who trusted their doctors than those who distrusted their doctors, an observation that has been supported by multiple studies. For instance, López et al.’s research proved that “the majority of Internet reviews (63% among all 712 samples) of primary care physicians are positive in nature.” (López et al. 2012).

While tagging reviews with high and low trust levels, sometimes we could see a potential correlation between certain themes. Although this article did not focus on investigating correlations among different domains
due to the small sample size, more studies regarding the correlations of domains for the Yelp reviews should be conducted in the future. In our studies, we observed some interesting pairs of themes which might have correlations. For instance, “Bedside Manner” and “Expertise/Competence” were always mentioned in the same review, falling into the doctors’ service and clinical skills, respectively. Approximately 30.5% of reviews about bedside manners of them also discussed the physicians’ expertise, indicating that there might be correlation between doctors’ service and clinical skills.

Moreover, some other “correlations” equated with common sense. For example, “pain” and “expertise/competence” were usually found in the same review. This was easy to interpret, since if the patient could feel the pain, it might show that the doctor did not possess excellent clinical skills and professional knowledge. However, some pairs of themes which were tagged together often seemed unexpected and random. One of the examples was that among all “Bedside Manner” reviews, roughly 35% also mentioned “Waiting Time/Scheduling.” This observation had not been discussed in the past studies. Since the sample size of this study was not large enough, some of the “random pairs” might not actually have correlations.

In our analysis, we failed to find correlations between
“Privacy/Confidentiality” and other themes. Among all 500 reviews, only 1 review with low trust level mentioned that he/she believed that the doctor/other staff had violated his/her privacy and had given his/her sensitive medical information to a third party. This review focused only on the privacy/confidentiality theme and did not mention any other themes. This result had been supported by multiple research work. For instance, Hall et al. also concluded that most of their subjects “were not concerned about how doctors in general protect confidentiality, and the variation in concern that exists is poorly correlated with responses to other trust questions” (Hall, Camacho, Dugan, Balkrishnan 2002). Similarly, Thom et al. pointed out that confidentiality “was rarely mentioned by patients and is more weakly associated with the concept of trust as defined by the other domains” (Thom, Hall, Pawlson 2004).

Conclusion

The article showed that even though most reviews were focused on the physicians’ clinical and service skills, interactions with other staff and procedures in the office could have positive or negative impacts on the patients’ trust levels. In some situations, even if the patients were satisfied with the doctors’ skills, other staff’s unprofessional behavior could
undermine their trust in the clinic office.

Although this study has provided new findings for evaluating the impacts of online reviews on physician performance and service, it still has limitations. First, the data we used for this study was limited to dental clinics in the New York metropolitan area. Therefore, results and conclusions from this research might not be generalizable to other geographical areas and other types of physicians or medical specialists. Ideally, a larger and more diverse sample drawn from various types of medical services would enhance the validity of the study. Second, some Yelp reviews are not necessarily related to the trust issue, but are still assigned to certain trust levels. Third, for the sake of data analysis and manipulation, the current methodology divided all reviews into two groups: high and low trust. There might be more reviews having neutral trust level. Those trust levels have not been totally reflected in this study. Future research should be conducted to address these concerns.
References


## Appendices

<table>
<thead>
<tr>
<th>ID</th>
<th>Trust level</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>4y2tw6AKgLSyPl65v2Zg</td>
<td>low</td>
<td>I have been coming to tribeca dental design for many years. I am writing this poor review because of an experience today that has forced me to switch dentists. I had an appt this morning which I cleared my work schedule for. Unfortunately a 30 minute commute turned into an hour commute due to subway delays. At one point I was in a spot where I had service and could send an email - at that point I thought I would be 10 minutes late. Unfortunately I ended up being 20 minutes late due to the subway simply stopping mid track and delayed with no service and no ability to contact them. I ran there once out of the subway to be told they could not help me. While I completely understand they have a schedule - I personally have sat in that waiting room for 15 minutes waiting for my appointment when they were delayed. Given that I had been coming there for many years and had a situation that was beyond my control I would have appreciated a little understanding. I mean a cleaning takes 15-20 minutes max. I was super frustrated and honestly it's made me decide to switch dentists.</td>
</tr>
<tr>
<td>2crU6oU0MHaZ2E9N085Kw</td>
<td>low</td>
<td>It may be just me, but I feel like since the move to the new location, the great service did not go along with it. Now do not get me wrong, I still love my dentist Dr. Haldey, but I noticed some of the things that made me love the practice so much no longer exist. For one, I remember at the old location, the dental assistant/hygienist would allow me to use Netflix on the computer for the duration of the cleaning. Also, the hygienist/dental assistant I had for my last procedure kept on hitting my teeth with her tools. Ouch! I do not feel that I received the same education or support that I did at my first annual. No new toothbrush, no paste, no care package LOL. Maybe I just like the freebies but every dentist I have been to has at least provided a toothbrush on your annual. I did receive some floss. What I do have to say has remained the same is the consistence of the front desk, they are very good about scheduling me for times that work for me, and give me ample notice if there is a conflict in Dr. Haldey's schedule. I really like Dr. Haldey as a dentist so I will continue to come to CitiDental, but as a business you cannot raise the bar so high, and then let your customers down when you can not maintain the precedent you have set.</td>
</tr>
<tr>
<td>yXHDB6GeZLYxZcfOfbwYow</td>
<td>low</td>
<td>After 6 months of getting fillings, two of them fell/come out. Won't be going back. I also told the dentist there was an abscess on my gums. He said it and said it I'll go away. Abscess on gums just don't go away because they are infections and the longer they stay the infections only get worse. I'm not sure what type of dentist doesn't know that.</td>
</tr>
<tr>
<td>f_X_U3QpFm6ekIDqJpzVA</td>
<td>low</td>
<td>My beloved Dr. Berger retired and good for him. However, I'm not a fan of the changes in office management. Now they require you to have a credit card on file and charge it without prior notification of the dollar amount. I'm uncomfortable with them charging the card without my knowing how much. The hygienist told me the amount for my gums and Dr. Hovovitz was just ok. Perhaps it's time to move on.</td>
</tr>
<tr>
<td>6SU68wPznh2gSbQkG9k9</td>
<td>low</td>
<td>This is a beautiful office with a team of capable dentists. I have a new technology and the best equipment at their disposal. I am excellent at personal touch. Almost to the point where I thought I was getting special treatment. Turns out they give that special treatment to everyone. So, I decided I would use the facility to get a tooth crowned. The dentist I was supposed to see had an emergency so they had another dentist see me, fine. I guess. She shot me up with so much numbing agent I was numb for two days and my eyes were twitching. She went over my options with me after depleting and temporarily filling my tooth. Either crown or a root canal and crown. I decided to go the more permanent route of doing both. I didn't know at the time that the prices were triple than I ever paid in my life. I would love to come here to my dental work but I just cannot afford it. The reason for 3 stars instead of 4 was because I was very disappointed during the last exchange I had with an employee there. He was misleading me. She was confirming my appointment and shot out a price through the text and asked if I agree to it. Nowhere in the text did it say what procedure was being done. It felt misleading. I am still stuck with half a job done by them that I have to try and find someone to finish. They never called me to follow up after cancellation.</td>
</tr>
</tbody>
</table>

Fig. 1 Sample reviews with "low trust level"
<table>
<thead>
<tr>
<th>ID</th>
<th>Trust level</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>2t0aq3Ka7gPYn7x9WnwZbw</td>
<td>high</td>
<td>Always have a great experience with this dental office. The doctors are informative and thorough. You don’t have to worry about wasting your time doing unnecessary procedures that are not needed like in other offices.</td>
</tr>
<tr>
<td>Uciz-uho9uWB0rv7xqyYA</td>
<td>high</td>
<td>I’ve been to this Dentist for a cavity filling and 2 cleanings with the hygienist and have always had a nice comfortable experience. Laura the hygienist is kind, gentle, and informative. Each time taking care of me, knows my dental needs and preferences. Each time I have arrived they’ve taken me on time. Recently I met Dr. Park who is new to Tribeca Dental, and she was very thorough and informative. I spoke to her about my concerns and she gave me good advice to remedy issues I was having. Great place!</td>
</tr>
<tr>
<td>KBHdjia8za-GGRHKz-gKZCw</td>
<td>high</td>
<td>Kay! I love this place! Super conveniently located on the same block as the subway station! It was nice and quiet. One other person in the waiting area with me. Nothing like those crowded clinics. My appointment began on time. The staff were super friendly and genuinely concerned about my personal comfort. I watched TV. The dental hygienist was FANTASTIC! I want to say that her name is Frances Estevaz. (Shortened version of a longer first name though) She was amazing and explained everything she was performing for me. Really sweet lady. Doctor Huang was cool and down to earth. Wasn’t upset when I began falling asleep during the treatment. I’ll be setting up another appointment soon. Oh, and some people are speculating that the reviews are staged or fake. Well, mine is not! Hopefully my next experience will be just as good.</td>
</tr>
<tr>
<td>dFxaYg2Yk80maly4p-OQ</td>
<td>high</td>
<td>Well I found my new dentist in NYC and went in this week for a cleaning and had a truly great experience. The office was clean and hip, the staff were all very friendly. Dina was my hygienist and did an awesome job. I recommend this place seriously and am looking forward to my next visit...is that weird?</td>
</tr>
<tr>
<td>wnT9k6DMQ4ufnQNGSwN6w</td>
<td>high</td>
<td>I just visited Dr. Han’s office for the first time today and I couldn’t be happier. My boyfriend (who’s very nervous at the dentist!) has been seeing them for about a year and has been very happy, so I figured that I’d be completely at home, seeing as I enjoy seeing the dentist, hahah! I visited one other dentist here, but was unhappy with them, so I’d been procrastinating another visit. Today, I called Dr. Han’s office on a whim and the adorable Australian receptionist asked if I wanted to come in right then. Why not? I made it downtown in about 20 minutes, filled out my paperwork on an iPad (cool) and was seen minutes afterward. Laura, my hygienist, was great and the dentist (also a young woman) was great. The last dentist I saw recommended that I have all of my wisdom teeth pulled because of a cavity in one—a procedure that I was reluctant to go for considering that a) the teeth are otherwise perfect and straight, b) it’s painful, and c) it’s expensive. The dentist at Dr. Han’s told me that they recommend just filling the cavity rather than removing all four teeth. I was extremely happy to hear that! Dr. Han’s office seems to be about actual prevention rather than just tacking on expensive services you don’t need. (Bonus points for super quick appointments and great availability. I’m going back Monday to get the filling done!)</td>
</tr>
</tbody>
</table>

Fig. 2 sample reviews with “high trust level”
Fig. 3. Word Cloud for high trust level

Fig. 4. Word Cloud for low trust level
Fig 5. low trust level reviews themes distribution

High trust level reviews themes distribution

- Front office experience/communication: 13%
- Billing/Cost: 4%
- Fraud/Upselling: 3%
- Waiting Time/Scheduling: 9%
- Clarity in Communication: 12%
- Bedside Manner (professional): 16%
- Expertise/Competence: 16%
- Office experience (cleanliness, design): 18%
- Technology: 3%
- Privacy/confidentiality: 0%
- Pain: 6%

Fig 6. high trust level reviews themes distribution

DISTRIBUTION OF DIFFERENT THEMES IN LOW/HIGH TRUST LEVEL REVIEWS

Fig 7. distribution of different themes in low/high trust level reviews
Public Pensions: A Crisis of Political Negligence

Christian Kincaid
Research Mentor: Hossein Abbasi

Christian Kincaid is a SURE Fellow in the Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: christiantkincaid@gmail.com.

Abstract

A ten-year economic recovery has failed to close the widening aggregate public pension funding gap in the United States. As savings accounts, asset prices, and employment levels have recovered since the 2008 financial crisis, a disparity has emerged in employee retirement plans: corporate pensions have recovered to a near fully-funded level while the funding gap in public pensions persists. This paper seeks to analyze the factors behind this phenomenon by answering two questions. First, have public pensions made unrealistic investment return assumptions? Comparing expectations to reported results will settle this question. Second, to what extent does investment performance influence the funded status of
public pensions? A regression analysis will assess the effectiveness of investment returns relative to their benchmarks as a predictor of pension funding levels.

Acknowledgements

This research was prepared under the guidance of Dr. Hossein Abbasi of the Department of Economics at the University of Maryland, College Park. Many thanks to Dr. Abbasi, Dr. Christina Elson, and the Ed Snider Center for Enterprise and Markets at the Robert H. Smith School of Business for supporting this research.

Introduction

Few financial structures are more important to the health of the American economy than the public pension system. Roughly 6,000 public retirement systems encompassing 14.5 million active working members and 10.3 million retirees operate in the United States. At a time when the personal savings rate in the United States is depressed relative to historical levels, the health of public pensions is more critical than ever. Teachers, firefighters, police officers, and public employees across the country expect the benefits from these plans to finance a comfortable retirement. But an
aging population, the 2008 collapse of asset prices, and fiscal recklessness on the state and municipal levels have created challenges for maintaining adequate pension funding status.

The measure that has the greatest effect on a pension plan’s finances is the investment return assumption. Since 1988, public pensions have earned about $7.5 trillion in total revenue. Investment earnings were by far the largest contributor to this total at $4.68 trillion earned, followed by employer and employee contributions at $1.97 trillion and $888 billion, respectively. Any shortfall in projected investment returns must be resolved by achieving higher returns on invested returns, increasing contributions, or cutting benefits to retirees. All else being equal, an investment return assumption that is too low artificially increases a plan’s unfunded liabilities, thus its cost to current taxpayers. Inversely, rates set too high would allow for the plan to appear funded and temporarily understate liabilities, allowing politicians to engage in fiscally irresponsible behavior over the short term.
This research seeks to empirically determine the degree to which forecasted (assumed) returns differ from realized ones and the importance of investment returns to a plan’s overall funded status.

**The Depth of the Problem**

No measure better captures the financial health of a pension than net cash flow, the difference between funds received and distributed. Positive net cash flow indicates the plan is receiving an adequate amount of funds (through investment returns or employer/employee contributions) to cover its promised obligations to retirees. Negative net cash flow indicates a plan is becoming underfunded because of a shortfall in funds received compared
to benefits paid. Plans with consistently negative cash flows are eventually forced to plug the gap by either cutting benefits to retirees or increasing contributions through politically hazardous budgetary decisions. But over the short term, states can use debt financing (usually through the issuance of bonds) to temporarily increase funding levels. This transient solution, by which debt is funded with more debt, is being adopted by some municipalities with low and deteriorating funded ratios.

Figure 2 shows the trend in net cash flows to state and local government pensions in aggregate over 88 years. Net cash flows peaked in 1997 at $70 billion, but have steeply declined since. Critically, cash flows turned negative in 2009 in the wake of the 2008 financial crisis, and the trend has only worsened despite the economic recovery that followed.
Any system which pays out more than it receives is unsustainable. Since 2009, pensions managed by state and local governments have failed to raise sufficient revenue to cover mounting obligations. Public pensions in the last decade have functioned as a transfer program funded by debt, rather than by contributions or investment returns.

As expected, aggregate funded status has deteriorated due to declining cash flows. Figure 3 plots aggregate pension assets versus liabilities from 2002 to 2016. While a gap has persisted for over 10 years, it has widened significantly since 2008. Liabilities nationwide stood at about $8 trillion as of 2016, compared to aggregate assets of only $3.8 trillion.
This leaves a current nationwide funding gap of $4.2 trillion – less than 50% funded.

The aggregate public pension funding gap in the United States – that is, the difference between plans’ existing assets and their promised benefits – is a crisis of unprecedented scale.

For roughly a decade, the United States has experienced an economic expansion accompanied by a bull market in most asset classes. Real estate prices, merger and acquisition activity, and domestic equity indices have all recovered from the depths of the financial crisis. The job market is strong, the labor market is tight, and inflation is low. These factors should be a boon to pension portfolio returns, yet the public pension funding gap continues to widen. My research seeks to explain this phenomenon.

**Literature Review**

The existing literature on public pensions is robust because of the issue’s importance to the financial system. However, the widening gap between public pensions’ assets and liabilities is a more recent phenomenon. The role of this literature review is to synthesize existing research and apply its insights to this modern problem.
In “State and Local Pension Reform Since the Financial Crisis,” Aubry and Crawford argue that public pensions in the United States are difficult to reform because of the varied and competing interests of the stakeholders involved. The authors conducted a comprehensive analysis of changes made to state and local pensions post-2008 that were designed to mitigate plans’ rising costs. Their analysis indicates that about three-quarters of states and more than half of localities have enacted some sort of pension reform since the financial crisis. They find the most common change to public pensions has been to increase employee contributions and reduce the cost-of-living adjustment (COLA) in decrements. Some plans also have increased age and tenure eligibility so that public employees must work longer before receiving benefits. The authors note that a number of long-standing legal protections prevent dramatic reform to the pension system.

In “State Public Pension Investments Shift Over Past 30 Years,” analysts for Pew Charitable Trusts examined the asset allocation trends of portfolios managed by public pensions. The analysts identify three main types of investments that make up pension portfolios:

- **Fixed Income Investments**: any investments for which returns are predictable and paid at designated times, often considered lower-risk
because they guarantee a specified return; include domestic or international bonds issued by governments or corporations

- **Equities**: stocks held by investors that represent ownership in a company, often considered riskier because they do not guarantee any return, but have the potential for much higher returns than fixed income; can be domestic or international, or be bundled as exchange-traded funds (ETFs), mutual funds, or index funds that track stock indices

- **Alternative Investments**: any investments other than the traditional asset classes of cash, fixed income, and equities. These investments often carry the highest risk along with the possibility of the highest rate of return; can include real estate, private equity, commodities, or hedge funds

These three types of investments comprise the sum total of invested pension assets. Historically, the authors write, public pensions invested most of their assets in fixed income, such as government bonds and highly rated corporate bonds. These investments were designed to provide a stable return while not risking the plan’s assets. While fixed income investments and equities continue to dominate pension portfolios, alternative investments in recent years have become more popular among plan managers. The trends in pension investment allocation are summarized below:
Pension Investment Allocation

<table>
<thead>
<tr>
<th>Investment Type</th>
<th>2006</th>
<th>2012</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Income &amp; Cash</td>
<td>28%</td>
<td>27%</td>
<td>-1</td>
</tr>
<tr>
<td>Equities</td>
<td>61%</td>
<td>50%</td>
<td>-11</td>
</tr>
<tr>
<td>Alternatives</td>
<td>11%</td>
<td>23%</td>
<td>+12</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

In the period during and directly after the financial crisis, traditional investments began to play a smaller role in pension investment portfolios. Alternative investments as a share of assets more than doubled from 2006 to 2012. This shift in asset allocation reflects plan managers’ desire to earn higher returns through hedge funds, real estate, distressed debt, and commodities. In the future, we expect portfolio returns to be even more volatile as plan assets become more concentrated in alternative investments.

In “Portfolio Allocation for Public Pension Funds,” Pennacchi and Rastad investigate the relationship between pensions’ investment return assumptions and the riskiness of their portfolios. Using data encompassing
125 state pension funds from 2000 to 2009, the authors state that “a pension fund’s portfolio allocation policy has first-order consequences for its funding status” because of the role asset allocation plays in setting the assumed rate of return. They go on to discuss the various methods of discounting pension liabilities.

The current standard for discounting pension obligations is the General Accounting Standards Board (GASB) approach, by which a plan’s future retirement payments are discounted using the assumed rate of return on its assets. Crucially, plan liabilities are *not* discounted according to the riskiness of their investments. The authors note that this approach is inconsistent with basic finance theory, in which a discount rate reflects risk. They explain that the current standard creates a moral hazard problem through what they call “accounting arbitrage:” plans are incentivized to invest in systemically risky assets to justify higher discount rates, which in turn reduce the actuarial value of liabilities. Pennacchi and Rastad argue that the current GASB standard is archaic because it “clouds the true market valuation of pension liabilities” and discourages plan managers from exercising adequate risk management.

The World Economic Forum’s 2017 white paper, “We’ll Live to 100 – How Can We Afford It?, explores the worldwide pension funding gap and offers several possible solutions to the problem. This paper offers
valuable insights into ways policymakers in the United States can improve pension funding levels and increase savings rates. The World Economic Forum identifies five policy recommendations for addressing the challenges facing public pensions. Among them, two stand out as particularly relevant to the subject of our research.

First, the report suggests that the normal retirement age for pension plans needs to rise in line with increasing life expectancies. In countries such as the United States, the United Kingdom, Canada, and Japan, future generations will have a life expectancy of around 100 years. To support a tenable ratio of workers to retirees, the analysts suggest gradually increasing the retirement age to 70 by 2050. This will ensure enough workers are contributing to the system to maintain adequate funding status.

Second, governments around the world should reform the pension system to make saving easier for workers. The report cites recent UK reforms in which 8% of each individual’s earnings are automatically contributed to a pension savings account. The policy has boosted savings rates among 22- to 29-year-olds and low-income workers since being implemented and is estimated to create $2.5 billion in increased pension savings per year. This proposal mirrors the “Save More Tomorrow” plan pioneered by Shlomo Benartzi and Nobel Laureate Richard Thaler and
could be implemented by public and private retirement systems in the United States.

If these recommendations (among others) are not implemented, the analysts warn that economic calamity will ensue. The World Economic Forum predicts the world’s six largest pension systems – the United States, United Kingdom, Japan, the Netherlands, Canada, and Australia – will face a $224 trillion gap by 2050. This gap is being driven by lower savings rates and longer lifespans, which threaten the long-term viability of these countries’ retirement systems.

The Pew Charitable Trust’s report published in 2018, “The State Pension Funding Gap: 2016,” examines the American public pension system on a state-by-state basis. Pew’s analysts found that state public pensions were well-funded until about 2000, when several states increased promised benefits without a corresponding increase in revenue. Courts in California and Illinois found that these obligations must be honored, creating funding problems for these states. Additionally, the bursting of the so-called “tech bubble” precipitated a collapse in investment returns. Many states had not recovered from the tech bubble when the 2008 financial crisis occurred, compounding losses and widening the funding gap. In their analysis, Pew identified Colorado, Connecticut, Illinois, Kentucky, and
New Jersey as the states with the most indebted pension plans. Only four states – New York, South Dakota, Tennessee, and Wisconsin – were over 90% funded.

**Expectations Versus Reality: An Analytical Approach**

To examine the relationship between investment return assumptions and realized investment returns, we analyzed the Public Plans Database (PPD). The PPD contains plan-level data encompassing 114 pensions administered at the state level and 66 administered locally. The data set covers 95% of public pension membership and assets in the United States from 2001 - 2017, making it a sufficiently representative sample for this analysis. In total, the data includes more than 3,000 observations of 269 variables, though for our purposes we have isolated these select plan features:

- Plan name
- State
- Fiscal year
- Investment return assumptions, used to discount promised future benefits
- Assets
- Liabilities
- Net Assets, defined as Total Assets – Total Liabilities (negative values indicate underfunded plans)
- Funded ratio, defined as \( \frac{\text{Total Assets}}{\text{Total Liabilities}} \)
- Funded status, defined as total assets as a percentage of total liabilities
- Pension portfolio investment returns over one, three, and five years
Plotting the mean investment return assumptions used by public pensions each year from 2001 through 2017, we detect a small but noticeable downward trend using a line of best fit approach (Figure 4). It is clear that return assumptions have decreased over time as fund managers revise down their expectations. By 2025, we expect the mean investment return assumption to have declined by 30 basis points from the 2017 mean. Figure 5 shows that 2017 investment assumptions currently follow a roughly normal distribution, with greatest frequencies clustered around a mean of ≈7.4%. We expect the mean of this distribution to shift leftward over time as this downward revision occurs.

Figure 4

Figure 5
Using a linear projection, we foresee a decline of 30 basis points (0.30%) in the mean investment assumption used by public pensions. While at first this trend appears small, any move in this assumption influences the rate at which liabilities are discounted. For the largest plans, a small decrease in the investment assumption can cause a multimillion-dollar increase in liabilities. Our analysis suggests that all else being equal, unfunded obligations will continue to swell as a result of decreased investment return assumptions.

The PPD includes plans’ investment performance over one, three, and five years. Since the investment return assumption functions to discount future liabilities, it is important for fund managers to accurately forecast returns. If actual returns for the year are lower than expected, liabilities swell. If excessively high assumptions are perpetually used, the pension becomes insolvent over time.

The obvious way to assess plan managers’ accuracy in setting investment return assumptions is to compare per-year performance to the investment assumption rate used. To do this, we measured the difference between the one-year investment return and the assumption used for each plan. The mean one-year realized investment return was 6.52%, while the mean return assumption used was 7.81%. This means that, from 2001 - 2017, pension plans underperformed their benchmarks by 1.29% on
average. This underperformance is statistically significant at the 95% confidence level and appears when considering three- and five-year average realized returns as well.

Figure 6: Public Pension Investment Performance, 2001 - 2017

Public pensions’ aggregate performance can best be observed by the visualization above. The x-axis plots return assumptions used by public pensions against the y-axis showing one-year realized returns. The horizontal line represents the mean return assumption used over this period (7.81%). Observations above this mean line represent plans that overperformed relative to their assumptions, with underperforming plans falling below. Clearly, while some plans overperformed, a majority of observations fall below this line, suggesting widespread underperformance over an almost two-decade period.
During our analysis, we discovered that while both realized and assumed returns followed roughly normal distributions, the realized return distribution was far flatter. In other words, realized returns were far more variable and dispersed from the mean than their benchmarks would suggest they should be. This is borne out by summary statistics of our data. The standard deviation of one-year realized returns was 10.78%, while return assumptions had a standard deviation of only 0.44%.

Public Pension Performance During the Financial Crisis and its Aftermath

In the wake of the financial crisis, the Federal Reserve embarked on an unconventional, two-pronged monetary policy: a lowering of interest rates to the zero bound and large-scale purchases of financial assets, including Treasury bonds, mortgage-backed securities, and collateralized debt obligations. Known as quantitative easing (QE), these policies had the effect of stimulating asset prices while driving bond yields down. As discussed in the Literature Review, riskier alternative assets are making up a greater percentage of plan assets than ever, motivated by plan managers’ desire to generate higher returns than were available through fixed income. Re-allocating pension assets toward alternatives like hedge funds or private
equity may increase returns over the short-term at the expense of investing employees’ contributions in riskier, more volatile portfolios.

Put simply, the Federal Reserve’s policies since the financial crisis have helped borrowers and hurt lenders, rewarding risk-takers and investors while punishing savers. It’s no wonder, then, that the American public pension system, one of the largest savings programs in the country, has struggled since 2008. The financial crisis caused realized returns to greatly diverge from their benchmarks, causing pension portfolios to suffer.

![Figure 7: Public Pension Investment Performance, 2008](image)

In 2008, the average return assumption was 7.91%, while the average realized return was 8.91%, yielding an average underperformance of 16.82%. That year, nearly every single public pension plan in the United States failed to meet its benchmark by a significant margin.
This analysis indicates the following: the decision-making bodies responsible for setting investment return assumptions failed to reflect the impact of financial shocks, whether domestic or exogenous, on portfolio returns. Plan managers and legislatures failed to account for the volatility of pension portfolio returns when setting investment return benchmarks.

**Predicting Funded Status Using Realized Returns**

In the previous section, we measured the accuracy of plans’ return assumptions by comparing the assumptions used to gains realized. Next, we examine the effectiveness of using investment performance and the accuracy of return assumptions as predictor variables of funded status. We define plans with more “accurate” assumptions as those that minimize the difference between actual and assumed returns. Underperforming plans report a negative difference while consistently accurate plans are closer to zero. Of the plans surveyed in the PPD, the mean difference between actual and assumed returns was -1.2%; that is, public pensions underperformed their benchmarks by over one percentage point on average over the 2001-2017 period.
Our first regression used five-year average investment returns as the independent variable and funded status as the dependent variable. Our analysis yielded the following linear relationship:

\[
Funded\ Status\ (%) = 53.41 + (2.65 \times Investment\ Return) + \epsilon
\]

This equation indicates that for every 1% increase in investment returns, a plan’s funded status rises by 2.65%.

Our second regression used the average difference between assumed and realized returns (that is, the average difference between benchmarks and actual returns) as the independent variable and funded status again as the dependent variable. Our analysis suggests the following linear relationship:

\[
Funded\ Status\ (%) = 76.21 + (3.62 \times Return\ Deviation) + \epsilon
\]

Thus, for every 1% difference between the investment return assumption and realized returns, a plan’s funded status changes by 3.62%. That is, if a pension overperforms its assumed return by 1%, its funding status will rise by this amount; if 1% underperformance occurs, its funding status declines by the same magnitude.
These findings demonstrate that public pensions must achieve robust
investment returns over time and must be more accurate in setting
investment return assumptions. Both variables play a key role in the long-
term viability of public pension plans.

Conclusion

The aggregate underperformance of American public pension plans
over the last two decades seriously threatens their ability to pay future
benefits. In the future, plan managers and legislative bodies must adopt
more modest return assumptions that properly reflect the composition and
volatility of their portfolios. Failure to do so will threaten public pensions’
ability to pay out benefits when retirees need them most.

Robert Prince, Co-Chief Investment Officer of Bridgewater
Associates, stated, “Because retirement outcomes unfold slowly over
decades, emerging problems are very hard to see and are virtually
unchangeable when they occur.” Unrealistic investment return assumptions
have allowed states and municipalities to obfuscate the scale of this crisis.
Unfortunately, the financial structure of public pensions makes it easy to
promise future retirement benefits and force the next generation of
politicians, taxpayers, and retirees to deal with it.
If the trend of mounting unfunded obligations is not corrected, its
effects will reverberate across the economy. Municipal bonds will carry a
higher default risk premium as investors question their soundness, raising
government borrowing costs in the form of higher interest rates. Politicians
and state retirement administrators will slash benefits or drastically increase
contribution levels to narrow the funding gap. The federal government,
its ين facing large deficits, may attempt to orchestrate a bailout to preserve
retirement benefits. And taxpayers, retirees, and public employees across
the country will pay the price.

Public pensions have never played a central role in a financial panic.
Unlike an asset price collapse or the failure of a systemically important
financial institution, they are unlikely to be the cause of the next recession.
However, they could accelerate a deterioration in economic conditions as
stakeholders realize that promised benefits don’t really exist, and, in fact,
started disappearing almost two decades ago.
Appendix A: Investment Return Assumptions Trend

We produced a line of best fit to determine the trend in investment assumptions used over the life of this data set. We used a linear model in order to plot the overall downward trend of investment assumptions, rather than overfit the trend to the data. Our linear trend output is summarized below, with investment return assumptions as a function of time:

Coefficients:

(Intercept)    year
0.8103835    -0.0003645

The equation is given by:

\[
\text{Investment Return Assumption} = 0.8104 - (0.0003645 \times Year) + e
\]

Estimates of future mean investment return assumptions used by public pensions are summarized in the table below:
Appendix B: Measuring the Difference between Expected and Realized Investment Returns

We conducted several t-tests to determine whether the difference between expected and realized investment returns is a statistically significant phenomenon. The t-test assumes that the tested data follow a normal distribution. To determine if our data met these criteria, we plotted both one-year realized returns and assumed returns from 2001-2017 as histograms.

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>7.52%</td>
</tr>
<tr>
<td>2018</td>
<td>7.48%</td>
</tr>
<tr>
<td>2019</td>
<td>7.44%</td>
</tr>
<tr>
<td>2020</td>
<td>7.40%</td>
</tr>
<tr>
<td>2021</td>
<td>7.37%</td>
</tr>
<tr>
<td>2022</td>
<td>7.33%</td>
</tr>
<tr>
<td>2023</td>
<td>7.30%</td>
</tr>
<tr>
<td>2024</td>
<td>7.26%</td>
</tr>
<tr>
<td>2025</td>
<td>7.22%</td>
</tr>
</tbody>
</table>
Both variables are normally distributed. Notably, the realized return distribution is far *flatter* than the return assumption distribution, indicating that realized returns are far more volatile than the assumed return would indicate. Assumed returns are clustered around a mean of 7.81%, while the
one-year realized return distribution has a mean of 6.52%. Our t-test answers whether this underperformance is statistically significant at a 95% confidence interval. Because we are determining the difference between a sample of values (one-year realized returns) and a target value (assumed return), we used a one sample t-test. Our null and alternative hypotheses are:

\[ H_0: \mu = m_0 \]

\[ H_1: \mu < m_0 \]

The null hypothesis states that realized returns approximately equal their benchmarks, and that no chronic underperformance exists. The alternative hypothesis states that average realized returns are less than the mean return assumption at the 95% confidence interval.

Our output is summarized below:

One Sample t-test (1-year return)

data: comparison$return_1yr
\[ t = -6.4764, \text{ df } = 2973, \text{ p-value } = 5.476 \times 10^{-11} \]

alternative hypothesis: true mean is less than 0.07805627

95 percent confidence interval:

\(-\infty, 0.0685018\)

sample estimates:

mean of x

0.06524771

Our output indicates a p-value of approximately zero, much smaller than = 0.05. Thus, we must reject the null hypothesis and conclude that there exists a statistically significant difference between expected and realized returns for public pension plans from 2001-2017. Specifically, the mean one-year realized return is significantly less than the assumed return over this period.

We conducted similar analyses using average three- and five-year realized returns. If anything, we would suspect these outputs to have higher p-values and be more likely to fail to reject the null hypotheses, as return assumptions are meant to reflect longer-term expectations. The output for the t-test using three-year realized returns follows:
One Sample t-test (3-year return)

data: comparison$return_3yr

t = -12.88, df = 2367, p-value < 2.2e-16

alternative hypothesis: true mean is less than 0.07805627

95 percent confidence interval:
   -Inf 0.065028

sample estimates:

mean of x

0.06311972

In this case, as with one-year realized returns, we must reject the null hypothesis and conclude that three-year average realized returns are statistically significantly less than the average benchmark used.
The output for the t-test using the five-year average realized return follows:

One Sample t-test (5-year return)

data: comparison$return_5yr
t = -18.344, df = 2780, p-value < 2.2e-16
alternative hypothesis: true mean is less than 0.07805627

95 percent confidence interval:

-Inf 0.06599956

sample estimates:

mean of x

0.06481156

Like the previous two tests, we must reject the null hypothesis and conclude that five-year average realized returns are also statistically significantly less than the average benchmark used.
Thus, whether using one-, three-, or five-year realized returns, we realize with 95% confidence that public pensions surveyed in the PPD underperformed their benchmarks over this period.

**Appendix C: Regressing Funding Status on Investment Performance and Accuracy of Investment Return Assumptions**

We sought to analyze the effectiveness of using realized investment returns to predict funded status. To do this, we regressed funded status (total assets divided by total liabilities expressed as a percentage) on five-year average investment returns. Our regression analysis using five-year average investment returns as the independent variable and funded status as the dependent variable is given below:

```
Call:

lm(formula = comparison$funded_status ~ comparison$return_5yr_percent)
```

Residuals:

```
  Min   1Q Median   3Q   Max
-51.32 -10.84  0.77 11.41 36.75
```
Coefficients:

|                      | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 53.4089  | 6.9336     | 7.703   | 1.81e-12 *** |
| return_5yr_percent   | 2.6491   | 0.9532     | 2.779   | 0.00616 ** |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.75 on 147 degrees of freedom

(29 observations deleted due to missingness)

Multiple R-squared: 0.04992, Adjusted R-squared: 0.04346

F-statistic: 7.724 on 1 and 147 DF,  p-value: 0.006161

Our findings are significant at the 95% confidence level.

In determining the relationship between accuracy of investment return assumptions and funded status, we used the difference between investment return assumptions and five-year average realized investment returns as our independent variable and funded status as our dependent variable. Our regression output is given below:
Call:

```
lm(formula = comparison$funded_status ~ comparison$difference_5yr_percent)
```

Residuals:

```
     Min 1Q Median 3Q Max
-54.740 -7.309 0.615 10.103 39.314
```

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 76.212      2.076  36.706  <2e-16 ***  
comparison$difference_5yr_percent -3.616      1.514  -2.388  0.0182 *  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 15.85 on 147 degrees of freedom

(29 observations deleted due to missingness)

Multiple R-squared:  0.03736, Adjusted R-squared:  0.03081

F-statistic: 5.705 on 1 and 147 DF,  p-value: 0.01819

Our findings are significant at the 95% confidence level.
References


Impacts of Data Analysis Pitfalls on Decision-Making

Sophie Lin

Foreword by Research Mentor, Dr. Courtney Paulson

Although businesses have been collecting and analyzing data to improve their processes for millennia, the dramatic rise in computing power over the past few decades has led to an explosion in the use of data and analytics at all levels of business. Previously this type of analysis was solely the domain of the Googles and Amazons of the business world, but now every company from a mid-size pizza chain to the local sports team is hiring teams of analysts just to compete.

While this represents a significant shift in the way businesses operate, most CEOs view data and data-driven decision-making as a necessary and positive shift. Companies are now collecting more data than ever before, and they are applying this data to build reliable models to inform and justify their choices. Because of this, many people are turning to online data science courses or self-directed Internet searches to explore resources to build increasingly complex data models.

As a budding data analyst from a business background, Sophie Lin approaches this phenomenon from a different perspective: just how dangerous is a little knowledge when it comes to making data-driven
decisions? Sophie’s passion for investigating the role of analytics in modern business shines through her work as she explores some of the most common errors that even well-meaning data analysts can make without realizing. As data analytics continues to grow, these errors can have wide-ranging repercussions for businesses’ bottom lines. Sophie not only examines these issues and their potential effects, but she also provides actionable insights for data scientists to help guide them through their analyses and arrive at rigorous, data-driven conclusions.

Sophie Lin is a SURE Fellow in the Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: lin.sophie7@gmail.com.

Abstract

Numbers don’t always speak for themselves. Decisions that CEO’s, executives, and managers make when introducing new ideas to the marketplace are data-driven, combined with business intuition. However, even when data results suggest one outcome, the implementation may show different results. This research will outline different ways data analysis pitfalls may affect analytics-based decision-making. The data analysis process is broken down into distinct steps, with in-depth discussion of the data methodology errors that can occur at each step. The paper also dives into the four types of data analytics and the industries that utilize them the most to make decisions. Finally, it explores case studies of data errors and suggests a method for more informed data-driven decision-making.
Keywords: business analytics, big data, decision making, pitfalls

I. Introduction

The 21st century saw a shift from the industrial revolution to a society focused on data and information technology. Businesses raced to integrate new information systems and advance the automation of business processes. “Big data,” “information technology,” and “analytics” became buzzwords that drove corporate conversations. Most notably, the influx of speculative Internet investments in the late 1990’s led to the burst of the dot-com bubble, a financial crisis that lasted approximately two years.

When the Information Age developed, decision-making cultures shifted from opinionated solutions to data-driven decisions. Time and time again, companies have shown improved performance when they effectively utilize the vast new flows of information available to them. In their paper *Big Data: the Management Revolution* Harvard researchers show that companies in the top third of their industry that use data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors (McAfee & Brynjolfsson, 2012). In 2001, a US airline conducted a data study which found that 10% of its flights had at least a 10-
minute gap between estimated versus actual arrival time, and 30% had at least a 5-minute gap. To fix this discrepancy, they implemented a new arrival estimate service (RightETA), which gathered mass amounts of data to generate more accurate airplane arrival times. The RightETA system gathered data about weather forecasts, flight times, and other proprietary information every 4.6 seconds to assist with the system’s pattern matching abilities. Using big data gathered over 10 years, RightETA has the ability to draw from past scenarios similar to a current flight in order to estimate highly accurate arrival times (McAfee & Brynjolfsson, 2012).

As big data becomes harder to handle, demand for data scientists is increasing exponentially. In their Harvard Business Review article, Davenport and Patil (2012) state data scientists are the “people who can coax treasure out of messy, unstructured data,” painting them as explorers with the power to discover key insights that drive society forward (Davenport & Patil, 2012). Particularly, data scientists are able to derive conclusions from data that help business leaders create solutions to their business problems. In selecting bright data scientists, researchers encourage hiring managers to look for people who can create stories and form “coherent insight[s]” when presented with one or many large datasets (Davenport & Patil, 2012). As of 2012, 2.5 exabytes of data are created every day, with the number doubling every 40 months (McAfee &
Brynjolfsson, 2012). As data advancements rapidly increase, a position as a data scientist has become the new hot job in the market. Successful data scientists can bridge the gap between those who are detached from the business world and those who are too fully immersed in business.

In this paper, we will discuss the pitfalls that can occur in the data experimentation and reporting process and how incorrect methods can affect data science results, and explore the decisions managers make consequently. In Section 2, we will explore the current decision-making landscape that employs big data and define pitfalls that can occur. Section 3 will develop the research methodology and define and connect the different types of analytics with the data experimentation and reporting process. In Section 4, we will dive into case analyses and real-world examples of data pitfalls that have affected the respective types of analytics.

II. Literature Review

Back when data was scarce, it was simple to turn to the most knowledgeable or experienced person in the room for answers. Now, the rise of the data revolution has allowed businesses to place more weight on data-backed solutions instead of the Highest Paid Person’s Opinion (HiPPO) (McAfee & Bryjolfsson, 2012). Businesses are asking themselves, “What do we know?” instead of “What do we think?”
However, there are two sides to the story. Along with the powerful advantage big data can provide, is the increased potential for pitfalls that may occur in business experimentation. In this literature review, we will identify and define the errors and pitfalls that arise during stages of the data analysis process. Real-world examples will be expanded upon in Section 4.

If data analysis is the key to decision-making, why do some data-backed solutions lead to failed business decisions? In an ideal scenario, a manager observes a set of significant data and subsequently implements a decision that leads to successful results. However, in the case of J.C. Penney, the correct data-driven decision consequently led to plummeting sales, increase in losses, and the firing of a new CEO. Harvard researchers Thomke and Manzi (2014) spent more than 40 years exploring improper business experimentations that led to greater costs for the respective companies. When introducing a new business idea or product, big data is largely used to look at customers’ past behaviors, when it should be used to predict future customer responses (Thomke & Manzi, 2014). Business experiments should follow the same rigorous experimental steps as the testing of a new drug or hypothesis. Because business experiments fall outside the traditional scientific domains doesn’t mean they shouldn’t follow the traditional scientific process.
Hence, we developed a checklist for proper business experimentation:

1. Determine the purpose of the experiment.

2. What is the buy-in? What changes will be implemented as a result of the experiment conclusions?

3. Is the experiment feasible?

4. Are the results reliable? What data errors could have occurred?

5. What is the best way to implement the experiment’s results so that it provides maximum value to the firm?

It’s always important to have the consumer in mind. In following these steps, companies can hope to avoid negative consumer reaction to ideas that were positively supported by data analysis.

**Errors and Pitfalls**

Big data cannot be a substitute for traditional data collection and analysis. The correct comprehension of all this data is just as important as analyzing it. For example, overfitting is an element of big data hubris that commonly occurs when analysts are too focused on big data methods and ultimately get non-generalizable results. When testing the results of a computer vision model, Sun et al. (2017) found that the performance of the model increased logarithmically with a larger size dataset. This means that
every additional piece of data would decrease the accuracy of the measured performance by 1/X (Sun, Shrivastava, Singh, Gupta, 2017). In Section 4, we will dive deeper into how this error caused a technology giant like Google to have such drastically different results in their algorithm’s flu predictions.

Even when data collection and analysis are performed correctly, the results may not always be communicated properly to the recipient of the data. For example, the principal-agent theory presents itself as an interesting obstacle when implementing data-driven decisions in education (Wohlstetter, Datnow, Park, 2008). In the principal-agent theory, the principal lacks time or expertise to do a task, therefore delegating it to the agent to complete. In education, the Board of Education (the principal) creates new ideas supported by data analysis, which they pass down to individual schools (the agent) to implement. Earl and Katz (2006) found that, in spite of system-level investments for decision making in education, teachers are not actively using data to guide planning and instructional decisions. Section 4 will further explore the ways data can be received through a fictitious chocolate study that went viral in media outlets.

Another issue seen with data-driven analysis is Simpson’s paradox, a statistical phenomenon in which a conclusion drawn from a subset of data may contradict a conclusion drawn from the overall dataset. Simpson’s
paradox always involves a third variable (the “lurking variable”) that is hidden when data analyses are performed, therefore misconstruing data conclusions (Simpson, 1951). Berman et al. (2012) show that this data anomaly has shown up in medicine, education, human resources hiring, and even e-marketing research.

But how can a statistical anomaly apply to business decision making? Simpson’s paradox can be a challenging situation in which decision-makers are presented with two equally valid and statistically correct conclusions that contradict each other. In their paper “Normative and Descriptive Analysis of Simpson’s Paradox in Decision Making” Shawn Curley and Glenn Browne (2001) ask the question, “When presented with a Simpson’s Paradox problem, should managers accept the aggregate analysis or subset conclusion?” Curley and Browne (2001) designed a variety of experiments to test respondents’ reactions to Simpson’s paradox cases. They found that Hiring and Production Tasks studies exhibited decisions more suggestive of aggregation. However, after accounting for different scenarios in Simpson’s paradox, Curley and Browne actually had inconclusive results. They encouraged future decision-makers to examine all angles of both sets of data and base their decisions on subjective analysis after the fact. Additionally, Berman et al. (2012) explores three real-life cases that reveal multivariate relationships that
indirectly affect statistical results. Unveiling these “lurking variables” can point to a clearer path for decision-making. Further analysis in Section 4 will propose ways to make decisions when faced with a Simpson’s paradox anomaly.

III. Methodology

Guiding questions for the research were:

1. What errors can occur in the data experimenting and reporting process?
2. What costs have resulted from incorrect data-driven decisions?
3. How can we mitigate these errors?

The research was broken into three areas of study:

- Steps in the data analysis process
- Types of analytics and what decision-making questions they address
- Pitfalls that can affect each type of analytics (to be addressed in the Case Analysis and Discussion section, Section 4)

Data Experimentation & Reporting Process

In the data experimentation process, we identify three main steps:

1. Gathering Data
2. Analyzing Data

3. Interpreting Data

After finalizing the analysis, analysts proceed to the additional concluding steps:

4. Publishing Data

5. Reception of Data

Steps in the process build on each other. If there is an error in an earlier step, it can have a cascading effect on later steps. More important, the reception of data is where decision-making mistakes frequently take place and can cause the most harm (to be addressed in Section 4).

Types of Data Analytics

Data analytics is usually broken down into four types: descriptive, diagnostic, predictive, and prescriptive analytics. The analytics increase in complexity with the type of data they describe and the questions they serve to answer.

Descriptive analytics answer the question, “What happened or is currently happening?” As the name suggests, these numbers or results describe the data and include key metrics derived from the raw data. Mean, median, mode, range, variance, and standard deviations from a data set are
all classified as descriptive analytics. Diagnostic analytics answer the question, “Why did something happen?” They focus on identifying patterns within datasets. The historical data gathered in descriptive analytics can be compared to other sets of data to look for common or irregular behaviors amongst variables. Predictive analytics answer the question, “What will likely happen in the future?” It uses both descriptive and diagnostic data to predict future trends to be used for forecasting. Prescriptive analytics are a combination of the other three analytics and answer the most important question, “What actions should be taken?” This is where data-driven decision making is commonly found. Prescriptive analytics focus on how to mitigate risks or take advantage of trends that were identified from predictive analytics. They are the most vulnerable to errors because they rely on all three other analytics types.

In integrating the four types of analytics with the first three components of the data analysis process, we create a model to illustrate how steps of the data analysis process can encompass different types of data analytics (see Figure 1).
First, decisions made through descriptive analytics lie at the intersection of data collection and data interpretation. This is the simplest type of analytics and involves the process of gathering data and interpreting the data through number summaries, in order to answer surface-level questions. Generally, no deeper analysis is performed in descriptive analytics. Second, diagnostic analytics-based decisions lie at the intersection of data collection and data analysis. Diagnostic decisions rely on deeper analysis of collected data. Decision-makers dig through collected data to find diagnostic results to explain why an occurrence happened. Third, the intersection of data analysis and data interpretation encompasses decisions that are made using predictive analytics. As previously noted, predictive analytics are generated to determine what will likely happen in the future. This comes from heavy analysis, creation of predictive models, and interpretation of models. Lastly, at the heart of all three steps of the
data analysis process lies prescriptive analytics, analytics that draw upon everything to make a final decision.

Although this framework demonstrates the overarching ideas and connections in data analytics, it is important to note the use of analytics is rarely this clean in practice. Because analytic types can build off each other, they are not mutually exclusive nor independent. Examples of pitfalls may bleed from one type of analytical decision into another, as we will explore in Section 4. Furthermore, the data analysis process is non-linear, as steps can go back and forth between data analysis, back to more data collection, and/or data interpretation, and back to data analysis.

IV. Case Analysis and Discussion

This section will explore real-world examples of data pitfalls and the impacts they had, or could have. Data collected from PricewaterhouseCooper’s 2016 Global Data and Analytics Survey is used to map which industries are most affected by which of the four types of analytics previously defined in Section 3 (PwC, 2016).
Flaw of Averages (FOA)

Type of Data Analytics Affected: Descriptive Analytics

Top Industries Using Descriptive Analytics in Decisions:

Communications, Government, Insurance (PwC, 2016)

Case Discussion:

Often, using only a fixed average to represent a set of data fails to consider the distribution of risk when making decisions. In 1994, Orange County, CA made the mistake of considering only an average interest rate when strategizing how to invest their Orange County Investment Pool funds. Although the portfolio was successful in previous years, the
volatility in the Federal Reserve’s interest rates in 1994 was an indicator that simple averages should never be used in determining risk (Savage, 2002). If county treasurer Robert L. Citron had considered the wide distribution of possible interest rates, he would have discovered there was a 5% probability of the fund losing over $1 billion. However, Citron failed to take that into account, and his assumption of an “average” interest rate cost Orange County $1.6 billion in losses, forcing the county to go into bankruptcy. Although there were many factors that caused the crisis, Citron’s key mistake was assuming a fixed average interest rate on strategies for the portfolio, without factoring in the volatility of those rates.

In a society overwhelmed with data, it can be easy to jump to quick conclusions. Especially in business, executives have the tendency to “plug in an average figure” to calculate their “average” bottom line (Savage, 2002). The prevalence of this trap has pushed even Generally Accepted Accounting Principles (GAAP) to require specific data inputs that eliminate uncertainties in the form of “averages.” David B. Hertz’s Monte Carlo simulation software is now widely used in a variety of industries (especially finance) to display the distribution of outcomes and their respective risk levels (Joseph, 2018). Increasingly, executives are considering a distribution of numbers, rather than a single-number summary, to get a more complete picture of data results before making decisions.
Overfitting

**Type of Data Analytics Affected:** Predictive Analytics

**Top Industries Using Predictive Analytics in Decisions:** Banking & Capital Markets; Asset Management; Insurance (PwC, 2016)

**Case Discussion:**

In 2013, Google Flu Trends incorrectly predicted two times the number of doctors’ visits caused by flu compared to the Center for Disease Control & Prevention (CDC). Lazer et al. (2014) describes the problem that occurred with Google Flu Trends (GFT) and its highly inaccurate flu predictions, caused in part by a larger issue called “big data hubris.”

Big data hubris occurs when analysts believe the greater the quantity of data, the more closely fit the results will be. Rather than using big data as a supplement to traditional data collection and analysis, these analysts are using it as a substitute. The problem is that in using a large quantity of data to fit specific predictors, the response variable becomes erroneously inflated. In the case of GFT predictions, Google used 50 million search terms to match only 1152 data points. After the mistake was discovered, Google Flu Trend developers re-analyzed and weeded out seasonal search
terms unrelated to flu, but strongly correlated to CDC data to get a more accurate representation of the true trends.

When information giants like Google knowingly collect exceedingly large amounts of data, they must be careful to use their analyses appropriately to ensure that results are generalizable to new data, rather than being too overfit to a particular small dataset. If possible, to minimize the risk of erroneous conclusions, analysts should try to hold out test data that generalizes to new, unseen data.

**Selective Testing**

**Type of Data Analytics Affected:** Predictive Analytics

**Top Industries Using Predictive Analytics in Decisions:** Banking & Capital Markets; Asset Management; Insurance

**Case Discussion:**

In 1994, the *Motley Fool* released an investment strategy, “Foolish Four,” which, from 1973 to 1993, had a reported portfolio average annual return of 25.5%. This was compared to an average of 11.2% annual return in the Dow 30 for the same time period. However, in their paper “Mining Fool’s Gold,” McQueen and Thorley (1999) discovered data mining pitfalls
that occurred during the creation of the Motley Fool’s Foolish Four strategy that caused the portfolio to display such highly inflated returns. This method of creation was arbitrary and based off back-testing historical data in multiple ways. The random steps established in constructing the portfolio strategy lacked coherent financial theory justifications. Some rules that contributed to the strategy were enforced only because they displayed a significance in returns after multiple other failed attempts. (For a more detailed description of the Foolish Four strategy, consult Appendix A).

Analysts dealing with strongly-correlated historical data like these should be careful to validate and justify their conclusions and recommendations. Creating portfolio strategies to back test financial data can be prone to analyst biases. One measure against such pitfalls is to test the portfolio strategy outside of the regularly tested data sample, to see if return results are still successful (McQueen and Thorley 1999). If analysts do not scrutinize the results of their strategies, financial institutions can leave themselves open to exploitation.

**P-hacking: “Statistical Significance”**

**Type of Data Analytics Affected:** Prescriptive Analytics

**Top Industries Using Prescriptive Analytics in Decisions:** Technology; Retail & Consumer (PwC, 2016)
Case Discussion:

In 2015, Harvard biologist and science journalist John Bohannon, along with researchers Diana Koch, Peter Homm, and Alexander Driehaus, created a fictitious experiment ripe with data analysis errors that claimed to prove the positive health benefit that eating chocolate has on weight loss. Bohannon’s goal was to measure how widely erroneous research findings can spread in the diet and journalism industry and impacts that might result. Bohannon did this by exploiting a commonly misunderstood phenomenon in statistics reporting: hacking results to get significant p-values (called “p-hacking”).

P-hacking is a serious issue in data science because of the importance the industry places on experiments showing “significant” p-values, a statistic which supposedly indicates that a result is analytically valid. Bohannon et al. (2015) illustrate the probability of a successful experiment with a “winning” p-value that is < 0.05, given n predictors and p level of significance, is

\[ P(\text{winning}) = 1 - (1 - p)^n \]

In his experiment, Bohannon et al. (2015) tested only 18 predictors. With a significance level of 0.05, this would come out to a 60% chance of a significant result (with a p-value < 0.05). If Bohannon had tested 100
predictors, the probability of finding a significant result would be 99.4%, almost a guarantee that some predictor will be significant. Bohannon used this to run a faulty study that eventually published in credible diet and health journals all over the world.

Although this chocolate case purposely used p-hacking, it is a data pitfall that researchers unknowingly practice when attempting to find significance in research and publish “meaningful” results. Bohannon’s chocolate hoax not only proved that p-hacking is an influential data issue, but exposed the lack of fact-checking and credibility in the diet industry. The ability for false information to spread so quickly and impact an entire marketplace serves as a lesson on the need for awareness in the data experimentation process and the decisions that result.

Any business or analyst team who wants to make decisions based on data experiments or rigorous statistical modeling should be wary of p-hacking. Statisticians are slowly moving away from using p-value as the holy grail method of determining significance due to its increasingly misleading results (Nuzzo, 2014). In addition, any business executive interpreting or using the results of analyses from a p-value significance benchmark should question the results to ensure the analysis is robust. If the analysis cannot stand up to scrutiny, analysts need to go back to the
drawing board and either collect more data or design a more appropriate analysis.

Simpson’s Paradox: Exploration of a Multi-Type Statistical Anomaly

Although Simpson’s paradox is an anomaly, not an error, the lack of consideration of this issue can nevertheless affect decision-making, thus warranting a more complete case discussion from multiple perspectives. It is important to note that due to the complexity of Simpson’s paradox, it can present itself in all types of data analytics, thus having an effect in all industries.

Type of Data Analytics Affected: Descriptive Analytics

Top Industries Using Descriptive Analytics in Decisions:
Communications; Government; Insurance (PwC, 2016)

Case I Discussion:

Perhaps the most famous case of Simpson’s paradox is in a University of California-Berkeley case, in which the university was sued for seemingly favoring males over females in graduate school admissions (Bickel, Hammel & O’Connell, 1975). Upon closer look, Bickel et al.
(1975) discovered a paradox (Figure 3). Results from the aggregate data of university-wide acceptance showed a positive difference between the Observed and Expected admit rates of men and a negative difference between the Observed and Expected admit rates of women. According to the aggregate data, more men than women were actually admitted compared to predictions of their admission rates.

*University-wide Admissions*

<table>
<thead>
<tr>
<th>Applicants</th>
<th>Observed</th>
<th>Expected</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Admit</td>
<td>Deny</td>
<td>Admit</td>
</tr>
<tr>
<td>Men</td>
<td>3738</td>
<td>4704</td>
<td>3460.7</td>
</tr>
<tr>
<td>Women</td>
<td>1494</td>
<td>2827</td>
<td>1771.3</td>
</tr>
</tbody>
</table>

*N = 12,763*

Fig. 3: (Bickel, Hammel & O'Connell, 1975)

However, consider the subsets of data which group Male vs. Female admissions by different Berkeley colleges in Figure 4. There was no
difference between the Observed and Expected admit rates of either men or women.

Departmental Admissions

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Applicants</th>
<th>Observed</th>
<th>Expected</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Admit</td>
<td>Deny</td>
<td>Admit</td>
</tr>
<tr>
<td>Admit</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Deny</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Department of Mathematics

Men

<table>
<thead>
<tr>
<th>Applicants</th>
<th>Observed</th>
<th>Expected</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Admit</td>
<td>Deny</td>
<td>Admit</td>
</tr>
<tr>
<td>Men</td>
<td>200</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Women</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Totals</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
</tbody>
</table>

Department of social warfare

Men

<table>
<thead>
<tr>
<th>Applicants</th>
<th>Observed</th>
<th>Expected</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Admit</td>
<td>Deny</td>
<td>Admit</td>
</tr>
<tr>
<td>Men</td>
<td>50</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>Women</td>
<td>150</td>
<td>300</td>
<td>150</td>
</tr>
<tr>
<td>Totals</td>
<td>200</td>
<td>400</td>
<td>200</td>
</tr>
</tbody>
</table>

Sure Journal 2019 | 328
<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>229.2</td>
<td>270.8</td>
</tr>
<tr>
<td></td>
<td>320.8</td>
<td>379.2</td>
</tr>
<tr>
<td></td>
<td>20.8</td>
<td>-20.8</td>
</tr>
<tr>
<td></td>
<td>-20.8</td>
<td>20.8</td>
</tr>
</tbody>
</table>

Figure 4: (Bickel, Hammel & O’Connell, 1975)

A major consideration in Simpson’s paradox cases is confirmation bias. When presented with multiple justified results, analysts must be careful to not select the data that unjustly favors their hypothesis. The Berkeley case became an infamous lawsuit centered around discrimination in university admissions. It wasn’t until later that closer analysis of the subsets of data revealed the tricky situation with Simpson’s paradox. An awareness in our confirmation biases can play a part in challenging us to dig deeper beyond what the descriptive data shows.

**Type of Data Analytics Affected:** Diagnostic Analytics

**Top Industries Using Diagnostic in Decisions:** Technology, Retail & Consumer (PwC, 2016)

**Case II Discussion:**

In their article “Simpson’s Paradox: a cautionary tale in advanced analytics,” Berman et al. (2012) present a price optimization program that delivered conflicting conclusions for a cosmetic manufacturer. The
business sought to determine whether the optimization models implemented through the program improved profit margins for the business without negatively impacting profit from each of their three store brands.

Results from the optimization program showed that the average profit margin for the overall business decreased by 0.3%, which deemed the program unsuccessful. However, when the results were broken down into profit margin by brand, the data showed an increase in profit margin of 20% to 25% after implementing the optimization program (see Figure 5) (Berman, DalleMule, Greene, & Lucker, 2012).

<table>
<thead>
<tr>
<th>Store Brand</th>
<th>Control Period</th>
<th>Test Period</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand A</td>
<td>10</td>
<td>12</td>
<td>20%</td>
</tr>
<tr>
<td>Brand B</td>
<td>5</td>
<td>6</td>
<td>20%</td>
</tr>
<tr>
<td>Brand C</td>
<td>4</td>
<td>5</td>
<td>25%</td>
</tr>
</tbody>
</table>

Fig. 5 (Berman, DalleMule, Greene, & Lucker, 2012)
It is important to keep in mind the lurking variable in Simpson’s paradox and the notion that this statistical phenomenon occurs due to the differing weighted averages of each group (i.e. each brand). As seen in Figure 6, inclusion of a hidden variable, “Mix of Brands,” shifts the weighted average between brands, thus illustrating Simpson’s paradox (Berman, DalleMule, Greene, & Lucker, 2012). In this case, the aggregate set of data showed an ineffective optimization program for the business as a whole, while the subset of data showed an effective optimization program for each brand.

<table>
<thead>
<tr>
<th>Store</th>
<th>Number of Stores</th>
<th>Mix of Brands</th>
<th>Margin</th>
<th>Number of Stores</th>
<th>Mix of Brands</th>
<th>Margin</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>Brand A</td>
<td>400</td>
<td>57%</td>
<td>10%</td>
<td>200</td>
<td>29%</td>
<td>12%</td>
</tr>
</tbody>
</table>
Fig. 6 (Berman, DalleMule, Greene, & Lucker, 2012)

The difficulty in resolving Simpson’s paradox is that it produces two valid statistical results, thus creating multiple branching choices for making decisions. The key is to go back to the problem at hand and determine if the hidden variable has a direct effect on that problem. For example, this case’s main issue is, “Should we implement the optimization program?” Figure 6 shows the hidden variable, Mix of Brands, was not constant and shifted through the optimization test period, thus creating the paradoxical results. This shows that changes in the distribution of brands in each store, not the optimization program, was the contributing factor to decreased profit margins. In this scenario, managers should use the results of the subset of data showing that the program was effective for each individual brand, to make subsequent business decisions.
V. Conclusion

This research has explored the data experimentation and reporting process, differing types of analytics that guide decision-making, and the data pitfalls that can result. Further research would encompass creating scenarios with purposeful data pitfalls integrated and measuring results against a benchmark of perfect data experimentation. Similar to the chocolate case, controlled error experiments would be useful to survey participants and test if a data error has a specific effect on a respective type of decision-making. In the world of such big data, we often seek “quick and easy” ways to summarize problems and develop successful solutions. However, it is important to be cautious of simple number summaries and realize that data results are not always what they seem. Analysts are responsible for presenting clear, well-reasoned analysis, but to create meaningful data-driven decisions that strengthen business strategies, business managers must be careful to challenge analysts’ conclusions.

Acknowledgements

I would like to thank my mentor, Dr. Courtney Paulson, for her continued guidance on this paper, and SURE program director, Dr. Christina Elson, as well as the SURE co-directors for their support of this project.
Appendix A

Foolish Four strategy

a) Dow Dividend: Dow Jones Industrial Average (DJIA) stocks sorted by dividend yields at the beginning of the year.

a. Step 1: Purchase an equally weighted portfolio of 10 highest-yielding stocks

b) Dow Five: Sort the 10 Dow Dividend stocks by price

a. Step 2: Purchase an equally weighted portfolio of 5 lowest priced stocks

c) Dow Four: Sort the Dow Five by price

a. Step 3: Drop the lowest priced stock

d) Foolish Four: Sort the Dow Four by price

a. Step 4: Eliminate the lowest priced stock and double up on the second-to-lowest priced stock
References


Hydroponics and Vertical Farming: A Study on the Multi-Level Perspective of Transitional Sciences in Global Agriculture

Eunisa Lu

Foreword by Research Mentor, Dr. Jose-Luis Izursa

When Eunisa and I met to discuss her research, it was clear her great interest about hydroponics as an alternative for sustainable farming. We know that due to urbanization and industrial development, we have been losing almost a third of our arable lands over the last 40 years and because the world’s population is constantly growing, we also know that feeding all the people is becoming a huge challenge. As many other people, we believe that unconventional agricultural practices using Controlled Environment Agriculture (CEA) technology such as hydroponics, aquaponics or even aeroponics can be the answer to this challenge.

Eunisa Lu examines in her research why hydroponics has been widely utilized in some countries overseas but did not find success in the United States. She came with the good idea to use the Multi-Level Perspective theory to understand and explain how hydroponics as an agricultural innovation can grow in the United States. Eunisa was able to produce an informative report on how some places in the United States may become good candidates to extensively use CEA. Working with Eunisa has been a
pleasure. I hope you enjoy reading this article that summarizes Eunisa’s findings.

Eunisa Lu is a SURE Fellow in the Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: eunisalu@gmail.com

Abstract

This study uses past research as a reference and looks into why hydroponics has not been successfully implemented in the agricultural industry. It also examines the best places to implement hydroponics and vertical farming into the United States, where it would have the largest impacts with the smallest displacement of current industries. First, a literature review studies why hydroponics has become so successful and vital to countries such as the Netherlands and China and compares the state of their agriculture industries with that of the United States. The study then looks further into suitable places in the United States, particularly areas that have long suffered from drought and irrigation problems, and at any recent agricultural or environmental policy that might hint at possible support for implementation. A holistic understanding of the sociopolitical factors that go into the adoption of disruptive innovations in areas such as agriculture, where there is heavy government involvement, can help unlock solutions for better and more efficient use of innovation.
Keywords: hydroponics, agriculture, vertical farming, Netherlands, China, United States, drought, irrigation, environmental policy, sociopolitical

Introduction

Although modern hydroponics systems weren’t invented until the late 20th century, growing crops without soil is not a completely new concept. One of the first instances of hydroponics was the Hanging Gardens of Babylon, in 600 B.C. Believed to have been built by Emperor Nebuchadnezzar, the garden was fueled by a continuous stream of water pulled from the Euphrates River and dispensed through a chain-pull system. Because Babylon was known for having a dry climate, experiencing little rainfall, this water-fed system was rather extraordinary. Modern hydroponics systems were perfected through various scientific experiments conducted through the 17th and 18th centuries. Scientists slowly learned about plant growth and its constituents, and finally in 1860, two German botanists found a method of dissolving nutrients in water and growing plants in that water. This was the origin of “nutriculture,” which identifies “nitrogen (N), phosphorus (P), potassium (K), magnesium (Mg), sulfur (S) and calcium (Ca)” as macro-nutrients, or elements required in the largest amounts for plants to grow (N). The earliest known implementation of
hydroponics was in the United States by the military during World War II. Several hydroponics systems were set up throughout the Pacific islands to provide fresh produce to soldiers. The widespread production of plastic in the 1970s spurred the development of the large PVC hydroponics systems seen today.

In recent decades, there has been a push for more sustainability in various industrial sectors, especially energy and urban planning. The food industry also has come under scrutiny, but much of the criticism has been directed toward the meat industry, because of the amount of waste it produces each year and for issues such as animal cruelty. Hydroponics has found its way into modern agricultural practices in the form of vertical farming. Vertical farming is the practice of stacking production of crops in layers, best accomplished through hydroponics, which is a method of growing plants without soil by allowing them to be submerged in constant running water. Hydroponics is praised as a greener alternative to traditional farming, as it requires 90% less water and less use of pesticides.

People with limited resources, particularly space, can have difficulty cultivating produce of their own. The majority of people are solely dependent on external marketers for their food, a situation which can pose a challenge for those wishing to know where their food came from (e.g. local or not) and how it was processed (e.g. with or without pesticides). In
addition to this, modern agriculture can be quite taxing on natural resources, particularly water and space, with approximately 51% of the United States’ land area dedicated to agricultural practices as of 2007 (Nickerson). Conventional agriculture, wherein crops are grown from soil, actually uses more water and far more space than hydroponic methods, wherein plants are grown in a nutrient solution (Barbosa et al.). Nevertheless, conventional methods remain dominant in the agricultural sector.

Vertical farming techniques, such as hydroponics and aquaponics, are becoming more relevant every day, thanks to a push for sustainable development around the world. Despite this attention, hydroponics has not found success in the United States and remains largely a novel idea with little implementation. It has, however, done very well in other agricultural powerhouses, such as the Netherlands and China. This research aims to analyze the factors of success hydroponics has had in these other countries and questions whether it is possible to implement on a larger scale in the United States. This study asks:

1. Why has hydroponics done so well overseas, and what economic, political, geographical, and technological factors have been key to this success?
2. Why has hydroponics failed to be as viable in the United States?

3. In what sectors of the food industry can hydroponics reasonably be implemented to access the most benefits?

Definitions

Hydroponics

Hydroponics is a type of hydroculture; a system that grows plants without the use of soil. Typically, plants are grown in water that contains dissolved nutrients.

Vertical Farming

Vertical farming, or VF, is a system in which plants, or even animals, are cultivated in a way that saves the most space. This is done by growing produce vertically in layers, along vertically inclined surfaces, or within buildings and other edifices that can be integrated into structures.

Multi-Level Perspective Theory (MLP)

This theory defines the transition process of technological innovations into society. The MLP theory differentiates three levels in transition processes: (1) the niche level, which involves a network of new technologies in the incubation or nurturing stage; (2) the regime level, which represents the engine of existing energy systems as it includes the institutions,
associations, and governance structures that exist within the dominant energy system; (3) the landscape level, including various external pressures that affect transition processes, such as oil price shocks. (Kern)

**Background**

Before conducting a literature review, it is necessary to understand exactly what a hydroponic system is, as well as the science behind modern-day industrial hydroponics. Simply put, hydroponics is a way of growing plants without the use of soil. Any hydroponics system will require the same elements that a soil-based plant culture does. There needs to be light, air, a medium for the plant, water, and certain nutrients. There are many ways to use water culture systems. Six unique techniques have been developed over the past few years. Of these 6 systems -- Wick, Deep Water Culture, Nutrient Film, Ebb and Flow, Aeroponics, Drip -- the most popular in commercial vertical farming is the drip system. The drip system is not used recreationally as they are relatively simple to operate on a large, industrial scale, but quite overbearing for smaller gardens. A reservoir of nutrients, with an air pump that constantly moves the solution around, sits under the plants. Another pump continuously sends nutrients up to the plants. This is called a drip system because the nutrient pump sits atop the top layer of medium and literally “drips” the nutrient solution onto the
plant. This pump is set to a timer, which makes it very easy to control the feeding and watering schedule.

![Figure 1: A drip hydroponics system (Singh, 2011)](image)

**Research Framework and Methodology**

**Theoretical Framework**

The theoretical framework of this study is the multi-level perspective, or MLP, theory of socio-technical transition, first coined by Professor Frank Geels, who described it as a heuristic model for looking at past technological transitions. Technological innovations occur constantly throughout history, and there are many models and frameworks that can be used to address them. The MLP is a popular model used to understand...
changes to socio-technical systems. Innovation systems are usually approached through the focus of production, but the MLP level combines science and technology and forces a more heuristic view of analyzing innovation through the societal functions it fulfills. It is necessary to use theories like MLP to develop a more sustainable future. Such existing systemic perspectives, approaches, and frameworks provide a good basis for developing an analytical framework for understanding the dimensions and dynamics involved in scaling processes (Wigboldus et al.) This technique is multi-dimensional and holistic, proposing three different levels that link innovation with success: the niche, regime, and landscape. MLP breaks down complex issues such as climate change and peak oil and their effects on entrenched technology, such as cars or radio. Simply put, MLP shows how technology is not a discrete entity. Rather, it is an operation interdependent on social, technical, and institutional forces within society.

Old technology is protected by these forces in what Geel describes as a regime. The regime level relates to the constellation or system of interacting practices and structures that have come to a certain relative stability and status quo. This may, for example, be the status quo in a sector. However, this stability may be disturbed (perturbed), e.g., as a result of new policies or changing environmental conditions (Wigboldus et al.)
There are three main actors in this regime: producers/suppliers, users, and infrastructure. Producers and suppliers of old technology are very possessive of their territory within this regime and do what they can to prevent disruptive technology from fully developing. In addition, financial institutions are wary of investing in disruptive technology because of the volatile nature of innovation.

Next is users. Humans are creatures of habit. People’s lifestyles, including certain expectations they have of their lifestyle, grow around old technology. Common, pervasive assumptions are developed on what certain technology should be. Culture and media contribute to this image, helping to prevent new technology from seamlessly blending into society. Even when people want to switch to new technology, old habits and routines connected to old technology make it difficult.

Finally, current infrastructure that has been built around old technology can makes it difficult for innovations to be successful. It is extremely expensive and difficult to change a built environment that has co-evolved with old technology. Just as it would be difficult to develop public transportation that is as convenient or widespread as highways, it is nearly impossible for all sorts of new technology to overcome structural difficulties to penetrate the market. In addition, public policy and regulation
that seek to provide certain industries create uneven playing fields between new and old technology.

How to change this regime are where the landscape and niche levels come in. Regimes will invariably come under pressure at the landscape level. When the landscape changes, outdated regimes need to adjust the way they interact with society. Some pressures that exist in the current landscape of the United States are increased awareness of climate change and environmental sustainability. These changes in the dominant values and beliefs create windows of opportunity for disruptive technology. They put pressure on existing regimes, such as energy, transportation, and agricultural industries. The technologies that can effectively take advantage of these windows of opportunity are typically niche innovation. Niches develop within regimes, in the protected spaces that are allowed by the regime. When the window of opportunity opens, they must be mature and ready to act, or they will not be successful. There are two options for niche technology in this case. The first is to replace regimes, such as when automobiles replaced horse-drawn carriages. The second is to be subsumed into the regime, so that they can reduce landscape pressure without completely toppling the regime. Even in the second case, niche technology eventually may be able to overcome the regime. The multi-level perspective is used mostly to analyze historical innovation, but it also can be applied to
the existing landscape to evaluate emerging niche technologies, such as hydroponics. Understanding society and technological transitions as a network of institutional factors can stimulate new ideas for a better future.

Fig 2: The MLP model for socio-technological innovation. This chart shows the three levels of technology, niche, regime, and landscape in order of size.

Methodology

This research utilizes the multi-level perspective, or MLP, model for socio-technological transitions in the context of the agriculture industry to understand how agricultural innovations, specifically hydroponics, can grow under the regime. The methodology of this study is primarily descriptive and includes defining and explaining the regime, landscape, and niche levels of agriculture in the United States.
To form conclusions on the nature of agriculture in the United States, it is necessary to first look at these factors separately. It compares these levels as they exist in the Netherlands and China, where hydroponics has become economically viable and an integral part of sustainable farming. This includes analyzing quantitative data and studies done on the geographical climate and the socio-political atmosphere of each country and comparing them to those of the United States. It then examines new opportunities created by windows in the regime where hydroponics could be implemented for success and sustainability. More specifically, the paper looks into drought problems in California as an opportunity for developing the hydroponic industry, combined with government support for environmental policy, opportunities for hydroponics in hemp and marijuana production, and the possibility of robotics…?

**Literature Review**

A study by doctoral student Austin Miller in 2011 titled “Scaling up or selling out? A critical appraisal of current developments in vertical farming” attempted to unpack the counter-hegemony discourse by proponents of vertical farming, as well as provide a holistic view of how various actors in agriculture attempt to navigate the dominant regime. It looked past the more normative challenges to ask more nuanced questions
about the compromise of political action and industrial hegemony. Rather than focusing on the technological impediments to hydroponics systems, the paper examined the delicate political balance among farmers, innovators, and government. It reveals aspects of political-social ecology and dives into why different actors will construct the nature of contemporary society in different ways. The study also highlights the disconnect experienced between the modern-day consumer and natural food systems. This is one of the first studies I read at that looked at the socio-political factors of vertical farming, rather than the technical aspects. It was important in leading me to find the proper theoretical framework to support my own research.

Another important study to consider is “Economic analysis of aquaponics and hydroponics production in the U.S. Midwest,” written by Professor Kwamena. Although this study focused more on the profitability of aquaponics in the Midwest, it provided a lot of insight into the nature of agriculture in the Midwest, an area highly dependent on their agriculture industry. The sources of data considered for this project came from three active aquaponics farms, a university greenhouse experiment, and other literature reviews. The two conclusions drawn from the study included two analyses comparing aquaponic profitability with hydroponic profitability. Hydroponics was deemed more profitable, with lower investment costs and
higher yields. The second analysis was on the optimal size for aquaponic farms, which was as large as possible to create a lower cost of production.

An article published in the journal *HortTechnology* introduces aquaponics, which combines hydroponic technology with aquaculture production of fish, creating a more sustainable agriculture system. Aquaponics, which uses the natural biological cycle to supply nitrogen to plants, is well suited to hydroponics and allows the system to minimize even more nonrenewable resources. The article explains the necessary components of aquaponics systems that allow it to successfully combine hydroponics and aquaculture. It also introduces not only the market for aquaponics but other hydroponics and vertical farming systems that compete with one another to become the most cutting-edge technology in the market.

One such competing system of farming is care farming. In a journal article published in *Sociologia Ruralia*, “Multifunctional Agriculture Meets Health Care: Applying the Multi-Level Transition Sciences Perspective to Care Farming in the Netherlands,” author Jan Hassink attempts to explain care farming through the MLP theory. It gave me a lot of insight into how to apply the MLP theory in my own research and successfully weave it throughout my paper. The theoretical framework under which a paper is
conducted must be ever-present throughout the data collection, analysis, and conclusion. This article helped me tie my different thoughts together.

Current studies of hydroponics vary widely and span many different social, political, and technical topics. Very few studies attempt to look at the topic holistically, to focus on the simplest state of agriculture in the United States and its relationship with new technology, such as hydroponics. This study attempts to provide a broad overview of the nature of agriculture in the United States, the spaces that hydroponics occupies, and what futures may exist for the niche technology.

The Regime

The US federal government spends about 20 billion dollars a year on farm subsidies. Roughly 40% of the 2.1 million farms in the nation receive subsidies (Andrzejewski). Most go to farms that produce cash crops, such as corn, soybeans, wheat, cotton, and rice. To keep food prices stable for consumers, the government tries to protect farmers from any changes in yield through insurance coverage and subsidies for marketing, research, and export sales. Aid for crop farmers is extremely comprehensive despite the fact that agriculture is not any riskier than most other industries. Many environmentalists believe that farm subsidies are costly not only to
taxpayers, but also harmful for the economy and the environment. In many cases, subsidies discourage innovation or diversification, preventing farming from becoming a truly competitive economy. These subsidies prevent innovative technology, such as hydroponics, from becoming truly viable. Hydroponics cannot compete with the low prices subsidies create for traditional farmers. Many farmers don’t see the point of switching to new technology when they are being paid exorbitant amounts to continue using traditional agricultural techniques.

US culture is deeply ingrained in farming and agriculture. Though only a tiny percent of the population consider themselves farmers, horticulture and gardening are very popular across urban and suburban areas in the United States. Recent estimates say that nearly a third of US households grow food, whether at home or in community gardens. The perception of farmers and traditional agriculture is well respected in the country, making it is difficult to imagine the public being able to embrace a whole new image of what farming means. Another problem comes with the organic status of hydroponics. The US Department of Agriculture (USDA) approves produce as organic, but the label has not yet been extended to plants not grown in soil. This has created a lot of stigma around hydroponics, and though the produce is the same, traditional farmers claim that soil-grown food is superior in taste and nutritional value. Without the
ability to be labelled as organic, hydroponics produce cannot overcome this stigma. Hydroponics produce is not labeled as organic in most other countries, but they do not value organic products as highly as Americans do. In fact, nearly 40% of the American public says that some part of their diet has organic food in it. The market for organic produce has grown by 10% each year and has been steadily increasing over the past few years ("Americans’ views about").

![Figure 3: Americans’ opinions on organic produce]("Americans’ Views About")
The Landscape

Although the current agricultural regime in the United States is relatively strong, there are certain windows of opportunity to be exploited. For one, there is a lot of criticism of traditional farming. People believe it is wasteful and misuses resources. The percent of arable land in the United States is roughly 17 percent. The United States is more than 200 times larger than the Netherlands and uses nearly 80 times as much of its land for farming. Given that the Netherlands closely follows the United States as the list of top food-producing countries, it is clear that the United States is wasting a lot of land to grow about as much food. Critics of government subsidies for agriculture believe subsidies are bad for the economy, setting prices in unfair ways. In addition, the food industry has a reputation of being unresponsive to more environmentally sustainable methods of food production, to the point of being extremely hostile toward possible competitors and disruptive technologies. But as public opinion has shifted in support of environmental policy and more efficient use of resources in all of America’s largest industries, certain niche technology and innovations have become more popular. The most prominent example of this shift in landscape has been the rise in popularity of renewable energy, such as solar panels and electric cars.
The Niche

One niche technology that has managed to grow under the regime is hydroponics. The market is expected to grow to USD 3,695.7 million by the end of 2025, with a Compound Annual Growth Rate (CAGR) of 20.3% between 2019 and 2025 (Kwamena). Hydroponics systems eliminate the need for pesticides as there is no soil for pests to live in. Crops such as lettuce, cabbage, tomatoes, peas, and other exotic and salad vegetables grow well in hydroponics systems. In the United States, these kinds of leafy vegetables hold the dominant market share, so there is definitely a market for crops that can be grown successfully in hydroponics systems. In addition, hydroponics systems are able to produce higher yields than soil-based agriculture, as every part of the crop’s intake, from water and temperature to nutrient levels and fertilization, is controlled. These factors are expected to push the market to grow more in the coming years. Leading manufacturers of hydroponics systems, such as AmHydro, General Hydroponics, Sinowell, Cropking, Hydrofarm Inc, are attracting new players by working tirelessly to develop their products to be even more cutting edge and economically viable (Jensen). Major challenges that the hydroponics market faces include lack of technical knowledge for the science as well as high transition and start-up costs. Despite this, a rising
demand for higher yield in crops, as well as pesticide-free food, is expected to drive the demand for eco-friendly technology like hydroponics.

Data Analysis

Netherlands

About two decades ago, the Dutch government made a national commitment to sustainable agriculture under the rallying cry “Twice as much food using half as many resources.” So far, it seems they have succeeded, as they are now the world’s top exporter of potatoes and onions and the second largest exporter of vegetables in terms of value (Treat). This change was not easy, although it was facilitated by the many social and cultural changes in the Netherlands in the past 20 years.

The Netherlands is considered to be a center of tolerance and liberalism and has long been touted for its flexibility and open-minded culture. It is the first country to legalize same-sex marriage and is one of more than 30 countries to have legalized prostitution. Amsterdam is famous for its “coffee shops,” where the sale and use of certain soft drugs have made the city a popular tourist destination. This open-mindedness stems from deep-rooted ideals of acceptance as well as pragmatism. The Dutch are very pragmatic people, a trait that has allowed them to be flexible to new ideas, especially new technology. In 2018, the Netherlands was ranked
second on the Global Innovation Index, which evaluates countries based on 80 factors ranked under pillars that include Institutions, Human Capital and Research, Infrastructure, etc. The Netherlands gained such a strong standing, in part because of their “strong, interlinked business sector that collaborates well with universities” (Cornell University). With all this in mind, it is easy to see how technological regimes in the Netherlands would be less than stable, given how flexible the landscape is. The Netherlands has cultivated an ecosystem of embracing disruptive innovation and supporting the emergence of new technology. Niche technology in the Netherlands is given many more windows of opportunity than in most other countries. How was it the windows of opportunity opened for agricultural innovation?

The Netherlands is known for having very little arable land. Though it does not suffer from drought, it receives much less rainfall than most people expect. The average rainfall for the Netherlands is around 700 millimeters, making it quite average, if not on the dry side, worldwide. The issue is not so much with the amount of rainfall, but how the climate behaves. Plants rely heavily on seasonal weather and do best in stable weather. The Netherlands is notorious for unpredictable rainfall and unexpected wet or dry days. All of these factors make it difficult for intensive agriculture to thrive. The Netherlands has a population of 17
million people and land size of 41,526 square kilometers. It is the most densely populated country in the European union and the 28th most densely populated country in the world. The Netherlands is also growing fast, with a net gain of one person every 11 minutes (Netherlands Population). As a small country, this poses a problem in providing enough food for the country. The Netherlands is not particularly fertile or suitable for crop production, but through successive agricultural innovation it has managed to become the horticulture capital of the world. Traditional agriculture was never a strong regime, making the potential for niche development much larger. Innovation and disruptive technology were never seen as a threat, but rather a problem solver.

Hydroponics has become increasingly widespread in the Netherlands. The largest agricultural sector of the country is in the Southern Westlands area, where 80% of the land is covered in greenhouses. Two decades ago, as the soil had become polluted and overused, Netherlands farmers were forced to take nearly all of their crops out of the soil and put them in vertical farming systems. Although the transition was not easy, it has paid off for the country. It is now is the second largest exporter of food after the United States, and they have accomplished this in much more efficient, resource-conserving ways. Since 2000, the government has managed to reduce dependency on water for key crops by 90% and
eliminated the use of chemical pesticides in greenhouses. This can be attributed to the widespread production of greenhouses and implementation of vertical farming. These greenhouses have incredible technology that is able to control nearly every aspect of the farming process. These climate-controlled farms “enable a country located a scant thousand miles from the Arctic Circle to be a global leader in exports of a fair-weather fruit: the tomato. The Dutch are also the world’s top exporter of potatoes and onions and the second largest exporter of vegetables overall in terms of value. More than a third of all global trade in vegetable seeds originates in the Netherlands” (Treat).

![Figure 4: Water footprint in tomato production (Treat, 2019)](image-url)
China

China is the most populated country in the world, with more than 1.42 people. The population is becoming increasingly urbanized. Over half of all farmers are over the age of 50, and young people do not want to farm. Meanwhile, a growing middle class seeks more varied produce. China is also losing land at an alarming rate compared to many other countries --its cultivated land has fallen to 40% of the world average. In response, the government has provided massive funding for vertical farming and sustainable agriculture initiatives, creating jobs for young people and higher yields for Chinese farms (‘Plant Factories’).

Chinese culture is heavily influenced by agriculture. The majority of Chinese citizens have family who still reside in the countryside, so it is rather uncommon to not have any ties to rural areas. Traditional values of simplicity and humility make agriculture a well-respected profession, despite being unpopular with a younger generation. At the same time, China has developed an intense fervor for rapid technological innovation in a battle for economic hegemony against the United States and other Western nations. This has allowed niche technology to be given attention it might not otherwise have received. One such technology thriving in China is hydroponics. Recently, jd.com and Mitsubishi partnered to create a factory of 11,000 square feet that is expected to produce not only more
produce, but more nutritious produce. The indoor farm is so well controlled it can control the growth factors of the produce so well that higher quality plants can be produced. Higher quality refers to higher micronutrient content, including spinach that contains 80% more folate, 25% more potassium, 32% more vitamin C, and 37% more phosphorus.

Unlike the Netherlands, China’s interest in hydroponics did not come solely from a place of necessity. Its goals of accessing dominance in all of its industries push it to new innovations. As a so-called second world country racing to surpass its competitors, China has created pocketed spaces for niche technology to develop. Because of outstanding cultural, political, and societal differences, it is difficult to compare China to the US or the Netherlands. What can be learned from the success of hydroponics in China, though, is that even nations still in development have found ways to access unique benefits provided by these fascinating systems. If China can exploit this innovative technology despite its diverse, climatic, and extremely fertile land, then the United States surely can open its regime to accept this new technology as well.

**Solutions**

The paper outlines several sectors in the US that would benefit from the implementation of hydroponics. These include areas that suffer from
drought, especially California, as well as in the growth and production of certain cash crops. Finally, I mention a couple of places that hydroponics can continue to innovate so that it may overcome certain socio-technical barriers that have prevented it from being successful in the past. These include more advanced systems that may employ robotics or more diverse systems.

**Drought Solution**

California is one of the top-producing agricultural states in the US and is known for high profile and high value products, such as grapes, citrus, and cotton. The state accounts for 13 percent of all crop revenue in the country, and California farms and ranches are said to have earned nearly $50 billion in revenue in 2017. However, many researchers believe California’s agricultural systems are unsustainable, relying too heavily on irrigation despite the arid climate and frequent drought. Three fourths of California cropland depends on irrigation. These croplands are concentrated within the Central Valley, including Sacramento Valley, known for citrus and nut orchards and San Joaquin Valley, known for their grape orchards and wineries. By conservative estimates, California’s fresh water reserves will be sufficient for only a couple of years (Morris and Bucini).
California’s average rainfall is around 540 millimeters, more than 200 millimeters less than the national average of 767 millimeters.

Fig 5: From this graph, it is apparent that exceptional cases of drought have become increasingly common since 2011, when there was very little drought (“California Percent,” 2017)

Compared to many other states, California is noticeably more open to sustainability; the state government already provides tax incentives for its citizens to use solar panels, as well as subsidies for solar production and installation companies. The most recent Renewables Portfolio Standard in California requires that 60% of electricity in the state come from renewable resources by 2030 and 100% by 2045. The California solar initiative, which
has been around for more than a decade, has helped solar power gain an edge on nonrenewable resources. These government policies have pushed Californians to be more open to environmental sustainability as well, including innovative technology that may help it reach the state’s goals.

California was number one on this year’s Bloomberg U.S. State Innovation Index, with 15 S&P 500 companies and many prestigious universities that push the state to be a hub for technology and innovation. Bloomberg also notes that California’s quest for water and power independence have made it the most innovative state in the country and a driving force for sustainability across the United States.

There are many ways hydroponics could potentially benefit California. Hydroponics is seen to be very effective in conserving water, with conservative estimates that it can save up to 70% of water compared to traditional farming. Given that Californians and the state government are willing to invest large amounts of money to achieve sustainability in energy, it is not absurd to think they might do the same for agriculture. The factors of drought, government policy, and landscape shift represent an opportunity to change the agricultural regime in the state of California, as well as the United States, that will help it plan for long-term functionality and sustainability. California’s drought can be seen as an opportunity to not only put landscape pressure on the regime, but to force farmers, users, and
policymakers to fully see the limits of natural resources and work to transform the restrictive agricultural policy framework that exists in the United States (Morris).

**Hydroponics and Cash Crops**

Due to the high overall investment costs of hydroponics, the best type of plants that they can produce are cash crops, especially that that require large amounts of water such as leafy greens and many exotic fruits. One unique crop that has thrived in hydroponic systems is cannabis or hemp. The cannabis industry, particularly cannabidiol (CBD) is becoming a large and lucrative market. Government regulation of cannabis is becoming increasingly lax, with 33 states having legalized medical marijuana. Most states have decriminalized medical and recreational marijuana, and 11 states have legalized it. According to *Entrepreneur*, CBD is becoming a billion-dollar market and is expected to grow to 22 billion dollars by 2022 (Leadem). It is a plant with value that will increase as it becomes more widely accepted in society and legally.

Over the past few years, growing marijuana through hydroponics systems has become popular. Many cannabis producers are forced to grow the plant indoors due to government regulation and strict rules on where and how much o can be grown. Hydroponics is particularly useful for
growing marijuana due to its high automation and stringent control of
different variables. The high cost of investment in hydroponics equipment
can be easily overlooked after considering the high value of cannabis.
Cannabis is only one example of a cash crop that can surpass the high
startup costs of hydroponics while taking advantage of the system’s many
benefits. Others include exotic fruits which have become increasingly
popular in the United States as well as staple fruits and vegetables, such as
tomatoes, potatoes, and greens.

**Robotics**

One of the main problems cited about hydroponics is its inefficiency
in terms of crop harvesting. While hydroponics is extremely labor efficient
in the germination and irrigation steps because of its automated process in
most modern greenhouses and indoor vertical farming systems, there are
not yet many efficient ways to automate the harvesting process. Where
traditional soil farms have extraordinary combine harvesters that can
harvest up to 150 acres of wheat per day and even perform advanced tasks
such as immediately shucking harvested corn, most hydroponics labs and
greenhouses still use traditional cutting methods to harvest crops. This
slows the harvesting process considerably. The tradeoff between vertical,
land intensive farming and delicate machines that cannot be run through
with combines makes hydroponics a less viable option for the traditional United States agriculture that relies heavily on staple and cereal crops that require huge areas of land.

A way to solve this problem is in the development of robotics to further automate the production cycle of soil-less greenhouses. One of the biggest issues plaguing farming today is a lack of labor. More than 3 billion dollars were lost in California last year because there were not enough people to handle the seedling and germination process. The average age of farmers is rising, and young Americans are becoming less willing to join the agriculture industry, preferring tertiary level jobs that are considered less labor-intensive. This is a perfect window of opportunity for hydroponics, which is a subset of hydroculture and a requires a lot of knowledge and experience. The problem then becomes that hydroponics cannot effectively take advantage of this window of opportunity, since it is not yet advanced enough to compete with the traditional agriculture industry’s massive yields.

A UK startup called Iron Ox has attempted to tackle this problem by unveiling its autonomous production farm in October of 2018. Within an 8000-square-foot space, Iron Ox engineers have been using a proprietary robot system to grow 30,000 heads of leafy greens in hydroponic vats. The robotic system uses laser cutters to harvest vats of plants, running them
across the tops of the heads of lettuce and other greens with extraordinary precision. CEO and co-founder Brandon Alexander explains that a major problem they were trying to solve was the absence of fresh produce in urban areas (Hurst). It can take days for harvested vegetables to reach restaurants and grocery stores. Through researching this issue, Alexander discovered other inherent problems with traditional farming, so he began to look into hydroponics and ways to make it more viable for the urban environment. Though Iron Ox has made tremendous strides in the automation of farming, humans still play a crucial role. The planting and packaging of final products are still done with human labor. Still, companies like Iron Ox prove that there is much potential for hydroponics to grow and expand. The technology itself will continue to grow under the regime of traditional agriculture and is already finding a place for itself in the current market.

Conclusion

In sum, the scope of this research draws 3 separate conclusions. The first is that hydroponics will probably never reach the same level of success in the United States compared to the Netherlands. When taking into account the political, social, and cultural factors of the agricultural regime in the United States, it is clear that traditional agriculture has a rather firm
hegemony of the market. Political support for hydroponics is unreasonable to expect when so much of the population still relies on traditional farming. It is also still quite a niche technology -- most people have never even heard of hydroponics. But the landscape of the United States is changing every day, and the future is very uncertain for agriculture around the world. It is clear that traditional agriculture will not be sustainable forever, but which new technologies will arise to take its place is still unclear.

The second conclusion is that though hydroponics may never be a disruptive innovation, it can definitely create a niche market for itself. To an extent, it already has. Through differentiating its produce from traditional crops and pushing to gain an organic status, hydroponic farmers must prove to traditional farmers that they are not taking away from the existing market, and can coexist with them. These niche markets may be concentrated in certain geographical regions, such as drought-ridden valleys of California, or for certain produce, like cannabis and other cash crops. The unsustainability of traditional agriculture has opened up many windows of opportunity for hydroponics, and careful surveying of the landscape can reveal the best places to implement this innovation.

The third and final conclusion is that hydroponics is far from a complete technology. There is still so much research that must be put into creating more efficient economies of scale. Whether this is through
automating the harvesting process to maximize return on investment or through development of more advanced systems that can cover more ground to compete with large scale traditional farming is up to hydroponic farmers to decide. Like all niche technology attempting to become disruptive innovations, hydroponics must seek out the people and places it can benefit and continue to grow and adapt until it can fit itself into those opportunities. The unsustainability of old technology is often seen as an alarming phenomenon, but innovators choose to see it as an opportunity to create new, more advanced, more sustainable technology that can move the world to better places.
References


https://www.forbes.com/sites/adamandrzejewski/2018/08/14/mapping-the-u-s-farm-subsidy-1-million-club/#3a4e6c433efc


China Population. (2019 April 1). Retrieved May 4, 2019, from
http://worldpopulationreview.com/countries/china/


https://www.pnas.org/content/112/13/3931/tab-article-info


Sure Journal 2019 | 375

doi:10.1016/j.techfore.2011.07.004


doi:https://curve.carleton.ca/system/files/etd/748f35ee-2799-4ff4-80b6de3032dd65f7/etd_pdf/a275c00b3631e12bc0a228ccf79bb6cd/miller-scalinguporsellingoutacriticalappraisalof.pdf


SEC Rulings on Dark Pool Transparency: A Case Study to Understand Its Effects on Price Discovery

Srivats Narasimhan and Eric Zhang
Research Mentor: Mark Flood

Srivats Narasimhan is a SURE Fellow in the Department of Computer Science and Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: snarasim@terpmail.umd.edu.

Eric Zhang is a SURE Fellow in the Department of Computer Science and Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: ericzhang98@gmail.com.

Abstract

This paper examines recent SEC rulings that demand increased transparency in Dark Pools and their effects on price discovery, using Cisco as a case study. A two-week and a four-week sample of two different liquidity measurements, Amihud’s measurement and Turnover ratio, were taken a month before the SEC rulings and a month after the implementation of the SEC rulings. These samples were converted to time-series, and a Granger Causality test was then performed on each separate group of time-series. Four of the five Causality
tests yield the same result of not causal. The fifth result yielded a causal relationship; however, this anomaly could possibly be attributed to exogenous variables in the market. Because there is no causal relationship present on a general level, based on the results of this study, it can be theorized that the SEC rulings did not affect price discovery. Yet, due to limitations of the case study itself, its results may not entirely indicate the implications of the SEC rulings on price discovery at a larger, more generalized level.

*Keywords: Dark Pools, Granger Causality, Time series, Amihud’s Measure, Turnover Ratio, Down sampling, Java, R*

**Introduction**

Dark pools have played a pivotal role in the financial markets in recent years. In fact, Rosenblatt Securities estimates that dark pools execute 14% of the U.S. equity volume (*Bain*). According to the CFA, Dark pools have surged in usage, accounting for an estimated 40% of all US stock trades in Spring 2017, compared to an estimated 16% in 2010 (*Dark Pools*).

The origin of Dark Pools can be traced back to the 1980’s. However, the recent book *Flash Boys*, written by Michael Lewis, has given dark pools public attention. The book described how many proprietary trading firms used tactics, such as ‘pinging’ dark pools, to unearth large orders that were hidden, then
engage in front-running and latency arbitrage to generate Alpha. This poses the following questions: What exactly are dark pools, and how do they impact financial markets? Although the term may colloquially sound rather ominous, it refers to a standard market practice. Dark Pools, or Alternative Trading Systems (ATS), are private exchanges intended for the trading of securities and are usually operated by larger financial institutions. It is important to note that Dark Pools are not readily accessible to the general public, but only to institutional investors with large amounts of capital. The primary facet of Dark Pools is the lack of transparency for general traders in the LIT markets. Typically, dark pools have limited or no pre-trade transparency, anonymity, and mid-quote pricing (Buti, Sabrina and Rindi, Barbara). The premise of Dark Pools is to prevent large amounts of orders from affecting the price discovery of certain assets in LIT markets.

The debate on the morality and legality of Dark Pools has sparked a recent general interest in whether they actually affect equity price discovery in the LIT markets. The theoretical premise of dark pools is to prevent the manipulation and saturation of LIT markets from affecting returns for institutional investors that typically make significant trades. However, as of July 18th, 2018 (Bain), the SEC ruled to increase transparency of dark pools. The ruling came into effect on October 9th, 2018.

The purpose of this paper is to investigate the SEC rulings to increase the transparency of Dark Pools and their potential impact on the price discovery and
efficiency of common equities. The SEC policy has tremendous implications on the financial industry, potentially enabling an increase in the amount of informed trades in LIT markets. The very premise of Dark Pools allowed investors to make informed trades without having to deal with the related implications that information would have on the general market. With recent SEC rulings, however, it is important to understand whether price discovery and efficiency actually increase for equities, because of newly disseminated information that previously wouldn’t have reached LIT markets.

This paper is an exploratory case study of Cisco and its stock value. To explore the possible impact of the ruling, this paper used two separate liquidity measures: The Turnover Measure and Amihud’s Measure. Time-series of these measures were generated, and Granger Causality tests were run to determine if there is a causal relationship between the time-series before and after the implementation of the discussed SEC rulings. The time frame of this study took place within the past year, since when the SEC rulings were implemented. This paper hopes to provide initial research into the impact of the SEC rulings on equities and to provide some insights into the debate how dark pools actually impact LIT markets.

**Case Study: CISCO**

This paper represents an exploratory case study, and is not to be misunderstood as an authoritative econometric analysis. As such, analyzing one
stock would effectively explore the paper’s intended goal, to understand how new SEC rulings on Dark Pools affect price discovery and efficiency. It would be increasingly more difficult to analyze price discovery and efficiency for more than one stock.

Certain criteria were essential in selecting a stock for this study. The stability of the firm itself is important, because it ensures there will be fewer idiosyncratic risks associated with small and high growth firms. These risks usually include high volatility in share price. The main measure of volatility in a stock is the Beta Coefficient, which represents how the share’s returns relate to those of the general market. The second important criterion is the nature of the firm itself. Choosing a larger, more established conglomerate would lessen the risk associated with highly specialized, single-industry firms. The third, and final, criterion is that the stock is traded on the S&P 500, to confirm the firm’s established nature and readily available data.

Based on these three criteria, the study chose to focus on Cisco (Ticker Symbol: CSCO). The Beta Coefficient of Cisco currently hovers around 1.06 according to Yahoo Finance; this implies that Cisco is a relatively stable firm whose returns tend to reflect that of the general market (i.e. it tends to move with the market). This is a relatively low Beta Coefficient, given the fact that Cisco operates within the software and networking hardware industry. However, because Cisco is a conglomerate and is diversified within the Technology Industry, it is less prone to concentration risk than its related counterparts and competitors. Finally, Cisco is
traded on the S&P 500, indicating that it is established and that data is readily available for retrieval and corresponding analysis.

**Methodology**

The basis of the analysis relies on high-frequency data. High-frequency data elucidates many essential measures to help understand what impacts price discovery. The main function of this data collection is to construct the time-series that are to be used in the study’s Granger Causality tests. The source of this data is the Wharton Research Data Services (WRDS) database; from there, access to the Trade and Quote Database (TAQ) provided relevant observations. It contains intraday transactions regarding trading and quote data for all securities exchanges, at a millisecond scale. The SEC rulings concerning increased transparency in dark pools were announced on July 18, 2018, and implemented on October 8, 2018. The studies’ data points were extracted from two intervals. The first type tested intervals that roughly spanned two weeks – those that took place approximately a month before the SEC rulings were announced and one month after the SEC ruling rulings went into effect. The second type of tested intervals spanned roughly four weeks -- these took place approximately one month before the SEC rulings were announced and a month after the SEC ruling rulings went into effect. The two-week span taken before the SEC rulings occurred from June 4 to June 18, 2018. The four-week span taken before the SEC rulings occurred from May 21 to June 18, 2018. The two-week span taken after the SEC rulings occurred was
November 5 to November 19, 2018. The four-week span taken after the SEC rulings occurred was October 22 to November 19, 2018.

**Down-sampling**

Down-sampling is critical to the functionality and analysis of the dataset. It is typically done to conserve memory from a computing standpoint and to reduce signal processing time when data has been oversampled. The high-frequency data that was utilized in this paper comes with data-points per millisecond. This study’s most relevant statistic concerns the volume-traded-per day. Down-sampling is the action of aggregating large data-sets into smaller, more focused intervals. For this specific case study, the data was down-sampled from a high frequency (data-per-millisecond) into a lower frequency (data-per-day). Through the usage of code written in Java, the data-per-millisecond was parsed and stored into a data-structure that contained daily totals in trading volume. This was then used to compute the liquidity measures and construct the corresponding time-series.

**Granger Causality**

A time-series is basically a set of quantitative observations arranged in chronological order. Specifically, with time-series, time is a discrete variable, meaning that when a time-series analysis is run, only a certain time interval can
be allocated for data-analysis. Usually, time-series help measure the forecast error, which is the difference between an observed value and its forecasted value. This study’s time-series consist of the measurements of liquidity (derived from high-frequency data) from the weeks specified above.

The specific analysis within our time-series is Granger Causality, which is a test of two different time-series used to determine the relationship between each time-series. The premise of a Granger Causality test is to determine if one time-series is causal for another time-series; in other words, a Granger Causality test helps show if one time-series can predict another.

From here, two separate time-series, X and Y, will be examined. Let \( I \) represent the specific data available at a certain time -- t. When translating this in terms of the whole time-series, \( X \) is, therefore, representative of the current and past data available for the time-series X at time t. The same would apply for Y, where \( Y \) would, therefore, be representative of the current and past values available for time-series Y at time t. When translating into a set equation,

\[
X_t = \{X_t, X_{t-1}, X_{t-K}\}, \quad \text{and} \quad Y_t = \{Y_t, Y_{t-1}, Y_{t-K}\}.
\]

The corresponding forecast error is denoted as \( \sigma^2 \).

There are explicit two separate measurements of Granger Causality: (1) Granger Causality and (2) Instantaneous Granger Causality. The following literature provides explanations of the two separate measurements:
(1) X is Granger Causal to Y if the following equation is satisfied:

$$\sigma^2(y_{t+1}|I_t) \lt \sigma^2(y_{t+1}|I_t - X_t).$$

In this equation, $X_t$ can be described as the lag data set, and the equation $y_{t+1}$ is used to denote the current time-series data set. This equation suggests that the future values of the time series Y, denoted by $y_{t+1}$, can be better predicted if the current and past values of X are used.

(2) X is instantaneous Granger Casual to Y if the following equation is satisfied:

$$\sigma^2(y_{t+1}|\{I_t, X_{t+1}\}) \lt \sigma^2(y_{t+1}|I_t).$$

This equation states that the future value of Y, denoted by $y_{t+1}$, can be better predicted if the current and past values of X, denoted by $X_t$, and if the future values of the time series X, denoted by $X_{t+1}$, are used within the equation.

Instantaneous Granger Causality is a part of the Granger Causality theoretical framework; however, this paper will not test for Instantaneous Granger Causality and will focus on Granger Causality.
Liquidity Measures

One of the most debated topics regarding the effects of dark pools on LIT markets involves the concept of liquidity. Liquidity can be understood as the matching of the demand and supply of a given asset with the least associated transaction costs (Gabrielsen, Marzo, and Zagaglia). More simply put, liquidity defines the idea that there exists a buyer for a seller in the market.

It is generally understood that the impact of liquidity on an asset has great implications to its pricing. From a conceptual standpoint, if an asset is illiquid, its rate of return must increase to compensate for the increased transaction costs that resulted from the illiquidity. In other words, if an investor cannot sell an asset that easily, they have to be fairly compensated as such. The Efficient Market Hypothesis (Farma) also dictates that, for an efficient market to exist, the market itself has to be large, liquid, and have easily accessible information for all market participants. Prices are therefore more inaccurate in illiquid markets.

Amihud’s Measure

\[ ILLIQ_{i}^{T} = \frac{1}{D_{T}} \sum_{t=I}^{D_{T}} \frac{|R_{i,T}^{t}|}{V_{t,T}^{i}} \]
One helpful calculation related to illiquidity is Amihud’s Measure. In this case, \( D_T \) is the number of days for which the data is actually available; \( V_{t,T} \) is the daily trading volume (the dollar amount of shares traded on that specific day); \( |R_{t,T}| \) is the return on that day -- for the given share. For this study’s scope, this liquidity measure is realistically calculable because of the availability of the related data. In other words, there exists high-frequency data that actually allows the calculation of this liquidity measure, therefore making it a candidate for the time-series that were used.

To compute Amihud’s Measure, several variables were factored in. The first calculation performed was the daily return of Cisco. Let \( X_t \) denote the daily closing stock price. The daily return is computed using the equation \( \left( \frac{X_{t+1}}{X_t} \right) - 1 \). The next variable that was essential to the calculation is the daily dollar value. This measure is simply computed by multiplying the daily stock price by the daily volume traded. Finally, Amihud’s measure is computed by taking the absolute value of the daily return and dividing it by the daily dollar value.

**Turnover Measure**

\[
TR_t = \frac{S_h_t}{N S h_t}
\]

The turnover measure, or ratio, is the second liquidity measure that was utilized in the case study. Essentially, because the equation entails the average
amount of shares traded over the total amount of shares outstanding (available to the market), it can act as a measure of how difficult it is to sell a share in the market. As such, time-series were created based on turnover measure calculations.

In order to compute the turnover measure, only two specific variables are needed. Let $N_{Sh_i}$ = the volume traded on a specific day. Let $Sh_i$ = total amount of Cisco shares outstanding. The latter was found within Cisco’s quarterly reports and was used as a constant for the calculation.

Each liquidity measure was calculated and compiled into a Bivariate time-series model.

Before the Granger Causality test was run, the null and alternate hypotheses were established based on time-series X: pPre SEC ruling, and time-series Y: post-SEC ruling.

- Null Hypothesis = X does not Granger Cause Y
- Alternative Hypothesis = X does Granger Cause Y
- If the resulting p value < 0.05, the Null Hypothesis is to be rejected. Rejecting the Null shows that X does Granger Cause Y, revealing that the values of Y can be predicted by X.
- If the resulting p value > 0.05, the Null Hypothesis is to be accepted. Accepting the Null hypothesis shows that X does not Granger Cause Y, revealing that the X cannot predict the values of Y.
Also, one lag variable was chosen when performing the Granger Causality test. One variable was ideal because of the limited sample size, where too many lag variables would have decreased the reliability of the results. If Granger Causality emerges after regulations, it says that informative trades are migrating into dark pools, and LIT markets are less informed, only learning what they learn after dark pools have completed their trades. Thus, if pre-SEC ruling liquidity measures Granger Causes post-SEC ruling liquidity measures, it can be said for certain that the SEC ruling does impact price discovery.

**Limitations**

There were a few key limitations to the study’s research scope. The main limitation involved lack of a substantial sample size. To perform an efficient and more reliable Granger Causality test, years and years of observations -- in other words, data-points -- are typically needed. However, for this study retrieving this data was not possible. The study was able to retrieve stock-related data for only two-week and four-week intervals before and after the implementation of the SEC rulings. Regardless of the fact that a Granger Causality test may not be as effective in cases with smaller data-sets, it doesn’t necessarily discount the resulting conclusions.

The team decided to approach this research project as more of a case study, mainly because there were too many potential data points if we considered more than one specific equity. Because this study was completed over the course
of only one semester, the data had to be reasonably sized. Personal computers
were used to conduct the data analysis, and these devices did not possess the
required computing power to handle large data-sets pulled from a general index
(many stocks). All of these limitations led to framing the case study to analyze
just one stock. After initial background research, it was found that Cisco (CSCO)
could provide a valid sample because it is a stable stock, with a Beta Coefficient
close to 1; this meant that general conclusions could potentially represent price
discovery on a higher level as well.

Another consideration that must not be forgotten is that investors are also
humans, that a psychological factor is involved in the investment process. In a
sense, transitive and anomalous behaviors might occur as a result
of the SEC rulings, which may affect the resulting data-sets that were utilized in
the study.

Data Analysis

Down-Sampling Code
Performed in Java

Working with high-frequency data is difficult, due to the sheer amount of
data points that must be manipulated. Even with just one share’s (CISCO) high-
frequency data, down-sampling had to be performed to create time-series for each
individual liquidity measure that was used in our Granger Causality tests.
To actually down sample the code, Java’s Excel API, Apache POI, was used to parse through the high-frequency data (which were stored in CSV files). A Linked-List data-structure was used to create individual nodes that stored trade volume on a specific day; a Linked-List was used because of its dynamic nature and ability to store data within each individual node. Because the data was pulled in millisecond intervals, it would have been ineffective and inefficient to use each trade-volume-per-millisecond metric as part of the time-series.

**Data Analysis Code**

**Performed in R (or RStudio)**

Essentially, in order to actually perform Granger Causality tests for the discussed liquidity measures, some sort of statistical package capable of performing it natively was required. Creating the actual computing backend of this statistical analysis is unnecessary for the purpose of this study. Thus, after using the down-sampled data, each individual liquidity measure was calculated in separate spreadsheets (through formulas in Excel), then imported the respective data into RStudio. There, using Granger Causality functions (found in a package downloaded for RStudio), the statistical analysis was performed.
Results

Amihud’s Measure

After the data for the Amihud’s Measure was computed in an Excel spreadsheet, the two-week periods before and after the SEC ruling were compiled into a Bivariate time-series model.

As seen above, the average Amihud Measure pre-SEC rulings was 0.000000000062227, and the average Amihud Measure post-SEC rulings was 0.0000000000152036. It can be noted that within this two-week period, the post-SEC rulings’ Amihud Measure nearly doubles its older counterpart. Subsequently, this can be visually noted in the time-series detailed above.

Fig 1: Time Series of Amihud’s measure from the two-week interval

As seen above, the average Amihud Measure pre-SEC rulings was 0.000000000062227, and the average Amihud Measure post-SEC rulings was 0.0000000000152036. It can be noted that within this two-week period, the post-SEC rulings’ Amihud Measure nearly doubles its older counterpart. Subsequently, this can be visually noted in the time-series detailed above.
The Granger Causality test based on the two-week sample of the Amihud Measure was run and yielded the following results:

<table>
<thead>
<tr>
<th>Res.</th>
<th>Df</th>
<th>Df</th>
<th>F-value</th>
<th>Pr (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>-1</td>
<td>0.192</td>
<td>0.6745</td>
</tr>
</tbody>
</table>

Fig. 2 Granger Causality results for Amihud’s Measure for the two-week interval

The following test revealed a P-Value of 0.6745. Since the P-value is greater than the 0.05 threshold already established, the Null Hypothesis could not be rejected. This reveals that within this two-week span, the SEC rulings were not Granger Causal.

The data was then expanded to include a larger sample size of 4 weeks. This increase in sample size was conducted to determine the consistency of the previous results and to reduce the effect of any exogenous variables that might affect the Amihud Measure. This data was then converted into another Bivariate time-series model.
Fig 3: Time Series of Amihud’s Measure from the four-week interval

Through a brief visual analysis of the previous model, there are a few important ideas to note. Both time-series appear to be relatively volatile, and, akin to the previous observation, it seems that there is an increase in the Amihud Measure after the SEC rulings. The average pre-SEC rulings value was 0.0000000000065084, and the average post-SEC rulings value was 0.0000000000136988. The same doubling effect observed in the two-week sample can also be observed in the four-week sample.
The Granger Causality was then run on the four-week sample of the Amihud Measure and yielded the following results:

<table>
<thead>
<tr>
<th>Res.</th>
<th>Df</th>
<th>Df</th>
<th>F</th>
<th>Pr (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>-1</td>
<td>0.0193</td>
<td>0.8912</td>
</tr>
</tbody>
</table>

*Fig. 4 Granger Causality results for Amihud’s Measure for the four-week interval*

This following data resulted in a p value > 0.05, thus leading us to accept the Null Hypothesis. This increase in sample size revealed the same conclusion as the two-week sample size, that the SEC ruling was not Granger Causal. Because the sample size was effectively doubled and the same result was achieved, a level of consistency is shown.

Both the time-series of the two-week sample period and the four-week sample period failed to reject the Null Hypothesis. This allows one to conclude that, in terms of Amihud’s Measure, there is no Granger Causality. The values of the pre-SEC ruling Amihud’s Measures do not Granger Cause post-SEC ruling Amihud’s Measures. Based solely on this, it seems that the SEC rulings did not affect price discovery and efficiency. To further justify these findings, the Turnover Measure was then used.
Turnover Measure

The Turnover Measure was the next liquidity measure to be considered in this study. Similar to Amihud’s Measure, the Turnover Measure was computed for before and after the SEC rulings. This liquidity measure was compiled over two-week and four-week interval, and was constructed into a Bivariate time-series.

The first two-week time-series was constructed based on the Turnover Measure, as detailed below:

![Time Series of Turnover Measure from the first two-week interval](image)

*Fig 5: Time Series of Turnover Measure from the first two-week interval*

The initial observations reveal an interesting trend between the two time-series. Both time-series seem to follow a similar trend, wherein they both increase...
significantly and decrease around the same time. The average Turnover Measure pre-SSEC rulings was 0.005935654, and the average Turnover Measure post-SEC rulings was 0.005461167. Contrary to Amihud’s measure, the average turnover measure was lower post-SEC ruling. When observing the graph, the pre-SEC ruling measure outperforms the post-SEC ruling except on day 10, when there seems to be a spike in trading in the time sample post-SEC ruling. This spike could potentially skew the average and the causality test.

A Granger Causality test was then performed on the time-series and yielded the following result:

<table>
<thead>
<tr>
<th>Res.</th>
<th>Df</th>
<th>Df</th>
<th>F</th>
<th>Pr (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>-1</td>
<td>17.252</td>
<td>0.004276</td>
</tr>
</tbody>
</table>

*Fig. 6 Granger Causality results for Turnover Measure for the first two-week interval*

The causality test revealed a p value of 0.004. This p value is < 0.05, meaning that the Null Hypothesis is to be rejected, and the Alternative Hypothesis would be accepted. Based on the results of the causality test for the turnover liquidity measurement over two weeks, the post-SEC rulings time-series is casual
to the pre-SEC rulings, and there is Granger Causality. This result seemed strange at first, because it deviates far from Amihud’s Measure.

To confirm this causal relationship, two alternative tests were run. A different two-week period, October 22 to November 4, 2018, was sampled. The turnover for this two-week period was therefore calculated, and a different time-series was generated; this is detailed below:

![Graph: Time Series of Turnover Measure from the second two-week interval](image)

*Fig 7: Time Series of Turnover Measure from the second two-week interval*

With this different two-week sample size, the average Turnover Measure pre-SEC ruling is 0.005006049, and the average Turnover Measure post-SEC ruling is 0.005537827. A slightly higher average in Turnover Measure is noted in the graph as well.
After conducting another Granger Causality test, the following results were yielded:

<table>
<thead>
<tr>
<th>Res.</th>
<th>Df</th>
<th>Df</th>
<th>F</th>
<th>Pr (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>-1</td>
<td>0.1536</td>
<td>0.7112</td>
</tr>
</tbody>
</table>

*Fig. 8 Granger Causality results for Turnover Measure for the second two-week interval*

The following revealed a p-value that is greater than 0.05 which leads to the conclusion that the pre-SEC rulings time-series does not Granger Cause post-SEC rulings time-series. This alternative result challenges the previous two-week interval time-series, which indicated a causal relationship.

A third test involving a four-week period sample size was tested, and the following time-series was constructed:
The Turnover measure pre-SEC ruling within this 4-week sample size was 0.005256364, and the average Turnover measure post-SEC ruling was 0.005756632. There is also an increase in the average turnover measure within this 4-week period.

The Granger Causality test yielded the following results:

<table>
<thead>
<tr>
<th>Res.</th>
<th>Df</th>
<th>Df</th>
<th>F</th>
<th>Pr (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The p-value for the four-week Turnover Measurement revealed a p value > 0.05, which leads to a failure to reject the Null Hypothesis. Based on the four-week data, this yet again leads us to the conclusion that pre-SEC ruling does not Granger Cause post-SEC ruling.

Although the first two-week test concluded a Granger Causal relationship, the next two-week test refuted this claim. From this we can conclude that, in terms of Turnover measurement, it is highly likely that there is no Granger Causality. The first test that yielded a Granger Causal relationship may be caused to exogenous variables in the market that could have temporarily affected the results of the Causality test.

**Conclusion**

After conducting Granger Causality tests for our two chosen liquidity measures, Amihud’s Measure and Turnover Measure, we are able to see that, on the whole, the time-series before are not Granger Causal to the time-series after; essentially, this portrays that, post SEC rulings, liquidity (and therefore price discovery) for CISCO stock has not changed in a significant manner. The general
purpose behind the study was to understand that if liquidity measures changed noticeably after the SEC rulings (or a Granger Causality emerged after regulations) and if informative trades were in fact migrating into dark pools, making LIT markets less informed beforehand. After the rulings, thus, a Granger Causal relationship would imply that LIT markets were more informed from decreased opacity in Dark Pools, thus increasing the price efficiency of equities.

While this may indicate a lack of impact on price discovery and efficiency, it is not necessarily a totally conclusive result. As indicated previously in the study, there are noticeable limitations in the methodology chosen for the study, as well as in the very assumption that CISCO represents LIT markets in general. Many exogenous variables may have affected the outcome of the study, including investor decision-making (a psychological factor), algorithmic trading decisions, and a general instability in the markets.

To perhaps extend this study, there are certainly a few steps that can potentially be taken. Having a greater amount of observations is greatly beneficial to the outcome of a Granger Causality test, as the premise of the test assumes that each time-series has an extensive amount of data behind it, thus making them more representative of general trends and movements. Furthermore, testing time-series for multiple equities would create a wider representation of LIT markets, thus increasing the validity of the results. Finally, studying a greater number of liquidity measures for various equities, both before and after the SEC rulings,
would further support a general conclusion; with more liquidity measures representing a general trend, a solidified conclusion is far easier to come by.

Acknowledgments

We would like to thank Dr. Mark Flood, Dr. Christina Elson, the Ed Snider Center for Enterprise and Markets, and the Robert H. Smith School of Business for supporting this research project.
Appendices

Appendix A: Java Source Code

Read Data Class

```java
package Research_Project;

import java.util.*;
import java.io.File;
import java.io.FileInputStream;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.IOException;
import java.sql.Time;
import java.util.Date;
import org.apache.poi.sl.usermodel.Sheet;
import org.apache.poi.ss.usermodel.Cell;
import org.apache.poi.ss.usermodel.Row;
import org.apache.poi.ss.usermodel.Workbook;
import org.apache.poi.xssf.usermodel.XSSFSheet;
import org.apache.poi.xssf.usermodel.XSSFWorkbook;

public class Read_Data {
    private static int inputSheetNum = 0;
    private static int outputSheetNum = 0;

    public static void main(String[] args) {
        int dataIndex = 0;
        int numOfNodes = 0;
    }

    public static void main(String[] args) {
        int dataIndex = 0;
        int numOfNodes = 0;
    }
}
```
// Opens the excel sheet
try {

    // Opens the input stream for reading
    FileInputStream fopen = new FileInputStream("C:\\Research\\Research_Data.xlsx");
    FileInputStream fwrite = new FileInputStream("C:\\Research\\Output_Data.xlsx");

    // Loads workbook for reading
    XSSFFWorkbook inputWorkbook = new XSSFFWorkbook(fopen);
    // loads workbook for writing
    XSSFFWorkbook outputWorkbook = new XSSFFWorkbook(fwrite);

    // Opens input sheet for reading
    XSSFSheet inputSheet = inputWorkbook.getSheetAt(inputSheetNum);
    // Opens output sheet for writing
    XSSFSheet outputSheet = outputWorkbook.createSheet("Sheet2");

    // finds the last value of the excel sheet
    int lastValue = inputSheet.getLastRowNum();

    // initializes the LinkedList with Volume_Data
    LinkedList<Volume_Data> filtered_Data = new LinkedList<>();

    // for loop to read into the data and write as well
    for (int i = 0; i < lastValue; i++) {
        // casts the date into an int and subtracts the 2018 year measurement
        int date = (int) inputSheet.getRow(i).getCell(0).getNumericCellValue() - 20180000;

        // finds the seconds of the data and rounds the value
        double volume = inputSheet.getRow(i).getCell(1).getNumericCellValue();

        // locates the trade volume executed at the specific millisecond interval
        Volume_Data trade = new Volume_Data(date, volume);
    }
}

// adds if the linkedlist is empty
if (filtered_Data.size() == 0) {
    filtered_Data.add(trade);
} else {
    // temporary variable for the current trade
    Volume_Data temp = filtered_Data.get(numOfNodes);
    // Checks if the dates's of the trade matches
    if (temp.getDate() == trade.getDate()) {
        // retrieves the current total
        double newTotal = temp.getVolume();
        // adds the new trade volume
        newTotal += trade.getVolume();
        // updates the trade volume
        filtered_Data.get(numOfNodes).setVolume(newTotal);
        // if the second of the trade is not the same, a
        // new linkedlist node is added
        // if the date's don't match, a new linkedlist
        // node is added
    } else {
        filtered_Data.add(trade);
        numOfNodes++;
    }
}
}
numOfNodes = 0;

for (Volume_Data dataPoint : filtered_Data) {
    // Creates the row for writing
    Row rows = outputSheet.createRow(numOfNodes);
    // Creates the cell for the date
    Cell dateCell = rows.createCell(0);
    // Writes the date cell value
    dateCell.setCellValue(dataPoint.getDate());
// Creates the cell for writing the volume
Cell volumeCell = rows.createCell(1);

// Writes the volume cell value
volumeCell.setCellValue(dataPoint.getVolume());
numOfNodes++;

// /24 -> Format Cells -> General -> h:mm converts from decimal to time
FileOutputStream write = new FileOutputStream("C:\Research\Output_Data.xlsx");
outputWorkbook.write(write);
inputWorkbook.close();
outputWorkbook.close();
System.out.println(numOfNodes);
System.out.println("Data Done");
} catch (FileNotFoundException e) {
    e.printStackTrace();
} catch (IOException e) {
    e.printStackTrace();
}
}

Volume_Data Class

package Research_Project;
import java.sql.Time;
import java.util.*;

public class Volume_Data {
    private int date;
    private double volume;
    public Volume_Data(int date, double volume) {
        this.date = date;
        this.volume = volume;
    }
    public Volume_Data(int date) {
        this(date, 0.0);
    }
    public Time getTime() {
        return new Time(date);
    }
    public double getVolume() {
        return volume;
    }
    public void setVolume(double volume) {
        this.volume = volume;
    }
    public int getDate() {
        return date;
    }
    public void setDate(int date) {
        this.date = date;
    }
    public void print() {
        System.out.println("Date: "+date);
        System.out.println("Volume: "+volume);
    }
    public static void main(String[] args) {
        Volume_Data volume = new Volume_Data(10, 15.0);
        volume.print();
    }
}
this.date = date;
this.volume = volume;
}
public int getDate() {
    return date;
}
public void setDate(int date) {
    this.date = date;
}
public double getVolume() {
    return volume;
}
public void setVolume(double volume) {
    this.volume = volume;
}
Appendix B: RStudio Source Code

installed.packages("lmtest")

library(lmtest)

ah2w<-
read.csv("c:/Research/Turnover_Causality/Amihud_Causality_2Weeks.csv",
header = TRUE, sep="","")

attach(ah2w)

Amihud_Causality_2Weeks.csv<-
read.csv(file.choose(), header=TRUE)
grangertest(AH_Post~AH_Pre, order = 1, data = Amihud_Causality_2Weeks.csv)

library(lmtest)

ah4w<-
read.csv("c:/Research/Turnover_Causality/Amihud_Causality_4Weeks.csv",
header = TRUE, sep="","")

attach(ah4w)

Amihud_Causality_4Weeks.csv<-
read.csv(file.choose(), header=TRUE)
grangertest(AH_POST~AH_PRE, order = 1, data = Amihud_Causality_4Weeks.csv)
installed.packages("lmtest")

library(lmtest)
to2w<-
read.csv("c:/Research/Turnover_Causality/Turnover_Causality_2Weeks.csv",
header = TRUE, sep="," )

attach(to2w)

Turnover_Causality_2Weeks.csv<-read.csv(file.choose(), header=TRUE)
grangertest(TO_Post~TO_Pre, order = 1, data =
Turnover_Causality_2Weeks.csv)

installed.packages("lmtest")

library(lmtest)

to4w<-
read.csv("c:/Research/Turnover_Causality/Turnover_Causality_4Weeks.csv",
header = TRUE, sep="," )

attach(to4w)

Turnover_Causality_4Weeks.csv<-read.csv(file.choose(), header=TRUE)
grangertest(TO_POST~TO_PRE, order = 1, data =
Turnover_Causality_4Weeks.csv)
References

A. Gabrielsen, M. Marzo and P. Zagaglia

Measuring market liquidity: An introductory survey


Available at SSRN: https://ssrn.com/abstract=1630499 or http://dx.doi.org/10.2139/ssrn.1630499


Proper Fit: Implications of Personalized Breast Cancer Screening

Carley Peyser

Foreword by Research Mentor, Dr. Henry C. Boyd III

Breast cancer remains a modern-day scourge. Sadly, too many young women die from this disease state, which is often preventable if caught early. There is a plethora of medical studies that underscore the importance of early detection. Hence, the prevailing mantra is that all women (those around ages 40-75) should routinely undergo breast cancer screening. This relatively simple, proactive behavior saves lives. Thus, we are forced to ask: “Why do so many women fail to comply with this important medical procedure?”

This research paper tackles the issue of personalized breast cancer screening. Over the last 20 years, the mammogram has been heralded as the gold standard in terms of breast cancer detection. Today, with the rapid advances of technology, physicians can take advantage of other screening modalities. Of particular note, doctors use personalized breast cancer screening which relies on a series of genetic biomarkers, family history, and other individualized indices.

The timeline for this project has been almost six months. In early December 2018, Carley Peyser reached to me as a possible faculty advisor. Given her background in biology, Carley was keen to look at the interplay of life
sciences and marketing. The overriding research questions she wanted to address were: (1) “What do young women currently know about personalized breast cancer screening?”, (2) “What impediments are working against compliance for those respondents representing Millennials or Generation Z?”, and (3) “What is the best way to approach such audiences (now and when they are of age to get screened) from a marketing standpoint?” I suggested Carley examine the relevant biomedical research pertaining to breast cancer screening. To get more insight, Carley (on her own initiative) collected primary data via Google survey. As her advisor, I can affirm that she brought enthusiasm and technical know-how to analyzing her dataset. Her exploratory findings shed some light on how best to approach young women of a certain socio-economic standing.

Carley Peyser is a SURE Fellow in the College of Computer, Mathematical & Natural Sciences at the University of Maryland, College Park, MD. Their email address is: cpeyser@terpmail.umd.edu.

Abstract

This exploratory research examines what a particular gender and age group, in this case, females ages 20-23, know about personalized breast cancer screening. This type of intervention to screening involves assessing someone’s risk for breast cancer based on genetic markers, family history, lifestyle, breast density, and other indicators to then recommend certain screening modalities and schedules for women (Onega et al., 2014). The purpose of this research is to identify exactly what these individuals know about personalized breast cancer screening, as well as to make some
reasoned predictions as to how they might take corrective action. First, the existing literature was explored to learn more about personalized breast cancer screening. Of note, this paper seeks to understand lifestyle and behavior choices of Millennial and Generation Z females (those ages 20-23).

Second, a Google survey was conducted to assess: (a) how engaged the respondents are with their health; (b) what they know about personalized breast cancer screening; (c) how likely they are to pursue screening in the future; (d) how they prefer to be educated about personalized screening moving forward. Interestingly, most of the respondents had no idea what personalized breast cancer entails. The key findings from the Google survey suggest that these individuals know very little about personalized breast cancer screening, so there are significant opportunities to rectify this situation. This study tells us there exist specific marketing opportunities or strategic windows to reach this group of women, now and when they are older.

**Introduction**

Breast cancer screening is a method of detection to spot the onset of cancer so that proper corrective measures can be taken. By doing so, a woman can extend the length and quality of her life. For many years,
mammography has been the standard method of breast cancer screening. It has been estimated that 64%-81% of persons eligible to be screened undergo the examination on a regular basis (Onega et al., 2014). Therefore, there is a sizeable percentage of women who routinely fail to take the proper steps to minimize breast cancer. The personalized approach involves taking into consideration genetics, lifestyle, breast density, and other personal factors to develop an individualized plan to get screened and manage risks associated with breast cancer (Onega et al., 2014). This approach leads to a plan that can recommend a proper screening regimen, i.e. when and how often to screen). In addition, this approach isolates which modalities (including mammograms), genetic tests, and/or imaging techniques can help women in detection and treatment (Onega et al., 2014). Doctors can analyze risk and direct the individual to the proper screening modalities. Today, there are also online risk assessment tools that individuals can access to learn more about which treatment options make sense (Keating & Pace, 2018). The personalized approach to breast cancer screening has garnered increased attention within the healthcare community. For example, there are ongoing studies, such the WISDOM study, which looks at the efficacy of a risk-based screening analysis versus screening by traditional mammograms (Keating & Pace, 2018). The United States Preventive Services Task Force recommends this personalized
approach especially when deciding how to go about getting mammograms for women 40-50 years of age. The taskforce also recommends getting biennial mammograms after age 50 (Keating & Pace, 2018). But these broad recommendations do not account for the potential individuality in breast cancer risk/ screening. That is where the personalized breast cancer screening can really help women who are at a higher risk for developing breast cancer.

Mammograms have been the standard for getting screened in the past, but, as the studies show, not everyone who should be getting screened is taking the necessary steps for their health. Concerns that people have with mammograms, include overdiagnosis, overtreatment, false positives, and financial burden (Esserman, 2017). However, it should be noted that there is an array of positive associations with mammography. For instance, it is common knowledge that a mammogram can detect breast cancer in its early stages, enabling possibilities for a better survival outcome. As explained in the literature review section of this paper, it is the author’s hope that with personalized screening more women will get screened and have a better attitude toward the entire screening process. Another focus of this research is to understand why people fail to take corrective action when it deals with their wellbeing. Staying healthy should be everyone’s agenda.
Therefore, the author remains confident that personalized screening can get more women engaged with their own health-related issues.

The respondents for this study were females 20-23 years of age. A market segmentation approach was used to better understand this group and how they might take action involving breast cancer screening when they are of the relevant age. This group is an interesting target, because it captures Millennials and Generation Z. The existing marketing literature has a lot to say about these two unique cohorts. This paper examines what implications these generations could have regarding targeted educational approaches which promote corrective action for breast cancer screening.

**Literature Review**

*Personalized Breast Cancer Screening*

In their paper, Onega et al. discuss some of the key implications of the personalized approach to breast cancer screening (Onega et al., 2014). Personalized breast cancer screening is an individualized approach to detect breast cancer. It involves examining the individual’s genetic markers, family history, lifestyle, breast density, and other factors to determine that person’s risk for breast cancer (Onega et al., 2014). Various genes can be tested for, but two genes which indicate a higher risk for getting breast cancer are BRCA1 and BRCA2 (Onega et al., 2014). Certain risk models
have been developed to determine steps women should take (Onega et al., 2014). Another paper reviewed describes other specific risk factors for breast cancer, including geographic location, age at menarche, age at menopause, age at first full pregnancy, diet, weight, consumption of alcohol, and taking exogenous hormones (such as oral contraceptives) (McPherson, Steel, and Dixon, 2000).

Additionally, multiple online resources have been established to allow women to calculate their own risk of getting breast cancer (Keating & Pace, 2018). Doctors also can use these tools to help assess patients’ risk for breast cancer and direct their patients to the next steps (Keating & Pace, 2018). Such tools include the National Cancer Institute’s Breast Cancer Risk Tool or the Breast Cancer Surveillance Consortium Breast Cancer Risk Calculator (Keating & Pace, 2018). The risk evaluation can then translate into deciding how a woman should go about getting screened, including with which screening modalities (i.e., the screening technology, including mammogram, ultrasound, MRI, or tomosynthesis) and how often to get screened (Onega et al., 2014). Also, this analysis can lead to further genetic tests, such as testing for the BRCA gene. The goal of this personalized approach is to maximize the benefits that screening can provide and to minimize the risks (Onega et al., 2014).
WISDOM Study

Researchers have been exploring the topic of personalized breast screening in recent years, as it has gotten increased attention. In her paper, Laura J. Esserman reviews the WISDOM (Women Informed to Screen Depending on Measures of Risk) Study, which involves a multidisciplinary study of physicians, policy makers, consumers, technology companies, and payers to explore the personalized breast cancer screening approach (Esserman, 2017). This study involves a clinical trial that compares risk-based (personalized) screening to the traditional annual breast cancer mammogram (Esserman, 2017). It has been going on since 2014, and its goal is to enroll 100,000 women between the ages of 40 and 74 (Esserman, 2017). In the personalized screening arm of the trial, the women explore multiple factors to determine whether they are high risk or low risk (Esserman, 2017). The high-risk individuals will get a different screening from the low risk individuals (Wisdom). Esserman states in her review that the "trial will determine whether screening based on personalized risk is as safe, less morbid, preferred by women, will facilitate prevention for those most likely to benefit, and adapt as we learn who is at risk for what kind of cancer” (Esserman, 2017). Additionally, this study is looking to gain insights that will help healthcare providers and women make the right
decisions about their breast health, instead of worrying about the debate regarding traditional screening (Esserman, 2017).

**Lifestyle of 20-23-Year-Old Females**

Various sources were consulted regarding the lifestyle and behavior of women aged 20-23. It is important to note that ages 20-23 are on the dividing line between the generational groups known as Millennials and Generation Z. Such information can help in market research and to prediction how these individuals might go about personalized breast cancer screening at the relevant time, most likely when these women reach their 40s.

When considering marketing to the different generations, Kaylene C. Williams and Robert A. Page discuss the different lifestyles of the different generations in the U.S. in their paper. They point out that Generation Z forms a group of individuals who are currently being educated (Williams & Page, 2011). Additionally, the authors state that both Millennials and Gen Z’ers value instant gratification, fitting in/ belonging, and connecting with their peers. These attributes can give insight into how these women might go about getting screened for breast cancer, as well as the best methodologies to educate these women. Additionally, both generations are considered “high-tech,” often use the Internet and social media, and use
multiple sources of information (Williams & Page, 2011). This high level of engagement with technology can definitely impact how women in this age group should be educated about screening, whether now or the years from now, when screening becomes more relevant to their life.

A survey of consumers to understand the future of healthcare, conducted by Accenture, was another source consulted for this research. That source said that the percentage of Gen Z and Millennials with primary care physicians is smaller than the percentage of older generations (Safant, Webbi & Kalis, 2019). The main reason for this is that the younger generations have issues finding primary care physicians who meet their standards of affordability and convenience (Safant, Webbi & Kalis, 2019). Additionally, the survey shows that a number of people in this age group prefer non-traditional healthcare services, such as wellness practices (i.e., yoga), walk-in clinics, virtual hatch options, digital therapeutics, and on-demand services (Safant, Webbi & Kalis, 2019).

A paper by Holly Korda and Zena Itani describes how social media can best be used for health promotion and behavior (Korda & Itani, 2013). People today use mobile devices to access health records (Korda & Itani, 2013). Other trends include looking up treatment information online and exploring internet-based health interventions (Korda & Itani, 2013). The authors also talk about how to leverage demographic profiles and
preferences to determine the type of tools and content related to health promotion on social media (Korda & Itani, 2013). This information can relate to education on breast cancer screening, because the behaviors and lifestyle of women in this paper’s target group (women ages 20-23) can impact how they are educated now and in the future.

**Research Methodology**

To answer the research questions generated by extensive background research and literature review, a Google survey was composed. The target demographic for this survey was females aged 20-23. The survey was sent out via Facebook and the University of Maryland network. Survey questions explored how often these individuals go to the doctor, how they interact with health information, what they know about breast health and breast cancer, what they know about traditional breast cancer screening involving mammograms, and what they know about the personalized approach to breast cancer screening. Additionally, questions measured how these individuals might want to be educated for breast cancer screening at the relevant time. The survey also assessed various demographic factors, including family income, ethnicity and education. The survey questions included the following:
● How engaged are you with posts, accounts, and advertisements on social media related to health and well-being?
● How often do you search the Internet for information relevant to medical conditions, symptoms, etc.?
● Within the last year did you get a medical physical?
● How often have you gotten a physical in the past 5 years?
● Have you heard of genetic testing?
● Would you (or have you) conducted such testing as 23 & Me?
● Have you ever performed a breast self-examination?
● Are you familiar with the BRCA gene linked to breast cancer?
● If you had the BRCA gene, would you be more likely to get screened when you are of the appropriate age?
● If your physician asked you questions about your lifestyle, would you be comfortable with them using such observations to determine if you were at risk for breast cancer?
● What is your knowledge of mammograms?
● If you did not answer "none" for the previous question, what is your perception of mammograms?
● What understanding do you have of the personalized approach to breast cancer screening?
● With the knowledge that your breast cancer screening could become more tailored to your family history, genetic markers, and lifestyle behaviors, as well as knowing there may be more advanced screening options available in the future, select the most appropriate response(s) based on how likely you would be to get screened for breast cancer when you are of age.
● In your opinion, what obstacles might stop you from taking corrective action after learning about health risks? Please select the response(s) that apply to your opinion.
● Please select response(s) below based on how you would want to be educated for breast cancer screening.
Results

Demographics

Fifty-two females between the ages of 20 and 23 responded. Originally, the target group was females between the ages of 20-23, but no women that were 24 or 25 responded. Therefore, this research shifted its focus to 20-23-year-old women, but this did not change the research process from the original plan. Sixteen respondents were 20 years old; 23 respondents were 21 years old; 10 respondents were 22 years old; and 3 respondents were 23 years old. Forty-one out of the 52 respondents (78.8%) are European-American or White, while the rest of the respondents are of various other ethnicities, including Latino or Hispanic, African-American or Black, Asian or Pacific Islander, Multiracial, and Mixed Race (Asian/European-American). In terms of the respondents’ household estimated yearly income while growing up, the results were mixed. Twenty-five percent of the responses included $100,000-$149,999, and 30.8% of the responses included greater than $200,000. When asked about the highest level of education they have completed or are currently pursuing, 32 responded that they have completed or are pursuing a bachelor’s degree. Sixteen responded that they have completed or are pursuing some college.
Not getting an even spread among the ages or other demographic factors are limitations to this research, but the option to look at more specific demographic segments can be future areas to explore.

*Engagement with Health*

When asked if within the last year they got a physical, 44 (84.6%) individuals responded that they have gotten a physical. Several individuals responded as to why they did not get a physical in the past year, including that they do not currently have a primary care physician, do not feel that they need it, do not have insurance, and do not trust the primary care physician. When asked how often they had gotten a physical within the past 5 years, 39 respondents (75%) said that they had gotten a physical within the past 5 years. Six respondents said they have gone every other year, and 5 respondents said that they have gone less than every other year. Some individuals shared the reasoning behind how frequently they get a physical. A handful of individuals discussed the fact that it is routine/recommended to go every year. Additionally, the survey asked if respondents ever performed a breast self-examination, and if so, how often. Twenty-seven stated they have performed a breast self-examination, while 20 said they have never done so. Five responded that they were unsure. When asked
how often these individuals perform a breast self-examination, a majority of the responses included a few times a year, once a year, and monthly.

One survey question asked the respondents about obstacles that might stop them from taking corrective action after learning about health risks. A set of responses were chosen from, and the results are summarized in *Figure 1*. The top responses include efforts/time commitment, fear of bad news and costs.

![Figure 1: This graph includes responses to the question “In your opinion, what obstacles might stop you from taking corrective action after learning about health risks?”](image)

Figure 1: This graph includes responses to the question “In your opinion, what obstacles might stop you from taking corrective action after learning about health risks?”
The survey also asked the respondents about their engagement with health information. Regarding how often they search the Internet for health-related information, there were overall mixed responses. Twenty-two responded with one to a few times a month; 11 responded with a few times a year; 10 responded with rarely; 5 responded with once a week; and 4 responded with more than once per week. When asked how engaged they are with health/well-being-related social media, there was a bell curve in the responses. Twenty-one said sometimes; 13 said often; 12 said rarely; 5 said always; and 1 said never.

**Familiarity with Health Terminology**

Forty-seven individuals responded that they have heard of genetic testing. When asked about whether they have/would conduct a genetic test such as 23 & Me, the responses were very mixed. Thirty individuals said that they have not/ would not conduct such tests, while 22 said that they have/would. When asked if they were familiar with the BRCA gene linked to breast cancer, 32 said yes, 17 said no, and 3 said maybe.
Knowledge/Perception of Mammograms

Survey participants were asked about their knowledge of mammograms.

Most respondents what they are. Some of the responses included:

- *I know they’re a screening for breast cancer.*
- *Once you are a certain age (40s?) you are supposed to get a mammogram, which is a radiology test for breast cancer. You are supposed to get them earlier/more often if you are high risk (e.g., BRCA positive).*
- *Very little. I know what they are but not how they work.*
- *I know that they test for breast cancer, but that's about it.*
- *Screenings to check for lumps in the breast*

Knowledge of the Personalized Approach to Breast Cancer Screening

When asked about knowledge of the personalized approach to breast cancer screening, 47 people said they had no knowledge. Five people responded with what they thought they knew about this screening methodology.

Here are the responses:

- *If you have a history of breast cancer in your family, you should go to a genetic counselor and get tested.*
- *Personal exams with your gyno*
- *Touching for lumps*
- *Not too familiar, but I do know that I am significantly more "at risk" because my mother's breast cancer diagnosis was before the age of 50.*
- *Partners often notice unusual lumps.*
Prediction Questions Based on Personalized Breast Cancer Screening

When asked if they had the BRCA gene, would they be more likely to get screened when they are of the appropriate age, 40 individuals said very much so, responding with a 5 on a scale from 1 to 5. Nine individuals responded with a 4, and 6 individuals responded with a 3. One individual responded with 1, meaning not at all.

Another question asked if the respondents’ physicians asked questions about their lifestyle, would they be comfortable using these observations to determine their risk for breast cancer? The responses ranged on a scale from 1 to 5, with 1 meaning “not at all” and 5 meaning “very much so.” Twenty-eight said 5; 16, said 4; 6 said 3; 1 said 2; and 1 said 1.

Without directly mentioning personalized screening, but giving a brief description of it, a question asked how likely the respondent would be to get screened. The responses are summarized in the Figure 2.
Figure 2: This graph includes responses to the question “With the knowledge that your breast cancer screening could become more tailored to your family history, genetic markers, and lifestyle behaviors, as well as knowing there may be more advanced screening options in the future, how likely would you be to get screened?"

Education

When asked to select response(s) from given choices, based on how the women would want to be educated for breast cancer screening, 49 respondents answered that they would want to be educated by word of
From their primary care physicians. The results to this survey question are represented by Figure 3.

**Figure 3:** This graph includes responses to the question “How would you like to be educated for breast cancer screening?”

**Conclusion**

A good amount of information can be drawn from both the literature analysis and survey results to answer the research questions of this paper. Based on the responses that many see their primary care physician once a year, it can be said that many do at least value their health to some extent.
Based on the research done before the survey, it was predicted that a higher percentage of individuals would not have seen a primary care physician within the past year. Thus, the results did not meet the expectations.

However, it is important to note that most of the responses came from high-income homes, so they might have the means to more frequently consult a primary care physician. This, as well as other responses, could have been different if the income range were more spread out. One factor that makes potential for this research interesting is that so many directions can be taken to explore different segments. Different and more specific demographic groups would be a good target for future research. Based on the responses that many individuals are so engaged with their health now, such as through physicals, self-examinations, and online engagement, it is the hope that they will continue to be engaged and get screened for breast cancer.

Additionally, the results show that most individuals have no idea what personalized breast cancer screening is, which is both predictable and understandable. Many of these women are not thinking about breast cancer screening at such a young age, but it might be a good idea to start thinking about this sooner to potentially increase the rates of women getting screened, as well as detecting breast cancer early. After briefly describing the nature of personalized screening in a question and asking women how likely they would be to get screened, many women said they would get
screened. Most people said that they would first want to learn more about the process and their options. Because people might be concerned that this type of screening may involve personal and sensitive information, they might be hesitant to go through all of this information. But it is important to also educate these individuals about how personalized screening can be helpful.

This research can lead to many different paths for personalized breast cancer screening. Regarding women aged 20-23 year, it can be said that there are multiple opportunities that will support them to get properly screened for breast cancer when they are of the necessary age. Education is the key component to giving these women a better understanding of the personalized approach to breast cancer screening and what their options include. Educational tools should be designed for primary care physicians to speak to these women about screening when they are of age. These tools should not be lengthy or complex (like detailed brochures), but should be brief and capture the attention of the audience. For example, these tools can even be digitized. Additionally, there is a good market opportunity for apps and social media initiatives to educate this age group, starting now and continuing to the time they should be getting screened. Also, there could be a market opportunity for a genetic test like 23 & Me, that directly can assess someone’s risk for breast cancer and give them a recommendation.
for screening modalities and scheduling. This genetic test could be paired with a questionnaire that asks about the individual’s lifestyle/behaviors and family history of breast cancer. With such tools made easily available, women might not have to leave their house to learn about their risk of getting breast cancer. Also, personalized health counseling might be a good option to educate women about their risk for breast cancer, as well as direct them toward the proper screening options.

Overall, it can be said that women who are currently 20-23 years old could definitely be engaged with the personalized breast cancer screening process when they reach the relevant age (most likely in their 40s). The key to their taking the proper action for their breast health will start with education that is tailored specifically to this age group and their needs. The key education opportunities are summarized in Figure 4. It is also important that education be tailored to other age groups and demographic segments. This would be an interesting area to explore in further research. The most important thing to consider is women taking the proper steps for their breast health, and that starts with getting screened.
Figure 4: This shows the key opportunities to educate women for personalized breast cancer screening.
Acknowledgements

I would like to thank Dr. Hank C. Boyd III, Esq. for supporting me throughout this research process. I would also like to thank Dr. Christina Elson and the SURE team for their guidance and support.
References


doi:10.1038/s41523-017-0035-5


https://doi.org/10.1177/1524839911405850


doi:10.1136/bmj.321.7261.624


doi:10.1002/cncr.28771


https://wisdom.secure.force.com/portal/WsdSiteStudy
Dispelling the Hype Over Hemp: How “Green” are the Production Opportunities for Hemp in U.S. Agriculture?

Elizabeth Thilmany

Foreword by Research Mentor, Paul Goeringer

The possibility of agricultural producers in the United States growing hemp again has become a reality with the passage of the 2018 Farm Bill. As the restrictions are lifted, more and more agricultural producers will consider growing hemp. At the same time, land grant universities have limited experience conducting research on hemp and currently can provide limited assistance to producers.

Elizabeth Thilmany has taken an initial look at one of the initial research related to hemp by considering an initial life cycle analysis. In the process of conducting this research, Ms. Thilmany determined that using hemp as a substitute for existing products may not be as environmentally friendly. I would like to thank the Snider Center's SURE program for this great opportunity of working with Ms. Thilmany. I feel privileged for the opportunity to work with her on this project.

Elizabeth Thilmany is a SURE Fellow in the College of Agriculture & Natural Resources and Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: elizabeth.thilmany@gmail.com.
Abstract

Since the production of hemp is a “new” endeavor for American farmers, there is an opportunity to address sustainability of production during the re-adoption by farmers, rather than switching to sustainable practices after establishment, as seen in the last two decades with major agriculture commodities, such as corn, soybeans and wheat. Then, a discussion is formed around the recent legalization of production of Industrial Hemp (hemp) and the potential opportunities for farmers through a legal review of the 2018 Hemp Farm Act, which establishes definitions and forms predictions around enactment. After which, the theoretical framework of Triple-Bottom Line Sustainability is used to analyze the positions of different stakeholders, specifically farmers, on hemp as a fiber product. As the hemp seed market develops, farmers could benefit from selling the seed on the market and using the fiber for on-farm applications, such as compost material, water management, or animal bedding. An Economic Input-Output Life Cycle Assessment (EIO-LCA) is used to evaluate hemp as a substitute for paper pulp in the production of paperboard. “Hemp”-board is estimated to require 73.86% more water than the equivalent volume of paperboard. Additionally, “hemp”-board has a 37.43% higher Global Warming Potential (GWP) than traditional
paperboard. From the models used in this research, hemp is not less costly to the environment when compared to the alternative fiber product used in this analysis, timber pulp. However, as this crop sector begins to be commercially grown and produced in the US, there is a need for data collection on hemp and further research.

**Background on Hemp: Legal and Research Considerations**

*An Overview of the “Hemp Farming Act of 2018” S.2667*

The Agricultural Marketing Act of 1946 had put hemp on the controlled substance list, prohibiting hemp production in the US. It was, however, legal to import hemp products. The Hemp Farming Act of 2018 amended that earlier ruling, making it legal to produce hemp. It is important to note that policy defined “industrial hemp” as the plant *Cannabis sativa* *L.*, including all parts of the plant, with a chemical composition that must not exceed 0.3% delta-9-tetrahydrocannabinol on a dry weight basis.

Sec. 297B lays the groundwork for the roll-out of hemp production via State and Tribal Plans. Ultimately, the state (through the states’ departments of agriculture) or tribe has been granted the legal power to develop, implement, and regulate the plan once it is approved at the federal level. Similar to the Clean Air Act of 1970, this process creates dual
federalism, where the federal government shifts power to the states to implement hemp production in their state under federal oversight of the plans. Given the underlying complexity of federal hemp legalization, with marijuana being legal in some states, State Plans allow state governments to decide how best to comply with the federal standards given where the current statute of a given state. Along the same lines, there is potential for state governments to limit, slow, or make hemp production illegal at the state level. A state not creating a State Plan does not prohibit production of hemp within the state, because the state will default to a set Federal Plan. Alternatively, the state could pass a state bill to make hemp production illegal. This has implications for farmers considering hemp production and for the development of the supply chain needed to process the raw plant material into useful products.

As mentioned earlier, the State or Tribal Plan is subject to set requirements. One example of a key piece of an implementation plan is testing and disposal. It is required that the Plan include a way of testing the delta-9 tetrahydrocannabinol concentration levels of hemp throughout production, as well as determining the most effective way to dispose of plant product that exceeds the 0.3% concentration level. Methods of disposal include burning and tilling the crop under. The states will need to assess the risks associated with different methods of disposal. For example,
burning a field of hemp in dry, fire-prone areas could put those communities at risk. Moreover, burning fields increases air pollutants, such as particulate matter and black carbon.

Enforcement of the Hemp Act is required when a grower or producer violates the implemented State or Tribal Plan. Types of enforcement include negligent violations, Corrective Action Plans, and accidental negligence. If a producer were to violate the plan three times within five years, they would not be allowed to grow hemp for five years. Enforcement comes across as forgiving because of uncertainty in the market. Given the complexity of growing hemp in the first years it is legalized, including no seed variety reliably able to not exceed the 0.3% threshold, the Hemp Act allows farmers to learn and improve through the process and not be penalized. However, the room for negligence and/or error could be a cause for concern for individuals taking advantage of the Hemp Act, because of a true intention to illegally grow marijuana.

Sec. 297C explicitly empowers the Secretary of Agriculture to issue further regulations and guidelines for hemp production if needed in the future. Additionally, there are funding opportunities for hemp research written into the Hemp Act: it was added to the Critical Agricultural Materials Act and the Supplemental and Alternative Crops in the National Agricultural Research, Extension, and Teaching Policy Act of 1977.
Beyond those opportunities, other funding may be targeted to initiate further research on hemp with Land Grant and other research universities, state departments of agriculture, and the United States Department of Agriculture. Depending on research findings and initial data analysis, policies can be written with greater accuracy and consideration of hemp’s unique production and marketing characteristics. Because there is a risk in starting hemp production, the federal government provides financial resources to alleviate the risks farmers are taking. Sec.5 states that hemp is added to Federal Crop Insurance, after sweet potatoes for the curious, meaning that hemp production can receive federal funding to compensate for risks in production. A waiver system is still needed to establish how hemp would fall under the Federal Crop Insurance Program.

Concluding the Act, Sec. 6 clearly removes hemp from the controlled substances list but maintains marijuana in that category. Additionally, Sec. 7 clearly states that there is to be no interference with the interstate commerce of hemp. It will be interesting to see if any states set out to make hemp production illegal and how that can coincide with interstate commerce laws.

The remainder of this project focuses on only hemp production of fiber as defined by the federal government. It is important to understand these legal definitions, new opportunities, and ever-changing constraints
with hemp production before evaluating the opportunities hemp can have for farmers and in conducting an EIO-LCA.

**Opportunities to be realized**

After reviewing a selection of research material on hemp, opportunities for analysis were realized. Three points of interest include: sustainable development through the initial adoption of the hemp industry in the US; discovering a lack of clarity on the environmental impact of the hemp-fiber industry; and realizing that the hype for hemp is derived from hemp seed production for CBD oil.

Switching to sustainable practices is costly, financially and because of the resources required of growers, producers, regulatory bodies and consumers. That said, analysis of hemp production will be conducted under a theoretical framework that provides an established definition of sustainability.

One point of contention found in previous research is whether industrial hemp should be considered a “green” substitute for fiber products. Online sites for consumers, such as blogs, social media, e-commerce sites, tout that hemp is a sustainable alternative to cotton, paper, construction material, and other fibers. One tool less utilized in the research is an Environmental Input-Output Life Cycle Assessment (EIO-LCA). An
EIO-LCA provides a framework to compile a data-driven, mathematical analysis of the environmental impact of hemp production for a given fiber product, which can be compared to potential substitute fiber products in a standardized way to evaluate estimated environmental damage. This article will present the EIO-LCA foremost to inform the potential sustainability of hemp, a in a form suitable for educating the public on the environmental impact of agricultural commodities.

In addition to fiber production, is cultivating hemp for seed from which hemp oil and further Cannabidiol (CBD) oil products are derived. Consumers are seeking the proposed health benefits that hemp oil can provide, whether in body creams or in a digestible form. Meanwhile, the supply chain enterprises, including farmers, processing facilities, distributors, and other business personnel, are attracted by the current high prices and potential profits of CBD oil. Since hemp is in the early stages of legalization and policy implementation, there is a lack of business management, decision-making information, and resources for farmers interested in adoption. Further, since there is not yet a mature marketplace for producers to sell into, systematic development of supply chain needs is necessary before successful wide-scale production of hemp crops for fiber and/or oil can be implemented. It also should be noted that, although Congress enacted the 2018 Farm Bill, there is still an opportunity for the
Food and Drug Administration to intervene in CBD oil because of its and
definition of how consumers use the product (i.e. ingesting as a pill or using
lotion, with hemp as an active ingredient). Because of the amount of
uncertainty and lack of price, crop, yield and demand data around hemp
seed to produce CBD oil, this research will focus on the production of
hemp for fiber.

**Theoretical Framework to Analyze the Sustainability of Hemp**

The recent passage of the Hemp Act of 2018 provides an opportunity to
analyze hemp and its potential to be substituted for other fiber crops that
are perceived to cause greater environmental damage. The definition of
TBL Sustainability comes from John Elkington and serves “to describe
economic, environmental, and social value of investment that may accrue
outside a firm’s financial bottom line (Elkington, 2004).” The best way to
visualize TBL Sustainability is through a Venn diagram with three
interrelated components: society, consumers, and the producer. The
convergence of the three circles can be considered the “golden area” that
allows one to jointly consider society, the consumers, and the
firms/producers and establishes the efficient production level.
In analyzing the position of producers in Figure A -- the farmers -- it is clear that there is uncertainty surrounding hemp production and supply chains. According to Kentucky’s Farmers and Agricultural Economy and the Department of Agricultural Economics at the University of Kentucky (UK), “Hemp fiber has great upside potential, but little of that potential has been realized. Hemp fiber production does not appear profitable relative to other crops given our assumed range of fiber prices and yields” (2013). Here, the price data used by UK stopped at 2011. Meanwhile, demand for hemp seed products has continued to increase, as quantified by the Congressional Report on hemp. “Researchers at the USDA and various Land Grant universities and state agencies have conducted several feasibility and marketing studies. More recently available market reports
indicate that the estimated gross value of hemp production per acre is about $21,000 from seeds and $12,500 from stalks” (Johnson 2018).

Additionally, as hemp develops, opportunities for dual cultivation may be realized, meaning that there could be a greater level of returns for the same amount of production, because both the seed and the fiber could be sold. Despite the value estimates, there are more established foreign producers and supply chains. “Hemp production in the United States also faces competition from other global suppliers. The world market for hemp products remains relatively small, and China, as the world’s largest hemp fiber and seed producer, has had and likely will continue to have major influence on market prices and thus on the year-to-year profits of producers and processors in other countries. Canada’s head start in the North American market for hemp seed and oil would also likely affect the profitability of a start-up industry in the United States” (Johnson 2018).

Therefore, the US hemp industry faces the challenge of establishing a supply chain while simultaneously competing against established, foreign hemp industries. The bottom line is that market data and development of the hemp industry are required before making an accurate judgement of profitability of hemp for producers.

In Figure A, consumers are identified as another stakeholder to consider in a cost-benefit analysis. As discussed earlier, consumers are
interested in hemp seed for health and food products. “Nevertheless, the U.S. market for hemp-based products has a highly dedicated and growing demand base, as indicated by recent U.S. market and import data for hemp products and ingredients, as well as market trends for some natural foods and body care products. Given the existence of these small-scale, but profitable, niche markets for a wide array of industrial and consumer products, commercial hemp industry in the United States could provide opportunities as an economically viable alternative crop for some U.S. growers” (Johnson 2018).

The last stakeholder in Figure A, completing the TBL analysis, is society. Society encompasses the collection of externalities that occur in the transaction between consumer and producer, in this case the hemp consumer and hemp producer. Given this definition, the environment is encompassed in the subject that is society. It is unclear what hemp production’s impact is on society. “Other studies highlight certain production advantages associated with hemp or acknowledge hemp’s benefits as a rotational crop or further claim that hemp may be less environmentally degrading than other agricultural crops” (Johnson 2018). This analysis seeks to clarify this uncertainty.
Methodology

Hemp grown as a farm resource and for dual-hemp cultivation

Before conducting an EIO-LCA comparing the environmental impact of “hemp”-board to paperboard, an understanding of how the crop grows and the inputs it requires to get the desired outputs is needed. The following analysis serves to compare hemp grown for farm resources and dual-hemp cultivation. The focus will be on what is needed for hemp to grow (seed through harvesting) for the EIO-LCA. It will be assumed that the transportation, processing and manufacturing will be comparable to other agriculture commodities used for fiber production (hemp, flax, tweed, etc.)

Hemp products that can be used as farm resources include, but are not limited to, green manure (nutrient management), animal bedding, compost agent, and water retention fiber. These products have in common their ability to stay on the farm, meaning their lifecycle is confined to the farm operation growing the hemp. Consider a permaculture approach to hemp production, using hemp as green manure. There is less need of inputs, such as water need, nutrient management and pest management, because the only purpose is to till the crop under the soil. This is important to understand through the EIO-LCA analysis. Dual-hemp cultivation, on the
other hand, is the concept of growing the hemp plant for more than one product, more specifically, for both fiber and seed.

Both hemp farm products and dual-hemp cultivation require planting hemp seed and beginning growth of the plant. Note that according to Cornell Cooperative Extension, the hemp plant will “produce greater mass for fiber production when planted early” and “have less stem length and mass when planted later for seed production” (2017). As seen in Figure B, the production and cultivation between the two will deviate when farmer uses nutrients on their crops. It should also be noted that the desired output needs to be considered before selecting which varieties to plant, as does timing of planting, as some varieties produce better quality fiber crop and others a better oilseed crop. Research and development on hemp varieties are needed to establish which varieties should be used to achieve the end goal for the intended hemp product.

Before planting, there are considerations that impact the EIO-LCA. When it comes to preparing a field and planting the seed, “Hemp production is a relatively simple process in that it requires a well-established seed bed, and planting can be accomplished with existing planting equipment, whether for seed or for fiber” (Thayer, Burley, Held, Parr 2017). This means farmers do not have to incur a cost associated with obtaining new planting equipment. A more unique factor of the hemp plant,
according to Cornell Cooperative Extension, is that the requirement for water is greater at the early stages of plant development, and as the plant grows past the early development stage, it becomes more drought resistant (2017). Therefore, climates with wetter spring and dryer summers may find hemp a less water-intensive crop than other grain or oilseed crops. It should be noted that the EIO-LCA will run a comparison using trans-continental data, therefore it is not possible to separate regions or states.

Nutrient management can be considered a “hot spot” along the supply chain for fiber production from crop plants, because of the requirements in producing the phosphorus and nitrogen fertilizers and the outputs that can lead to eutrophication and other environmental problems when the fertilizers interact with groundwater and other ecosystems.

In designing a hybrid model, it is necessary to understand how hemp assimilates other crops in its production, since there is no current reliable data on hemp as a crop. In terms of nutrient management, “Hemp is very similar to corn in its nutrient requirements. A lack of nitrogen greatly decreases mass for fiber production” (Thayer, Burley, Held, Parr 2017). Other similarities hemp has to corn are the preferred humidity levels, soil conditions, cross-pollination, and the use of monecious plants for crop production. Depicted in Figure B, notice that the nutrient requirements are not an input to be considered when growing hemp as a green manure crop.
Again, the use of hemp as a crop affects the environmental impact of hemp by having the EIO-LCA model adjust for the inputs required for the desired output.

One potential way for farmers to reduce the environmental impact of nitrogen fertilizers is to use manure from CAFO’s. The capabilities of the EIO-LCA model, it will not be possible to factor in the use of manure as a fertilizer, but it is important to understand how sustainable farming practices can reduce the environmental damage sourced from crop production. Framed by the EPA’s EnviroAtlas, “The high nutrient (nitrogen and phosphorus) and organic matter content of manure makes it a good organic fertilizer. Use of manure also recycles some of the nutrients that otherwise would be lost to the air or to aquatic ecosystems” (2014). The importance of sourcing fertilizers for crop production to reduce environmental impact is further solidified by Gonzalez-Garcia, Hospido, Feijoo, Moreira, “The use of Nitrogen-based fertilizers (ammonium nitrate for hemp and calcium ammonium nitrate for flax) is an important source of nature of nutrient related emissions. In some situations, the type of fertilizers used is the main driver of the emissions at the whole farm level, and changing the type of fertilizers could reduce emissions and therefore, the environmental impact” (2010).
There are differences in the harvesting processes of hemp grown for seed and hemp grown for fiber, as seen in Figure B. For seed harvesting, existing farm equipment generally can be used. “Harvesting hemp seed will require some modification of existing combines so as to harvest the tops of the plants and leaving the bulk of the stalk” (Robbins, Snell, Halich, Maynard, Dillon, Spalding 2013). Meanwhile, the harvesting of seed may require new or custom machines, or variations of other farm equipment. “Harvesting of the stalk will likely require a sickle bar mowing machine as opposed to the disk mowers whose blades rotate. The stalks will be allowed to dry to the point that any leaves will shatter when the crop is baled” (Robbins, Snell, Halich, Maynard, Dillon, Spalding 2013). Dual-cultivation cropping will require further work in analyzing how to best harvest the seed, the more valuable part of the plant, without destroying the fiber structure necessary for other manufacturing needs. However, using hemp for composting, water management, hay, animal bedding, etc. does not require the same level of preservation of the fiber structure, meaning that a farmer could harvest for seed, then use the leftover stock for on-farm fiber uses.

Storage is another element of hemp cultivation that needs to be considered in the EIO-LCA. Continuing to follow the trend of assimilating other crop processes, hemp seed storage resembles other oilseed storage
requirements as examined by UK. “Storage of hemp seed will be essentially the same as for any other oil seed crop” (Robbins, Snell, Halich, Maynard, Dillon, Spalding 2013). Again, hemp fiber requires different processes than seed production. For example, when being processed for farm-use, the hemp requires baling that involves different consideration. “The current process for fiber separation prefers a longer stalk produced by the round bales as opposed to square bales. After baling, the bales will need to be protected from the weather” (Robbins, Snell, Halich, Maynard, Dillon, Spalding 2013). Dual-cultivation would require storage processes and facilities for the seed and the fiber. The requirements of dual-cultivation of hemp reveal that there are greater equipment, facilities, energy, and labor requirements for the same-sized field.

From this point the seed would be sold on a market for processing, and the fiber would be used on the farm or be sold to a processing facility that would decorticate the hemp. The processes and requirements to grow hemp, now evaluated, will be translated into the EIO-LCA model. Further, the processing, manufacturing, transportation, and other components of the supply chain will be added.
Cataloging Industrial Hemp’s Inputs

**Inputs**
- Land
- Harrow Fields
- Treated Seed*
- Rainwater or Irrigation*
- Sunlight
- Nutrients*
- Weed Control* (chemical or mechanical)
- Farm Equipment
- Fuel/ Energy
- Labor

**Outputs**
- Oxygen & Green Manure
- Hemp Seed
- Hemp Husk (fiber)

*Optional treatments: varied uses depend on the farmer’s purpose for growing the plant. These inputs may require additional inputs that may not be accounted for in the LCA.

Figure B: Collecting the inputs and outputs of hemp production
Figure B catalogs the inputs and outputs of hemp production. The solid lines act as sum lines that result in an output/product of hemp, with the inputs above the line accounting for all of the inputs required for the output to result. The intention of the producer of the hemp results in the different output, thus different input requirements, as illustrated in Figure B. For example, farmers planning a trial run with hemp may plan to use hemp as green manure. The flowchart in Figure B can be a tool to understand the inputs needed to produce the output, green manure in this case. For the EIO-LCA, hemp as a fiber product will be evaluated and therefore requires all of the inputs listed, with the most likely marked with an asterisk.

Defining the EIO-LCA Model

An EIO-LCA is used to evaluate hemp as a substitute for paper pulp in constructing paperboard, a manufactured fiber product used primarily to construct cardboard boxes and pallets. The EIO-LCA calculates the direct and indirect environmental impacts related to the other sectors.

Producer data model, rather than the purchaser data model, was chosen, because the economic matrices data is used to evaluate the difference in environmental impact between “hemp”-board and paperboard. The major difference between producer and purchaser data is that taxes,
subsidies, and profit margins, among other components, can be found in the purchaser price but not in the producer prices. For the sake of an environmental analysis and comparison, the key difference between the producer and purchaser data is transportation of the finished good. The purchaser data would include the impact of the transportation to the final destination and, possibly, the warehousing of that good. Although this is an important element in comparing hemp v. paper board, producer data was chosen because “hemp”-board is a theoretical product at the time of this analysis, therefore an accurate estimation of the difference in the environmental impact of the last stages -- the final mile -- of the good is unlikely. Additionally, the difference in environmental damage of transporting “hemp”-board vs. paperboard to the consumer is predicted to be marginal.

In selecting the amount of economic activity for the sector, $1 million was used. Because the model analyses are standardized to one value for the economic activity, the value used will not make a difference in comparing the environmental impact of hemp board and paperboard, as the units will cancel out.

After cataloging the inputs and outputs for the given product, the inputs will be translated into total industry output vectors represented by:
\[ \underline{x} = [(I - A)^{-1}]f \]

(1)

The parameters are defined as: \( I \), the diagonal identity vector; \( f \), the final demand vector; and \([(I - A)^{-1}]\) the total requirement matrix, also known as the Leontief Inverse.

Each industry that has an output associated with the production of the fiber product being evaluated will have an \( \underline{x} \) value. For example, water required for the growing of hemp would be \( x_{\text{water}} \).

Once the total industry output vector is calculated, the model will calculate the vector of total environmental outputs associated with a given industry’s output and will be depicted as:

\[ \underline{r} = E_{\text{dir}} \underline{x} = E_{\text{dir}}[(I - A)^{-1}]f \]

(2)

With \( \underline{r} \) representing the overall sustainability impacts per unit of final demand (\( \underline{x} \)), given that \( E_{\text{dir}} \) is given an industrial sector’s direct environmental impact per dollar of output from the firm.

Resulting in the final step, taking the product of \( E_{\text{dir}} \) and \([(I - A)^{-1}]\) to create a multiplier matrix.
Creating the Hybrid Model for “Hemp”-Board

The difference between the Paperboard Mill and hypothetical “Hemp”-Board Mill was that the Paperboard Mill requires inputs from logging and forestry services, whereas the “Hemp”-Board Mill requires inputs from grain, oilseed and all other crop farming. This is the source of the justification used for this analysis: comparing “hemp”-board to paperboard.

The Green Design Institute of Carnegie Mellon University has a US 2002 Benchmark Model for paperboard that is “built upon the inter-sector input-output transactions of the US economy as compiled by the Bureau of Economic Analysis (BEA) of the US Department of Commerce (BEA 2008)” (Weber, Matthews, Venkatesh, Costello, Matthews 2010).

Since hemp assimilates corn for nutrient and water management and also assimilates oilseed farming for production purposes, a decision had to be made on how to distribute the direct economic demand value from the tree production sectors to most appropriately estimate hemp. Table C depicts different weights applied to the demand oilseed and grain sector demand vector. The Global Warming Potential (GWP) is the output and
was generated by the hybrid EIO-LCA model that was adjusted with the changed demand vectors.

The lowest GWP value was 2,510,000 CO₂ equivalents, which is when the oilseed sector demand was at the lowest. As the demand was increased by the weighting parameter, GWP increased. The GWP calculated by the 2002 Paperboard model is 1,950,000 CO₂ equivalents, meaning that no matter the adjustments to the oilseed and grain demand vectors, paperboard has a smaller GWP.

The final version of the “hemp”-board model used was with $\theta=0.5$, for two reasons. First, the GWP of 2680000 was the average of all the GWP’s calculated. Second, hemp resembles the grain industry for early stages of plant production, including input requirements. Then hemp more resembles oilseed for harvesting and processing of the plant material, seeming that the grain sector resembled half of the hemp plant’s life and the oilseed sector resembling the other half.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$(1-\theta)$</th>
<th>$\chi_{oilseed}$</th>
<th>$\chi_{grain}$</th>
<th>GWP in CO₂ equivalents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>3.9704</td>
<td>241.6011</td>
<td>2510000</td>
</tr>
<tr>
<td>0.25</td>
<td>0.75</td>
<td>64.1789</td>
<td>181.3926</td>
<td>2590000</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>124.3874</td>
<td>121.1841</td>
<td>2680000</td>
</tr>
</tbody>
</table>
Table C depicts the change of the weighting parameter applied to the economic demand vector inputted into the EIO-LCA model and the resulting GWP from the EIO-LCA, with no other variables (demand vectors) changed.

\[
(\theta \sum_{k=0}^{1M} \Delta x) + x_{\text{grain}} = x_{\text{grain}_0}
\]

\[
((1 - \theta) \sum_{k=0}^{1M} \Delta x) + x_{\text{oilseed}} = x_{\text{oilseed}_0}
\]

With \(0 \leq \theta \leq 1\)

\(\theta\) is used as a weighting parameter and, in the context of building the hybrid model, is the variable to adjust to estimate the proportion of the aggregate deleted values attributed to the oilseed sector and grain sector.

\(\Delta x\) represents the changes, in this case deletions, to units of demand.

Therefore, taking the sum represents the aggregate deletions to the other direct economic sectors.
\( \mathbf{x}_{\text{name of sector}} \) is the demand vector for the given sector originating from the 2002 Producer Paperboard model.

\( \mathbf{x}_{\text{name of sector}_0} \) is the resulting demand vector to be used in the “hemp”-board Hybrid Model based on the 2002 Producer data.

Greenhouse Gases is an abbreviation for Global Warming Contribution attributed to the “hemp”-board’s entire EIO-LCA, measured in kilograms of CO\(_2\) equivalents based on the Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts (TRACI) Impact Assessment developed by the Environmental Protection Agency (EPA). The GWP is a calculation of carbon dioxide and carbon dioxide equivalents: methane and nitrous oxide over a defined period of time. Best summarized by the EPA’s User’s Guide, “TRACI 2.0 allows the quantification of stressors that have potential effects, including ozone depletion, global warming, acidification, eutrophication, tropospheric ozone (smog) formation, human health criteria-related effects, human health cancer, human health noncancer, ecotoxicity, and fossil fuel depletion effects” (2012). For a further analysis and to see the calculations used to create the TRACI model, reference the TRACI version 2.1 User’s Guide. Evaluating GWP as an output of the EIO-LCA corresponds with the theoretical framework of sustainability, because TRACI accounts for the time dimension of the air pollutants and creates a common unit of
measurement for environmental damage among an array of air pollutants for evaluation purposes.

As seen below, the changes are factored into the entire supply chain. With hemp being a new product lacking standardized data over time, the changes made were limited and are made with a justifiable purpose. Justifications are listed below.

- Logging changed from $146.2 million of direct economic impact (demand) to $0 million of direct economic impact (demand).
- Sawmills and wood preservation changed from $63.391 million of direct economic impact (demand) to $0 million of direct economic impact (demand).
- Forest nurseries, forest products and timber tracts changed from $31.243 million of direct economic impact (demand) to $0 million of direct economic impact (demand).
- Grain farming changed from $3.9704 million of direct economic impact (demand) to $244.8044 million of direct economic impact (demand).
- Oilseed farming changed from $0.76711 million of direct economic impact (demand) to $244.8044 million of direct economic impact (demand).
<table>
<thead>
<tr>
<th>Sectors with adjusted value from the Paperboard Model</th>
<th>Changes in Direct Economic Impact in million dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paperboard Mills</td>
<td>1000000</td>
</tr>
<tr>
<td>Oilseed farming</td>
<td>120417</td>
</tr>
<tr>
<td>Grain farming</td>
<td>120417</td>
</tr>
<tr>
<td>Logging</td>
<td>-146200</td>
</tr>
<tr>
<td>Forest nurseries, forest products and timber tracts</td>
<td>-31243</td>
</tr>
<tr>
<td>Sawmills and wood preservation</td>
<td>-63391</td>
</tr>
</tbody>
</table>

*Table C illustrates the changes made to the paperboard model to create the estimate “hemp”-board model.*
Results

With the already established paperboard producer 2002 model and a hybrid “hemp”-board model based on the producer 2002 data, the EIO-LCA was run, and results were generated for water usage, air pollutants and TRACI CO₂ equivalents.

Figure D shows the breakdown of water withdrawals required for the complete production of “hemp”-board. According to the CMU model, “Water footprint is a measure of direct and indirect water used by each sector. The EIO-LCA model uses the United States Geological Survey (USGS) data to estimate direct water withdrawals for each consumption category, such as power generation, irrigation, industrial, livestock and aquaculture, mining, public supply, and domestic water use” (Blackhurst et al., 2010; CMU, 2002). Approximately 75% of water withdrawals for “hemp”-board is sourced from the grain farming sector. The methodology of creating the hybrid model for hemp, grain farming and oilseed farming was used to estimate the environmental effects of the hemp sector. The second largest user of water for “hemp”-board is the paperboard mills sector, meaning the water required in the final process of turning the pulp into the “hemp”-board product. As seen for the paperboard product, the largest water withdrawal comes from the paperboard sector. Only the plant product was changed between the hybrid model and the paperboard model,
therefore the paperboard mills sector water withdrawal is the same between “hemp”-board and paperboard. The supply chain still needs to develop before knowing if this is an accurate prediction, however, the evidence supports that hemp pulp and paper pulp do not have a significant difference in processing into the final “board” product.

The most important conclusion from Figure D is that hemp requires 73.86% more water than traditional paperboard. The difference comes from the water requirements for the cultivation of hemp, accounted for in the grain sector. It should be noted that adjusting the weighting parameter to have hemp align more with the oilseed sector, say 0.75 to oilseed demand vector and 0.25 to grain sector, may result in less water withdrawal, but it would still be more significant than the traditional paperboard water requirement.
Figure D is a graphic comparison of the water requirements of “hemp”-board compared to paperboard broken down by the top seven water-demanding sectors.

The next output of the EIO-LCA to evaluate are the air pollutants, as seen in Figure E. The air pollutants are greater for “hemp”-board, except for NO\textsubscript{x}, which is 7.4% lower than paperboard. Understandably, NH\textsubscript{3} is higher for hemp because burning the fields will be required if the field exceeds the THC threshold backed by the “2002 US Benchmark Version of the Economic Input-Output Lifecycle Assessment (EIO-LCA) Model” documentation: “Methane emissions due to field burning were associated
by the EPA with specific crops, which were matched to NAICS sectors” (2010). However, interpreting the TRACI results is more productive than analyzing each category of air pollutant, because it links the environmental impact to the product.

*Figure E illustrates the different air pollutant emittances between “hemp”-board and paperboard.*

**Conventional Air Pollutants Comparison: "Hemp-Board" and Paperboard**

<table>
<thead>
<tr>
<th>Categories of Air Pollutants</th>
<th>&quot;Hemp-Board&quot; Total for all sectors</th>
<th>Paper-Board Total for all sectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>13.6</td>
<td>8.94</td>
</tr>
<tr>
<td>NH3</td>
<td>0.408</td>
<td></td>
</tr>
<tr>
<td>NOx</td>
<td>5.73</td>
<td>6.19</td>
</tr>
<tr>
<td>PM10</td>
<td>9.52</td>
<td></td>
</tr>
<tr>
<td>PM2.5</td>
<td></td>
<td>2.26</td>
</tr>
<tr>
<td>SO2</td>
<td>7.05</td>
<td>6.74</td>
</tr>
<tr>
<td>VOC</td>
<td>3.13</td>
<td>2.58</td>
</tr>
</tbody>
</table>

Moving to the graph of TRACI outputs for “hemp”-board v. paperboard in Figure F, again, paperboard has lower levels of environmental damage then “hemp”-board. For GWP, “hemp”-board is
37.43% higher than paperboard’s GWP. TRACI, combined with the EIO-LCA model, shows that the overall damage from various pollutants and accounted for over time of “hemp”-board’s lifecycle is greater than the damage from the paperboard’s lifecycle. Bringing it back to the TBL discussion, hemp as an alternative to tree pulp for the production of paperboard is not the socially optimal choice.

Figure F: TRACI results for paperboard and “hemp”-board
The model was based on a 2002 source and had broad sectors. It seemed appropriate to do an analysis of hemp farming compared to logging with the 2007 producer price data sourced from the EPA. A comparison between timber and raw forest products (timber for short) sector and fresh soybeans, canola, flaxseeds, and other oilseeds (agriculture for short) sector allows for a “zoomed-in” look at the differences between the two sectors. Additionally, there is the benefit in checking the results of the hybrid model based on estimates and older data with an established model based on the EPA’s 2007 agriculture data.

The air pollutant emissions are displayed in Figure G. When looking at whether the agriculture sector or timber sector had greater air pollutants, there are two differences: CO is greater for the timber sector, and NO\textsubscript{x} is approximately the same between the sectors, with the agriculture sector being slightly greater.

Running an EIO-LCA of the timber sector and agriculture sector and comparing the results show that the agriculture sector, overall, is releasing more air pollutants.
**Figure G: Air pollutant comparison of the timber and agriculture sector**

The bottom-line is that the EIO-LCA “hemp”-board model was built on estimations. The values of air pollutants, water usage, and global warming impacts should be evaluated in comparison to paperboard rather than examined at face-value. In doing so, hemp pulp causes more environmental damage than tree pulp in the creation of a paperboard product.

**Concluding Remarks**

As discussed in the methodology section, the data on hemp production is extremely limited within the US. In building the “hemp”-board EIO-LCA model, the demand vectors were configured on estimates.
of hemp’s use of resources and the economic impact on those given sectors. Therefore, the results should be interpreted understanding that they are estimates on what is known of hemp and how the model was constructed to interpret “hemp”-board production.

The data used is also a decade or older. This is less of a concern for this analysis of an agriculture commodity processed for fiber production compared to timber products, because the EIO-LCA was done in a way that standardized the processing and manufacturing, meaning increases in efficiency in manufacturing would be realized, proportionally, by both industries. If there have been changes in the sustainability of logging operations since the data collected in 2002 and then 2007, there could be differences in the results of the analysis.

Only pulp products were being compared in this analysis, but the results could be considered for other fiber products for which hemp is being considered an alternative, including flax and other oilseed crops for linen, cotton for fabric in clothing, and hardy fibers for yarn and rope. Separate EIO-LCA’s should be completed to compare hemp to these products to determine whether hemp is a more sustainable alternative. Future research also is required for plant breeding to improve seed reliability, meaning the seed does not produce a plant that crosses the THC threshold, and that it yields plants that lead to high quality products, especially for dual-purpose
hemp cultivation. The development and stabilization of the hemp seed and fiber market. Once a more established market has formed, further analysis will need to be done to evaluate the profitability of hemp as a crop. Another suggestion is to shift the focus to recycling and repurposing fiber materials, such as pulp, rather than switching/adding hemp to paperboard production or other fiber products.

The bottom-line is that there are opportunities for farmers as hemp production in the US becomes legal. It is important to consider sustainable development of hemp and its production, because the entirety of the supply chain will need to be rebuilt. An EIO-LCA is a powerful model for analyzing the environmental impacts of production for a specific good. The EIO-LCA results comparing hemp to tree pulp for paperboard show that hemp is not a more sustainable alternative because “hemp”-board is estimated to require 73.86% more water than the equivalent volume of paperboard. Additionally, “hemp”-board has a 37.43% higher Global Warming Potential (GWP) compared to traditional paperboard.
Appendices

The Data Table results for the “hemp”-board EIO-LCA model

Appendix T: “hemp”-board air pollutants data table from EIO-LCA

<table>
<thead>
<tr>
<th>Top Ten Emitting Sectors</th>
<th>CO</th>
<th>NH3</th>
<th>NOx</th>
<th>PM10</th>
<th>PM2.5</th>
<th>SO2</th>
<th>VOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain farming</td>
<td>5.14</td>
<td>2.44</td>
<td>0.546</td>
<td>4.34</td>
<td>1.12</td>
<td>0.157</td>
<td>0.6</td>
</tr>
<tr>
<td>Paperboard Mills</td>
<td>4.88</td>
<td>0.142</td>
<td>3</td>
<td>1.04</td>
<td>0.753</td>
<td>3.89</td>
<td>1.71</td>
</tr>
<tr>
<td>Oilseed farming</td>
<td>2.21</td>
<td>0.76</td>
<td>0.456</td>
<td>3.82</td>
<td>0.757</td>
<td>0.128</td>
<td>0.357</td>
</tr>
<tr>
<td>Natural gas distribution</td>
<td>0.52</td>
<td>0</td>
<td>0.023</td>
<td>0.002</td>
<td>0.001</td>
<td>0.007</td>
<td>0.023</td>
</tr>
<tr>
<td>Truck transportation</td>
<td>0.208</td>
<td>0</td>
<td>0.219</td>
<td>0.063</td>
<td>0.011</td>
<td>0.005</td>
<td>0.023</td>
</tr>
<tr>
<td>Oil and gas extraction</td>
<td>0.198</td>
<td>0</td>
<td>0.144</td>
<td>0.001</td>
<td>0.001</td>
<td>0.01</td>
<td>0.201</td>
</tr>
<tr>
<td>Paper mills</td>
<td>0.156</td>
<td>0.004</td>
<td>0.12</td>
<td>0.03</td>
<td>0.021</td>
<td>0.188</td>
<td>0.052</td>
</tr>
<tr>
<td>Power generation and supply</td>
<td>0.135</td>
<td>0.006</td>
<td>0.975</td>
<td>0.136</td>
<td>0.11</td>
<td>2.17</td>
<td>0.01</td>
</tr>
<tr>
<td>Other industrial machinery</td>
<td>0.09</td>
<td>0</td>
<td>0.004</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.005</td>
</tr>
<tr>
<td>manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household goods repair and</td>
<td>0.089</td>
<td>0</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.007</td>
</tr>
<tr>
<td>maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total for all sectors</td>
<td>13.6</td>
<td>3.34</td>
<td>5.73</td>
<td>9.52</td>
<td>2.78</td>
<td>7.05</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Emissions of air pollutants to air from each sector (in metric tons) from top 10 emitting sectors
Appendix U: “hemp”-board water withdrawal data table from EIO-LCA

<table>
<thead>
<tr>
<th>Water withdrawals in thousands of gallons for top ten sectors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain farming</td>
<td>160000</td>
</tr>
<tr>
<td>Paperboard Mills</td>
<td>34900</td>
</tr>
<tr>
<td>Power generation and supply</td>
<td>14900</td>
</tr>
<tr>
<td>Other</td>
<td>5200</td>
</tr>
<tr>
<td>Cotton farming</td>
<td>2590</td>
</tr>
<tr>
<td>Oilseed farming</td>
<td>1140</td>
</tr>
<tr>
<td>Pesticide and other agricultural chemical manufacturing</td>
<td>293</td>
</tr>
<tr>
<td>Paper mills</td>
<td>190</td>
</tr>
<tr>
<td>Other basic organic chemical manufacturing</td>
<td>184</td>
</tr>
<tr>
<td>All other basic inorganic chemical manufacturing</td>
<td>171</td>
</tr>
<tr>
<td>Sand, gravel, clay and refractory mining</td>
<td>169</td>
</tr>
<tr>
<td>Total for all sectors</td>
<td>215000</td>
</tr>
<tr>
<td>Top Ten Polluting Sectors</td>
<td>Glob Warm kg CO2e</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Paperboard Mills</td>
<td>975000</td>
</tr>
<tr>
<td>Power generation and supply</td>
<td>525000</td>
</tr>
<tr>
<td>Grain farming</td>
<td>419000</td>
</tr>
<tr>
<td>Oilseed farming</td>
<td>268000</td>
</tr>
<tr>
<td>Oil and gas extraction</td>
<td>78100</td>
</tr>
<tr>
<td>Fertilizer Manufacturing</td>
<td>58300</td>
</tr>
<tr>
<td>Petroleum refineries</td>
<td>33600</td>
</tr>
<tr>
<td>Coal mining</td>
<td>33300</td>
</tr>
<tr>
<td>Other basic organic chemical manufacturing</td>
<td>28600</td>
</tr>
<tr>
<td>Pipeline transportation</td>
<td>26200</td>
</tr>
<tr>
<td>Total for all sectors</td>
<td>2680000</td>
</tr>
</tbody>
</table>
### Data Results Table for Paperboard EIO-LCA’s

**Appendix X: air pollutant data table for paperboard EIO-LCA**

Emissions of air pollutants to air from each sector (in metric tons) from top 10 emitting sectors

<table>
<thead>
<tr>
<th>Top Ten Emitting Sectors</th>
<th>CO</th>
<th>NH3</th>
<th>NOx</th>
<th>PM10</th>
<th>PM2.5</th>
<th>SO2</th>
<th>VOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paperboard Mills</td>
<td>4.88</td>
<td>0.142</td>
<td>3</td>
<td>1.04</td>
<td>0.753</td>
<td>3.89</td>
<td>1.71</td>
</tr>
<tr>
<td>Forest nurseries, forest products and timber tracts</td>
<td>0.619</td>
<td>0</td>
<td>0.58</td>
<td>0.068</td>
<td>0.066</td>
<td>0.06</td>
<td>0.088</td>
</tr>
<tr>
<td>Logging</td>
<td>0.499</td>
<td>0</td>
<td>0.141</td>
<td>0.015</td>
<td>0.013</td>
<td>0.026</td>
<td>0.08</td>
</tr>
<tr>
<td>Natural gas distribution</td>
<td>0.493</td>
<td>0</td>
<td>0.022</td>
<td>0.001</td>
<td>0.001</td>
<td>0.007</td>
<td>0.022</td>
</tr>
<tr>
<td>Truck transportation</td>
<td>0.227</td>
<td>0</td>
<td>0.24</td>
<td>0.069</td>
<td>0.012</td>
<td>0.005</td>
<td>0.025</td>
</tr>
<tr>
<td>Agriculture and forestry support activities</td>
<td>0.215</td>
<td>0</td>
<td>0.003</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.018</td>
</tr>
<tr>
<td>Oil and gas extraction</td>
<td>0.171</td>
<td>0</td>
<td>0.124</td>
<td>0.001</td>
<td>0.001</td>
<td>0.008</td>
<td>0.173</td>
</tr>
<tr>
<td>Paper mills</td>
<td>0.156</td>
<td>0.004</td>
<td>0.12</td>
<td>0.03</td>
<td>0.021</td>
<td>0.188</td>
<td>0.052</td>
</tr>
<tr>
<td>Grain farming</td>
<td>0.151</td>
<td>0.072</td>
<td>0.016</td>
<td>0.128</td>
<td>0.033</td>
<td>0.005</td>
<td>0.018</td>
</tr>
<tr>
<td>Power generation and supply</td>
<td>0.128</td>
<td>0.006</td>
<td>0.922</td>
<td>0.129</td>
<td>0.104</td>
<td>2.05</td>
<td>0.009</td>
</tr>
<tr>
<td>Total for all sectors</td>
<td>8.94</td>
<td>0.408</td>
<td>6.19</td>
<td>2.26</td>
<td>1.21</td>
<td>6.74</td>
<td>2.58</td>
</tr>
</tbody>
</table>
Appendix Y: water withdrawals data table for paperboard EIO-LCA

<table>
<thead>
<tr>
<th>Water withdrawals in thousands of gallons for top ten sectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paperboard Mills</td>
</tr>
<tr>
<td>Power generation and supply</td>
</tr>
<tr>
<td>Grain farming</td>
</tr>
<tr>
<td>All other crop farming</td>
</tr>
<tr>
<td>Cotton farming</td>
</tr>
<tr>
<td>Paper mills</td>
</tr>
<tr>
<td>Other basic organic chemical manufacturing</td>
</tr>
<tr>
<td>Sand, gravel, clay, and refractory mining</td>
</tr>
<tr>
<td>All other basic inorganic chemical manufacturing</td>
</tr>
<tr>
<td>Synthetic dye and pigment manufacturing</td>
</tr>
<tr>
<td>Total for all sectors</td>
</tr>
</tbody>
</table>
## Tools for Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI) for the top ten sectors

<table>
<thead>
<tr>
<th>Top Polluting Sectors</th>
<th>Glob Warm kg CO2e</th>
<th>Acidif Air kg SO2e</th>
<th>HH Cit Air kg PM10e</th>
<th>Eutro Air kg Ne</th>
<th>Eutro Water kg Ne</th>
<th>OzoneDep kg CFC-11e</th>
<th>Smog Air kg O3e</th>
<th>EcoTox (low) kg 2,4D</th>
<th>HH Cancer (low) kg benzene eq</th>
<th>HH NonCancer (low) kg toluene eq</th>
<th>EcoTox (high) kg 2,4D</th>
<th>HH Cancer (high) kg benzene eq</th>
<th>HH NonCancer (high) kg toluene eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paperboard Mills</td>
<td>975000</td>
<td>6670</td>
<td>2270</td>
<td>150</td>
<td>14.8</td>
<td>0.544</td>
<td>80800</td>
<td>6.07</td>
<td>81</td>
<td>28600</td>
<td>7.75</td>
<td>971</td>
<td>128000</td>
</tr>
<tr>
<td>Power generation and supply</td>
<td>496000</td>
<td>2850</td>
<td>564</td>
<td>41.5</td>
<td>0.007</td>
<td>0</td>
<td>22900</td>
<td>5.76</td>
<td>8.44</td>
<td>4750</td>
<td>5.78</td>
<td>25.2</td>
<td>25000</td>
</tr>
<tr>
<td>Oil and gas extraction</td>
<td>67200</td>
<td>106</td>
<td>6.5</td>
<td>5.51</td>
<td>0</td>
<td>0</td>
<td>3710</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Coal mining</td>
<td>32400</td>
<td>77.1</td>
<td>26.7</td>
<td>3.37</td>
<td>0.022</td>
<td>0</td>
<td>1910</td>
<td>1.3</td>
<td>5.28</td>
<td>2110</td>
<td>1.3</td>
<td>76.1</td>
<td>79900</td>
</tr>
<tr>
<td>Truck transportation</td>
<td>27400</td>
<td>195</td>
<td>83.7</td>
<td>10.7</td>
<td>0</td>
<td>0</td>
<td>6050</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Petroleum refineries</td>
<td>26400</td>
<td>63.3</td>
<td>13.2</td>
<td>1.17</td>
<td>0.021</td>
<td>0.003</td>
<td>667</td>
<td>0.066</td>
<td>0.731</td>
<td>176</td>
<td>0.089</td>
<td>1.99</td>
<td>211</td>
</tr>
<tr>
<td>Paper mills</td>
<td>26100</td>
<td>295</td>
<td>78.5</td>
<td>5.8</td>
<td>0.194</td>
<td>0.021</td>
<td>3170</td>
<td>0.213</td>
<td>2.45</td>
<td>805</td>
<td>0.272</td>
<td>23.6</td>
<td>3520</td>
</tr>
<tr>
<td>Other basic organic chemical manufacturing</td>
<td>26000</td>
<td>95.5</td>
<td>19.7</td>
<td>2.71</td>
<td>0.094</td>
<td>0.188</td>
<td>1530</td>
<td>0.123</td>
<td>2.18</td>
<td>271</td>
<td>0.427</td>
<td>10.1</td>
<td>668</td>
</tr>
<tr>
<td>Pipeline transportation</td>
<td>24400</td>
<td>5.55</td>
<td>0.302</td>
<td>0.31</td>
<td>0</td>
<td>0</td>
<td>263</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Natural gas distribution</td>
<td>16500</td>
<td>24.5</td>
<td>4.04</td>
<td>0.982</td>
<td>0</td>
<td>0</td>
<td>649</td>
<td>0.001</td>
<td>0.003</td>
<td>0</td>
<td>0.007</td>
<td>0.029</td>
<td>0</td>
</tr>
<tr>
<td>Total for all sectors</td>
<td>1950000</td>
<td>12600</td>
<td>4350</td>
<td>323</td>
<td>15.4</td>
<td>1.24</td>
<td>163000</td>
<td>53.3</td>
<td>209</td>
<td>95300</td>
<td>57.4</td>
<td>1530</td>
<td>1050000</td>
</tr>
</tbody>
</table>
References


University of Kentucky, Department of Agricultural Economics. 2013. “Economic Considerations for Growing Industrial Hemp: Implications for Kentucky’s Farmers and Agricultural Economy.” July.
The Great Recession: Impact on Commercial Real Estate Leasing Markets

Shreyas Vaidya

Shreyas Vaidya is a SURE Fellow in the Department of Computer Science and Department of Economics at the University of Maryland, College Park, MD. Their email address is: shreyasy@terpmail.umd.edu.

Acknowledgements:

I would like to thank my research mentor, Dr. Susan White, from the Finance Department of the Robert H. Smith School of Business, as well as the entire Ed Snider Center for Enterprise and Markets for assisting me throughout this research process.

Abstract:

The aim of this research paper is to analyze the effects of the Great Recession with specific respect to commercial real estate leasing markets and loan pricing in relation to commercial real estate in the United States.
The Great Recession created a huge dent in commercial real estate markets during this time, when a large percentage decrease occurred. However, in contrast, the loan prices during the Great Recession rose sharply and declined sharply after the end of the Great Recession. This may have been caused by the issue of overextended mortgage debt throughout the Great Recession, along with many other factors that led to a decline in loan pricing after the Recession. Using the data provided throughout this study, we can make a proper analysis about the impacts during and after the Great Recession, as well as predict when the next recession could occur. We also could make an analysis of what to do when a recession occurs, and how to get out of the recession. Throughout the research study, we notice trends that can be both easily and not easily hypothesized, sparking further discussion about commercial real estate markets and their impacts during recessions, past, present, and the future.

*Keywords: Great Recession, economics, commercial real estate, loan pricing, leasing markets*

**Introduction:**

The Great Recession is notable for being one the most severe periods of economic decline in recent history. While the recession impacted the
entire world, the greatest severity of the recession occurred in the United States, the focus of this study. The recession was the largest disruption of economic history, largely caused by the real estate market in the United States. During the period between 2007 and 2009, significant job loss and drop in consumption were among many disastrous impacts on the American economy. One of the major causes of this recession was the failure of economists to predict that the recession would occur; therefore, many Americans were not prepared for the quick crash, creating a larger effect than should have occurred.

There were many causes of the Great Recession. Overextended mortgage debt is often the most notable cause, because of the dramatic crisis that followed. Overextended mortgage debt led to a shutdown of the expansion of consumer credit, the major factor for demand expansion during this era. There were other significant causes and similarly grandiose results, such as the trade impacts on the United States. For example, O’Rourke details in his paper:

“The spread of manufacturing across the world also had implications for the regional composition of the Great Trade Collapse of 2008-9, as compared with the trade collapse of the interwar period. The volume of exports fell in an almost identical manner in both advanced and emerging economies after 2008, although this average
trend for the emerging economies disguises some regional variation, with exports from Africa and the Middle East, and especially Latin America, falling less than elsewhere” (O’ Rourke).

Throughout this study, commercial real estate markets and the impact of loans and leasing for the commercial real estate sector in the United States were analyzed. Commercial real estate is often land used for gaining profit, and is non-residential. Leasing was an important factor in the causes of the Great Recession, and significant studying of leasing prices specifically for commercial real estate markets needs to occur. The Great Recession had a significant impact on commercial markets, and therefore the real estate leasing behind business.

**Literature Review:**

To completely understand the effects of the Great Recession, it is crucial to understand what led up to it, including previous recessions and the impacts of these recessions, as well as the Great Moderation, a period of reduction of volatility in business growth between the mid 1980’s until the start of the Great Recession, in 2007.
The Great Depression is often known as one of the largest depressions in the United States, along with the recent Great Recession. The Great Depression is taught in every American history class because of the significant impact it had on the American economy and society. The depression occurred from 1929-1933, largely caused by a stock market crash. While the ways the economy worked in the 1930s and the 2000s were vastly different, the Great Depression and the Great Recession and had quite a few similarities.

To understand the differences between the Great Depression and the Great Recession, it is crucial to understand the difference between a recession and a depression. According to the Federal Reserve Bank of San Francisco, a recession is a decline in economic growth for a mild amount of time, while a depression is a more severe detriment to economic growth (“What is the difference between a recession and a depression?”). We know that the Great Depression had a more severe impact on American life than the Great Recession, which has held generally true. However, it is important to note that the causes and impacts of the Great Depression and the Great Recession were vastly different. The cause of the Great Depression was primarily a crash in the stock market, while the cause of the Great Recession was primarily overextended mortgage debt. The impacts of
the Great Depression, including a severe loss of jobs and a higher rate of unemployment, were far more serious than the impacts of the Great Recession,

*Recent Recessions:*

Between the Great Depression and the Great Recession, no severe recessions occurred. However, the graph below (Cynamon et.al.) details the important recessions after the Great Depression, and the drastic impact the Great Recession had on employment rates.

*Figure 1: Recent Recession Employment Rates Compared to Great Recession*

A good example from the case study above would be the recession that occurred in the early 2000’s in the United States
Many historic events, such as the .com boom and the 9/11 terrorist attacks, took place during this time. However, the biggest economic impact in this era took place in the manufacturing industry in the United States. The Bureau of Labor Statistics details:

“The upswing in information technology manufacturing (especially semiconductors) turned into massive layoffs in 2001, with the slump extending into wholesale trade and computer and data processing services. The falling output of IT goods also forced a cutback in investment in new production capacity—a trend that became generalized in manufacturing and led to significant job losses in industrial machinery” (Langdon et. al.)

Figure 2, from the Bureau of Labor Statistics, outlines the significant dip in manufacturing for the employment industry in America:
Great Moderation:

The final important time frame before the Great Recession was the Great Moderation period, from the 1980’s until 2007, which marked the beginning of the Great Recession. The Great Moderation was a period when there was neither growth nor decline in the American economy. Because of the inactivity of the economy during the Great Moderation, many economists did not predict the onset of the Great Recession. According to the Federal Reserve Bank of New York, the major reasons for the Great Moderation were luck and “good policy,” such as vigilance on inflation (“The Great Moderation, Forecast Uncertainty, and the Great
Recession”). The Great Moderation was a time for the United States to recover from the 2001 recession, but it ended with the beginning of the 2008 recession (Hakkio). Figure 3 (shown below) details the impact of the Great Moderation in terms of GDP growth:

Figure 3: Great Moderation In Terms of GDP Growth

Along with understanding other recessions prior to the Great Recession, it is also crucial to know about the specific area of this study in relation to the recession --commercial real estate markets and leasing.

Commercial real estate is defined as “property that’s owned for the purpose of making money” (“Commercial Real Estate - Definition from KWHS”). A large portion of the real estate market in the United States is commercial real estate, therefore is a major cause of the Great Recession.
This study focuses on how specifically leasing commercial real estate was affected during this time period and how leasing was affected in the years following the Great Recession. Leasing is defined as the act of providing a good for a limited time (such as rent) (“Lease”). Therefore, we can define leasing in this study’s context as specifically providing rent for commercial real estate in the United States.

The events from the 1900’s on were crucial for determining the causes as well as the impacts of the Great Recession, and how commercial real estate markets were severely affected by the Great Recession.

**Theoretical Framework and Research Methodology**

This research study is based on data collected by different Federal Reserve Banks across the United States, most notably the Federal Reserve Bank of St. Louis, which collects economic data from the U.S. Bureau of Economic Analysis.

To build the framework for this study, it is crucial to create a prediction of what happened during the Great Recession with loan prices of commercial real estate banks. It is also imperative to study the impacts of the Great Recession, as well as predict whether there will be a recession sooner than we believe, so that, unlike the lack of preparedness of the Great Depression and the Great Recession, we are ready.
There are several variables used in this study which help to quantify the loan pricing difference throughout the study:

- Percent Change of Pricing (Primarily form the year prior)
- U.S. Dollars
- Employment as a percentage of level at the beginning of the recession
- Number of months since beginning of recession
- Time (primarily in months or years).

These variables are crucial to understanding how the Great Recession affected Americans from 2007 to 2009.

This study has both qualitative and quantitative aspects. The quantitative aspect of this research is acquired from the Federal Reserve Bank, as mentioned earlier. This study uses different variables to specify the amount of loan pricing difference for commercial real estate markets in the United States. Most of the quantitative data will range from 2006, which was one year before the recession, to the current year of this research study being conducted, 2019. The qualitative analysis examines the significant economic effects of the Great Recession.
Data Analysis

There are multiple graphs from the Federal Reserve Bank of St. Louis to analyze, in regards to commercial real estate leasing markets and leasing trends.

Figure 4: Commercial Real Estate Prices in the United States

The first graph, above, from the Federal Reserve Economic Data (FRED) database details commercial real estate prices from 2007 to 2017. In the graph, the gray area defines the crux of the Great Recession, which throughout the data, we will define to be from the beginning of 2008 through the end of 2009. The two variables used in the graph were time (represented by the y-axis) and percent change from the previous year.
(represented by the x-axis), which means the average price change of typical commercial real estate in the United States from the previous year.

As we can see in the graph, in 2007, just prior to the Great Recession, commercial real estate prices were steadily increasing, with an average growth of 17%. However, as soon as the recession hit in 2008, we see a total percent decrease of around 15% from 2008 till 2009, reaching a percentage decrease of 35% by 2010. Compared to the rest of the graph, this is an extreme decrease in pricing of commercial real estate markets in the United States. This decrease in pricing led to the harsh downfall of employment in this sector. After the recession, in just one year, the percent change jumped back up to a 10% increase, compared to the percent change in 2007. After 2010, we notice that the percentage increase of commercial real estate pricing from year to year is staying relatively the same. Therefore, even though there was a major hit during the recession, percentage increase trends came back quickly, less than a year after the end of the recession.
Figure 5: Real Estate Loan Prices for Commercial Real Estate Banks

The second graph from the FRED database (Figure 5) details the prices of commercial real estate loans from all commercial banks across the United States. The time period for the graph above ranges from the beginning of 2005 through the beginning of 2018, with the shaded area marking the Great Recession (As in the previous graph, the Great Recession for this data set starts from early 2008 to the end of 2009.) The variables in the graph are time (represented by the y-axis) and billions of United States dollars (represented by the x-axis). The x-axis portion of this graph, which represents the billions of dollars of commercial real estate loans, shows the total dollar price of loans given out in the United States for the specified year.

There are a lot of surprising trends to notice. Before the Great Recession, we can see a steady increase of commercial real estate loan
During the Great Recession, we also see an increase in the commercial real estate loan pricing. The first surprise from the data above comes from June 2008, where we see a jump in the price of billions of dollars. After this jump, we notice a steady decline, not only just until the end of the Great Recession, but until June 2012, three years after the recession ended. This is the biggest surprise throughout the graph. At the beginning of 2013, the price of commercial real estate loans steadily increases, and by 2016, reaches the same price peak during the Great Recession. After 2016, there is a sharp increase of commercial real estate pricing till the dataset of the graph finishes in 2018.

**Conclusion:**

Based on the findings of this study, we are able to start creating predictions of when a recession might occur. Through the graphs shown in the data analysis, we can see trends that occur during recessions and can start making assumptions about what can occur in future recessions.

The first graph in the data analysis, which showed us commercial real estate prices in the United States, resembled many hypotheses about how the real estate market would be affected during and after the recession. However, the Great Recession created such a large impact that the percent
change from the previous year was exaggerated, due several conditions, especially overextended mortgage debt. Therefore, the results from the first graph in the data analysis were predictable.

The second graph in the data analysis, which showed commercial real estate loan pricing, was much less predictable and did not follow many predictable economic hypotheses. Commercial real estate loans continued to rise steadily until after the recession, when the major dip occurred between 2009 and 2012, a deviation from the norm of the dip occurring during the recession. The likely cause of this trend would be the post-recession effects of the Great Recession, when commercial real estate pricing had long term effects that directly affected loan pricing through the graph.

To conclude this research, we can use the data to help predict when a recession may occur, as well as provide insight into how to get out of a recession analytically. In fact, many economists predict that a recession could occur within the next few years of this study being released. It is crucial to note that, if done properly, there are many ways to ensure the United States does not go through another major recession like the Great Recession. We can use the Great Recession as an example of how to fix an unexpected recession and not cause a recession in the future.
References:

Board of Governors of the Federal Reserve System (US), Real Estate Loans: Commercial Real Estate Loans, All Commercial Banks [CREACBM027NBOG], retrieved from FRED, Federal Reserve Bank of St. Louis;


Commercial Real Estate - Definition from KWHS. (n.d.). Retrieved from

https://kwhs.wharton.upenn.edu/term/commercial-real-estate/


https://www.federalreservehistory.org/essays/great_moderation

International Monetary Fund, Commercial Real Estate Prices for United States [COMREPUSQ159N], retrieved from FRED, Federal Reserve Bank of St. Louis;


Do Voices of General Employees Matter? The Importance of Their Involvement and Impact in Designing and Implementing a Human Resource Information System (HRIS)

Liuxin Yan  
Research Mentor: Hui Liao

Liuxin Yan is a SURE Fellow in the Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: lyan56@terpmail.umd.edu.

Abstract

The study aims to examine the important role of general employees besides Human Resources (HR) professionals in the human resource information system (HRIS) development and implementation process. Overall, study results indicated that it is necessary to integrate certain self-performance management functions into the HRIS to help employees increase their strategic decision-making capacity. Among all relevant functions, study results indicated that employees showed a strong tendency
to demand functions that can assist them in monitoring their own performance and that can help them better understand what make some employees more efficient than others. Furthermore, survey results indicated that there is a positive correlation between general employees’ involvement and their willingness to use the HRIS, as well as their overall user satisfaction. While most existing research focused on investigating how HR personnel perceive and influence the development and implementation of the HRIS, the study can make unique contributions by highlighting the role of general employees in increasing usefulness of the HRIS, thus fully realizing the strategic potential hidden behind human capital within an organization.

**Keywords:** Human resource information system, HRIS, strategic human resource management, information systems

**Acknowledgement**

This research is conducted under the guidance of Dr. Hui Liao and Dr. Christina Elson from the Robert. H. Smith School of Business, University of Maryland. I would like to thank the *Ed Snider Center for Enterprise and Markets* for supporting and supervising this research project.
Introduction

In the recent decade, the role of technology in all aspects of business operations has become increasingly important. As the business environment changes rapidly in the age of globalization, it is important for an organization to develop a high level of agility to the external environment and make timely strategic decisions to absorb any unexpected changes. This makes the use of information systems more important than ever. However, the use of technology in the human resource management (HRM) area receives relatively less attention, as HR processes are less strategically aligned and receive less IT support. However, in the recent decade, human capital has become one of the most valuable assets for a company to develop competitive advantages to ensure the long-term success. This emphasis on human capital and intangible knowledge puts heavy pressure on HR practitioners in today’s dynamic business environment. As strategic potential hidden within human capital is hardly realized by traditional HRM practices, a transformation from traditional HRM to electronic human resource management (e-HRM) to implement a human resource information system (HRIS) is necessary in today’s complex and dynamic business environment (Maier, Laumer, Eckhardt, & Weitzel, 2013). A successful HRIS is able to serve as a distinct and supporting tool for the HR
department, and to develop competitive advantages for the company as a whole (Hendrickson, 2003).

However, despite increasing use and perceived importance of HRIS, the current research on the development of HRIS is surprisingly scarce (Stone, & Dulebohn, 2013). There is a lack of research on how to develop a HRIS that takes both information technology (IT) and HRM factors into consideration. This is due to people’s general misconceptions that HRIS is in the domain of information technology alone while ignoring the equally important role of strategic human resource management (Stone, D., & Dulebohn, J., 2013). However, not only should elements of a successful information system be evaluated, the incorporation of different HRM factors also should be carefully analyzed.

From another perspective, the contemporary HRIS allows employees to track their performances, thus make more efficient efforts toward targeted goals (Aral, Brynjolfsson, & Wu, 2012). There is an extreme dearth of research on this self-performance management component of the contemporary HRIS design. To bridge this knowledge gap, my research will focus on (1) studying necessary self-performance management functions that should be integrated as a part of the HRIS package, to better assist general employees besides HR professionals in making strategic decisions; (2) addressing the importance of general employees’
involvement in the initial development and implementation process of HRIS within an organization.

Literature Review

2.1 The definition and evolution of HRIS

The use of human resource information system (HRIS) began to emerge in the late 1990s. It was initially defined as “a system used to acquire, store, manipulate, analyze, retrieve, and distribute pertinent information about an organization’s human resources” (Tannenbaum, 1990). The form of an HRIS is not fixed, as the design varies depending on specific needs of each organization. It can be very informal, as with payroll records and compensation management. It also can be very formal, as in a huge central database collecting all relevant information necessary for making human resource decisions (Kovach, & Cathcart, 1999).

In the early years when the HRIS first emerged, it was used purely as a tool to cut HR costs by automating some administrative tasks, such as gathering basic information and filling out routine reports. As the technology matured, the HRIS became a tool to help HR managers, non-HR managers and general employees make better decisions (Kovach & Cathcart, 1999).
The HRIS has become much more strategically concentrated. It is no longer a simple tool to automate routine procedures and cut administrative costs. The HRIS has become a tool that adds value to an organization by helping users make better decisions. A contemporary HRIS is defined as a dynamic database of demographic and performance information about each employee within an organization. The HRIS will collect information on recruitment, hiring, organizational structure, performance evaluation, leadership, etc. (Lippert & Swiercz, 2005). It is able to make the HR department a strategic player by assisting it in making better decisions in hiring, promoting, organizing and allocating resources. The implementation of HRIS shifts major functions of the HR department from finishing administrative tasks to making strategic decisions, thus creating an environment full of innovation and creativity (Maier, Laumer, Eckhardt & Weitzel, 2013). Moreover, the comprehensive HRIS also provides a platform for employees in other departments to track their performances and meet targeted goals more efficiently (Aral, Brynjolfsson, & Wu, 2012).

Due to technological and practical limitations, early research in the HRIS era focused only on the user side. The framework was relatively simple. Research conducted in 1997 studied conditions required for a successful information system. The research summarized these conditions into three categories -- individual, organizational and system. While no
correlations between conditions and system usage level were found, the research found strong positive correlations between these conditions and the overall user satisfaction (p<0.05). The user satisfaction was positively correlated (p<0.05) with independent variables, including education level, prior work experiences, in-house training, quality of documentation, flexibility, and ease of use (Haines & Petit, 2017). Haines’s research investigated basic elements of a successful HRIS and set a foundation for further research to improve usefulness of the system. The research implied that many different demographic variables have a significant impact on individuals’ acceptance level of the HRIS, and that these factors should be taken into consideration while designing the system.

2.2 The human resource management component of HRIS

We cannot discuss the contemporary HRIS if we separate information system design and human resource management strategies. Both should be taken into consideration while designing a comprehensive HRIS. Aral’s research drew a three-dimensional model on three complementarities that interconnect information technology, human resource analytics and performance pay. The research indicated that a company can gain more benefits by using HRM software if it
simultaneously implements a system that allows HR managers to closely monitor employee performances and provide feedback to help employees track and improve their performances (Aral, Brynjolfsson & Wu, 2012). The research suggested that when a company has adopted a performance pay system, HCM and HR analytics are parts of the complementary system; when a company has adopted HR analytics, performance pay and HCM are also parts of the complementary system. It confirmed the hypothesis that adopting a full set of HRM components in the HRIS can significantly boost an organization’s productivity (Aral, Brynjolfsson, & Wu, 2012).

The study highlights the importance of developing an integrated and comprehensive HRIS, which can better assist HR managers in making strategic decisions by combining a cluster of HR tasks, including recruiting, monitoring, organizing, allocating resources and making promotion decisions. The study also points out an important limitation, which is that different industries have different demands and preferences for making HR-related decisions. For example, some industries might value human capital more than other industries. This factor should be taken into consideration when developing an HRIS for a specific company.

2.3 The strategic importance of contemporary HRIS
As information technology developed dramatically during the past decade, the application of HRIS evolved to be much more sophisticated. The research began to focus on how HRIS can help managers make better strategic decisions. The early HRIS was purely a tool to acquire and store existing information to automate some HR tasks to increase job efficiency and reduce costs. The contemporary HRIS is able to retrieve a new set of information to help HR managers better interpret existing knowledge, thus make more strategic decisions (Tansley & Newell, 2007). In other words, while HRIS was viewed as a tool to collect and store information in earlier years, the contemporary HRIS aims to produce a new set of information based on existing data. Therefore, the contemporary HRIS is able to help managers make more informed decisions, involve more critical thinking, and perform their jobs more efficiently. Therefore, the use of HRIS became especially important for organizations that view human capital as an important competitive advantage.

A study conducted in 2007 highlighted the increasing popularity and strategic importance of the HRIS in today’s business environment. By sending questionnaires to 450 HR managers and analyzing their responses, the research suggested that companies, regardless of the company size, have become more reliant on HRIS to support strategic decision-making. The use is considered to be strategic for activities that include HR planning,
salary advice, employee benefits, industrial relations, assessment and training needs, and recruitment and performance management, etc. (Hussain, Wallace & Cornelius, 2007). SME and large companies all show a tendency to increase their usage of HRIS in support of making strategic decisions (p<0.01). The size of the company does have an impact on its degree of usage in making recruiting and performance management decisions, but the gap is rapidly narrowing as more SME companies undertake HRIS or other e-HRM software (Hussain, Wallace & Cornelius, 2007). This study demonstrated the usefulness and developmental potential of a comprehensive HRIS in the current business competitive landscape. More importantly, it highlighted the importance of updating the existing HRIS and including inputs from a more diverse class of users.

2.4 Performance management

In essence, the ultimate objective of implementing a HRIS is to help a company better manage its employees and increase their performance more efficiently. This is a process generally known as performance management. Performance management is defined as “a continuous process of identifying, measuring, and developing the performance of individuals and teams and aligning performance with strategic goals of the organization” (Aguinis, 2007). A study conducted in 2006 highlighted the
important role of employees in a given company’s performance management cycle. The study suggested that a well-designed performance management system should focus on soliciting continuous feedback from general employees to more accurately determine measurable and realistic performance goals (Piskurich, 2006). Therefore, the study supported our hypothesis that integrating a self-performance management component into the HRIS will be beneficial to the individual employee’s performance, as well as to the overall performance management cycle of an organization.

2.5. The impact of user feedback and involvement on information systems

When implementing a HRIS within an organization, employees’ attitudes are inevitably influenced, as this is perceived as an organizational change for employees. It is important to note that perceived attitudes of employees have a significant impact on system usage and their overall job satisfaction. Over time, if employees hold negative attitudes toward organizational changes and have lower job satisfaction, the job turnover rate tends to be higher. One way to improve employees’ attitudes toward the new system is to increase employee involvement during the system development process. Study results from research conducted in 1997 indicated that if the HRIS is internally developed, and employees are more engaged in the process, there is a higher system usage and greater
satisfaction ratings (Haines & Petit, 1997). More recent research conducted in 2017 confirmed the hypothesis that if employees are more engaged during the system development process, and if the company provides more ongoing technological support, employees will have higher user satisfaction opinions of the system and can receive more benefits by using the system (Bano, Zowghi, & Da Rimini, 2017). Therefore, ignoring user feedback and not involving employees during the system development and actual system implementation process can cause severe consequences, reversing benefits of the new system (Maier, Laumer, Eckhardt & Weitzel, 2013).

2.6 The role of employees other than HR professionals

Although people generally believe that HRIS is designed primarily for the HR department, there are, in fact, three classes of users of the system. They are HR personnel, managers in functions areas and general employees. (Hendrickson, 2003). In other words, the HRIS is not only a tool for HR personnel, it is also a powerful platform for employees in other departments to make better decisions. As most existing research focuses on feedback from HR personnel, the primary users of the HRIS, the role of general employees has been long ignored.

As the contemporary HRIS has become a tool for general employees to track their performance and make more efficient efforts towards targeted
goals, it is important to design a comprehensive HRIS to fully address employee needs. Therefore, the implementation of HRIS should also be reviewed and adjusted based upon feedback from general employees, the secondary users of the HRIS. A 2009 study provided the supporting evidence for this hypothesis. The study proposed that a two-way communication system is necessary to increase the strategic capacity of a HRIS. A two-way communication system allows information to flow back and forth between information senders and recipients. It not only gives HR personnel a channel to distribute information to employees, it gives general employees opportunities to classify information and return feedback, adding potential value to HR practices (Stone & Lukaszewski, 2009). In this way, a HRIS can provide important supportive activities for management functions, such as planning, leading, controlling and organizing. To make the controlling component more effective, the HRIS will communicate results back to employees and allow them to provide feedback (Ngoc Duc, Siengthai & Page, 2013). The study confirms the hypothesis that including feedback from general employees is positively related to employees’ attitudes towards the e-HRM implementation (Stone & Lukaszewski, 2009). The study highlighted the importance of increasing user involvement and collecting user feedback of both HR personnel and general employees.
I. Hypothesis and Methodology

Though the use of technology seems to be prevalent in today’s business world, IT in the human resource management (HRM) era receives less attention than other aspects of business operations. There are few studies on theoretical frameworks for designing a comprehensive HRIS to improve strategic decision-making capacity of the HR department, as well as the entire organization. An integrated HRIS allows HR managers to become strategic players by unfolding hidden potential within a company’s existing human capital. The objective of this study is to help individual organizations increase the usefulness of the HRIS.

3.1 Study Objectives

The current HRIS research focused much more on the role of HR professionals during the development and implementation of HRIS. Most research ignores the role of general employees in unfolding strategic potential through the self-performance management functions provided by a HRIS. To bridge this knowledge gap, we conducted this research to study whether taking general employees’ opinions into consideration will make a difference in strategic decision-making power of a HRIS within an organization.
The contemporary HRIS is able to provide employees with a platform to help keep track of their own performance, get timely feedback from supervisors or peer workers, evaluate their performance critically, and make more efficient efforts toward directed goals. Therefore, we proposed that including opinions of general employees will add value to the HRIS by improving its strategic decision-making capacity for a more diverse class of users. Certainly, the development of HRIS needs to focus on opinions of HR personnel, who are the primary and daily users of the HRIS. But the study also will collect feedback from general employees in other departments to ensure that the HRIS can address needs of different parties within an organization.

To be more precise, the research focuses on the effects of general employees’ involvement during two phases -- the system development process and the actual system implementation process. For the system development process, the first objective is to study what self-performance management functions should be integrated into the HRIS package to assist general employees in making strategic decisions. By allowing employees to add their inputs about which functions are more critical and beneficial, employees are more involved in the development process, and the final HRIS product can be tailored more accurately to meet employees’ needs. The second objective is to study whether employees’ participation matters
in terms of increasing their willingness to use the system. For the system implementation process, the objective is to study whether employees’ involvement increases their overall user satisfaction.

3.2 Perceived importance of different HRM functions

To achieve the objective of helping employees become better strategic players, it is essential to integrate appropriate self-performance management functions to help them keep track of their own performance and make timely adjustments. Therefore, we want to study which specific functions should be integrated into the HRIS, as different functions might have different perceived importance.

The performance management process is generally operationalized as communications between managers and subordinates, goal setting, and problem solving. It is also very closely connected to procedures that include performance review and employee discipline (Piskurich, G, 2006). Based on these theories, we designed five self-performance management functions -- compensation management, goal-setting, progress feedback, performance report, and peer evaluation. (a) The compensation management function is by far the most common and typical employee self-serve service function in a HRIS package. It allows employees to update their personal information and benefits records, approve attendance records, review pay scales, and
change retirement benefit programs. (b) The goal-setting function is designed to foster personal growth in the workplace by allowing employees to set their short-term and long-term objectives. By setting goals on a regular basis, employees can compare their actual performance with their initial goals to generate plans for improvements in the future. (c) The progress feedback function will provide employees with a self-appraisal tool to keep track their performance through different indicators, such as feedback from their supervisors and KPIs, as most employees rely on these types of immediate feedback to make more efficient efforts toward directed goals. (d) The performance review function will analyze performance data and generate reports to help employees better understand what makes some employees more effective than others, thus make improvements. (e) The peer evaluation function will allow employees to give ratings and feedback to their teammates. It also will allow employees to evaluate their supervisors to create a two-way communication between managers and employees. To analyze perceived importance of these five functions, we designed a survey that asks participants to read a detailed description of each function, then rate their perceived importance for each function on a scale from 1 to 7.
3.3 The impact of individual involvement

To study the role of individual employees’ involvement, we employed the information system user participation level model proposed by Muneera Bano1, Didar Zowghi and Francesca da Rimini in their study published in 2017. To measure the correlation between user participation and user satisfaction, the model conceptualized elements affecting user participation levels into five categories: level of participation, effectiveness of participation, user representation, communication between users and developers, and top management support. As the target participants for this study are primarily general employees besides HR professionals, user representation is excluded from our study.

Then, based on four other elements, we designed four hypothetical scenarios for the HRIS development process and another four hypothetical scenarios for the HRIS implementation process. Each hypothetical scenario represents a condition with an increasing participation level. We designed a survey that asked participants to rate their willingness to use the system and their user satisfaction under each hypothetical scenario, on a scale from 1 to 7, after reading detailed descriptions of each scenario. We deem each hypothetical scenario as an independent variable and deem each individual participant’s rating as a dependent variable to study the correlation between them.
IV. Results

4.1 Data collection and participant demographics

Aiming to cover a diverse study population, we distributed the survey in the form of anonymous online links through a variety of channels. Qualified participants are individuals who have at least one year or more full-time work experience in the United States. We received 41 responses, of which 35 responses are deemed appropriate for data analytics after eliminating incomplete responses and unqualified participants.

The results showed that 31.43% of all participants obtained a bachelor’s degree; 40% obtained a master’s degree; 2.86% obtained an MBA degree; and 25.71% obtained a PhD. Predominantly, 48.57% of all participants are 19-30 years old. 17.14%, 22.86% and 11.43% of all participants are 31-40 years old, 41-50 years old, and 51-60 years old, respectively (See figure 4.1.1).

Survey participants included 41.67% currently working in the education industry and 11.11% of participants working in the consulting industry. The remaining participants come from a variety of industries, including biology, retailing, real estate, government, and software development. Among all participants, work experience is evenly spread
out: 31.43%, 28.57%, and 28.57% have 5 to 10 years, 10 to 20 years, and more than 10 years of work experience, respectively (See figure 4.1.2).

![Figure 4.1 <Participant Demographics>](image)

4.2 Participation and user satisfaction

Sixty percent of participants indicated that their current companies have implemented a HRIS with different degrees of complexities. However, only 28.57% indicated that the HRIS employed by their
companies included a self-performance management component. At the same time, 72.68% of participants strongly agreed that it is necessary to integrate a self-performance management component into the HRIS, and that they can benefit from this component by improving their strategic decision-making capacity.

To better understand the relationship between individual participation levels and user satisfaction, we broke the overall HRIS implementation process into two phases -- the system development process and the actual system implementation process.

4.2.1 The system development process

4.2.1.1 The perceived importance of different HRM functions

As a human resource information system is proposed to be a tool to assist employees in making strategic decisions, the objective of the system is to help employees identify their areas of weakness and make necessary adjustments in a more timely manner. The key is to integrate appropriate self-performance management functions, which can help employees track their own performance and make timely adjustments in a more efficient manner. Therefore, we want to study which specific functions should be integrated into the HRIS design, as different functions might have different
perceived importance. We asked participants to rate their perceived importance of five proposed HRM functions. The results indicated that the performance review function received the highest rating among all five functions. The ranking of their perceived importance, in descending order, is performance review, progress feedback, compensation management, peer evaluation, and goal setting (See figure 4.2.1.1).

We then categorized results based on different industries, as employees working in different industries have different demands for performance management. Across all but the retail & wholesale industry, the performance review function received the highest average rating. Despite slight differences in the ranking, participants across all industries tend to value performance review and progress report more than three other functions (See figure 4.2.1.2). This suggests that employees want to be able to monitor their own performance, get timely feedback, and understand what make some employees more efficient than others.
Figure 4.2.1.1 <Perceived Importance of HRM Functions>

<table>
<thead>
<tr>
<th>HRM Function</th>
<th>Perceived Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Review</td>
<td>6.029</td>
</tr>
<tr>
<td>Progress Feedback</td>
<td>5.543</td>
</tr>
<tr>
<td>Compensation Management</td>
<td>5.143</td>
</tr>
<tr>
<td>Peer Evaluation</td>
<td>4.886</td>
</tr>
<tr>
<td>Goal Setting</td>
<td>4.857</td>
</tr>
</tbody>
</table>

Figure 4.2.1.2 <Perceived Importance of HRM Functions by Industry>

4.2.2 The involvement and willingness to use the system
Allowing employees to rate their desired self-performance management functions is one way to give them opportunities to add their inputs to the HRIS during the system development process. There are many other ways to involve employees during this stage. Therefore, we designed four hypothetical scenarios (See appendix 6.1), each representing a condition with an increasing participation level. The independent variable is the degree of participation level represented by each hypothetical scenario, and the dependent variable is each participant’s rated willingness to use the system. We first averaged participants’ rated willingness to use the system, on a scale from 1 to 7. The results indicated that as employees become more involved in the system development process, they are more willing to use the system (See figure 4.2.1.3). To further validate the result, we performed a regression analysis, indicating that there is positive correlation ($\beta=0.85$) between user participation and their initial willingness to use the system during the system development process ($p< 0.001$).
4.2.1.3 *The relationship between participation level and the willingness to use the system*

4.2.3 *The involvement and overall user satisfaction*

The second phase we studied is the relationship between individual employee involvement and overall user satisfaction during the actual implementation process. The degree of involvement during the
implementation process is determined by the company’s frequency in collecting employees’ feedback, resolving ongoing technical difficulties, and updating the system based on user feedback. We designed four hypothetical scenarios (See appendix 6.1), each representing a condition with an increasing participation level. The independent variable is the degree of participation level represented by each hypothetical scenario, and the dependent variable is each participant’s rated user satisfaction. We first averaged the participants’ rated user satisfaction under each of these hypothetical scenarios. The results indicated that as employees become more involved in the system development process, they are more satisfied with the system (See figure 4.2.1.4). To further validate the result, we performed a regression analysis, indicating that there is a positive correlation ($\beta=0.93$) between individual employee involvement and overall user satisfaction ($p<0.01$).

In conclusion, if employees feel that their opinions are more valued, and they are more involved during the HRIS implementation process, employees will be more satisfied and motivated to use the HRIS to improve their performance.
Figure 4.2.1.4 <The relationship between participation level and user satisfaction>
V. Discussion

5.1 Implication

The study results indicated that the involvement of general employees besides HR professionals has a significant impact on the HRIS development process and the actual system implementation process. Previous studies suggested that HRIS also can be used by employees in other departments to track their performance and meet targeted goals more efficiently (Aral, Brynjolfsson, & Wu, 2012). The study confirmed the strategic importance of integrating a self-performance management component into the current HRIS and allowing general employees besides HR personnel to get more involved during this process. Therefore, to fully realize the strategic potential of human capital, HRIS software development should take this into consideration to design more beneficial HRIS packages for organizations.

5.2 Study Limitation

Although the study can explain and verify the correlation between employees’ individual involvement and their willingness to use the system, as well as overall user satisfaction, some limitations do exist. Insufficient data sample size (n=35) may lead to some inaccurate results, especially
when participants’ demographics are not diverse enough. Most participants (41.67%) are working in large-sized companies in the education industry. As different industries have specific demands for the development of a HRIS, participants from more diverse industries are preferred. Future studies may be conducted with participants from a variety of industries with more diverse demographics.

5.3 Suggestions for future studies

Although our study has confirmed the strategic importance of taking general employees’ opinions into consideration by integrating a self-performance management component into the HRIS, many knowledge gaps remain.

First, if general employees are neither IT or HR professionals, it is difficult for them to transform their direct feedback into practical and executable suggestions to design or implement the HRIS. Future research can evaluate if their feedback is truly valuable and study how to practically take advantages of the feedback.

Second, the current study indicated a strong preference for employee performance management functions that can help them keep track of their performance and understand what make some employees more effective.
than others. Future studies can investigate how to measure and quantify employee performance in an accurate and meaningful manner.

Third, different industries have different demands and different performance measurement indicators. Companies of different sizes also have distinct requirements to balance costs and benefits of implementing a HRIS. Future studies can investigate the impact of industry and company size on the HRIS development to fully address needs of various organizations.

Finally, the HRIS is an extremely expensive investment for a company due to initial capital requirement, user training costs, and ongoing maintenance costs. More research should be conducted to help organizations weigh benefits versus costs of implementing a HRIS. The current research primarily studied how the self-performance management component of a HRIS can increase strategic the decision-making capacity of employees. Future research can study other potential benefits that could be brought about by the HRIS from different perspectives.
Appendix

6.1 System development process survey

Scenario 1:

The company implements a new HRIS by making an authoritative announcement without asking for any of your opinions and forces all employees to start using it.

Please rate your willingness to actively use the system at a scale from 1 to 7.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(not willing at all)</td>
<td></td>
<td></td>
<td>(neutral)</td>
<td></td>
<td></td>
<td>(extremely willing to)</td>
</tr>
</tbody>
</table>
Scenario 2:

The company collects some feedback before the actual implementation, but it implements the system anyway without taking your opinions into consideration.

Please rate your willingness to actively use the system at a scale from 1 to 7.

<table>
<thead>
<tr>
<th>1 (not willing at all)</th>
<th>2</th>
<th>3</th>
<th>4 (neutral)</th>
<th>5</th>
<th>6</th>
<th>7 (extremely willing to)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scenario 3:

After implementing the HRIS arbitrarily, the company holds various training sessions to help you better understand the importance of HRIS and help you develop skills of using the system.

Please rate your willingness to actively use the system at a scale from 1 to 7.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>
Scenario 4:

Before the HRIS project starts at all, the company asks employees for opinions and attitudes towards the system. It has a long and smooth process of introducing the HRIS and allows employees at all levels to add their inputs during the system development process.

Please rate your willingness to actively use the system at a scale from 1 to 7.

<table>
<thead>
<tr>
<th>1 (not willing at all)</th>
<th>2</th>
<th>3</th>
<th>4 (neutral)</th>
<th>5</th>
<th>6</th>
<th>7 (extremely willing to)</th>
</tr>
</thead>
</table>

6.2 System implementation process survey
Scenario 1:

The company does not continuously monitor system maintenance and does not actively provide technical support for employees.

Please rate your degree of perceived user satisfaction at a scale from 1 to 7.

<table>
<thead>
<tr>
<th>1 (extremely unsatisfied)</th>
<th>2</th>
<th>3</th>
<th>4 (neutral)</th>
<th>5</th>
<th>6</th>
<th>7 (extremely satisfied)</th>
</tr>
</thead>
</table>

Scenario 2:

The company periodically holds training sessions to help solve ongoing technical problems encountered by employees.

Please rate your degree of perceived user satisfaction at a scale from 1 to 7.
Scenario 3:

The company continuously involves employees during this process by collecting their feedback about the system usage for further improvements.

Please rate your degree of perceived user satisfaction at a scale from 1 to 7.

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>(neutral)</td>
<td></td>
<td></td>
<td>(extremely satisfied)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scenario 4:

The company not only holds training sessions periodically, but also continuously involves employees by collecting their feedback about the system usage for improvements.

Please rate your degree of perceived user satisfaction at a scale from 1 to 7.

<table>
<thead>
<tr>
<th>1 (extremely unsatisfied)</th>
<th>2</th>
<th>3</th>
<th>4 (neutral)</th>
<th>5</th>
<th>6</th>
<th>7 (extremely satisfied)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


