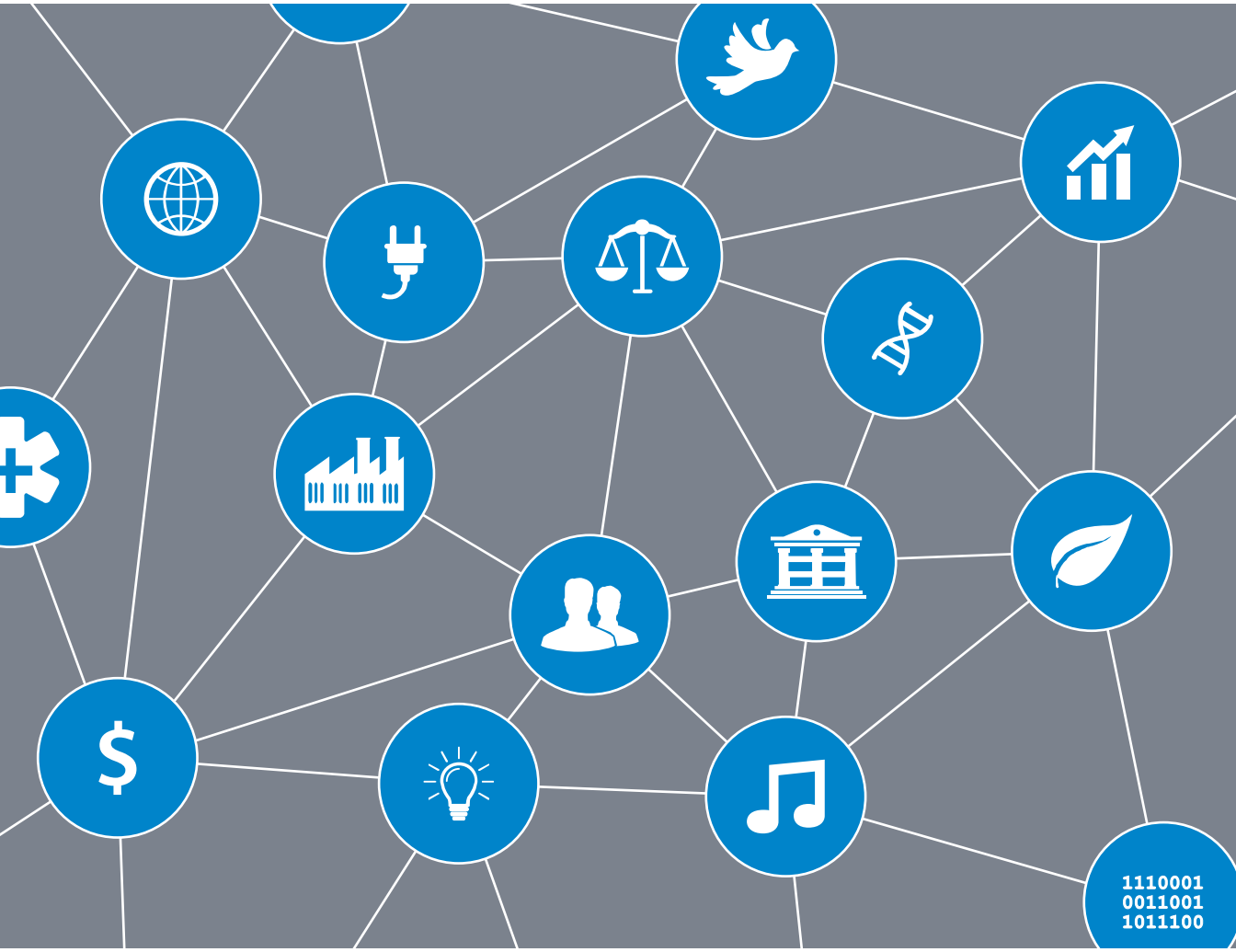
FUSIO | The Bentley Undergraduate Research Journal

Volume: II Fall 2017 Issue: I



Integrative and multidisciplinary undergraduate research at the leading edge of the Arts and Sciences and Business

Published by the Honors Program at Bentley University

BENTLEY UNIVERSITY

FUSIO | The Bentley Undergraduate Research Journal

Business and Editorial Correspondence:
Bentley University
Honors Program
175 Forest Street
Waltham, MA 02452

fusio@bentley.edu
www.fusiojournal.org

FUSIO | The Bentley Undergraduate Research Journal

VOLUME II FALL 2017 ISSUE I

ARTICLES

Aims and Scope

Established following Bentley University's mission of creating impactful knowledge within and across business and the arts and sciences, *Fusio* is a multidisciplinary undergraduate journal committed to the dissemination of original, high-quality undergraduate research. The journal is published by Bentley University's Honors Program and edited by both students and faculty across disciplines. *Fusio* encourages submissions from undergraduate students, with an emphasis on articles that span both business and arts and sciences topics as well as multidisciplinary topics. The journal is currently open only to undergraduate students at Bentley, and will consider original research by students as well as student/faculty joint work. All submissions undergo a blind peer review process.

Fusio is published twice yearly, in the fall and spring. For subscription information, please contact the Honors Program at Bentley University, or the journal office at fusio@bentley.edu. Online open access to these articles may also be found at www.fusiojournal.org.

Acknowledgements

We gratefully acknowledge the contributions of our reviewers in supporting Fusio and the authors in ensuring the highest standards of academic quality. The authors have benefitted tremendously from their time and insightful comments, so we thank you.

The Effect of the Medicaid Provision in the ACA on Insurance, Health Utilization, and Health

By Emily Hunt*

In 2010, President Obama enacted the Patient Protection and Affordable Care Act (ACA) in order to address problems facing the U.S. healthcare system. This study analyzes the impact of a key provision of the ACA, the Medicaid expansions, on three separate tiers of outcomes relating to (1) insurance status, (2) recipient health care utilization and (3) recipient health. Previous research has been limited and inconsistent on addressing this issue in a wide-scale manner. My study addresses some of the knowledge gaps in the literature while utilizing one of the largest and most recent datasets.

Economic theory indicates that expanding Medicaid eligibility should raise insurance coverage as more newly-eligible individuals' take-up insurance. Increased insurance coverage would reduce health care costs to individuals and therefore increase healthcare utilization among the newly insured. Consequently, physical and mental health of the recipients is expected to improve. I test the following hypotheses by applying multivariate regression analysis using differences-in-differences models to compare changes in outcomes among states that expanded Medicaid to those that did not.

My estimates suggest that the Medicaid expansions provided insurance coverage to nearly 3 million previously uninsured individuals. In addition, I find that the expansions significantly reduced out-of-pocket costs and increased the likelihood that low-income adults receive preventative care. I do not find any significant effects on physical or mental health, though impacts on health likely take 1-2 years to be materialized in data. Although the impact on health remains inconclusive, my study finds that the Medicaid Provision has both reduced inequality and improved accessibility in the U.S. healthcare system.

Keywords: Affordable Care Act, Medicaid Provision, Insurance, Low Income, Health.

I. Introduction

Health care spending has been one of the fastest growing economic sectors within the United States, growing from 5% of GDP in 1960 to 17.4% of GDP in 2013. Growth has been driven by increased demand for health care, largely due to the aging popula-

* Email: huntemily6@gmail.com. I would like to thank Professor Dhavel Dave for his consistent help and support throughout both my time spent working on this paper as well as my time at Bentley.

tion as well as increases in personal income, lifestyle-based illnesses and health insurance (Catlin et al. (2015)). In 2013, Organisation for Economic Co-operation and Development (OECD) reported that the average global healthcare expenditure in developed nations was 8.9% of GDP. In addition, per capita spending in the United States was $8713 per person; $5360 more than the global average for developed nations (OECD, 2015). However, despite above average healthcare expenditures in the U.S., individuals have not experienced superior health when compared to those around the globe. For example, the average life expectancy in the U.S. remains below the OECD average by a year (OECD, 2013). That being said, the added resources into health care (costs) were not producing higher health (benefits). Overall the rising costs indicate an inefficient system; although health expenditures have rapidly increased, personal health has remained largely unaffected. This has led to the system being criticized for being wasteful and distortive as well as leading to unnecessary cutbacks in other public programs.

In addition to potential quality problems, the rising costs increasingly left low income Americans without access to affordable healthcare. In 2012, 16.5% (or 12.45% unweighted) of Americans said there was a time in the past 12 months where they needed to see a doctor but could not due to cost (BFRSS). If a family of four made a salary at 138% of the federal poverty line, average per capita costs would be equal to over 100% of their earnings. By the early 2000's, the rise in healthcare costs were far outpacing increases in income leading many to neglect health problems because they could not afford to treat them. Additionally, prior to the ACA, public funding contributed to 48% of healthcare expenditures, significantly below the OCED average of 73% (OECD, 2015). The problem was strikingly obvious, Americans living in poverty could not afford to access basic health care. Not surprisingly, throughout the first decade of the 2000's, large public concern demanded a more accessible and affordable health care system.

Although the country struggled to agree on the best fit solution for some time, by 2010 Congress passed the Affordable Care Act. The law was enacted to expand insurance coverage, reduce inequality and make health care expenditures more affordable. The primary components to reduce the uninsurance rate were (1) the individual mandate, which required essentially every individual to have creditable health insurance, and (2) the Medicaid Provision, which greatly expanded the previous Medicaid program.

The ACA led to the largest reduction in the uninsured population in four decades. Over the past five years, the ACA has increased the insured population by 16.4 million, worked to fight health care inequality, and reduce uncompensated care in hospitals by an estimated 6.2 billion (2014 estimate) (Dranove (2017)). In addition, the ACA was successful in curbing the growth of health expenditures, as growth is now slower than the growth experienced over the last two decades prior to the 2008 recession, peaking at 10.9% annual growth in 2009. Declines in growth are associated with cost sharing and Medicare updates found in ACA. In 2014, national healthcare expenditures grew

by only 5.8%, largely due to the ACA's major expansions, stronger than expected economic growth, and the aging population (Centers for Medicare and Medicaid Services (2015)). Although the ACA was largely successful in achieving its goals, controversy specifically surrounding the Medicaid provision has remained notable largely due to increases in public funding. Therefore, it is important to examine and understand how the Medicaid expansions have impacted individuals the provision was aimed at helping.

Prior to the ACA, the nation lacked a universal Medicaid Program which increasingly left low-income individuals with little access to affordable health care. The first Medicaid bill, passed in 1965, required states to provide health insurance coverage to specific groups of people deemed unable to provide it for themselves. The program was jointly funded by the Federal and State governments in order to increase health insurance coverage and aggregate health. By the early 2000's, federal Medicaid regulations expanded Medicaid to include a variety of individuals determined "in need"; including pregnant women, young children with family incomes below 133% of the Federal Poverty Line, older children with family incomes below 100% of the Federal Poverty Line and adults that met certain requirements. Although federal Medicaid regulations governed the minimum requirements that states were legally bound too, many states began providing more generous Medicaid policies, creating an unbalanced health care system largely dictated by state regulation. This system was not only inequitable for individuals in differing states but it was also inefficient as it increased moral hazard in states that had relatively more generous programs.

To mitigate inequality and unaffordability in the health care system, the Medicaid Provision was implemented in the ACA. The provision intended to increase the Medicaid income eligibility to 138% of the poverty line for every citizen by 2014. However, critics argued that the new program would increase public budget deficits and decrease employment. Controversy surrounding the new legislation cumulated in 2012 with the Supreme Court Case *NFIB* vs. *Sebelius*. That year, the court ruled that the Medicaid Provision in the ACA was unconstitutional because the Federal Government could not withhold Medicaid funding from states that do not enact the provision (Focus on Health Reform (2012)). Instead, the ruling allowed states to maintain discretion over their Medicaid program.

Although the Medicaid Provision was never fully enacted, 32 states to date have adopted the policy, expanding the income eligibility for low-income adults living in the state to receive fully or partly subsidized health insurance. States that chose not to adopt this provision have less generous Medicaid policies and insure less people through the program.

In 2015, the Medicaid program insured 76 million people, and by June 2016, Medicaid enrollment had grown by over 27% compared to before it was enacted in 2014. However, increases in individuals utilizing Medicaid have led to increases in federal spending as the program has been largely financed by the Federal government. Currently, health care spending accounts for 17.5% of total public spending and is expect-

ed to grow to 20.1% by 2025. In 2015, federal spending on major health care programs (Medicare, Medicaid and the Children's Health Insurance Program) exceeded spending on Social Security for the first time. The increases were largely driven by Medicaid outlays, which increased by 14% in 2014 and 16% in 2015. The large increases in government health outlays have fueled controversy regarding the current Medicaid program and raise an important question; has the expansion been effective in decreasing the uninsurance rate and improving health in beneficiaries? Although the long-term effects of the Medicaid expansion Provision are unclear, the immediate effect on insurance coverage, healthcare utilization and overall physical and mental health should be researched at this time. Therefore, this paper examines the effect of the Medicaid Provision on beneficiaries and draws conclusions about the Medicaid expansions impact on insurance coverage, health care utilization, and overall health.

II. Theoretical Model

Economic theory suggests that the Medicaid expansion will be effective in decreasing the amount of uninsured individuals and ultimately increase the newly insured individual's health. Mainly, the increase in health is derived from insured individuals being more apt to access medical care than those not insured, for cost reasons. Therefore, as the government increases the availability of health insurance to low-income individuals, overall national health will increase. Specifically, the "Grossman Health" economic model shows a direct correlation between increased insurance and improved health.

The Grossman Model states that demand for health is derived from an investment model, meaning the model views spending money on health as a stock that will appreciate over time. The model explains that individuals invest in health care to attain better health because good health allows individuals to work and earn an income. Therefore it is assumed an individual will make more money from work than what they invest in their own health, indicating that investing in health will produce a greater return over time. Insurance reduces the amount an individual spends on health, or the "out-of-pocket costs" to the individual, as the government subsidizes some or all of the costs. The decrease in out-of-pocket spending decreases the elasticity of individual demand for health care as consumers become less price sensitive. In general, ex-post moral hazard increases simultaneously with demand for health care as individuals with insurance are more likely to go to the doctor than those who do not have insurance. In other words, insurance increases demand for health care due to the decreased cost to individuals. This change in cost sharing features due to increased insurance should result in a higher demand for and use of preventative services. The increase in demand, and use of, healthcare services should correlate with improved individual health.

III. Literature Review

A number of prior studies have provided early evidence of the Affordable Care Act on insurance coverage and health care utilization, and, in some limited respects on health outcomes. There is a strong consensus in this literature that the ACA has significantly expanded insurance coverage and has narrowed racial disparities in health care access (Courtemanche (2016), Works (2016)). Some recent studies have also examined the impact of the ACA Medicaid expansions and found increased health care coverage and health care utilization in association with these expansions in public insurance as well as some reduction in financial strain (Simon et al. (2016), Hu et al. (2016), McMorrow et al. (2016), Baicker et al. (2013), Blendon et al. (2016)). Given the fact that most of the provisions of the ACA did not go into effect until 2014, this literature is still emerging and several gaps remain. First, due to the lags with which national data becomes available, many of these studies were based on earlier time periods (usually up to 2013 or 2014). Effects of the ACA may generally take time to materialize given the lags from obtaining coverage to obtaining care to realizing effects on mental and physical health (Colman and Dave (2015)). Many individuals may not realize they are eligible for Medicaid coverage under the new expansions, and take-up rates for Medicaid typically increase over time. This study utilizes data up to 2015, taking advantage of Medicaid expansions that occurred in 2014 along with an additional year of data that may capture some of these lagged and longer-term effects on insurance and utilization. While effects on health may yet require longer periods of time to detect, if there are such effects, this study assesses whether there were any discernible effects on broad measures of physical and mental health within one to four years of the ACA Medicaid expansions (6 states expanded their Medicaid programs in 2010 and 2011).

Second, while most studies find that insurance coverage and health care utilization increased, there is some variation in the magnitudes of these effects, which may reflect differences in sample sizes, time periods, and the number of states that had expanded their Medicaid programs during the sample period under consideration across the various studies. This study, by utilizing data up to 2015, is able to capture expansions in all 32 states that have currently expanded their Medicaid programs under the ACA (see Table 1); the remaining states have opted not to expand Medicaid.

Furthermore, since the individual insurance mandate went into effect in 2014, this study is able to capture two full years of post-ACA data with respect to this and other provisions of the ACA (for instance, the community rating, income subsidies, etc.). Third, many prior studies separated their focus. For example, some studies looked at only insurance coverage, or others looked at both insurance coverage and health care utilization, or a few others focused on measures of health.

TABLE 1
STATE MEDICAID EXPANSIONS

State	Date	State	Date
Alaska	14-Jan	Michigan	14-Apr
Arizona	15-Jan	Minnesota	14-Jan
Arkansas	14-Jan	Montana	16-Jan
California	10-Jul	Nevada	14-Jan
Colorado	10-Jul	New Hampshire	14-Aug
Connecticut	10-Jul	New Jersey	11-Apr
Delaware	14-Jan	New Mexico	14-Jan
District of Columbia	10-Jul	New York	14-Jan
Hawaii	14-Jan	North Dakota	14-Jan
Illinois	14-Jan	Ohio	14-Jan
Indiana	15-Feb	Oregon	14-Jan
Iowa	14-Jan	Pennsylvania	15-Jan
Kentucky	14-Jan	Rhode Island	14-Jan
Louisiana	16-Jul	Vermont	14-Jan
Maryland	14-Jan	Washington	11-Jan
Massachusetts	14-Jan	West Virginia	14-Jan

This study provides a comprehensive analysis of the ACA expansions on important sets of outcomes related to insurance coverage, health care utilization including preventive care use, and broad measures of mental and physical health based on the same data set. This makes it feasible to compare across outcomes since all models are estimated for the same sample and over the same time period, based on the same specification, and thus inform whether any improvements in health care use and preventive medical care are consistent with the expansions in insurance coverage. Finally, while this study is not the first to use data from the Behavioral Risk Factor Surveillance System (BRFSS), use of these dataset, which is nationally-representative and provides a sample size of between 2.6 and 4.8 million observations, complements all of these contributions that this study makes to the literature and helps to maximize the statistical precision of the estimates. As effects on some outcomes, for instance downstream effects on preventive care or health, may be small, it is important to have a large sample size in order to improve statistical power. All of these questions that this study undertakes are framed by economic theory - drawing upon the Grossman model for the demand for health capital - and conclusions are drawn based upon the weight of the evidence from the various models and specifications.

Impact of the ACA on Insurance Coverage

Recent studies have found that the ACA was effective in increasing insurance coverage and reducing income and racial inequalities within the health care system. For example, (Courtemanche et al. (2016), Works (2016)) found that in 2014 insurance

coverage for adults between the ages of 18 to 64 increased by 5.9 percentage points in states with full ACA adoption, indicating the ACA was successful in decreasing uninsurance rates. In addition, the study found that the ACA had the greatest impact on reducing uninsurance rates for low income, non-whites, young adults, and unmarried people, indicating that the ACA was effective in reducing income and racial disparities. The paper specifically refers to the three-leg stool, consisting of (1) the federal exchange program, which improves non-group insurance markets for those who do not have access to employee sponsored insurance, (2) the individual mandate, and (3) the Medicaid expansion, as primary drivers of the overall increase in insurance coverage. The study primarily relied on data from the American Community Survey (ACS) between the years of 2011 and 2014.

In addition, Works (2016) confirms Courtemanche et al. (2016) findings by reporting the ACA increased insurance coverage by 5.9 percentage points in states that enacted the full ACA compared with the 3.0 percentage point increase in states that did not expand Medicaid.

Impact of Medicaid expansion on Insurance Coverage

Not surprisingly, previous literature almost unanimously has found that the Medicaid expansion alone has also contributed to an increase in insurance coverage (Cohen (2014), Kaestner et al. (2016), Courtemanche et al.(2016), Frean et al. (2016)). This is expected as the Medicaid expansion increased the Medicaid eligibility threshold for fully subsidized health insurance from 100% of the FPL to 138% and increased partial subsides up to 400% of the FPL, allowing more adults to be eligible to receive both partly and fully subsidized health insurance. Both federal and private research have found that the 2014 Medicaid expansions have effectively decreased uninsurance rates in post-expansion states. The National Center for Health Statistics reported that in 2014 adults were less likely to be uninsured in post-expansion states. Specifically, states that adopted the Medicaid expansion had a reduction in uninsurance rates from 18.4% in 2013 to 13.3% in 2014, compared to states that did not enact the expansion where the uninsurance rate decreased from 22.7% in 2013 to 19.6% in 2014 (Cohen, 2014).

In addition, private studies, including Kaestner et al (2016), Courtemanche et al. (2016) and Frean et al. (2016) also found that Medicaid expansions were associated with increases in coverage and corresponding decreases in the uninsured population.

Specifically, Kaestner et al. finds that Medicaid expansions were associated with 50% increases in Medicaid coverage. The study uses data from the ACS and the Current Population Survey and limits results to (1) low-income childless adults and (2) low-educated adults. Overall, it finds that both groups saw up to 50% increases in Medicaid for both samples, while childless adults saw slightly higher increases. In addition, Medicaid expansions decreased the proportion of uninsured by a slightly smaller amount due to some crowd out of private insurance. Courtemanche et al. supports the

findings reporting a 5.9 percentage point increase in insurance coverage for states that fully enacted the ACA compared to 3.0 percentage points in states that did not expand Medicaid, indicating that the Medicaid expansion has led to further decreases in the uninsurance rate.

Additionally, Frean et al. expands upon the Kaestner et al. and Courtemanche et al. study, providing a greater understanding of the impact of Medicaid on insurance. Frean et al. finds that 60% of the ACA's coverage gains were attributed to the Medicaid expansions while 40% were attributed to exchange premium subsidies. In addition, the study also found that the Medicaid expansion led to increases in insurance coverage both through newly-eligible individuals (by roughly 9 percentage points in 2014 and 14 percentage points in 2015) and additionally through the "woodwork effect" (by roughly 2.6 and 2.4 percentage points respectively). The woodwork effect accounts for those who were previously eligible for Medicaid but did not apply for Medicaid coverage or subsidies until after the full ACA was implemented due to increased awareness, streamlining the application process and the individual mandate. The data relies on the ACS and CHIP information for low income adults between 2012 and 2014. Sample includes non-elderly adults residing in U.S. except for Massachusetts.

Impact of Medicaid Expansion on Financial Strain

Further studies have found that Medicaid has reduced financial strain related to health care expenses for low income adults, likely due to the increased insurance coverage. For example, Hu et al. (2016) found that the Medicaid expansion reduced debt collections on an average of $69 among the top quartile of low income adults, those most impacted by Medicaid. In addition, McMorrow et al. (2016) study found that increases in Medicaid eligibility between 1997 to 2009 led to a reduction in out-of-pocket spending for low income adults. Lastly, the Oregon Experiment also found reduced out-of-pocket spending due to the Oregon Medicaid expansions (Baicker et al. (2013)). Overall, the literature suggests that increasing the Medicaid income eligibility threshold by any amount will reduce out-of-pocket spending on health related expenditures as a greater amount of individuals can receive full or partly subsidized health insurance.

Impact of Medicaid expansion on Health Care Utilization

Additional studies have found that the Medicaid expansions have increased health-care utilization (Simon et al. (2016), Baicker et al. (2013), Blendon et al. (2016)). Simon et al. (2016) finds that preventative care increased for low-income childless adults in states that expanded Medicaid. The study relies on the BRFSS survey data for years 2012 to 2014. Specially, the paper finds that the Medicaid expansions increased dental visits by 20%, breast exams by 14% and monograms by 16%, indicating a reduction in health-related income disparities. In addition, the paper finds that the Medicaid expan-

sions lead to modest improvements in self-assessed health and decreases in the number of work days missed due to poor health. Additionally, the Oregon experiment reports increases in the probability of receiving cholesterol checks, blood tests, mammograms, and Pap tests for adults that received Medicaid (Baicker et al. (2013)). In general, it is not surprising that the literature suggests increased preventative care due to decreases in out-of-pocket spending for Medicaid participants. Lastly, Blendon et al. (2016) used a differences-in-differences approach to compare healthcare utilization between states that expanded Medicaid and those that expanded private option insurance. Data was collected from 3 states (Arkansas, Kentucky and Texas) and controlled for statistical biases. The study found that both Kentucky's Medicaid program and Arkansas's private option expansion led to increases in outpatient utilization, preventative care, and improved health qualities.

Impact of Medicaid Expansions on Health

Studies have been largely inconclusive on their findings on physical and mental health. This is likely because the expansion was recently enacted and health takes time to improve. Most notably, the Oregon Experiment, found that expanding Medicaid had no impact on physical health. In 2008, Oregon enacted a Medicaid expansion system based on lottery drawings from a waiting list. This allowed the researchers to collect in-person data from approximately 11,000 individuals eligible for Medicaid. The lottery was randomly chosen, so it allowed researchers to compare the health effect of Medicaid on individuals eligible for Medicaid with the control group of those who were eligible but did not receive Medicaid. Overall, the study found that an expansion in Medicaid coverage did not correlate with significant improvements in physical health. However, Medicaid expansions were correlated with lowered rates of depression, indicating a positive relationship between Medicaid and improved mental health. Some of the limitations include the relatively small sample size, the state focus as opposed to national scale, and statistical error in the fact that people who seek Medicaid through a lottery might have different characteristics than those who seek Medicaid through coverage mandates (Baicker et al. (2013)).

Conversely, other studies have found improvements in specific aspects of physical health. For example, Sommers et al. (2012) finds that the Medicaid expansions in Medicaid significantly reduced mortality and improved self-reported health. This study compared New York, Maine, and Arizona, three states that have expanded Medicare post 2000, with neighboring states that have not expanded Medicaid. It collected data from the Compressed Mortality File of the Centers for Disease Control and Prevention, along with individual insurance coverage, costs of care and self-reported health through both the Current Population Survey and the BRFSS between the years of 1997 to 2007. Although the study was done prior to the 2014 Medicaid expansions, it suggests that adults in states that have expanded Medicaid have a lower probability of death and higher probability of reporting "very good" or "excellent" health than states that did not expand Medicaid.

Conclusion

Overall, the previous literature lacks a comprehensive study on the actual effect of the Medicaid expansion on insurance coverage, health care utilization, and overall health. In addition, much of the previous studies are limited to less than five years and use smaller, more restricted sample sizes when compared to this study. My study provides comprehensive results to support the diverse research on the Medicaid expansions.

IV. Data and Variables

This study uses data collected from the Behavioral Risk Factor Surveillance System (BRFSS) and the Henry J. Kaiser Family Foundation for the years of 2005 to 2015 (Kaiser Family Foundation (2016)). The combination of the two data sets allow the study to compare the effect of the Medicaid Provision on insurance, health utilization and health status through a wide variety of health factors at the national scale.

The BRFSS contains health, insurance, and demographic data for 4.9 million U.S. residents living within the 50 states and Washington, D.C. The data is collected through annual telephone surveys issued by state health departments and the District of Colombia. The BRFSS aims to collect 4,000 interviews per state every year. Once respondent data is collected, it is forwarded, complied and analyzed by the Centers for Disease Control and Prevention (CDC). Survey questions are largely standardized and contain data of U.S. resident risk behavior, preventive health practices, and overall health. The questions include standard core questions asked by all states every year, rotating core questions asked by all states every-other year, optimal modules which are a set of optimal standardized questions, and state added questions.

The samples in the BRFSS datasets include: (1) household sample – which accounts for 80% of participants and (2) cellular telephone sample –which accounts for 20% of participants. Disproportionate stratified sampling (DSS) is used for telephone surveys while the cellular telephone sample is randomly generated based on confirmed area codes. The data set accounts for weighting variables in order to remove known statistical biases. Overall the data set is national in scale and provides a variety of variables that can be used to understand the basic health behaviors of individuals in the sample.

This study uses the data to estimate linear probability models on a variety of binary variables controlled for education, income, gender, and state with the key variable being Medicaid. The following models limit all samples to individuals under the age of 65.

MEDICAID

All variables are regressed against the variable *Medicaid* as it is the key variable in the analysis. *Medicaid* is a binary variable where 1 indicates that states had enacted the full Medicaid expansion and 0 indicates states have not. Information about the Medicaid expansions was gathered from the Henry J. Kaiser Family Foundation and is summarized in Table 2.

TABLE 2
SUMMARY STATISTICS

Name	Obs	Mean	Standard Deviation
Insurance (Covered by any health insurance plan)	4,805,879	0.89	0.31
Medicaid (Resides in a state that expanded Medicaid at the time the individual is interviewed)	4,737,468	0.11	0.31
Medcost (Individual reported that they could not access healthcare due to cost reasons)	4,809,085	0.12	0.32
Test HIV (Individual was tested for HIV)	3,742,866	0.32	0.47
Pneumonia Vaccine (Individual received a pneumonia vaccine)	4,276,170	0.36	0.48
Flu Vaccine (Individual received a flu shot)	4,628,848	0.45	0.50
High Cholesterol (Individual was diagnosed with high cholesterol)	2,652,784	0.62	0.49
Physical Health (Number of days in the past 30 days that the individual reported their physical health was not good)	4,716,238	4.31	8.83
Mental Health (Number of days in the past 30 days that the individual reported their mental health was not good)	4,737,324	3.38	7.70

INSURANCE

The study first assesses the impact of Medicaid on insurance status as insurance status is expected to increase health care utilization and health. *Insurance* is a binary variable where 1 indicates the individual has insurance and 0 indicates the individual does not have insurance. Insurance includes any form of coverage, including prepaid plans such as HMOs, government plans such as Medicaid and Medicare, and Indian Health Services plans.

The study uses the following categorical sociodemographic variables to analyze and compare the impacts of insurance, health utilization and health status for the following sociodemographic characteristics: *gender*, *income level*, *marital status*, *race*, *age*, and *education* level.

Health Care Utilization

Next, the study uses a variety of binary variables to examine the impact of Medicaid expansion on health care utilization including: *Medcost*, *Checkup*, *Aids Test*, *Pne Vaccine*, *Flu Vaccine*, and *High Cholesterol*. *Medcost* indicates the percentage of people who could not see a doctor in the past twelve months due to cost. *Checkup* indicates the percentage of people who had a doctor checkup in the past 12 months. *Aids Test* is used to indicate the percentage of adults who have been tested for HIV at some point in their life. *Pne Vaccine* is used to indicate the percentage of adults who have ever had a pneumonia vaccine in their lifetime. *Flu Vaccine* is used to indicate the percentage of adults who have ever had a flu vaccine in the past twelve months. *High Cholesterol* is used to indicate the percentage of adults who have been told by a doctor, nurse, or any health care professional that they have high cholesterol.

Health

Lastly, this study uses the following variables to examine the impact the Medicaid Expansion had on overall health: *Physical Health* and *Mental Health*. *Physical Health* indicates the percentage of adults who have experienced poor physical health in the past 30 days. *Mental Health* indicates the percentage of adults who have experienced poor mental health in the past 30 days.

In general, the models are often restricted to individuals with incomes less than $35,000 as they are expected to be the most impacted by the Medicaid expansions

V. Empirical Methodology

The objective of this analysis is threefold: (1) to examine the effect of the Medicaid expansion on insurance status, (2) to examine the effect of the Medicaid expansion on health care utilization and (3) to examine the effect of the Medicaid expansion on overall health. To perform this analysis we formulate the following multivariable regression equation to estimate these effects:

$$Y_{ist} = \hat{B}_0 + \hat{B}_1 \, (Medicaid_{st}) + X_{ist} \, \alpha + E_{ist}.$$

In the model, for objective 1, Y is a binary indicator if the individual has insurance or not. For objective 2, Y represents a multitude of variables for *Medcost*, *Test Aids*, *Pne Vaccine*, *Flu Vaccine* and *High Cholesterol*. For objective 3, Y represents a

measure of physical or mental health measured as number of days in the past month individuals had bad physical or mental health. The subscripts for *Y* denotes individual *i* who resides in state s in year *t*.

Medicaid is a binary indicator that captures when a state expanded Medicaid and it is a state and year variable. Variable *X* represents a variety of sociodemographic variables for *Age*, *Education*, *Income*, *Sex*, *State*, *Year*, and *Race*.

The coefficient of interest is $\hat{B}_1$, the main coefficient representing the effects of the Medicaid expansions on *Y*. When *Y* is insurance we expect $\hat{B}_1$ to be positive because more people should have insurance in states that expanded Medicaid. Likely, when *Y* is health care utilization variables, we expect $\hat{B}_1$ to be positive as more individuals have insurance, thus reducing the cost of health care, and increasing the use of preventative care. Lastly, when *Y* is physical or mental health we expect $\hat{B}_1$ to be negative as increased health care utilization should led to a decrease in bad physical and mental health days. Results were as expected. Although the results for physical and mental health indicate negative numbers, not enough time has materialized to show fuller, more statistically significant, effects on overall health.

This methodology is known as a "difference-in-differences" framework, wherein trends in outcomes for the "treatment group" which comprises states that expanded Medicaid in a given year are compared against trends in outcomes for the "control group" which comprises states that have not yet expanded Medicaid. These models control for state binary indicators, that account for all stable state-specific factors, and year binary indicators, that account for all national trends (such as economic cycles and other confounding policy changes).

VI. Results

Impact of Medicaid on Insurance

The primary purpose of the Medicaid Provision was to expand insurance coverage for low income individuals. Because Medicaid has only been expanded in 32 states, we are able to directly compare the impact of the expansion on insurance through comparing states that have expanded Medicaid relative to those that did not. The first model offers an overview of the effect on insurance coverage on the overall population. Model 1 indicates that for adults under 65 years of age, insurance coverage increased by 1.5 percentage points in states that expanded Medicaid relative to those that did not (see Table 3, column 2). The increase in insurance coverage, by corollary, represents a reduction in uninsurance rates among the population. Specifically, prior to the ACA, 14.5% of the population was uninsured. The 1.5 percentage point gain represents a 10.49% decrease in the uninsured population, suggesting that uninsurance rates for adults decreased from 14.5% to 12.73% due to the Medicaid Provision.

TABLE 3
EFFECTS OF THE MEDICAID EXPANSIONS ON INSURANCE COVERAGE

	No Income Limit	Income<=$35,000	Income<=$25,000
Medicaid	0.01517192**	0.04506304***	0.04897797**
Age			
18-24	(base)	(base)	(base)
25-34	-0.00888834	-0.05749013	-0.07134399
35-44	0.02079839***	-0.04251018	-0.05221843
45-54	0.04489486***	0.00551334	-0.00246421
55-64	0.08137527***	0.08245306***	0.07760696***
Education			
Did not Graduate High School	(base)	(base)	(base)
Grad High School	0.09983445***	0.07237669***	0.06016357***
Attended College or Tech School	0.13604484***	0.10756716***	0.0947405***
Graduated College or Tech School	0.16453091***	0.12803846***	0.10581459***
Missing	0.10999464***	0.0322556*	0.01603017
Income			
<$15,000	(base)	(base)	(base)
$15,000-25,000	-0.00724217	-0.00044477	0.00184955
$25,000-35,000	0.09793429***	0.10894252***	
$35,000-50,000	0.17422605***		
$55,000+	0.23883485***		
Missing	0.13091638***		
Sex			
Male	(base)	(base)	(base)
Female	0.02890134***	0.05789482***	0.06175869***

Asterisks indicate variable statistically significant at the 1% (***), 5% (**), and 10% (*) significance level

Additionally, the model suggests trends in relation to insurance coverage among specific sociographic variables, including age, education and income are consistent with the literature. Overall, insurance coverage for age represent a U-Shaped curve, declining up to a certain point and then increasing. Additionally, the model shows that in general insurance rates increase with both income and education. However, because the expansions have a larger impact on low income adults, these specific relationships are examined further for individuals with less than $35,000 of annual income below.

While, Model 1 presented estimates for individuals across the entire income distribution, it is important to underscore that the Medicaid expansions under the ACA specifically targeted lower-income individuals. As noted earlier, states which expanded their Medicaid programs now covered all individuals (100% subsidies) with income levels under 138% of the federal poverty level (FPL). Furthermore, individuals who signed up for health insurance under the individual insurance exchanges also receive partial subsidies up to 400% of the FPL. Thus, Model 2 specifically assesses how the Medicaid expansion impacted "low-income" individuals, defined as those with an annual income < $35,000 (see Table 3, column 3).

As expected, the Medicaid expansions have a much stronger impact on insurance coverage among lower-income households. The estimate is statistically significant and suggests that insurance coverage for low-income individuals increased by about 4.5 percentage points post-expansion among states that expanded Medicaid relative to states that did not. This is a net increase in insurance coverage, and thus, by corollary, also represents a net reduction (4.5 percentage points) in uninsurance rates among this population of low-income adults. For instance, prior to the ACA, uninsured prevalence for this group of individuals was 28.33%. The estimate from the model suggests that the percent of uninsured adults thus decreased from 28.33% to 23.83% (change of 4.5 percentage points), a decrease of 15.88%. In comparison, insurance coverage for those individuals with incomes above $35,000 only increased by 0.07 percentage points.

Like in Model 1, the estimates from the other variables are consistent with the literature. The following data suggests coverage trends within the defined population in respect to age, education, gender and income are consistent with those found across the entire income distribution. With respect to age, the model suggests a U-shaped relationship. Insurance coverage tend to decline up to a certain age, and then increase. Specifically, the probability of coverage is about 5.7 percentage points lower among adults ages 25-34 (relative to the reference category of adults ages 18-24). Older adults between the ages of 35-44 are about 4.3 percentage points less likely to be insured (relative to the base group), though when compared with those ages 25-34, they are about 1.4 percentage points more likely to be covered. The probability of coverage further increases with age. This is likely a result of the fact that as people age their healthcare demand increases and therefore insurance becomes more valuable.

With respect to education, there is a consistent and monotonic increase; higher-educated individuals are more likely to have insurance coverage. It is possible that higher educated individuals have a greater awareness of the benefits of insurance and therefore are more likely to attain it. The data shows that relative to those with less than a high school education, college educated individuals are about 13 percentage points more likely to be covered. The effects across the age and education distributions are generally statistically significant.

Similarly, there is a consistent and monotonic increase in respect to incomes above $25,000. Individuals that have annual incomes up to $25,000 have a similar probability of being insured. However, individuals with incomes above $25,000 are more likely to

be insured if they have higher incomes. This could be due to the fact that education and income are closely tied, therefore people with higher incomes are often more educated and therefore more likely to buy insurance.

Additionally, the data indicates that females are more likely to be insured than males. Specifically, for adults within the population, females are 5.7 percentage points more likely to be insured than males. The results are statistically significant and could be attributed to the fact that on average females spend more on personal health care than males, thus making health insurance more valuable to them. In 2102, the Centers for Medicaid and Medicare found that females spent 22% more on personal health care than males (Centers for Medicare and Medicaid Services (2015)).

Insurance trends in respect to age, education level, gender, and income are generally consistent throughout the following models.

While Model 2 provides general information regarding insurance coverage in the population, the following models examine the effect of the Medicaid Provision on specific sociographic groups within the defined population of individuals making less than $35,000 per year.

Overall, for adults with income less than $35,000, Medicaid expansion affected males and females similarly. This is expected as the expansions did not intend to close insurance gaps between genders.

In addition, low-income Whites had a greater percentage point increase in post-expansion insurance rates than low-income Blacks. Specifically, the model suggests that insurance coverage for Whites increased by 6.0 percentage points post-expansion among states that expanded Medicaid relative to states that did not, while Blacks saw a 5.1 percentage point increase. The results suggest that uninsurance rates for White adults decreased from 28.33% prior to the expansion to 22.3% in states that expanded Medicaid, representing a 21.35% decrease in uninsurance rates. Meanwhile, the model suggests that post-expansion insurance coverage among Blacks within the defined population increased by 5.06 percentage points, indicating a decrease in uninsurance rates from 30.67% to 25.61%, or a post-expansion decrease in uninsurance rates of 16.49%.

Lastly, the expansion led to a larger increase in insurance coverage for those between the ages of 25-55 as opposed to adults between the ages of 18-24 and 55-64. This could be due to the new regulation which allows individuals younger than 26 to remain on their parents insurance and the already high likelihood that adults between the ages of 55-64 were insured prior to the expansion.

In summary, the data suggests that the Medicaid Expansion had a larger impact on low-income individuals than the entire population. In addition, for adults with incomes less than $35,000, the Medicaid expansion had the largest impact on Whites and adults between the ages of 25-55.

The final set of insurance models further test the impact of the Medicaid Provision on even lower income individuals (see Table 3, column 4). In 2015, the Medicaid eligibility threshold for full coverage (100%) was $16,243. Therefore, the following

models are restricted to "very-low income" individuals defined as those who make equal to or less than $25,000, which is slightly above the amount needed to qualify for full coverage. It is not surprising that the Medicaid Provision had the greatest impacts on this population. Specifically, the results show that insurance coverage increased by 4.9 percentage points post-expansion among states that expanded Medicaid relative to states that did not. This suggests that the Medicaid Provision increased insurance coverage from 65.23% to 70.13% in states that adopted the expansion. Overall, this reduced the uninsurance rate for the defined population by 14.1%, compared to a reduction in the uninsurance rate for all incomes of 10.49%.

In general, gender, race, and age sub-groups showed similar trends when compared to a $35,000 income restriction. However, the listed variables indicated greater percentage point differences between insurance coverage among states that have expanded Medicaid compared with those that have not. This contributes to the conclusion that the Medicaid Provision had a larger impact on low-income individuals in all sociographic groups. In general, the closer an individual's income is to the expanded Medicaid threshold (138% of the FPL), the greater increases in probability we would expect to see in insurance coverage in post-expansion states.

Impact of Medicaid on Health Care Utilization

Prior to the enactment of the ACA, cost increasingly restricted individuals from receiving access to basic health care. However, given the decrease in uninsurance rates, as demonstrated in the post-expansion states, individuals in post-expansion states are less constrained by out-of-pocket costs to receive health care. That being said, we expect these individuals will be more likely to utilize health care. The following models look at a variety of preventative care variables to examine the impact of health care utilization of individuals in post-expansion states (see Table 4).

Table 4

Effects of the Medicaid Expansion on Healthcare Utilization

	Entire Pop	White	Black
Income <= 35,000			
Medcost	-0.03395034	-0.03885283	-0.04880457
Checkup	0.01247193	0.01580523	0.01699612**
Aids Test	-0.00197479	0.0062671	-0.00590728
Pne shot	0.00823681**	0.00759122*	0.0163193**
Flu Shot	-0.00559367	-0.01166973	-0.01101428
High Cholesterol	0.01194364***	0.01362566***	0.01899825***
Income <= 25,000			
Medcost	-0.03896932	-0.04665938	-0.05499337
Checkup	0.01614215	0.02033066*	0.01779176**
Aids Test	-0.00486211	0.0006553	-0.00089568
Pne shot	0.00553107	0.0034853	0.01456926**
Flu Shot	-0.00261558	-0.00814631	-0.01650187
High Cholesterol	0.01133417***	0.01306414***	0.01729921**

Asterisks indicate variable statistically significant at the 1% (***), 5% (**), and 10% (*) significance level

First, we examine the effect of the expansion's impact on cost. One of the major goals of the ACA was to increase the affordability of health care. Prior to the ACA, 15.14% of adults were not able to see a doctor in the past 12 months due to cost. Not surprisingly, this percentage was larger for low-income adults. Specifically, for adults with incomes less than $35,000, 30.82% were not able to see a doctor in the past 12 months due to cost. Additionally, adults with incomes less than $25,000 the percentage was 34.57%. The results suggest the Medicaid expansion was able to reduce the likelihood that adults would not be able to see a doctor due to cost. Specifically, the results indicate a 3.4 percentage point decrease in the probability adults with incomes less than $35,000 and 3.9 percentage point decrease for adults with incomes less than $25,000. Additionally, Blacks saw larger decreases in inability to see a doctor due to cost when compared to Whites. For Blacks with incomes less than $35,000 there was a 4.9 percentage point decrease in their likelihood that they could not see a doctor due to cost in states that had enacted Medicaid compared to those that did not. Meanwhile, Whites had a 3.9 percentage point decrease. This data further indicates that the Medicaid Provision has been effective in increasing affordability and equality in the health care system.

Next, we examine if the decreased out-of-pocket costs led to increases in healthcare utilization through looking at a variety of preventive care variables. The regression models indicate with statistical significance that adults with incomes less than $35,000 were more likely to have received a pneumonia shot and a pap smear. Specifically, in states that expanded Medicaid, adults under 65 with incomes up to $35,000 were 0.823 percentage points more likely to have a pneumonia shot. Pneumonia is an infection in the lung and currently causes 15% of all deaths in children under 5 years old. For older children, although often less fatal, pneumonia is the number one cause of hospitalizations. In addition, adults are also susceptible to the disease. Overall, the American Thoracic Society reports that pneumonia was one of the top ten most expensive conditions of inpatient hospitalizations, accounting for $10.6 billion in costs in 2011 (American Thoracic Society (2015)). An increase in the probability that individuals will receive a pneumonia shot as a result of the Medicaid expansion will likely decrease pneumonia related hospitalizations and deaths.

Additionally, adults with incomes less than $35,000 were 1.2 percentage points more likely to be diagnosed with high cholesterol in states that had expanded Medicaid compared to those that did not. This represents a 3.3 percent increase in the likelihood of being diagnosed with high cholesterol. Currently, the CDC reports that 31.7% of adults in the U.S. have abnormal levels of low-density lipoprotein, also known as "high" cholesterol. People with high cholesterol are at twice the risk for heart disease compared to people at ideal levels of lipoprotein. In addition, high cholesterol does not have many symptoms, so people without a doctor diagnosis are often unware they are at heightened risk for heart attacks. Without an early diagnosis, people with high cholesterol are likely to need more costly and less effective treatment such as bypass surgery and additionally have a higher risk of dying of heart failure (Centers for Disease Control and Prevention (2015)).

Lastly, while not statistically significant, there is some likelihood that adults were more likely to receive a checkup, Pap smear and aids test in states that enacted Medicaid. Overall, the results suggest that the Medicaid expansion has increased preventative health care utilization. This is important because preventative care is both less expensive and more effective in treating diseases when compared with secondary and tertiary types of care. For many illnesses, such as cancer or heart disease, identifying the health problem early can be the difference between severe health problems and relatively easy treatment.

Medicaid Impact on Health

Overall, being able to identify and diagnosis diseases earlier through preventative care should led to an increase in both physical and mental health. In addition, it is likely that providing needy individuals with affordable and accessible health care will decrease stress and improve mental functions. The last set of regression models assess the impact of the Medicaid expansions on both physical and mental health (see Table 5). Although our data did not find substantial evidence that the Medicaid expansions improved physical or mental health, we have good reason to expect that this would be significant in the future as health is a stock that takes time to improve. With the Medicaid Provision only being enacted 1 year prior to the data, it is likely that the impact on self-reported physical and mental health has not become apparent just yet. However, given the preliminary evidence regarding insurance and health care utilization, we expect this to improve in the future.

Table 5
Effects of the Medicaid Expansion on Health

	Entire Pop	**White**	**Black**
Income<=35,000			
Physical Health	-0.00285262	-0.00352065	-0.00588732
Mental Health	-0.00232956	-0.00124013	-0.00725472
Income<=25,000			
Physical Health	-0.0036503	-0.00942421	-0.00577629
Mental Health	-0.00416804	-0.00586452	-0.01244556

Asterisks indicate variable statistically significant at the 1% (***), 5% (**), and 10% (*) significance level

VII. Conclusions

The results indicate that the Medicaid expansions have effectively decreased the uninsurance rate, providing close to 3 million low-income individuals with health insurance. Prior to the ACA, rising healthcare costs increasingly left low-income adults

without healthcare options. However, the increase in insurance coverage has decreased out-of-pocket costs for recipients and expanded their treatment options. Overall, health care became more affordable and accessible to low-income adults and this will likely translate into better overall health in the future.

Moreover, the Medicaid expansions have shown to decrease income inequality in the health care system. For example, lower income individuals were more positively affected by the Medicaid expansions. This is especially timely as health care income inequality has risen in the past decades. In 1977, when Medicaid and Medicare were first implemented, the poor were receiving 14% more care than the wealthy. This was an expected difference as the poor were likely to be sick than the wealthy. However, since then, the trend has eroded and in 2012 the wealthy were going to 40% more doctor visits than lower income Americans. The shift in income trends within the health care system is likely due to the rising health care costs, which were largely shifted onto patients (Woolhandler (2016)). The Medicaid Provision was able to offer a partial solution to this problem, as it has decreased out-of-pocket spending and increased health care utilization for Medicaid recipients.

In addition, although Whites saw greater decreases in uninsurance rates, Blacks saw greater increases in access to health care when compared to the Whites. Current racial disparities in average life expectancy between Whites and Blacks indicate racial inequalities in the U.S. health care system. Although racial inequalities have been trending downwards, from a 7.6 year difference in 1970 to a 3.8 year difference in 2010, the problem still resides. In general, the disparities are mainly attributed to Black's higher death rates at younger ages due to heart diseases, diabetes, and cancer as well as higher risks for HIV infection, homicide, and infant mortality (Ayanian (2016)). The Medicaid expansions indicate that Blacks have seen greater decreases in healthcare costs and increases in health care utilization when compared to Whites, which will likely led to further decreases in racial inequality.

Despite the successes of the Medicaid expansion, the ACA and Medicaid Provision have faced fierce political criticism, mainly from the Republican Party largely due to ideological disagreements. Healthcare was a prominent topic of discussion and debate throughout the 2016 election cycles. Debates have recently cumulated in a rejection of the "Better Care Reconciliation Act"; the Republican led bill aimed at addressing concerns regarding the ACA. Primary components of the bill included removing the individual mandate and employee mandates, loosening restrictions on insurance companies and reducing federal funding for Medicaid. Specifically, the bill reduced federal funding from 95 percent federal funding for the expansion population to 85 percent by 2021. The bill stated that gradual decreases between 2021 to 2024 would ultimately led to the standard state matching rate for the expansion population (between 50-75 percent depending on enroll). The Congressional Budget Office (CBO) reported the bill would reduce federal spending for Medicaid by $772 billion and would decrease Medicaid enrollment by 16% (H.R. 1628).

In July 2017 the bill passed in the House of Republicans but was stopped at the

Senate by a close vote. Although the nation will maintain all of the current ACA regulations for the time being, health care will likely remain a topic of debate and an area for improvement. However, as the nation searches for ways to improve the current healthcare system, it is important to understand the impacts of the Medicaid expansion. If an appeal were to take place, health insurance could be taken away from millions of individuals, making it harder for low-income and minority residents to attain health care. In addition, decreases in preventative care will likely lead to higher health care spending in the long-run. For example, a reduction in Medicaid eligibility would likely increase uncompensated hospital care and overall healthcare costs. This is an area for further study as the long-term impacts of cost savings have not yet become apparent. Although we cannot predict what changes will take place or how they will impact the American people, this paper demonstrates the Medicaid expansions were successful in helping millions of needy Americans afford and access health care.

References

American Thoracic Society (2015). "Top 20 Pneumonia Facts." Thoracic.org. Web 2015.

Ayanian, John (2016). "The Costs of Racial Disparities in Health Care." *Harvard Business Review*, 05 Feb. 2016. Web. 21 Dec. 2016.

Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein (2013). "The Oregon Experiment — Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine* 368, no. 18, pp. 1713-1722.

Blendon, R. J., Orav, E. J., & Epstein, A. M. (2016) "Changes in utilization and health among low-income adults after Medicaid expansion or expanded private insurance." *Journal of the American Medical Association Internal Medicine*, 1 Oct 2016. Web. 5 Sept. 2017.

Catlin, Aaron C. and Cathy A. Cowan (2015). "History of Health Spending in the United States, 1960-2013" Historical NHEP, November 19, 2015. Web Sept 5 2017.

Centers for Disease Control and Prevention (2015). "High Cholesterol Facts." CDC.com, 17 Mar. 2015, www.cdc.gov/cholesterol/facts.htm

Centers for Medicare & Medicaid Services (2015). "National Health Expenditure Fact Sheet.". Web. 21 Dec. 2016.

Cohen, Robin, Michael Martinez (2014). "Health Insurance Coverage: Early Release of Estimates from the National Health Interview Survey". Division of Health Interview Statistics, National Center for Health Statistics. Web. 2014.

Colman, Greg and Dhaval Dave (2015). "It's About Time: Effects of the Affordable Care Act Dependent Coverage Mandate on Time Use". *NBER Working Paper Series*, November 2015.

Congressional Budget Office (2017). "H.R. 1628, Better Care Reconciliation Act of 2017." CBO.gov, 26 June 2017, www.cbo.gov/publication/52849.

Courtemanche Charles, James Marton, Benjamin Ukert, Aaron Yelowitz, Daniela Zapata (2016). "Impacts of the Affordable Care Act on Health Insurance Coverage in Medicaid Expansion and Non-Expansion States". *NBER Working Paper Series*, April 2016.

Dranove, David (2017). "The Impact of the ACA's Medicaid Expansion on Hospitals' Uncompensated Care Burden and the Potential Effects of Repeal." The Commonwealth Fund, Commonwealth Fund, 3 May 2017.

Frean, Molly, Jonathan Gruber, Benjamin Sommers (2016). "Premium Subsides, the Mandate, and the Medicaid Expansion: Coverage Effects of the Affordable Care Act". *NBER Working Paper Series*, April 2016.

Hu, Luojia, Robert Kaestner, Bhaskhar Mazumder, Sarah Miller, and Ashley Wong (2016). "The Effect of the Patient Protection and Affordable Care Act Medicaid Expansions On Financial Well-Being". *NBER Working Paper Series*, April 2016.

Garrett, Bowen, Anuj Gangopadhyaya (2016). "Who Gained Health Insurance Coverage Under the ACA, and Where Do They Live?" Robert Wood Johnson Foundation, December 2016. Web. 23 Aug 2017.

Kaestner, Robert, Bowen Garrett, Anuj Gangopadhyaya, and Caitlyn Fleming (2015). "Effects of ACA Medicaid Expansions on Health Insurance Coverage and Labor Supply." *NBER Working Paper Series*, December 2015.

Kaiser Family Foundation (2012). Focus on Health Reform, August 2012. "A Guide to the Supreme Court's Decision on the ACA's Medicaid Expansion". The Kaiser Family Foundation. 21 Dec. 2016.

Kaiser Family Foundation (2016) "Status of State Action on the Medicaid Expansion Decision." Kaiser family Foundation - Health Policy Research, Analysis, Polling, Facts, Data and Journalism. 14 Oct. 2016. Web. 21 Dec. 2016.

McMorrow, S, G. M. Kenney, S. K. Long, and D. E. Goin (2016). "Medicaid Expansions from 1997 to 2000 Increased Coverage and Improved Access to Mental Health Outcomes for Low-Income Parents". Health Policy Center, Urban Institute, Washington, D.C. January 2016.

OECD (2013). "Life expectancy in the US rising slower than elsewhere, says OECD". *OECD Health at a Glance*. OECD, 2013. Web. 21 Dec 2017.

OECD (2015). "OCED Health Statistics 2015". Country Note: How Does Healthcare Spending in the United States Compare. OECD.org, 7 July 2015. Web. 21 Dec 2016.

Simon, Kosali, Aparna Sonia, John Cawley (2016). "The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the 2014 ACA Medicaid Expansions". *NBER Working Paper Series*, May 2016.

Sommers, Benjamin D., Katherine Baicker, and Arnold M. Epstein (2012). "Mortality and Access to Care among Adults after State Medicaid Expansions." *New England Journal of Medicine* 367, no. 11.

Trump, Donald (2016). "Healthcare Reform." Trump Pence Campaign, Web. 21 Dec. 2016.

Woolhandler, Steffie (2016). "Health Care Inequality On The Rise." *The Huffington Post*. TheHuffingtonPost.com, 5 Aug. 2016. Web. 21 Dec. 2016.

Works, Richard (2016). "The Effect of the Affordable Care Act's Medicaid Expansion on Health Insurance Coverage in 2014." Bureau of Labor Statistics. June 2016. Web. 5 Sept. 2017.

ESG Rating Accuracy: The Case of Greenhouse Gas Emissions

By Cyrus Woodman*

Investments made in accordance with environmental, social and governance (ESG) criteria have grown to represent nearly $8.72 trillion in the United States, approximately 1 of every 5 dollars under professional management. Research has shown, on the aggregate large universe level, that the bulk of the ESG scores generated by ESG rating agencies are inaccurate in measuring true firm sustainability while exhibiting low predictive power in explaining firm performance. Although, by analyzing an isolated ESG related event and corresponding ESG metrics for firms exposed to that event, one might get better accuracy and predictive power from those metrics. By leveraging key regulatory developments from recent international meetings such as the 2015 Paris Agreement and 2016 Kigali Deal, we establish a specific ESG related event that theoretically posed a material impact on publicly traded securities. The valuation impacts realized by firms around these events allow us to test the correspondence and accuracy of ESG ratings. Accordingly, we perform OLS regression analysis of stock price performance and ESG scores for firms that should have been impacted by these regulatory developments to make determinations on the predictive power and accuracy of ESG scores.

Keywords: ESG Ratings, GHG Emissions, Paris Agreement, Kigali Deal, event study.

I. Introduction

The purpose of this research endeavor is to leverage two recent events: the 2015 Paris Agreement, and the 2016 Kigali Deal (amendment of the Montreal Protocol) to assess the accuracy of ESG rating data in tracking the potential impacts of these developments on the stock price performance of greenhouse gas (GHG) emitters. Quite specifically, the recent 200 nation agreement on eliminating hydrofluorocarbons (HFCs) has an extremely relevant and material impact on air-conditioner, and refrigerator

---

* Email: cyruswoodman@gmail.com. I am extremely grateful to have had the opportunity to work with Professor Otgontsetseg Erhemjamts, Associate Professor of Finance at Bentley University. Her guidance, commitment and encouragement were essential to my success over the last 14 months on this research endeavor. In particular, the effort that Professor Erhemjamts put forth during the revision phase was what allowed me to communicate this work in a clear and meaningful way. I also thank Professor Dave Szymanski, Associate Professor of Natural and Applied Sciences at Bentley University, for being an inspiration to students like myself and stimulating the drive to be curious, study and achieve. Additionally, I applaud Professor Aaron Jackson, Professor of Economics at Bentley University for his dedication to our university's Honors Program and this undergraduate research journal, *Fusio*

manufacturers.[1] With growing international consensus that not mitigating GHG emissions (e.g., continued reliance on HFCs) could lead to global temperature increases of between 3.7º C and 4.8º C (between 6.7º F and 8.6º F) by 2100, governments have acted decisively to phase out their relevant products.[2] In turn, this has fundamentally shifted the profit strategy and financial model for many air-conditioner, refrigerator and chemical manufacturers. With the cheaper HFC emitting technology being restricted, companies' in these industries have faced the impact of regulatory constraint in driving innovation for the sake of sustainable gains: "The recent amendment of the Montreal Protocol on the use of HFCs highlights a growing trend in our industry and across many industries: the need to innovate and continue to develop more environmentally sustainable technologies" (Kevin McNamara, SVP of LG Electronics USA Air Conditioning Business). Since sustainable investing tools attempt to integrate corporate environmental performance and strategy into fundamental security analysis, the developments surrounding HFCs offer an opportunity to analyze the effectiveness and capability of this integration process.

While sustainable investing has seen tremendous growth in practice as of late, the successful implementation of ESG strategies will depend on the accuracy and timeliness of ESG ratings. Sustainable investing in the finance industry has traditionally been centered on the idea that corporate environmental performance is materially relevant in determining firm risk and free cash flow (Busch et al., 2015). In a study of the world's 500 largest companies, Harvard Business School and Calvert Investments determined that "ESG performance has a high correlation with strong valuation, expected growth and lower cost of capital" (Serafeim et al., 2015). However, it is not entirely clear how well the leading ESG rating systems have been able to measure a firm's underlying sustainability through "ESG performance". In two different white papers produced in 2015 & 2016 by MIT and Boston Consulting Group, investor surveys indicated that only 36% of respondents believe that a firm being included in a sustainability index is an important thing. The report also highlighted that indices and ranking agencies typically vet corporate environmental reports for completeness rather than accuracy. And with nearly 50,000 companies involved in ESG reporting, being evaluated by 150 different rating systems on a total of 10,000 different performance metrics, the complexity of these rankings and indices have been questioned by those in both the scientific and investment communities. Other research also shows that while 45% of asset managers and owners claim to integrate ESG issues in to fundamental analysis, it is still unclear how this integration is done and to what effect ESG rankings are systematically included in security analysis (Slager, 2014). Ultimately, these issues underscore the need to further investigate ESG rating systems to better understand their responsiveness to global developments in the field of sustainability.

1. For example, the cost to India of phasing out HFCs by 2050 could range, depending on the mitigation plan, from around $13 billion to $38 billion, according to a study by the India-based Council on Energy, Environment and Water (https://www.nytimes.com/2016/10/13/world/asia/india-air-conditioning.html?mcubz=0).
2. http://www.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_SPM.pdf

Having established a need for more investigation in to ESG rating systems, it is important to emphasize the importance of this inquiry to the field of investing and more broadly, the field of sustainability. According to the United States Sustainable Investing Forum (US SIF), assets invested with inclusionary strategies now exceed those invested with exclusionary approaches.[3] For investors, this means that the success of their sustainable investing strategies is closely tied to the ability to adequately leverage ESG data and ranking system results. This is somewhat similar to the importance of ESG data and ranking systems to sustainability advocates, scientists and policymakers. Research has found that in order for sustainable investing to spur truly sustainable business development, ESG metrics and rankings must take a long-term view on firms that is focused on underlying sustainability strategies that impact material aspects of a firm such as their business plan, human resource policies and approach to governance (Slager, 2014). These factors are what make this research endeavor so appropriate. By focusing on the developments around HFC's and the Montreal Protocol, we identify a scenario whereby the interests of sustainability minded investors and advocates intersect with the performance of ESG data and ranking products.

This paper attempts to assess the performance of prevailing ESG ratings in correctly tracking the impacts of international HFC regulations. Investigating ESG metrics of HFC emitters before the Paris Agreement as well as the Kigali Deal, and checking them against equity price reactions to these events helps determine how strong of a proxy these metrics were for the readiness of HFC emitters for phasing out HFC and introducing new and more expensive technology. In turn, this will allow us to inform the sustainable investing field on how different ranking systems and methodologies performed in light of changes in international regulations. It will simultaneously allow us to better understand how different ranking systems are accounting for underlying firm sustainability, because developments in HFC regulation are materially relevant to firms in the long run (business plan, manufacturing etc.). Our research question is as follows: "With the explosive growth of ESG factor integration in fundamental financial analysis and equity research, how accurately have ESG rating entities tracked the trends in sustainable development and regulatory impact for publicly traded companies? More specifically, looking at the Paris Agreement of 2015 and Kigali Deal of 2016, have GHG emitters (including HFC emitters) been accurately assessed by ESG rating engines on their ability to manage sustainable corporate strategy around this change?"

At the center of our examination of ESG rating accuracy and responsiveness to HFC regulatory developments are the actual ratings themselves. At a very high level, ESG raters aim to produce quantitative ratings (scores) on various environmental, social and governance performance indicators as a means to provide ESG minded inves-

3. Negative/exclusionary screening means the exclusion from a fund or plan of certain sectors or companies involved in activities deemed unacceptable or controversial (e.g., fossil fuels, nuclear power, tobacco, gaming, pornography, military weapons, etc.). Positive/best-in-class or ESG integration means the systematic and explicit inclusion by investment managers of positive ESG factors into traditional financial analysis.

tors with descriptive information that adds transparency as to the degree that a firm's actions are in fact "responsible" (Chatterji et al., 2009). While some "ESG minded" investors seek company ratings to better align their individual values with their investment strategy, the practical application of company ESG data is for asset managers to evaluate a firm's financial risks and opportunities stemming from company ESG decisions (Chatterji et al., 2009). Over the course of the last 20 years, as sustainable investing and ESG integration strategies have grown in practice and monetary value, so too has the granularity of ESG ratings and breadth of security and market coverage. The growth of the ESG analysis field has thus led to a number of mergers, acquisitions and partnerships that have resulted in a more condensed group of leading ESG raters that exists today. In this paper, we utilize company level ESG data and reporting information from four leading ESG data providers: MSCI ESG Research, Thomson Reuters, CDP, and Sustainalytics.

II. Recent International Climate Change Accords

Paris Agreement (COP 21)[4]

On December 12th, 2015, parties to the United Nations Framework Convention on Climate Change (UNFCCC) agreed to aggressively combat climate change and ramp up efforts to pursue and facilitate a low carbon global economy. The agreement's core purpose is to establish a long-term temperature limit, aiming to keep global temperature increase below 2º C. To achieve this goal, COP 21 parties established "global peaking" through Article 4 of the Paris Accord, where different countries agree to stop GHG emissions growth after certain deadlines depending on the developed or developing nature of their economies (UNFCCC). Through the use of "Nationally Determined Contributions" or NDCs, the Paris Agreement establishes binding commitments by the parties to COP 21 to communicate their contribution efforts with clarity every 5 years.

Based on the United States published "intended nationally determined contribution (INDC)", the plans behind this domestic GHG emissions strategy seem to be quite material for U.S. equities. By 2020, the U.S. INDC expects to reduce emissions "in the range of 17% below 2005 level in 2020" (US NDC). An additional target of achieving a further 9-11% GHG emission reduction based on 2005 levels by 2025 is also disclosed in this same INDC. To reach these targets, the U.S. INDC states that the U.S. will utilize existing regulatory authorities and reduce emissions from all economic sectors, "target reflects a planning process that examined opportunities under existing regulatory authorities to reduce emissions in 2025 of all greenhouse gases from all sources in every economic sector" (US INDC). Furthermore, according to the U.S. INDC, the United States does not intend to utilize market mechanisms to facilitate these reductions.

4. COP stands for Conference of the Parties, referring to countries that have signed up to the 1992 United Nations Framework Convention on Climate Change (UNFCCC). The COP in Paris is the 21st such conference.

For our research purposes, we believe that these U.S. INDC specifics highlight the materiality and significance of COP 21 to publicly traded firms in the U.S. As the INDC states, the United States "intends to achieve an economy-wide target of reducing its greenhouse gas emissions by 26-28%". This economy wide targeting is what we expect efficient markets, aware of material impacts on firm operations, to price in to securities after the COP 21 agreement. In particular, the horizon of these emission reductions (2020 & 2025) is short term, meaning that firms will begin to face challenges soon if they are poorly mitigated relative to the direction of regulatory efforts included in the U.S. INDC for COP 21.

Kigali Deal (MOP 28) – Montreal Protocol Amendment[5]

On October 15th, 2016, at the 28th meeting of the parties to the Montreal Protocol (MOP 28) in Kigali, Rwanda, 170 countries successfully agreed to phase out hydrofluorocarbons (HFCs) from refrigerant chemicals. Due to their potent global warming potential (GWP), which is approximately 1,000 times more powerful than conventional carbon dioxide, HFC's represent a significant challenge to the 2016 Paris Agreement goals of limiting greenhouse gas emissions to avoid a 2º C increase in global temperatures that would likely have dire effects on the planet. HFC's are anthropogenic compounds that served as a replacement for chlorofluorocarbons (CFC's) and hydrochlorofluorocarbons (HCFC's) that were previously phased out by the Montreal Protocol in 1989 (Hurwitz et al., 2015). CFC's and HCFC's both contain chlorine (Cl) and or bromine (Br) and when they are synthesized in the stratosphere, they deplete ozone (O_3) via catalytic cycle reactions (Hurwitz et al., 2015). As a result, the Montreal Protocol phased these chemicals out of refrigerant and food products, in order to halt stratospheric accumulations that were creating ozone losses capable of allowing harmful levels of UV-B radiation to reach the surface of the earth. UV-B radiation is of course, the leading cause skin cancer and is naturally screened by proper levels of ozone in the stratosphere. Since HFCs do not contain chlorine or bromine and therefore have very small ozone depleting potential, they were favored as replacements to CFC's and HCFC's.

After HFC's became the industrial substitute to CFCs and HCFCs, their increasing accumulation in the stratosphere alarmed climate scientists and government policymakers. According to the 2015 study completed by NASA that has been cited in a number of academic inquiries, "as atmospheric concentrations of HFC's increase, so will their global climate impact" (Hurwitz et al., 2015). The study goes on to say that, gram for gram, HFC's have much higher heat trapping potential when compared to CO_2, "Of concern to global climate is that HFCs are particularly strong radiative forcers: many have large global warming potentials (GWPs), the ratio of their climate impact per unit mass…1 kg of HFC 125 provides the radiative forcing equivalent to 3,450 kg of CO2" (Hurwitz et al., 2015). The results of the meeting in Kigali, Rwanda are two separate

5. MOP stands for Meeting of the Parties to the Montreal Protocol. MOP in Kigali, Rwanda is the 28th such meeting, and the meeting was on substances that deplete the ozone layer.

phase out schedules, one for developing countries (Article 5) and one for developed countries (Article 2) (Tashobya, 2016). Since HFC alternative products are more expensive, developing nations were in conflict with developed nations over the phase out schedule. The agreement will require Article 2 countries to agree to baseline 2013 levels of HFC use with cuts in HFCs beginning in 2019 (Tashobya, 2016). Article 5 nations will be given more lenient terms and instead will be held to 2024-2026 baseline levels with a freeze date of 2028 (Article 5 group 1), or they will be held to 2020-2022 baseline levels with a freeze date of 2024 (Article 5 group 2) (Tashobya, 2016). As a result, the developments in HFC regulation agreed to in Rwanda will apply real pressure on air cooling manufacturers to develop, market and produce HFC alternative cooling technology that is profitable (Vidal, 2016).

III. Literature on the Use of ESG Rating Data

Disagreements among ESG Rating Providers

Academic inquiry into the broader merits of sustainable, responsible, and impact (SRI) investing for investors, stakeholders and corporate parties is fairly rich, but for this study, we are particularly interested in the data collection and ESG rating generation that assessed GHG emissions. In order to understand how accurately ESG raters assessed the impact that regulation would pose on manufacturers of HFC emitting products following the Kigali Deal for example, research on the scope of ESG data and rate generation can provide insight on if and how rating agencies processed this development. Chatterji et al. (2009) found that firm ratings from different agencies, overtime, tend to deviate from one another or disagree with each other on the scope and weight of certain ESG rating elements. In parsing this disagreement, Chatterji et al. (2016) found that low convergence between ESG raters is not only driven by different theorizations in terms of what ESG constructs to measure, but also by low commensurability in terms of how those constructs are measured. That same study analyzed correlations between pairs of rating agencies' ESG data, finding very low correlation between most agency pairs and some negative correlations that indicate "extreme disagreement" on ratings. Most of this disagreement and variation seems to originate from raters using different methods and variables to measure the same construct. In Chatterji et al. (2016), they highlight that some ESG ratings measure environmental performance by assessing a firm's ESG management process on a key issue, while others assess the same performance by concentrating on a firm's environmental outcomes. Since our research endeavor seeks to analyze ESG ratings around HFC regulation for air-cooling manufacturers for a period before (MOP 28), these findings provide useful background. ESG data between firms could potentially disagree on firm management or preparation for HFC developments, especially if Sustainalytics' or MSCI have low commensurability in terms of whether they assess HFC product risks with process or outcome based scopes. Based on this research, we conjecture that the rating agency

with better ability to measure management processes around the development of HFC alternatives preceding the amendment would likely offer a more accurate depiction of the risks posed to a particular firm by the impending agreement.

Accuracy of ESG Ratings

In addressing the accuracy of ESG ratings, the scope and theorization of process and outcome dimensions shed light on rating agency discrepancies. In a principle component analysis of ESG ratings from KLD, Trucost and SAM, Delmas et al. (2013) found that environmental performance cannot be reduced to one specific dimension that explains the overall ratings.[6] Instead, they found that two dimensions (processes/practices, outcomes) explain 80% of the variance between ratings from different raters. They also found that contrary to what might be expected, across the three rating agencies, "two distinct dimensions implies that process and outcome, at least as it pertains to environmental impacts, are much less linked than what we would perhaps expect". It seems as though ESG rating data is still somewhat lacking in terms of linking processes to specific outcomes. For instance, in Delmas et al. (2013), a hypothetical scenario is drawn up based on their findings, essentially emphasizing how companies may "excel at reporting, governance and the utilization of environmental performance systems, but still emit substantial amounts of pollution" (Delmas et al., 2013). These points are important for our endeavor, because it could be possible that some air-cooling manufacturers might have bolstered their process measures and management strategies in anticipation of HFC regulation, but not have actually realized positive corresponding outcomes. In turn, this could potentially be picked up in a company's stock performance, which will be part of the two-pronged research approach we will pursue.

As Delmas et al. emphasize, in the event that a company's security prices have appreciated in relation to favorable ratings of process and strategy development around a key ESG issue, major market corrections can be expected if the outcomes or results of those processes turn out to be insignificant. For our research, we will examine stock price reactions to see if in fact one rater's assessment is overly favorable for process measures beforehand, but not accurate in terms of predicting or indicating a corresponding firm outcome around that key ESG issue. Checking market corrections in security prices for air-cooling manufacturers after the Kigali deal would be one such analysis.

While the research discussed thus far reveals some expectations and cautions for the scope and breakdown of ESG data, other research can offer us a more in depth understanding of which rating assessments may be more accurate than others. In Chatterji et al. (2009), statistical analysis revealed a low mean inter-correlation of .01 among the KLD strength sub-scores and 0.17 among the KLD concern sub-scores. This find-

6. Kinder, Lynderberg, and Domini (KLD) is one of the ESG data pioneers. It was originally developed in 1991, but got acquired by RiskMetrics in 2009, and RiskMetrics was acquired by MSCI in 2010. Trucost (www.trucost.com) also got acquired by S&P Dow Jones Indices in October 2016. SAM (www.robecosam.com) was founded in 1995, got acquired by Robeco in 2006, and became RobecoSAM in 2013.

ing indicates that examination of sub-scores individually might be better for assessing a particular ESG performance than using the overall/aggregate score. These results are useful for our study, because based on this assessment of rating accuracy and predictive power, we may want to pull specific data on certain (accurate) metrics if possible. Interestingly enough though, the Total KLD Strengths and Total KLD Concerns indices were correlated in the same study at 0.25 (Chatterji et al., 2009). This indicates that to an extent, firms with more concerns may also be exhibiting more strengths, or vice versa. In terms of accuracy and tendency, Chatterji et al. (2009) found that incremental KLD strengths "have no large or statistically significant relationship with our environmental performance metrics, but environmental concerns do". This could mean that proactive air-cooling manufacturers may not deviate from each other on their absolute ratings, since strength assessments are less impactful overall and similar concern assessments in the background could drive fairly even ratings.

Stock Price Reactions to ESG Related Events

In this study, we leverage stock price data from air-cooling manufacturers to assess how well Sustainalytics and MSCI ESG Research were able to gauge their preparedness for the impending regulation. In a study of additions and deletions to KLD's Domini Social 400 (DS400) index, Ramchander et al. (2012) found that when firms are added to the index as a result of stakeholder-related CSR efforts they realize significantly positive cumulative abnormal returns. In fact, on average, these firms realized daily cumulative abnormal returns of positive .85 percent over an event window of plus/minus three days (-3,+3). In addition, Ramchander et al. (2012) find that rival firms experience negative cumulative abnormal returns after their rivals are added to the DS400 index. In their study of the reverse scenario, when firms were negatively assessed by KLD and deleted from the KLD DS400, firms experienced significantly negative abnormal returns. In their plus/minus three-day window of analysis, negative abnormal returns were -1.3 percent on average for the firms studied. Based on their analysis of these dynamics for different industries, Ramchander et al. (2012) also found that industries which are more "information opaque" tend to be affected at a different magnitude by ESG data announcements.[7] For instance, firms that are more "information opaque" tend to realize much higher abnormal returns from DS400 additions. On average, this "opacity premium" resulted in a 67 basis point addition to abnormal returns and in deletion scenarios, the negative premium averaged to about 317 basis points.

In a similar study, Kappou and Oikonomou (2016) reproduced some of the results from Ramchander et al. (2012), but instead found the only material effects on stock prices to come from index deletions, rather than additions. In their summary, Kappou

7. Ramchander et al. (2012) define firms that sell intangible products and/or carry intangible assets as informationally opaque. Examples include engineering services, tourism, and legal services, because the attributes of service-related products are oftentimes difficult to grasp ahead of consumption. For similar reasons, industries that are research and development (R&D) intensive, such as high technology manufacturing, are likely to suffer from information opacity because the value of these intangible investments is intrinsically hard to determine, and therefore difficult to convey to outside investors.

and Oikonomou (2016) state that there is "no significant short-term price pressure related to stocks added to the social index". They also show that firms that were deleted from social indices experienced abnormal returns of -1.61% in a two-week post-event period. This study showed that the negative abnormal returns occurring after deletion event can be longer-term than one might think. In fact, they found that negative abnormal returns can accumulate for several months after an index deletion, reaching -14% some six months after the deletion. The results of this study are important, because they caution against the expectations established in Ramchander et al. (2012) that stock price effects from positive ESG events may be significant. However, Kappou and Oikonomou (2016) do reinforce similar findings from other work that negative stock price effects from index deletions have significant and noticeable effects on firm stock prices. Ultimately, these studies show that in our analysis of stock price reaction to the (MOP 28), we may not realize significantly positive stock price reactions for firms that are strategically prepared for the regulation. In fact, we may only observe negative performance for those firms that were not sufficiently prepared for the results of (MOP 28), none the less, an important nuance to bear in mind for our study findings.

IV. Data and Methodology

ESG Ratings from MSCI, Thomson Reuters, Sustainalytics, CDP

In order to perform firm-level regression analysis that tests the predictive power and accuracy of ESG ratings in relation to stock price reaction to global regulatory developments, we used ratings from four leading ESG data providers: Morgan Stanley Capital International (MSCI), Thomson Reuters (THOM), Sustainalytics (SUST) and Carbon Disclosure Project (CDP).[8] From each of these sources, we collected desired scores for our universe of firms on an annual basis for 2015, since these scores are not published or updated on a more regular (monthly, quarterly or semi-annual) basis. The reason for choosing 2015 is that we wanted to see how these databases rated the GHG emitters prior to Paris Agreement and Kigali Deal. Collectively, these ESG scores represent the independent variable inputs that were utilized in our regression tests.

MSCI ESG Research provides tools for asset managers by releasing annually a set of positive and negative binary indicators on a number of Environmental, Social and Governance (ESG) issues. This dataset is collectively known as the MSCI ESG KLD STATS database and includes approximately 60 indicators that group separately in to the three ESG umbrellas: environmental, social and governance. For fiscal year 2015, MSCI ESG KLD STATS included a total of 23 "strength" and "concern" metrics on *environmental* issues alone (Appendix A). While our initial approach was to

8. MSCI's ESG database is called MSCI ESG KLD STATS (which is accessed through http://wrds.wharton.upenn.edu). Thomson Reuter's ESG database is called Asset4 (https://www.thomsonreuters.com/content/dam/openweb/documents/pdf/tr-com-financial/fact-sheet/esg-data-fact-sheet.pdf). As for Sustainalytics (http://www.sustainalytics.com/) and CDP (https://www.cdp.net/en), we accessed them through Bloomberg terminals at Bentley University.

specifically look at "Env_Con_C", an indicator dedicated to concerns around firm involvement in "ozone depleting substances", this indicator was discontinued by MSCI in 2014. Therefore, we collected all of MSCI's available environmental strength and concern scores and normalized them to derive an overall adjusted environmental score. By subtracting the aggregate number of environmental concerns (ALL_CON) from the aggregate number of environmental strengths (ALL_STR), we first created a Net Score for each firm in our universe (NET_STR). Then, to normalize and adjust the net score, we subtracted the minimum observed net score from each firm's individual net score and divided that figure by the difference between our observed maximum MSCI net score and minimum MSCI net score.[9] Our adjusted environmental score (MSCI ENV ADJ) is thus a relative measurement of a firm's net environmental strength for fiscal year 2015.

Thomson Reuters distributes their ESG scores through their Asset4 database for nearly 5,000 publicly traded firms on approximately 400 different ESG issues that group in to the following umbrella categories: economic performance, environmental performance, social performance and corporate governance performance. Environmental category alone includes 245 datatypes, often including multiple data types on each issue (see Appendix B for emissions related variables from Asset4 database). Since Asset4 coverage is extremely thorough and granular compared to MSCI ESG KLD STATS database, we collected Thomson Reuters' aggregate environmental score, "ENVSCORE", as well as their score on emission reduction of F-gases called "ENERO06S" for each firm in our universe via a Bloomberg subscription. The ENVSCORE is Thomson's holistic score that assesses a firm's measurable impact on natural systems and their ability to use best practices in order to avoid environmental risks and capitalize on environmental opportunities. The ENERO06S score is more finite in scope, assessing how the company reports on initiatives to "recycle, reduce, reuse or phase out fluorinated gases such as HFCs (hydrofluorocarbons), PFCs (perfluorocarbons) or SF6 (sulphur hexafluoride)". By including ENVSCORE data, we are able to compare Thomson's overall environmental score with corresponding rankings from other agencies.

Sustainalytics, much like MSCI and Thomson Reuters, scores companies based on approximately 70 core and industry-specific indicators. Different from other rating agencies though, Sustainalytics varies the weight of indicator importance uniquely between industries in order to assess policies, management systems and performance outcomes surrounding the most material issues for 6,500 companies worldwide. Sustainalytics scores asses company ESG performance by evaluating ESG issues in four dimensions: Preparedness, disclosure, qualitative performance and quantitative performance. Collectively, all environmental, social and governance indicators are compiled separately to determine an overall environmental, social and governance percentile

9. Simply counting the number of strengths and number of concerns for each firm, and calculating the Net score resulted in a score that ranges from -2 to 6. By doing the transformation we described, our normalized score ranges from 0 to 1. It was our effort to make the score comparable to those of Thomson Reuters (which ranges from 0 to 100) and Sustainalytics (which also ranges from 0 to 100).

score for each firm. These relative percentile scores are standardized in order to assess one firm's sustainability performance relative to the field. For our research, we used a Bloomberg database subscription to collect the Sustainalytics Environmental Percentile Score for each firm in our universe for fiscal year 2015. While Sustainalytics has more granular level of details within each category, our subscription only allows us to see overall environmental, social and governance percentile scores for each firm they cover. However, we can still compare the environmental percentile score from Sustainalytics with overall environmental scores from MSCI and Thomson Reuters.

The final set of environmental scores that we collected for our analysis comes from the Carbon Disclosure Project (CDP), an NGO that collects company-reported raw environmental metrics for now up to 5,600 firms. CDP's approach to data collection is different from traditional ESG rating agencies, focusing solely on environmental performance in three areas: climate, water and forests. Within their climate focus, CDP compiles self-reported firm data on a number of different emissions, including greenhouse gases, ozone depleting chemicals and regulated pollutants. All of CDP's metrics are derived based on company responses to the annual CDP "questionnaire". For our research, we collected CDP raw emissions data on GHG and HFC emissions (See Appendix C). For GHGs specifically, we utilized CDP's primary raw emissions dataset "CDP_SCOPE_1_EMISSIONS_GLOBALLY". This "scope 1" category measures any direct GHG emissions from sources that are owned or operated by a firm and reports them in millions of metric tons of carbon dioxide equivalent (mtCO2e). For HFC emissions, we utilized CDP's "CDP_HFC_EMISSIONS_CO2_EQUIV" dataset, which reports the amount of hydrofluorocarbons emitted by a company measured on an equivalent basis (mtCO2e). These scores were only available for 205 of the 573 firms in our universe, but in order to use them in our regression analysis, we performed separate tests on a refined universe of these 205 firms with intersecting data from each agency.

To create a universe of firms that would allow for comparative analysis of ESG scores, we first captured all available firm-level ESG data from MSCI, Sustainalytics and Thomson Reuters. Based on our research of COP 21 and MOP 28, we performed our query to include only North American firms, as the horizon for phase out schedules pursuant to GHGs and HFCs was most immediate for developed economies. With each agency offering a varying degree of firm universe coverage, we performed a merge operation in Excel to identify company codes that were common amongst all agency ESG datasets. This operation yielded 578 North American firms with overlapping coverage from each rater (MSCI, THOM, and SUST) and gave us our primary universe to work with for our analysis of these events (Universe A). After collecting our desired CDP raw emissions data for all available firms, a secondary universe was created by merging this new dataset with Universe A, yielding 206 firms with overlapping coverage from all four raters (Universe B). While our initial thinking was to use a number of NAICS code industries to isolate firms that were theoretically much more involved in HFC related businesses, we found this to be quite difficult. The

HFC producer segment is small, with only a few firms that are more strictly engaged in fluorogas operations.[10] In addition, many of the predominant HFC manufacturers are either wholly owned subsidiaries of large multinationals, or smaller divisions hidden underneath big corporate umbrellas. As a result, using the NAICS code "325120 Industrial Gas Manufacturing" for example, yields a list of firms that includes a few HFC producers, though these producers are all invested in HFC production to varying degrees and many firms in the industrial gas segment may not produce HFC at all. It was also difficult to consistently obtain environmental scores for these firms that we identified to be HFC producers. Therefore, we abandoned out attempt to isolate HFC firms with NAICS codes and instead looked at all firms that are covered by the major ESG data providers. Based on the universe of firms that are covered by all major ESG data providers, we can then classify them into high and low emission firms based on their carbon emissions, HFC emissions, etc.

In order to perform more comprehensive analysis of our ESG ratings, we used CDP and Thomson Reuters data to create emissions intensity scores for our secondary universe of firms. In particular, taking CDP's "CDP_SCOPE_1_EMISSIONS_GLOBALLY" metric for each firm and dividing by that firm's fiscal year 2015 total assets gave us a Scope 1 emissions per dollar of assets measure for all 206 firms in the universe. This standardized metric allowed us to score firms with Scope 1 emissions intensity values above the median with a 1 and below the median with a 0 (i.e., high vs. low emitters). Since just 29 of these 206 firms in our secondary universe had listed scores for the "CDP_HFC_EMISSIONS_CO2_EQUIV" metric, we created a binary score that gave a firm 1 if they had a reported value for this metric or 0 if they did not (i.e., HFC emitter vs. non-emitter).[11] We also used Thomson Reuters ENERO06S score to assign firms with a 1 if their ENERO06S score was above the average among the aggregate secondary universe or a 0 if their score was below the average (high F-gas reduction vs. low F-gas reduction). In sum, by creating these binary metrics from different intensity scores, we were able to create both high and low emitters, or high and low emission reducers segments for each of these universes that could be used in different regression permutations.

Methodology

With the end goal of assessing the predictive power and accuracy of ESG ratings relative to the valuation impacts for publicly traded firms after COP 21 and MOP 28, we use an event study analysis to generate variable inputs for our OLS regression analysis. Our dependent variable for this approach is the 3-day cumulative abnormal return (CAR) realized on a post-event basis for both COP 21 and MOP 28 (CAR3 and

10. For example, some of the relevant industries for our study are Household Refrigerator and Home Freezer Manufacturing industry (NAICS 335222) and Air-Conditioning and Warm Air Heating Equipment and Commercial and Industrial Refrigeration Equipment Manufacturing industry (NAICS 333415).
11. If a particular firm did have HFC emissions, but not reported in CDP, this may create some measurement error (misclassifying some emission-intense firms as non-intense) and bias the results towards 0.

CAR3.1 respectively). The independent variables utilized are the measures of firm size, industry dummies and ESG scores.

A. Event Study

To assess the valuation impact of both events (COP 21 and MOP 28) on each firm, we follow the standard "event study" methodology of Brown and Warner (1985). Pursuant to the data input requirements of an event study, we utilized a Capital IQ subscription to collect daily closing stock prices for 120 days before and 3 days after both COP 21 and MOP 28 agreement dates for each firm in our universe (i.e., t = -121,….,+3). We also collected daily closing prices for the S&P500 index for the previously mentioned timeframes and fiscal year 2015 total asset values and SIC codes in order to control for each firm's size and industry in our regression analysis. Capital IQ, like Bloomberg, is a tool produced by S&P Global, giving equity analysts easy access to firm level financial and non-financial data. Collectively, these datasets enabled us to execute the event study for both events (COP 21 and MOP 28), resulting in cumulative abnormal return (CAR) measurements that were used as the dependent variables in our regression tests.

Based on our findings of the announcement timing for both COP 21 and MOP 28 events, we determined that the time zero (t = 0) for our event studies were 12/14/2015 for COP 21 and 10/17/2016 for MOP 28.[12] To calculate cumulative abnormal returns for every firm after these t = 0 dates, we first compute historical daily holding period returns (HPRs) for all the securities in our primary universe (Universe A). By adding capital gains yield and dividend yield for each day, we obtain daily holding period return:

$$HPR_{it} = \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}} + \frac{P_{it}}{D_{it}}.$$

Repeating this calculation for 120 days before our t = 0 event dates and for 3 days after our t = 0 event dates gave us a total of 123 HPR observations for each of the securities. Next, we estimate one factor (market) model over the 120-day estimation window and record intercept (alpha) and slope (beta) values for each security:

$$R_{it} = \alpha_i + \beta_i \times R_{mt} + \varepsilon_{it},$$

where R_{it} is the return on stock i at time t, R_{mt} is the return on S&P500 index at time t, and t = -121,…,-1. By using the predicted alpha and beta coefficients from the market

12. While the COP 21 took place from Nov 30 to Dec 11, 2015 (Friday) in Paris, the announcement (press release) was on Saturday, Dec 12, 2015. So Dec 14, 2015 is the Monday following the announcement. Similarly, Kigali Deal was announced on Sunday, Dec 15, 2016, so we picked the following Monday, Dec 17 as t=0 for our event study.

model, we estimate the expected (predicted) return in the event window using equation:

$$\hat{R}_{it} = \hat{a} + \hat{\beta} \times R_{mt},$$

where $t = +1, +2, +3$. Comparing the predicted return ($\hat{R}_{it}$) against the actual return (R_{it}) during the event window will yield some deviation between the two (+/-) that is "abnormal". After calculating the abnormal return (AR) for each day in our three-day estimation windows for each security ($AR_{it} = R_{it} - \hat{R}_{it}$), we summed these abnormal returns on a cumulative basis throughout 3-day post-event estimation windows (+1,+3). These three-day "cumulative abnormal returns" or CARs then became the base independent variables for our OLS regression testing.

Since our primary universe of firms include all firms with available ESG data in MSCI, THOM, SUST, and CDP, without regard to their industry membership, we created two industry dummy variables to track the stock price reaction of firms in the oil & gas industry and chemical industry. These two industries, "Oil & Gas" and "Chemicals", by the nature of their business operations, should have been the easiest or most obvious of firms for an efficient market to price in valuation changes in our post-event window. We create these dummy variables by scoring securities with a 1 if they are associated with selected two digit SIC codes corresponding to oil & gas or chemical industries, or a 0 if they are not.[13] Given the material impact of the Paris Agreement and the Kigali Deal on the high GHG emitters, we expect average CAR to be negative for high GHG emitters (i.e., oil & gas firms, and chemical companies) and positive for low emitters. In other words, for heavily polluting firms, these international agreements should be negative news, and the opposite for firms that are taking appropriate steps to innovate and stay ahead of the curve.

B. Regression Analysis

Merging ESG scores from MSCI, THOM, and SUST for 578 firms with cumulative abnormal returns, firm size, and industry dummies results in 573 observations (5 firms were dropped for not having enough observations to get reliable estimates for the market model) from Universe A. Similarly, Universe B drops in size from 206 to 205 observations. Finally, we estimate the following regression equation to see whether ESG scores from various agencies prior to COP 21 or MOP 28 have any explanatory power for the 3-day cumulative abnormal returns:

$$CAR3_i = \alpha + \beta_1 \times ESG\_Score_i + \beta_2 \times Log(Total\ Assets_i) + \beta_3 \times Industry\ Dummies_i + \varepsilon_i.$$

13. 2-digit SIC code 13 is for oil and gas extraction, 29 is for petroleum refining and related industries, so firms in these two 2-digit SIC code industries are classified as oil & gas companies. As for chemicals, 2-digit SIC code 28 is for chemicals and allied products.

If ESG scores were accurately reflecting readiness of the GHG emitters for the Paris Agreement and Kigali Deal, we expect coefficient β_1 to be positive. In other words, high GHG emitters should have been scored low by ESG rating agencies prior to these agreements, and should have experienced negative cumulative abnormal returns after these agreements were announced. In contrast, low GHG emitters should have been scored high by ESG rating agencies, and should have experienced positive cumulative abnormal returns after these agreements were announced. As for coefficient β_3, we expect it to be negative, since firms in these industries should be high GHG emitters relative to other firms in the sample.

For our primary universe (Universe A) of 573 firms, we ran 6 different permutations of this regression equation: 3 for Paris Agreement (using CAR3) and three different ESG scores (MSCI, THOM, and SUST), and 3 for Kigali Deal (using CAR3.1) and three different ESG scores. In each of these 6 permutations, we included the log of fiscal year 2015 total assets for each firm to account for firm size as well as the two industry dummy variables (Oil & Gas and Chemicals).

For our secondary universe (Universe B) of 205 firms, we utilized our emissions intensity variables to perform subsample analysis wherein the "high emitters" were tested separately from the "low emitters". Using three separate emissions intensity variables (GHG emissions, HFC emissions and F-gases), we created 6 different subsamples (high GHG emitters, low GHG emitters, high HFC emitters, low HFC emitters, high F-gas emitters and low F-gas emitters). Therefore, running 9 subsample regressions for each of the events (Paris Agreement and Kigali Deal) means we collected a total of 18 regression outputs for our secondary universe.

V. Results

Descriptive Statistics

In Table 1, we present descriptive statistics for variables we used in our analysis. On the dependent variable side, it is important to highlight summary statistics for CAR3 and CAR3.1. The mean and median values of CAR3 are negative (-1.44% and -1.40%, respectively), suggesting that following the Paris Agreement, most of our 573 equities experienced a negative stock price reaction. This is somewhat along the lines of what we would expect, as many of our 573 firms should in theory be affected to some varying degree from regulatory decisions to seriously reduce greenhouse gas emissions. CAR3.1 instead had a positive mean and median values (0.73% and 0.50%, respectively). Given the material impact of the Kigali Deal on the high HFC emitters, we expected average CAR3.1 to be negative for high HFC emitters (i.e., oil & gas firms, and chemical companies) and positive for low emitters. Given our expectations, positive CAR3.1 means that low HFC emitters may be dominating our sample. Not all of these 573 firms are high HFC emitters due to our sample construction. This result further highlights the importance of subsample analysis, dividing the sample into high and low emitters.

Table 1
Descriptive Statistics

Variable	N	Mean	Median	Min	Max	StDev
CAR3	573	-1.44%	-1.40%	-21.22%	26.18%	3.61%
CAR3.1	573	0.73%	0.50%	-10.72%	19.21%	2.62%
MSCI_ENV_ADJ	573	0.40	0.38	0.00	1.00	0.18
THOM_ENV_SCORE	573	60.91	72.04	10.22	95.15	30.16
SUST_ENV_PCTL	573	44.03	42.86	0.00	100.00	28.77
Total Assets ($mil)	573	56,413	15,518	635	2,417,121	190,352
Oil & Gas Dummy	573	0.06	0.00	0.00	1.00	0.24
Chemicals Dummy	573	0.08	0.00	0.00	1.00	0.28
F-Gas Emission Reduction in THOM	205	0.75	1.00	0.00	1.00	0.44
HFC Emissions in CDP	205	0.14	0.00	0.00	1.00	0.35
GHG Emissions in CDP	205	0.50	1.00	0.00	1.00	0.50

On the independent variable side, our results for each agency's environmental score show that MSCI_ENV_ADJ has a mean of 0.40 and median of 0.48, and ranges from 0 to 1 by construction. THOM_ENV_SCORE has a mean of 60.91, median of 72.04, and ranges from 10.22 to 95.15. SUST_ENV_PCTL has a mean of 44.03, median of 42.86, and ranges from 0 to 100. In unreported correlation analysis, we see that the correlation between THOM_ENV_SCORE and MSCI_ENV_ADJ is 0.5671, the correlation between THOM_ENV_SCORE and SUST_ENV_PCTL is 0.5672, and the correlation between MSCI_ENV_ADJ and SUS_ENV_PCTL is 0.5653.

Out of 573 firms in our primary universe, only 36 firms qualified as "Oil & Gas" firms (2-digit SIC codes 13 and 29) while 48 firms qualified as "Chemicals" firms (2-digit SIC code 28). The firm size measured by total asset values from fiscal year 2015 has a mean value of $56.4 billion, median value of $15.1 billion, and ranges between $635 million and $2.42 trillion (the largest asset value belongs to JP Morgan). The summary statistics for our emissions dummy variables in terms of F-Gases, HFCs and GHGs (which are equal to 1 if they are available in CDP and 0 otherwise) show how coverage for these scores was varied. While more than half of the 205 firms in our secondary universe were scored on F-gas emissions and GHG emissions, the opposite was true for HFC's. This observation reflects the fact that, on the aggregate level, more firms are involved in F-gas and GHG related activities than HFC related activities. The fact that only 14% of the 205 firms have HFC emissions data available helps explain why we saw positive CAR3.1 for our primary universe after the Kigali Deal.

Discussion of Regression Results

Table 2 presents results from six different regressions, three for the Paris Agreement, three for the Kigali Deal. The results from our primary universe regression outputs for the Paris Agreement (COP 21) did not indicate any statistical significance for each of our three environmental scores: MSCI, SUST, THOM (Table 2, first column). Our hypothesis was fairly straightforward, predicting that high environmental scoring firms, on the aggregate level, would have either more positive or less negative CAR reactions following COP 21 (i.e., positive coefficient for the environmental scores). Essentially, we predicted that if a firm had a high environmental score, they were relatively more prepared to mitigate risks to their operations posed by regulation limiting GHG emissions. However, without any statistical significance linking high scores to relatively better valuation impacts following COP 21, we cannot confirm that hypothesis. The only result from our primary universe regression outputs for COP 21 that show any statistical significance are those for the Oil & Gas dummy. The coefficient for this dummy variable was negative and highly significant (p-value = 0.000). Since the average CAR3 for our primary universe was -1.44%, this indicates that firms in the oil and gas industry realized more negative valuation impacts relative to other firms in the sample.

The same regression outputs from all three environmental scores were generated for our primary universe of 573 firms around the Kigali Deal (MOP 28), resulting in reduced R^2 overall at approximately 0.06 (Table 2, second column). Of our three environmental scores (MSCI, THOM, SUST), MSCI and SUST both yielded significance at the 5% level. However, since the average CAR3.1 valuation impact for our primary universe of firms was 0.73% for MOP 28, the negative coefficients for both of these regression coefficients indicated that high scoring firms realized less positive valuation impacts. Again, like our hypothesis for COP 21, this is counter to what we expected based on our assumption of environmental score accuracy and predictive power. However, not realizing our hypothesized results was not entirely unexpected as the COP 21 only impacted high GHG emitting firms, and the MOP 28 only impacted those firms engaged in activities related to HFC emissions and production. In contrast, our sample construction was done my merging all firms covered by THOM, SUST, and MSCI, which means there could be firms that have low GHG profile in our primary universe.

Similarly, the coefficient on the chemical SIC code dummy was positive and significant for each of the three regressions for MOP 28, implying that chemical firms realized positive valuation impacts following MOP 28, counter to our expectations. Again, it could be due to the fact that not all chemical companies produce products involving HFC. We found that identifying firms involved in the production or procurement of HFC products was inherently difficult. In addition, HFC producers are often just marginal business units underneath multi-national conglomerate umbrellas, and not pure-play companies, making the valuation impact hard to see through an event study method. Ultimately, since our universe includes only U.S. firms, it could be pos-

sible that domestic chemical firms were in fact well prepared to mitigate ESG risks related to regulation from MOP 28. This would explain our positive coefficient results for the chemical dummies.

TABLE 2
CROSS-SECTIONAL CUMULATIVE ABNORMAL RETURN (CAR) REGRESSIONS ON ESG RATINGS FOR THE PRIMARY UNIVERSE

	Event: COP 21 Dependent Variable: CAR3 (N=573)		**Event: MOP 28** Dependent Variable: CAR3.1 (N=573)	
Panel A. MSCI	coefficient	*p*-value	coefficient	*p*-value
MSCI_ENV_ADJ	-0.005	0.556	-0.021	0.001***
Log Assets	-0.003	0.249	0.004	0.030**
Oil & Gas Dummy	-0.051	0.000***	0.015	0.001***
Chemical Dummy	0.005	0.351	0.013	0.001***
R-squared	0.12		0.06	
Panel B. THOM	coefficient	*p*-value	coefficient	*p*-value
THOM_ENV_SCORE	0.000	0.572	0.000	0.149
Log Assets	-0.004	0.162	0.004	0.059*
Oil & Gas Dummy	-0.050	0.000***	0.018	0.000***
Chemical Dummy	0.004	0.401	0.012	0.002***
R-squared	0.12		0.05	
Panel C. SUST	coefficient	*p*-value	coefficient	*p*-value
SUST_ENV_PCTL	0.000	0.465	-0.000	0.004***
Log Assets	-0.003	0.289	0.004	0.026**
Oil & Gas Dummy	-0.051	0.000***	0.017	0.000***
Chemical Dummy	0.004	0.388	0.011	0.003***
R-squared	0.12		0.06	

***, **, * Significance at the 1%, 5%, and 10% levels, respectively.

Based on counter-intuitive findings reported in Table 2, we further try to distinguish between high and low GHG emitters by using CDP data. If we can identify the high or low emitters, we may be able to see our hypothesized coefficients. So in Table 3, we summarize results from 18 different regressions for COP 21 by only listing the coefficient on the ESG score. First column reports results from HIGH subsamples (based on Scope 1 emissions intensity, F-gas emission reductions, and HFC emissions), and the second columns reports results from LOW subsamples. These regres-

sions were run with one ESG score at a time, with log of total assets as the only control variable. Use of CDP database means we are now using the secondary universe of 205 firms (Universe B).

TABLE 3
SUBSAMPLE ANALYSIS: RELATIONSHIP BETWEEN CUMULATIVE ABNORMAL RETURNS AND ESG RATINGS FOR THE COP 21

	Subsample HIGH Dependent Variable: CAR3 (N=205)			Subsample LOW Dependent Variable: CAR3 (N=205)		
Panel A. Scope 1 Emissions	Coeff.	*p*-value	R-Sq.	Coeff.	*p*-value	R-sq.
Reg 1 Indep Var: MSCI	-0.034	0.025**	0.37	-0.001	0.944	0.04
Reg 2 Indep Var: THOM	0.000	0.233	0.35	0.000	0.845	0.04
Reg 3 Indep Var: SUST	0.000	0.260	0.35	0.000	0.452	0.04
Panel B. F-Gas Reduction	Coeff.	*p*-value	R-Sq.	Coeff.	*p*-value	R-sq.
Reg 1 Indep Var: MSCI	-0.025	0.039**	0.30	-0.008	0.708	0.25
Reg 2 Indep Var: THOM	0.000	0.787	0.28	0.000	0.550	0.25
Reg 3 Indep Var: SUST	0.000	0.552	0.28	0.000	0.718	0.25
Panel C. HFC Emissions	Coeff.	*p*-value	R-Sq.	Coeff.	*p*-value	R-sq.
Reg 1 Indep Var: MSCI	-0.030	0.143	0.09	0.004	0.748	0.01
Reg 2 Indep Var: THOM	-0.001	0.002***	0.32	0.000	0.706	0.01
Reg 3 Indep Var: SUST	-0.000	0.014**	0.21	0.000	0.524	0.01

***, **, * Significance at the 1%, 5%, and 10% levels, respectively.

For our secondary universe of 205 firms analyzed around the COP 21 valuation event, the only statistical significance we observed came from regression outputs for the high emitter subsets. For high GHG emissions firms, as determined by a firm's level of CDP_SCOPE 1 emissions (scaled by total assets) being greater or equal to the median value for the entire universe, the coefficient on the MSCI's environmental score was significant at a 5% level (*p*-value=.0246). However, the coefficient was negative (-0.034), indicating that high MSCI environmental score firms realized more negative valuation impacts relative to the entire universe. This is again counter to what we expected.

The same regression test for the subset of firms from our secondary universe with good reduction of F-gases (high ENERO06S scores from THOM) resulted in significance at the 5% level from MSCI's environmental score (*p*-value = 0.0393). Like previous results though, the coefficient from this permutation was also negative and with the average CAR3 for all firms in this subset of the primary universe at -1.67%, this indicates that high MSCI environmental scoring firms realized worse valuation impacts post COP 21. MSCI ENV score did not show any significant relationship with CAR3 when we generated regression outputs for the high HFC emitter subset of our secondary universe, while THOM and SUST both were significant at a 5% level (p-values at 0.0016 and 0.0139, respectively). However, the regression outputs for high HFC emitting firms for THOM and SUST both yielded negative coefficients, implying once again that firms with higher Thomson Reuters and Sustainalytics environmental scores realized more negative valuation impacts post COP 21. These results are certainly different than our expectations. We hypothesized that firms with higher environmental scores from ESG rating agencies, would react either more positively or less negatively to COP 21.

It is also worth highlighting that for each of these secondary universe, COP 21 regression permutations, the R^2 value was above 0.2 and was as high as 0.37 in the case of the test using the high CDP_SCOPE 1 emissions subset of firms. This implies that by refining our secondary universe based on high (1) versus low (0) raw emissions values, we were able to increase the fitness of our data. This R^2 result potentially validates our theory that by analyzing environmental scores at a sub-universe level for prescreened firms in a select time window, we can potentially increase the predictive capability and value of environmental score data.

The results of our regression tests for the high (1) emitting subset of firms within our secondary universe for MOP 28 (Table 4) did yield statistical significance at the 10% and 5% level in a number of different cases. For high SCOPE_1 emissions firms, MSCI and SUST scores both yielded p-values of 0.0885, but negative coefficients from these regression outputs indicate the opposite valuation impact than what we were looking for. In the case of high scoring ENERO06s (F-gas mitigation) firms, MSCI and SUST scores were both significant at a 5% level with *p*-values of (0.02 and 0.0018) respectively. Without seeing positive coefficients from these score tests though, we were not able to observe that better F-gas reducing firms with higher aggregate environmental scores had more positive valuation impacts. Finally, in the case of high (1) HFC emitting firms for our secondary universe after MOP 28, only MSCI's score yielded significance at a 10% level (p-value = 0.0837). With the coefficient for that output again negative at (–0.05), the valuation impact was opposite of our predicted positive coefficient.

Table 4
Subsample analysis: Relationship between cumulative abnormal returns and ESG ratings for the MOP 28

	Subsample HIGH Dependent Variable: CAR3.1 (N=205)			Subsample LOW Dependent Variable: CAR3.1 (N=205)		
Panel A. Scope 1 Emissions	Coeff.	*p*-value	R-Sq.	Coeff.	*p*-value	R-sq.
Reg 1 Indep Var: MSCI	-0.021	0.088*	0.13	-0.015	0.445	0.03
Reg 2 Indep Var: THOM	0.000	0.383	0.11	0.000	0.469	0.03
Reg 3 Indep Var: SUST	-0.021	0.088*	0.14	0.000	0.182	0.04
Panel B. F-Gas Reduction	Coeff.	*p*-value	R-Sq.	Coeff.	*p*-value	R-sq.
Reg 1 Indep Var: MSCI	-0.025	0.024**	0.07	0.028	0.401	0.09
Reg 2 Indep Var: THOM	0.000	0.347	0.05	0.000	0.363	0.09
Reg 3 Indep Var: SUST	0.000	0.002***	0.10	0.000	0.565	0.09
Panel C. HFC Emissions	Coeff.	*p*-value	R-Sq.	Coeff.	*p*-value	R-sq.
Reg 1 Indep Var: MSCI	-0.051	0.084*	0.19	-0.015	0.161	0.03
Reg 2 Indep Var: THOM	0.000	0.422	0.11	0.000	0.955	0.02
Reg 3 Indep Var: SUST	0.000	0.114	0.17	-0.000	0.049**	0.04

***, **, * Significance at the 1%, 5%, and 10% levels, respectively.

For each of the regression permutations we generated based on the low (0) emitting subset of firms for our secondary universe after both COP 21 and MOP 28, we observed significance in just one of these tests. While the "low" emitter, COP 21 regression tests for this secondary universe did not yield any statistical significance, the "low" HFC emitter MOP 28 regression test did yield a 5% level of significance with SUST environmental scores. This result would have been interesting, except for the fact that the coefficient variable at –0.0002 suggests that high SUST environmental score firms with relatively lower raw HFC emissions realized less positive valuation impacts. This observation is somewhat difficult to explain, particularly when considering that regardless of any environmental score, a lower HFC emitting firm should have theoretically been much less impacted by HFC regulation.

VI. Conclusion

Based on the results from our OLS regressions, we are not able to confirm our hypotheses and as a result we cannot make definitive statements about the accuracy and the predictive power of ESG ratings. Following the announcements of the Paris Agreement and Kigali Deal, we expected that market participants would be able to price in valuation impacts of these regulatory phase downs of GHG emissions on publicly traded companies based on their ESG ratings. In addition, we assumed that aggregate environmental ratings from Sustainalytics, Thomson Reuters and MSCI would have some ability to capture a firm's relative level of preparedness and mitigation for the impacts discussed. However, without observing the hypothesized coefficients, we cannot conclude that these aggregate scores had some capability to accurately predict the how stock prices will react to news of these agreements.

There are a number of reasons that can explain why rating agency scores have low correlation in general, but our comparison of rater scores around one environmental issue highlight how significant these disparities are. It has been fairly well documented, based on our literature review, that aggregate ESG scores from different rating agencies exhibit low correlation primarily due to differences in methodology. While all of the raters we studied generate scores based on proprietary frameworks or methodologies that incorporate similar themes around preparedness, disclosure, performance and momentum; the relative weights of these metrics tend to deviate substantially from agency to agency. For instance, while Sustainalytics' ESG performance framework is heavily reliant on quantitative and qualitative performance measures, MSCI's framework leverages "controversy assessments" in order to assess firm level exposure to environmental issues. Some assign percentile scores that range from 0 to 100, while others rate companies only on an indicator variable (i.e., whether a company has a problem with certain issue or not). Component variables that go into each aggregate score differ as well. As a result, rating agency opinions deviate at the aggregate level. However, from a comparison standpoint, this research demonstrates that even on a narrower ESG issue such as GHG emissions, rating agencies still disagree. Ultimately this observation allows us to understand how level of coverage on each particular issue can differ between rating agencies.

While our raw emissions intensity indicators did help us increase the general fit of our ESG rating data and CAR valuation data for particular subsets of firms, we were not able to realize more accuracy from the ratings with this method. Overall, these subsets of firms from our universe marked "high" and "low" in separate emissions categories did yield many more observations of statistical significance between the aggregate ratings and valuation impacts than our initial primary universe tests. However, the issue of disagreement between our coefficient values and our hypothesized valuation impacts was not resolved in any of our permutations, including both the primary and secondary universe.

We can conclude from these findings that aggregate environmental ESG ratings are likely, in their current state, unable to accurately predict valuation impacts that are realized in relation to environmental regulatory efforts to reduce emissions. This finding could be due to the fact that our universe included only those firms in the U.S. equity markets. As mentioned previously, the relative time horizon for firm impacts from environmental regulation from the Paris Agreement and the Kigali Deal is closer for firms in developed economies. However, it is also likely that domestic firms may be better prepared to mitigate the risks imposed by the Paris Agreement and the Kigali Deal. For example, DuPont Company, once a pioneer in HFC gases, is also now a pioneer in R&D efforts to produce low GWP alternatives. In addition, HFC related business units for chemical firms are often hard to identify and analyze as they are often grouped within a multi-national conglomerate umbrella. This inevitably makes the hypothesized valuation impact that much harder to observe.

Our research also revealed that none of these rating agencies, with the exception of CDP, produce emissions data for HFC gases alone and it seems their coverage of firms is not big enough to produce meaningful insight. Having said that, Thomson Reuters ESG data seems most detailed, given the fact that they have the "ENERO06S" data item which assesses the following: "Does the company report on initiatives to recycle, reduce, reuse or phase out fluorinated gases such as HFCs (hydrofluorocarbons), PFCs (perfluorocarbons) or SF6 (sulphur hexafluoride)?". It is therefore inherently difficult, based on the data that exist, to get insights around which firms are relatively better or more prepared to mitigate the risks associated with impending climate agreements.

For further study, we would encourage a similar study on international firms. Since many domestic firms could be fairly prepared to deal with emissions reduction legislation, looking at international firms could offer more variability and therefore more insight into relationship between their ESG ratings and stock price reactions. Based on the literature we reviewed, there is a great deal of room for this type of analysis as it can be used not only with aggregate ESG scores, but with granular, issue-specific scores as well. We also see a room for a study where one looks at changes in ESG ratings following the climate agreements, and see if rating agencies updated or changed their ratings. Ultimately, we still believe there is a pressing need for more studies assessing rating accuracy. After all, to the ESG conscious investor, scores are relevant in the sense that they, like fundamental analysis, can potentially offer additional degree of insight on firm valuation and investment return.

References

Barry, Ellen, and Coral Davenport, 2016. "Emerging Climate Accord Could Push A/C Out of Sweltering India's Reach." nytimes.com. *New York Times*, 12 Oct. 2016. Web. 3 Dec. 2016.

Brown, Stephen J., and Warner, Jerold B., 1985. "Using Daily Stock Returns." *Journal of Financial Economics* 14: 3-31.

Busch, Timo, Friede, Gunnar and Alexander Bassen, 2015. "ESG and financial performance: aggregated evidence from more than 2000 empirical studies." *Journal of Sustainable Finance & Investment 5.4*: 210-33. Business Source Premier. Web. 10 Mar. 2017.

Chatterji, Aaron, David I. Levine, and Michael W. Toffel, 2009. "How Well Do Social Ratings Actually Measure Corporate Social Responsibility?" *Journal of Economics & Management Strategy 18(1)*: 125-169. Accessed October 20, 2016.

Chatterji, Aaron, Rudolphe Durand, David I. Levine, Samuel Touboul, 2016. "Do Ratings of Firms Converge? Implications For Managers, Investors and Strategy Researchers." *Strategic Management Journal* 37: 1597-1614. Accessed October 10, 2016. EBSCO.

Delmas, Magali A, Dror Etzion and Nicholas Nairn-Birch, 2013. "Triangulating Environmental Performance: What Do Corporate Social Responsibility Ratings Really Capture?" *Academy of Management Perspectives* 27 (3): 255-267. Accessed November 26, 2016. EBSCO Business Source Complete.

ESG Ratings Methodology, 2015. www.msci.com. May 2015. Accessed October 10, 20. https://www.msci.com/documents/10199/123a2b2b-1395-4aa2-a121-ea14de6d708a

Hurwitz, M.M, E.L Reming, P.A Newman, F.U. E Miawer, K Cady-Pereira, and R Bailey, 2015. "Ozone depletion by hydrofluorocarbons". *Geophysical Research Letters*. September 30, 2015. 42.8686-8692, doi:10.1002/2015GL065856. November 20, 2016.

Johnston, Chris, Oliver Milman, and John Vidal, 2016. "Climate change: global deal reached to limit use of hydrofluorocarbons." theguardian.com. *The Guardian*, 15 Oct. 2016. Web. 20 Oct. 2016.

Kappou, Konstantina, and Ioannis Oikonomou, 2016. "Is There a Gold Social Seal? The Financial Effects of Additions to and Deletions from Social Stock Indices." *Journal of Business Ethics* 133(3): 533-552. *Business Source Complete*, EBSCOhost (accessed December 5, 2016).

MSCI ESG Research, 2015. "MSCI ESG Research: Overview and Products." www.msci.com. December 2015. Accessed November 20, 2016. https://www.msci.com/documents/1296102/1636401/MSCI_ESG_Research_Factsheet.pdf

MSCI ESG, 2015. "MSCI ESG KLD STATS: 1991 - 2014." www.msci.com. June 2015. Accessed October 10, 20. https://www.msci.com/documents/10199/123a2b2b-1395-4aa2-a121-ea14de6d708a

Ramchander, S., Schwebach, R. G. and Staking, K., 2012. "The informational relevance of corporate social responsibility: evidence from DS400 index reconstitutions". *Strategic Management Journal*, 33: 303–314. doi:10.1002/smj.952

Research Methodology, 2016. ESG Research Methodology. July 2016. Accessed December 01, 2016. http://www.sustainalytics.com/research-methodology.

Sustainalytics contributors. *Sustainalytics' Research Methodology*. Company ESG Research, Executive Summary, July 2016. http://marketing.sustainalytics.com/acton/attachment/5105/f-078f/1/-/-/-/-/Sustainalytics%20Methodology.pdf

Eccles, Robert, Ioannis Ioannou, and George Serafeim, 2014. "The Impact of Corporate Sustainability on Organizational Processes and Performance." *Management Science* 60(11): 2835-2857. Http://www.hbs.edu/faculty/Publication. Web. 1 Mar. 2017

G. Serafeim, E. Kaiser, J. Linder, I. Naranjo, K. Nguyen-Taylor, and J. Streur, 2015. "The Role of the Corporation in Society: Implications for Investors," September 2015, https://www.calvert.com/calvert-serafeim-series-report.php

Slager, R., 2014. "SRI Indices and Responsible Corporate Behavior: A Study of the FTSE4Good Index." *Business & Society* 54(3): 386-405. Sage Publications. Web. 10 Feb. 2017.

Tashobya, Athan, 2016. "Montreal Protocol: Nearly 200 Countries Adopt Kigali Amendment to Phase out HFCs." www.newtimes.co, October 15, 2016. Accessed November 20, 2016. http://www.newtimes.co.rw/.

United States, 2016. Environmental Protection Agency. *United States Intended Nationally Determined Contribution*. Vol. 1. Washington DC: EPA, 2016. Print.

United Nations, 2015. "Framework Convention on Climate Change." Environment/Paris Agreement - Paris Agreement under the United Nations Framework Convention on Climate. Vol. 1. Paris: n.p., 2015. Unfccc.int/resources. United Nations, 12 Dec. 2015. Web. 5 Jan. 2017.

G. Unruh, D. Kiron, N. Kruschwitz, M. Reeves, H. Rubel, and A.M. zum Felde, 2016. "Investing For a Sustainable Future," *MIT Sloan Management Review*, May 2016.

Vidal, John, 2016. "Kigali Deal on HFCs Is Big Step In Fighting Climate Change." www.theguardian.com, October 15, 2016. Accessed October 25, 2016. https://www.theguardian.com/environment/2016/oct/15/kigali-deal-hfcs-climate-change.

Appendix A. MSCI ESG KLD STATS 2015 Data Set: ENV Indicators

ENV Indicators	MSCI ESG KLD STATS - 2015 Description
ENV-str-A	Environmental Opportunities - Environmental Opportunities in Clean Tech
ENV-str-B	Pollution & Waste - Toxic Emissions and Waste
ENV-str-C	Pollution & Waste - Packaging Materials & Waste
ENV-str-D	Climate Change - Carbon Emissions
ENV-str-G	Environmental Management Systems
ENV-str-H	Natural Capital - Water Stress
ENV-str-I	Natural Capital - Biodiversity & Land Use
ENV-str-J	Natural Capital - Raw Material Sourcing
ENV-str-K	Climate Change - Financing Environmental Impact
ENV-str-L	Environmental Opportunities - Opportunities in Green Building
ENV-str-M	Environmental Opportunities - Opportunities in Renewable Energy
ENV-str-N	Pollution & Waste - Electronic Waste
ENV-str-O	Climate Change - Energy Efficiency
ENV-str-P	Climate Change - Product Carbon Footprint
ENV-str-Q	Climate Change - Climate Change Vulnerability
ENV-str-X	Environment - Other Strengths
ENV-con-D	Toxic Emissions and Waste
ENV-con-F	Energy & Climate Change
ENV-con-H	Biodiversity & Land Use
ENV-con-I	Operational Waste (non-hazardous)
ENV-con-J	Supply Chain Management
ENV-con-K	Water Stress
ENV-con-X	Environment - Other Concerns

Appendix B. Thomson Reuters Asset4 Data Set: Emissions Related ENV Variables

ENV Variables	Variable Name
ENVSCORE	Environmental Score
ENER	Emission Reduction
ENERDP0051	Emission Reduction Processes/Policy Emissions Reduction
ENERDP0161	Emission Reduction Objectives/Targets Emissions Reduction
ENERDP023	CO2 Equivalents Emission Total
ENERDP024	CO2 Equivalents Emission Direct
ENERDP025	CO2 Equivalents Emission Indirect
ENERDP027	Cement CO2 Equivalents Emission
ENERDP033	NOx and SOx Emissions Reduction Initiatives
ENERDP034	NOx Emissions
ENERDP035	SOx Emissions
ENERDP036	VOC Emissions Reduction Initiatives
ENERDP040	VOC Emissions
ENERDP058	Water Pollutant Emissions
ENERDP068	Emissions Trading
ENERDP096	CO2e Indirect Emissions Scope 3
ENERDP123	Estimated CO2 Equivalents Emission Total
ENERO01S	Score - Emission Reduction/Biodiversity Impact
ENERO02S	Score - Emission Reduction/Biodiversity Controversies
ENERO03S	Score - Emission Reduction/Greenhouse Gas Emissions
ENERO04S	Score - Emission Reduction/Cement CO2 Emissions
ENERO05S	Score - Emission Reduction/CO2 Reduction
ENERO06S	Score - Emission Reduction/F-Gases Emissions
ENERO07S	Score - Emission Reduction/Ozone-Depleting Substances Reduction
ENERO08S	Score - Emission Reduction/NOx and SOx Emissions Reduction
ENERO09S	Score - Emission Reduction/VOC Emissions Reduction
ENERO10S	Score - Emission Reduction/Waste
ENERO11S	Score - Emission Reduction/Waste Recycling Ratio
ENERO12S	Score - Emission Reduction/Hazardous Waste
ENERO13S	Score - Emission Reduction/Discharge into Water System
ENERO14S	Score - Emission Reduction/Waste Reduction
ENERO15S	Score - Emission Reduction/Innovative Production

ENERO16S	Score - Emission Reduction/Environmental Partnerships
ENERO17S	Score - Emission Reduction/Environmental Management Systems
ENERO18S	Score - Emission Reduction/Environmental Restoration Initiatives
ENERO19S	Score - Emission Reduction/Transportation Impact Reduction
ENERO20S	Score - Emission Reduction/Spills and Pollution Controversies
ENERO21S	Score - Emission Reduction/Spill Impact Reduction
ENERO22S	Score - Emission Reduction/Climate Change Risks and Opportunities
ENERO23S	Score - Emission Reduction/Environmental Compliance
ENERO24S	Score - Emission Reduction/Environmental Expenditures
ENPIDP029	Fleet CO2 Emissions

Appendix C. CDP Data Set: Emissions Related ENV Variables

CDP Variable Label	CDP Variable Definition
CDP_SCOPE_1_EMISSIONS_GLOBALLY	Total global amount of scope 1 emissions emitted by the company, measured in millions of metric tons of carbon dioxide equivalent (mtCO2e). Scope 1 emissions are direct GHG (greenhouse gas) emissions from sources that are owned or operated by the company.
CDP_SCOPE_2_EMISSIONS_GLOBALLY	Total global amount of scope 2 emissions emitted by the company, measured in millions of metric tons of carbon dioxide equivalent (mtCO2e). Scope 2 emissions are indirect GHG (greenhouse gas) emissions that are caused by the company through the consumption of imported heat, electricity, cooling, or steam. Also known as "Purchased Electricity".
CDP_SF6_EMISSIONS_CO2_EQUIV	Amount of sulfur hexafluoride (SF6) emitted by the company, measured in millions of metric tonnes of carbon dioxide equivalent (mtCO2e).
CDP_N2O_EMISSIONS_CO2_EQUIV	Amount of nitrous oxide (N2O) emitted by the company, measured in millions of metric tonnes of carbon dioxide equivalent (mtCO2e).
CDP_CH4_EMISSIONS_CO2_EQUIV	Amount of methane (CH4) emitted by the company, measured in millions of metric tonnes of carbon dioxide equivalent (mtCO2e).
CDP_PFC_EMISSIONS_CO2_EQUIV	Amount of per fluorocarbons (PFCs) emitted by the company, measured in millions of metric tonnes of carbon dioxide equivalent (mtCO2e).
CDP_HFC_EMISSIONS_CO2_EQUIV	Amount of hydro fluorocarbons (HFCs) emitted by the company, measured in millions of metric tonnes of carbon dioxide equivalent (mtCO2e).

Downside Deviation and Volatility Target Strategies: A Comparative Study of Two Risk Management Approaches to Investment

By Harry Chengzhe Yao*

In this paper we will introduce Downside Deviation Target strategy (DDT) and Volatility Target strategy (VT), two commonly used risk management approaches. Both strategies use an ex post dynamic rebalancing of a portfolio to meet a certain pre-specified risk target. However, each approach, although similar in purpose, offers different characteristics and exposures through different risk metrics.

Using numerical simulations, we use historical data to compare the strategies. Using R as the programming language and historical S&P 500 data, we simulate the performance of both investment strategies using four distinct cases where we observed periods of increasing, decreasing, cyclical, or crashing markets. Using historical data, we provide a comparative analysis of the performance of both DDT and VT in each of these market environments.

We found that although DDT and VT share many similarities, especially in rebalancing mechanism, DDT offers a better way to control the risk-return tradeoff. In practice, this flexibility can offer investors significant upsides but with additional risk.

Keywords: Investment Strategies, Risk targets, Risk controls, Downside Deviation, Volatility, Volatility Target, Dynamic Portfolio Allocation, Portfolio Allocation.

I. Introduction

Investors' main objective function is to maximize their returns. But as Markowitz (1952) has said in his paper 'Portfolio Selection,' investors must cope with the issues of risk and return (Markowitz 1952). Markowitz (1952) presented those risks and returns as a form of trade-off. This was the basis of the mean-variance portfolio management models. In these models, variance is used as a measure of risk; and investors, according to Markowitz (1952), should try to maximize the return for the least amount of variance.

Variance, in finance, is more often contemporarily referred to as volatility, which is a measure of deviation. Further, when volatility is mentioned in finance, it is often

* Email: abcyao9@gmail.com. Many thanks to my advisor, professor Victoria Steblovskaya.

seen as symmetrical, meaning that it takes into account the whole set of objects and calculates the mean deviation of each object to the mean of that set. So, high volatility implies high probability of both gains and losses.

In the financial markets where the end states of portfolios, at some time in the future, are uncertain, investors need the ability to control their exposure to that source of uncertainty. In this paper, we will explore two ways of controlling that risk and the caveats that come with each method: Downside Deviation Target Strategy (DDT) and Volatility Target Strategy (VT).

We will present the formal structure for both DDT and VT in the Theoretical Implementation section of the paper. In brief, the DDT strategy utilizes the same rebalancing algorithm in the VT that was originally derived from the framework of Albeverio, Steblovskaya, and Wallbaum (2013) where they presented a VT dynamically rebalancing portfolio allocation method (Albeverio, Steblovskaya, and Wallbaum 2013). Both VT and DDT uses a risk metric as the decision criterion for rebalancing the portfolio's risky and riskless allocation. The DDT borrows the above strategy written by Albeverio et al.; the integration of the downside deviation will be explained in greater detail in later sections.

The risk metric for the DDT strategy, downside deviation, is the measure of the distance a conditional set of returns is away from the minimal acceptable return. This format is quite different from volatility, although it may sound similar to standard deviation. We will explain how to calculate such a metric, as well as the theoretical implementation, in the Theoretical Implementation section of this paper.

The rationale behind the DDT strategy is to give investors a way to optimize their risk exposure via monitoring of downside returns instead of the volatility of the entire portfolio. This format offers a more adaptive portfolio that is at times comparable to the VT strategy. In this paper, we will first introduce the theoretical framework for downside deviation and the VT strategy in the literature review. Then we remark on the theoretical implementation of both strategies, to finally arrive at the comparative numerical analysis of both strategies in four distinct cases.

In each of our historical cases, we provide a breakdown of the performance of both VT and DDT strategies, commenting on the performance of both strategies during specific periods of increasing, decreasing, cyclical, or crashing markets. During an increasing market, we compare the returns and downside deviation, and comment on the behavior of the strategy, to arrive at some comparative conclusions. Based upon those conclusions, we then compare the strategies in times of decreasing markets, particularly checking to see if VT or DDT offers a particularly more appealing risk-reward trade-off. Lastly, cyclical markets and crashing markets offer a particularly important testing using recent data from the last eight years, where we experienced a recession in addition to high cyclicality and volatility of volatility.

Next, we examine the impact of various targets on both DDT and VT strategies. We compare the trade-off that each strategy makes in terms of returns volatility, maximum drawdown, and returns for a stricter or looser risk target. And finally, we comment on the implementation of DDT and VT in industry.

II. Literature Review

In 'Investment instruments with volatility target mechanism,' Albeverio, Steblovskaya, and Wallbaum (2013) constructed the framework for an investment portfolio with a VT. After their theoretical framework, they examine this mechanism by conducting a historical study on a portfolio containing an equity index that uses this mechanism. The VT mechanism proposed by Albeverio, et al. (2013) is a rule-based dynamic rebalancing mechanism.

Regarding this mechanism, Albeverio, et al. (2013) reached four major conclusions. First, the VT mechanism provides a better risk-return than a pure equity index in a high volatility depreciating market or a low volatility appreciating market. Second, the rebalancing period of the VT mechanism needs further exploration. In their numeric example, they used a monthly scheme in order to minimize transaction cost; however, new schemes could generate different results. Third, the performance of the VT portfolio can also vary depending on the historical volatility estimation horizon. Finally, the VT mechanism can also be adapted with other risk metrics. My capstone will address that last conclusion and explore the effects of using different metrics on the performance of a VT portfolio.

Variance, or volatility in Albeverio, et al.'s (2013) paper, is a risk measure that was described in Markowitz's (1952) paper entitled 'Portfolio Selection' (Markowitz 1952). In their book *Statistical Models and Methods for Financial Markets*, Lai and Xing (2008) described variance as the deviation of the set of end states, particularly the set of total returns. Therefore, according to Lai and Xing (2008), variance is a measure of risk that Markowitz's model seeks to optimize for the greatest return (Lai and Xing 2008, 67-69). Of course, there exist many other forms of risk measures; and according to Artzner, Delbaen, Eber, and Heath (1999) in their paper 'Coherent measures of risk,' risk is the variation of the future value of a positon that is caused by some form of uncertainty (Artzner, Delbaen, Eber, and Heath 1999, 210). Therefore, working off of that definition, variance, calculated as volatility, measures the standard deviation of a set of a certain cardinality. The cardinality of the set is the horizon in which standard deviation is calculated. What standard deviation seeks to capture is the variation of the set from its mean; and practitioners use this measure to estimate the magnitude of future deviation, which will impact the end state of their portfolio.

In his book *Portfolio Selection*, Markowitz (1959) made an important observation regarding standard deviation (Markowitz 1959). He said, "A distribution is *symmetric* if reflecting it about E [its mean] reproduces exactly the same distribution…The variance of a distribution is always the same as the variance of its reflection" (Markowitz 1959, 190). He notes that, given a non-symmetric distribution and its reflection around the set's mean X, their respective variances are equal. Hence, a set that has a number of small negative returns followed by a large positive return can have the same variance as a set with one very large negative return followed by a number of very small positive returns. Although the variances of the two sets are the same, the prospect of one

very large negative return is less desirable to investors than a very large positive return. In other words, variance is low when the set of returns has a narrow range. However, since investors are concerned about preventing losses, they need a way to calculate the possible downside risks. To illustrate the disadvantages of variance, refer to the following example:

Note in this example, E[.] is the expectation of a set and V[.] is the variance of a set.

Let $R = \{-0.05, -0.05, 0, 0.05, 0.15\}$ *be the set of returns of a stock.*

Then $E[R] = 0{,}02$, $E[R^2] = 0.006$,

$V[R] = E[R^2] - E[R]^2 = 0.006 - (0.02)^2 = 0.0056$.

Now let R^r *be the reflection of R around* $E[R]$ *or 0.02.*

Then $R^r = \{0.09, 0.09, 0.04, -0.01, -0.11\}$, *and note that* $V[R^r] = 0.0056$.

Therefore, the variances for R and R^r *are equal.*

Markowitz (1959) addressed this situation by using a metric called semivariance, referred to in other works as downside deviation (Markowitz 1959, 194). In this paper, we will refer to semivariance as downside deviation.

Downside deviation only measures the deviation of the values below a threshold within a portfolio. The formula we will use is from more contemporary authors, Rachev, Stoyanov, and Fabozzi (2008), who described in their book *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization* the formula for downside deviation (Rachev, Stoyanov, and Fabozzi 2008, 177).[1] As a risk measure, variance considers both very high and very low returns undesirable, whereas downside deviation considers only the very low returns undesirable (Markowitz 1959, 188-194). Although Markowitz (1959) used downside deviation in a portfolio optimization context, it certainly can also be a balancing criterion in Albeverio, et al.'s (2013) VolTarget framework. Theoretically, the integration of downside deviation into the VT model is straightforward. If a portfolio were to have a downside deviation that is over the targeted amount, then, to reduce it, we would rebalance the portfolio and incorporate more riskless assets. As we do so, the calculation of downside deviation is more computationally intensive (Markowitz 1959, 193).

Using downside deviation as a rebalancing metric is a relevant method for to limiting the probability of losses while simultaneously increasing the probability of gains. This metric, and many other asymmetrical metrics such as Value at Risk, may offer a way for investors to construct a portfolio with potentially desirable characteristics. Finally, these metrics may provide additional depth to Albeverio, et al.'s (2013) VT framework by addressing one of their key conclusions.

[1] See the Fusio Journal website for an appendix containing the coding.

III. Theoretical Implementation

For the purpose of theoretical implementation of both DDT (Downside Deviation Target) and VT (Volatility Target) strategies, we will first introduce the calculation of downside deviation which, as seen from the literature review, is of a different nature from standard deviation. Then we will introduce the framework for the VT mechanism, which in our paper will form the VT strategy comparable to the DDT. After this, we review the implementation of DDT as well as the theoretical advantages and disadvantages of the DDT when compared with the VT.

Calculation of Downside Deviation

Downside deviation measures the deviation of a subset of returns from the minimum acceptable return, for a given historical horizon. In order to calculate downside deviation, we will need a set of returns denoted H. This will be our estimation set. The cardinality of the set H, denoted *n*, is the size of the set of returns, which represents the historical window. We also need the minimum acceptable return denoted *m*. The minimum acceptable return is the key threshold value, in which a return below this value would be included within the calculation and a return above it would be set to 0. Finally, let *S* denote downside deviation; and please note that, for the remainder of this paper, S is not variance as it normally is represented but is downside deviation. According to Markowitz (1959), one must take the following step to calculate downside deviation. First, given a set of returns H of cardinality *n* and a minimum return *m*, we separate the returns within H into two groups: group 1 contains all returns greater than *m*, and group 2 contains all returns less than or equal to *m* [Step (1)]. Secondly, we set all returns within group 1 to 0 [Step (2)]. Then we combine the groups back together to form a new set of returns; call it G. And finally, we calculate the average distance of all returns in G from *m* [Step (3)]. The average distance is our downside deviation for the set H [Step (4)]. So, in general, if x represents all individual returns within the set of returns H and *n* represents the cardinality of H, the formula below represents the generalized method for calculating downside deviation (Markowitz 1959, 188-194):

$$\text{Downside deviation} = S = \sqrt{\frac{\Sigma_{i=1}^{n}(\min(x-m,0))^2}{n}} \qquad [1]$$

For example, say we have a set of four returns {0.1, -0.1, 0.2, -0.01}, that is H = {0.1, -0.1, 0.2, -0.01}, and our minimum acceptable return, *m*, is 0; then we calculate the downside deviation by first to separate two groups of returns as per step (1) above: group 1 = {0.1, 0.2} and group 2 = {-0.1,-0.01}. Then do step (2), thus group 1 = {0,0}. Doing step (3) gives us: set G = {0,0,-0.1,-0.01}. Now we calculate the average distance of each point in G from *m*. The work below is equivalent to using the generalized method shown in equation [1]:

$$\text{Downside deviation} = \sqrt{\frac{(0)^2 + (0)^2 + (-0.1)^2 + (-0.01)^2}{4}}$$

$$= \sqrt{0.002525} = 0.050249$$

THEORETICAL IMPLEMENTATION OF VT AND DDT

Now that we have presented how to calculate downside deviation, we will move on to how to construct VT and DDT strategies. Below we will first present the guidelines for VT taken from Albeverio et al.'s (2013) paper, then we will integrate this structure with downside deviation as the metric to form the DDT strategy.

In Albeverio et al.'s (2013) paper, the VT mechanism will function to balance the portfolio to a certain level of targeted volatility periodically. To construct such a portfolio, in step (1), the investor needs to determine a horizon for calculating their historical portfolio volatility and set a target volatility. In step (2), the investor constructs a portfolio with the initial volatility at the target volatility allocating the ratio of the historical volatility to the target volatility into the risky asset and 1- that into the riskless asset. At the end of each rebalancing period, in step (3), the investor recalculates historical volatility for their portfolio using the most recent data and rebalances as in step (2). If the portfolio deviated from the volatility target, then the investor will rebalance the portfolio with a mixture of risky and riskless assets, in order to rebalance the portfolio's volatility towards the target (Albeverio, Steblovskaya, and Wallbaum 2013, 1520). By following these three steps, an investor is able to construct a portfolio with a VT mechanism.

Using the above set of instructions, we can change this format by changing the risk metric at which the strategy targets and rebalances around. Before we go further, let us first introduce some variables. Again we will denote downside deviation as S, which is not variance as usual. Then we will denote the target downside deviation as g, and the historical downside deviation calculated at time $t = 0$ as S_0 . When we compare g with the downside deviation calculated at time t, S_t, we assume both are calculated annualized to the whole period of the historical estimation horizon (i.e., if we have one month for the estimation horizon, then both g and S_t are scaled to one month by multiplying the daily with $\sqrt{21}$). The portfolio weight that we invest in risky assets at time t is denoted by α_t, and the risk-free weight is denoted as β_t. Since these are percentage allocations of the total portfolio value, the sum of α_t and β_t should always equal 100%. Now let the value of the portfolio at time t be denoted as v_t. Then the value of the risky allocation at time t is $v_t * \alpha_t$ and risk-free allocation equal to $v_t * \beta_t$ and value of the portfolio is $v_t = v_t * \alpha_t + v_t * \beta_t$.

Expanding upon steps (1), (2), and (3) above, while also integrating the calculation of downside deviation, gives the following list of instructions: Following step (1), the investor should define initial investment v_0, define which assets to invest (risky and riskless), define the historical window (denoted as τ), the target downside deviation g,

and the rebalancing window. Now following step (2), the investor is constructing the portfolio using the following modified equation from step (2) above.

$$\alpha_0 = \min\left[\frac{S_0}{g}, 2\right], \qquad [2]$$

where α_0 denotes the risky allocation, and β_0 will be calculated using the equation below:

$$\beta = 1 - \alpha_0. \qquad [3]$$

The investor then will rebalance the portfolio every rebalancing period by doing the above calculation again and constructing new weights which he or she will invest into. Note that the amount of risky and riskless assets invested were explained above. The major modification we conducted was to change the risk metric in the weight calculation equations [2] with downside deviation instead of volatility. In doing so, we preserve the way in which this portfolio reconstructs its allocations, but test new and different risk metrics. Further, we have some important observations regarding the operation of this rule-based allocation strategy. First, there is a leverage limit placed in calculating the allocation multiple to ensure that we do not over-leverage or invest an unrealistic amount in either risky or riskless assets, much like what Albeverio et al. (2013) did. Secondly, this strategy allocates based on the difference between the target downside deviation and the historical downside deviation of the risky asset, which functions like the VT strategy. However, since we have changed the risk metric, we have some different behavior of how the DDT strategy rebalances, and thus we have arrived at its advantages and disadvantages.

Theoretical Disadvantages and Advantages

By rebalancing using downside deviation (DDT), our portfolio exposure to the risky asset depends on the deviation of the set of negative returns. If the downside deviation is lower than the target, then the negative returns have a smaller spread than the target, and we will increase our allocation to risky assets. Therefore, when we compare DDT to VT, we have certain advantages of using downside deviation over volatility. Given an initial position and a downside deviation target, our risky exposures will not decrease when volatility increases from upside price shocks. Therefore, we gain additional potential for upside gains while not scaling back on the increased volatility. Starting at the same initial position and a downside deviation target, our risky exposure will decrease when volatility increases from sufficiently large downside price shocks. Hence this strategy reduces risky allocation as the downside deviation increases and as more downside price shocks take place. And finally, in highly volatile markets, this

strategy performs better compared to straight market exposure and volatility-based dynamically rebalancing. This is due to the calculation of downside deviation where less returns have an incremental impact when compared to volatility calculation. We will see this in the historical simulations. These advantages are due to how downside deviation is formulated. If we have, for a given historical estimation window, an additional positive return, irrelevant in magnitude, then the additional return will lower the current observed downside deviation, thereby increasing our risky allocation. However, the negatives of the downside deviation strategy occur when we have to handle additional negative returns.

The disadvantage of DDT when compared to VT is how it handles negative returns. Since downside deviation is the average distance of a conditional set of returns from the minimal acceptable return, i.e., returns less than the minimal accepted return, an additional negative return, depending on the return magnitude, can actually decrease, increase, or have a negligible impact on downside deviation.[2] Hence the DDT strategy does not perform well when the market is trending downward, performing especially badly if the market is cyclically trending downwards with low volatility. In this situation, downside deviation does not increase as much as it should and therefore we over-allocate in risky assets. VT in this case will have a stricter allocation to risky assets since volatility is bi-directionally impacted. And as with most dynamically rebalancing strategies, we are observing the historical marketplace, assuming that future marketplace movements can be predicted by past movements. Further, it is also possible, given a consistently increasing stock market with no negative returns, that we will actually liquidate our risky allocation, where instead we should be investing more. This is because, in that case, our downside deviation is 0 and thus we are underweight on risky assets.

Given large negative price shocks, such as market crashes, the DDT strategy will lose more when compared to the VT strategy, since the DDT strategy generally allocates more to the risky assets due to the way in which downside deviation is calculated. This can be addressed by choosing a good target downside deviation; however, an extremely strict target downside deviation will essentially behave just like VT and therefore lose its advantage. Using the historical cases, we will illustrate these advantages and disadvantages of DDT when compared to VT. For an investor, whose objective with these strategies is to maintain a market allocation while also enacting some sort of risk control to prevent excess loss, the DDT strategy offers a desirable attribute in the way in which it allocates to the risky assets. However, this attribute, in conjunction with the ex post style in which it rebalances the portfolio after large negative price shocks, means that the investor can take on larger individual losses when compared to the VT strategy.

[2] This is true since an additional negative return that is far away from 0 will result in an increase in downside deviation, whereas an additional negative return that is near 0 will decrease downside deviation. And an additional negative return that is close to the current downside deviation will have no impact on the downside deviation.

IV. Historical Simulations

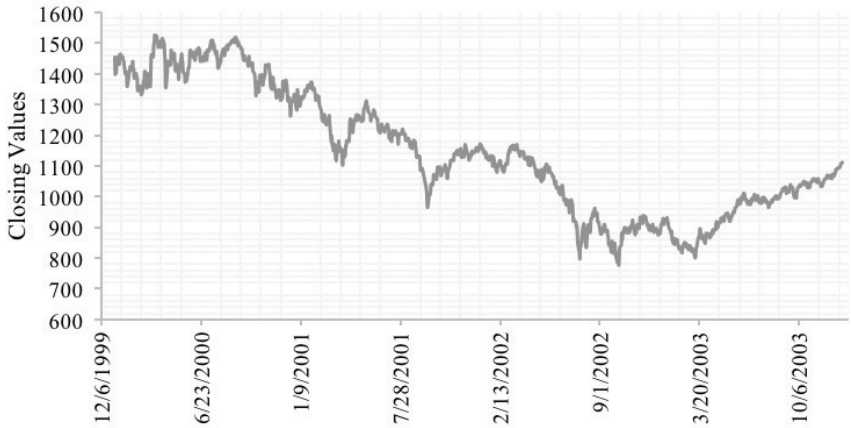
A primary way in which we can evaluate a strategy is through back-testing, and thus with historical simulation we can observe how DDT will perform in differing market situations. Further, we will conduct a comparative analysis of relative performance against VT. Before we introduce the cases, we first note that these simulations are intended to comment on the general performance of the strategy, and not on the optimal settings for the rebalancing frequency, target downside deviation, or historical estimation set size. After presenting the cases, we will examine the relationship between target downside deviation and volatility and their respective resulting portfolio characteristics. Secondly, throughout all our tests, we used a historical estimation set size of twenty trade days and a rebalancing frequency of five trade days, with target downside deviation and volatility both 0.02 (monthly). To simulate our market portfolios, we wrote a software in the data analysis language r that simulates the operation of this strategy. In our software, we do have a dataset of the U.S. national debt in the form of T-bonds; however, for our simulations, we will assume a flat interest rate as if the capital were sitting in a savings account.

We will examine each of these cases with the following protocol. For each case, we have selected a period of distinctly different risky asset performance. In our analysis, we will first review the risky asset performance, and then we will review the summary metrics on both downside deviation and the volatility control. For all our back tests, we will use the S&P 500 index as our risky asset (the SPY ETF offers a good way of getting exposure to the S&P 500 index and can be used in implementation).

Case 1: Historical Simulation between 2000 and 2003

This period was selected due to the relatively flat performance of the S&P 500. As seen from the Figure 1 below, we have a cyclical movement of the index which helps us point out in an empirical study the disadvantages that we remarked upon in the previous section of this paper. The monthly volatility for this period is 4.7896%. To analyze this case, we will first compare the final performance during this period with the index's, VT's and DDT's performances. We have the following equity curve for the VT rebalancing mechanism and DDT rebalancing mechanism. In this period of cyclicality, we see the DDT strategy slowly underperforms the VT strategy. In this particular historical situation, we have the cyclicality of the S&P 500 index weighting down on the performance of the DDT strategy.

Figure 1
S&P 500 Closing prices between 01/03/2000 and 12/31/2003



To see how this cyclicality is affecting downside deviation, we will observe how our downside deviation has changed over this case. Referring to both Figures 1 and 3, we can conclude that Downside deviation will not increase in periods of slow and steady decrease in the marketplace, resulting in more losses compared to VT due to over-exposure in the marketplace. This can be seen at around 5/24/2002. In periods of large negative returns, downside deviation will spike up and increase, e.g., as seen at around 4/19/2001, 11/5/2001. Therefore, in this case DDT and VT are similar; the risky allocation will be decreased. And in fact, we can see volatility spiking in both of those cases. After periods of large negative returns, when the market has reached its troughs, downside deviation decreases to low levels.

Figure 2
Equity Curves of DDT and VT Portfolios for Case 1

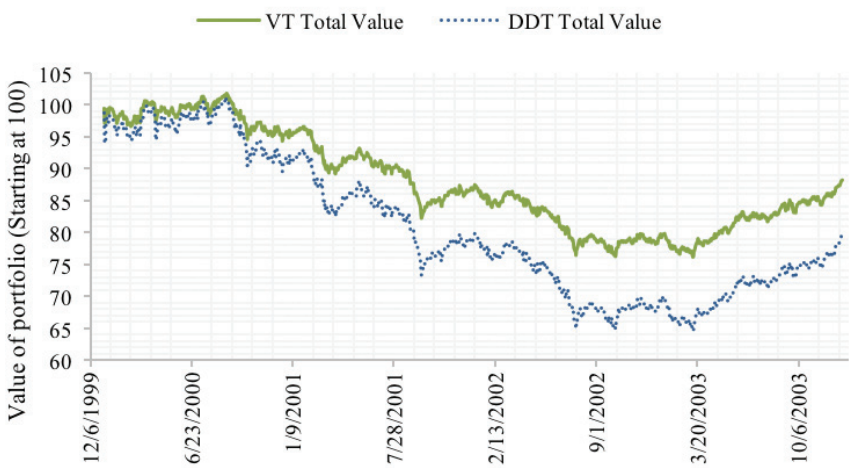
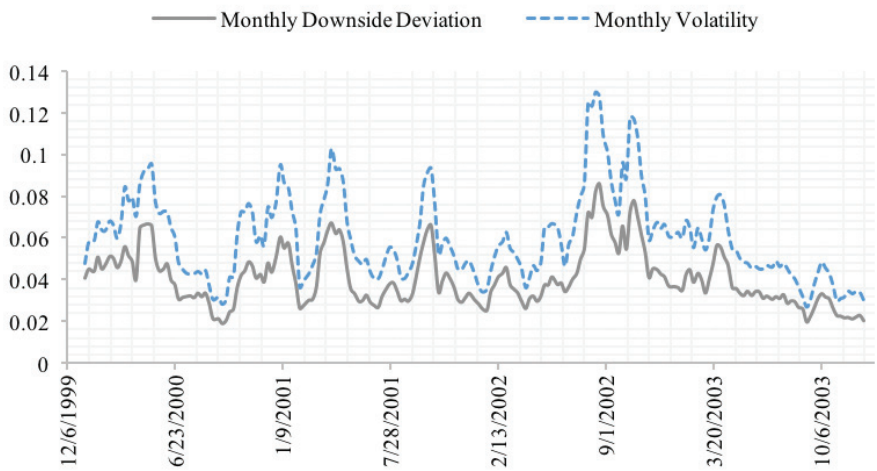


FIGURE 3

1 MONTH DOWNSIDE DEVIATION AND VOLATILITY (RECALCULATED EVERY WEEK), CASE 1



Therefore, downside deviation is in fact quite responsive when there are relatively volatile changes in prices, especially large negative decreases; however, when there are less volatile negative cyclical trends returns, as seen on 5/24/2002, downside deviation does not increase as much as volatility. This fact results in larger losses due to more risky asset exposure. In this particular time period, DDT performance is far worse when compared to VT, and close to raw S&P 500 exposure. Again, this confirms our theoretical remarks on the strategy such that it will not perform well in periods of stable losses, since it over-weights the risky assets in these times by not adjusting the measurement accordingly. Below are the return statistics for VT and DDT compared to the general market returns.

TABLE 1

TABLE OF KEY STATISTICS FOR CASE 1

Strategies	Return (% for the whole period)	Monthly Volatility	Maximum Drawdown
VT	-11.814%	2.173%	-2.048%
DDT	-20.327%	3.294%	-2.970%
S&P 500	-23.591%	6.336%	-5.828%

In this case, we have a decreasing market where the market decreases slowly to a local minimum and then increases again, only to fall below the previous local minimum; VT gave a vastly less negative return when compared to DDT. This is due to the aforementioned fact that, when markets decrease by relatively similar rates of change, downside deviation actually stays relatively constant whereas volatility will change more dramatically. Furthermore, notice that in this case DDT's worst daily return is at

-2.970%, which is worse than VT but much better when compared to the marketplace. And all the above observations point to the fact that the additional losses are due to a higher risky allocation, which is due to a lower downside deviation that results from this negatively trending S&P 500.

Case 2: Periods of High Volatility and Extreme Returns, 2007 – 2012

During the recession of 2008, we saw a significant drop within the marketplace. Since such an extreme event can happen any time in the future, it is necessary to observe the performance of DDT and VT during that period of time and compare it to the market. As seen in Figure 4 below, during the latter half of 2008 the S&P 500 decreased significantly.

This significant decrease in S&P 500 at around the end of 2008 can give us a way to test how our strategy would function during crashing markets and how long it would take for the strategy to recover. Below in Figure 5 shows the equity curves of both the DDT strategy and the VT strategy.

Figure 4
S&P 500 Close Between 2007 and 2012

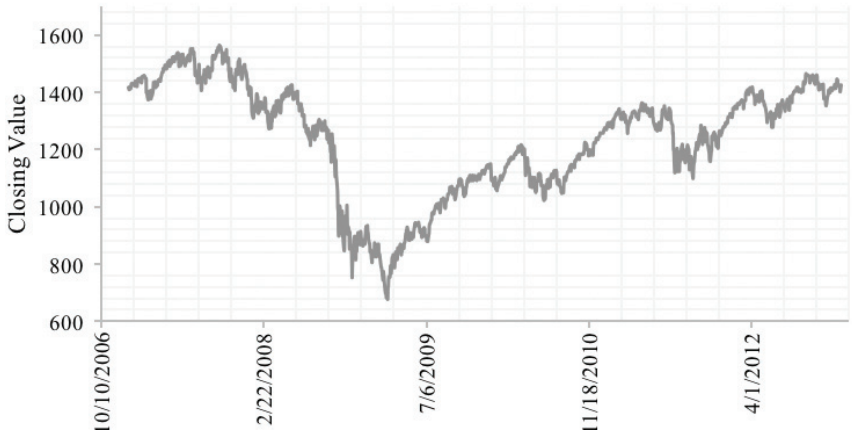
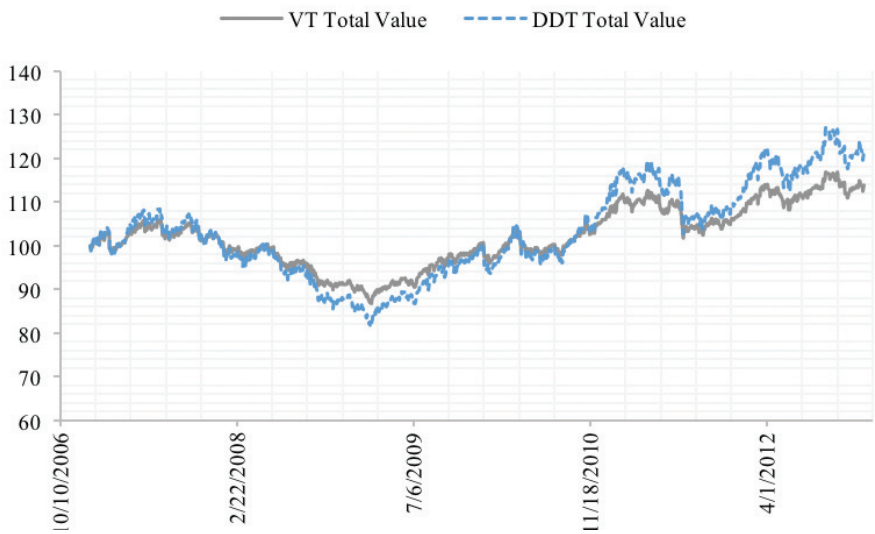
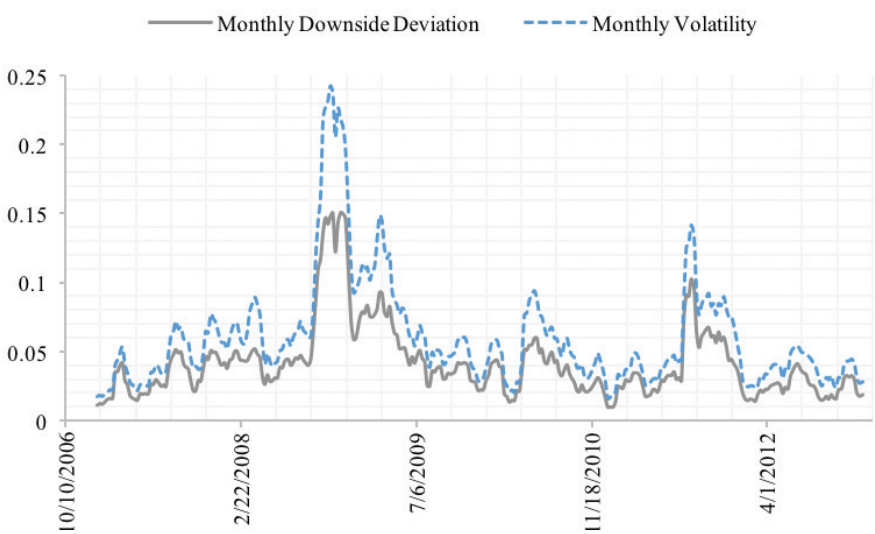


FIGURE 5
EQUITY CURVES OF THE DDT AND VT STRATEGIES FOR CASE 2



Comparing this to case 1, we see that the downside deviation follows monthly volatility closely. But downside deviation peaks less than volatility while also decreasing at relatively the same rate. Therefore, from the graph in Figure 6 below, we can conclude that during the crash DDT maintained a higher risky allocation when compared to VT, due to the lower peak downside deviation.

FIGURE 6
MONTHLY DOWNSIDE DEVIATION (CALCULATED EVERY REBALANCING PERIOD) FOR CASE 2



This is mainly because volatility is stricter when it comes to its risky allocations, since both positive and negative movements affect it. In order to analyze this in greater detail, we will review how downside deviation behaved during this case. By referring to Figures 5 and 6, we can conclude that in periods of cyclically increasing stock prices, as seen from 2010 until the end of this case, DDT generates a higher return when compared to VT. Further, note that during the same period, when stock prices decrease, DDT and VT both reduce their respective portfolio's risk exposure. In periods of low volatility and with cyclical non-trending stock prices, as seen in the beginning of the case until 3/1/2008, DDT had comparable returns to VT with both volatility and downside deviation low. The divergence between the VT and DDT took place at around 11/18/2008, at which point downside deviation spiked downward, drastically increasing the risky allocation, whereas VT did less of since volatility was still much higher. From the observations above, we can see that, given the formulation of downside deviation, it follows the same general pattern as volatility. However, it is a much smaller number since it only considers the deviation of downside returns and as such will allocate more risky instruments. Over this case, the key performance indicators for the VT, DDT, and the market benchmark are displayed below.

TABLE 2
COMPARISON OF KEY STATISTICS FOR CASE 2

Strategies	Return (% for the whole period)	Monthly Volatility	Maximum Drawdown
VT	13.952%	2.241%	-3.07%
DDT	21.884%	3.231%	-4.52%
S&P 500	0.677%	7.185%	-9.03%

Referring to Table 2 above, we can observe that both DDT and VT yielded a lower volatility and a higher return in this highly volatile market condition, with a monthly volatility of 7.185% for the S&P 500. Further note that, although DDT suffered a larger decline (observed in the Maximum Drawdown) during the market crash, it recovered faster when compared to VT, ending the case with highest returns. The trade-off in this case is that DDT has more volatility but, since we are not balancing on volatility, it is a given that volatility would be higher.

CASE 3: PERFORMANCE IN AN UPWARD TRENDING MARKET, 1987 – 1995

From the two cases above, we observed a downward trending market in case 1, and an extremely crashing market in case 2. Now for case 3 we will analyze the performance of DDT compared with VT in an upwards trending market with relatively low volatility. Below are Figures 7 and 8, showing the S&P 500 and equity curves for DDT and VT, respectively.

From case 1 and case 2, we can observe that DDT invests in more risky assets when compared to VT, since downside deviation does not take into account upside gains. This means that the DDT strategy is more beneficial in an upwards market environment as seen from Figure 8. So, in this upwards trending marketplace, we see that DDT generates a much higher return when compared to VT. However, we should also note that (1) there is one period in which the S&P 500 experienced a sudden price decline (around 10/26/1987); and (2) during that period we can observe that the gains attributable to excess risky allocation, when compared to VT, are quickly lost and the two strategies' values are relatively similar. This is because there was not enough time for the DDT strategy to build up enough returns to buffer the impact from that sudden drop.

FIGURE 7

S&P 500 CLOSING PRICE BETWEEN 01/05/1987 AND 12/31/1995

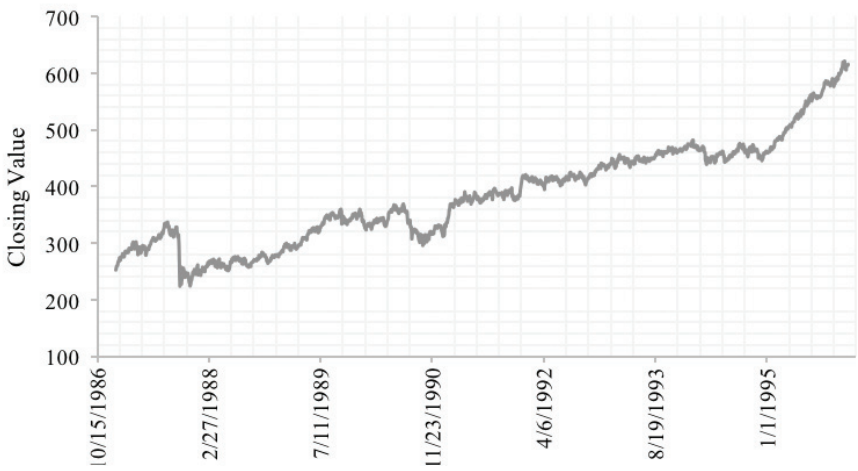


FIGURE 8

EQUITY CURVES FOR THE DDT AND VT STRATEGIES FOR CASE 3

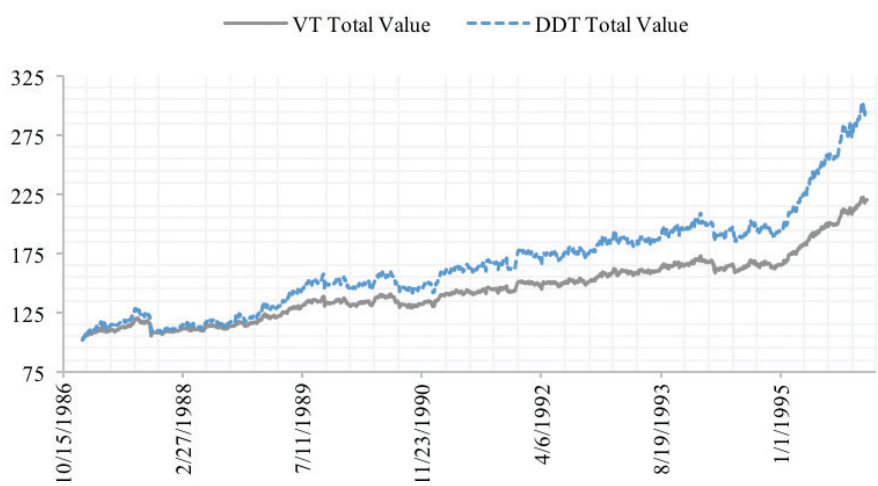


TABLE 3
KEY RETURNS STATISTICS FOR CASE 3

Strategies	Return (% for the whole period)	Monthly Volatility	Maximum Drawdown
VT	120.752%	2.314%	-6.238%
DDT	197.156%	3.410%	-8.946%
S&P 500	144.233%	4.590%	-20.467%

Referring to the maximum drawdown, we notice that DDT's largest drawdown or negative return is greater than VT's. Further, in this case, we noticed that VT generated a lower return than even the marketplace whereas, in all of the previous cases, VT generated relatively higher returns. As we can observe from Table 3 above, and in all of the previous cases, DDT allocates more in risky assets when compared to VT; so, naturally, the largest negative return for DDT is bigger than VT. But the most valuable case to test would actually be the most recent case, since it is characteristic of current market behavior.

CASE 4: PERFORMANCE IN RECENT MARKET CONDITIONS, 2010 – 2016

During the period from 2010 to 2016, our markets have been trading in a manner such that the actual average volatility changes very frequently. This dataset and case is the most relevant case of all, since it embodies the recent market environment (see Figure 9). Therefore, DDT's performance in this environment would best help us determine its benefits to the investor when compared to VT and pure market exposure.

On initial review, we observe that DDT performs better in upwards cycles while trading off to larger losses when markets decline suddenly. So again, from Figure 10, we notice that the same observations from cases 1, 2, and 3 all apply to this case as well. And in general, with the market conditions as sampled in this case, DDT seems to offer superior performance when compared to VT but less than pure market exposure. Yet at the same time, during periods of large negative returns, we note that DDT drawdowns are larger when compared to VT. Table 4 below shows the key statistics for return and volatility; and they offer mostly similar conclusions as above, with one caveat.

FIGURE 9
S&P 500 CLOSING VALUE BETWEEN 01/01/2010 AND 12/31/2016

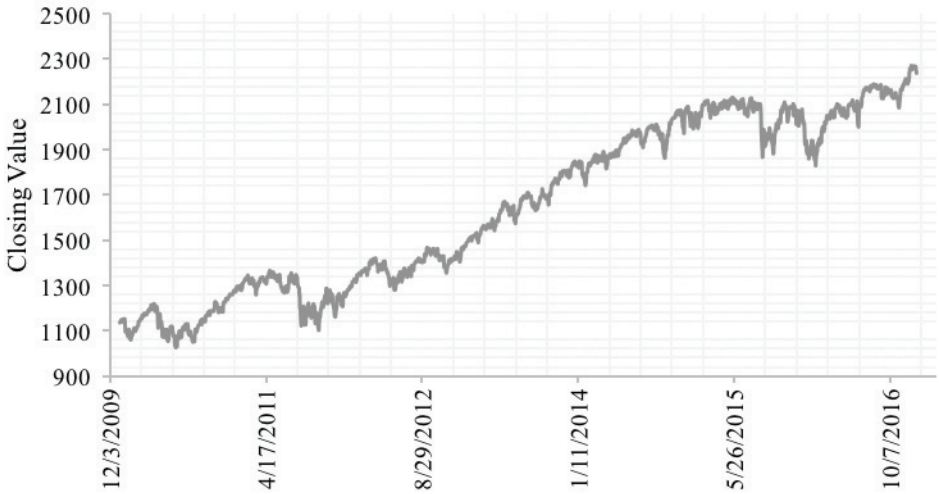
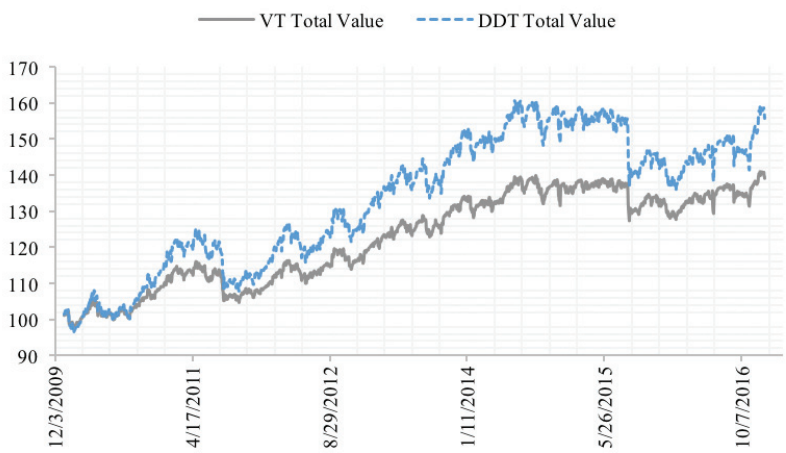


FIGURE 10
EQUITY CURVES FOR DDT AND VT STRATEGIES FOR CASE 4



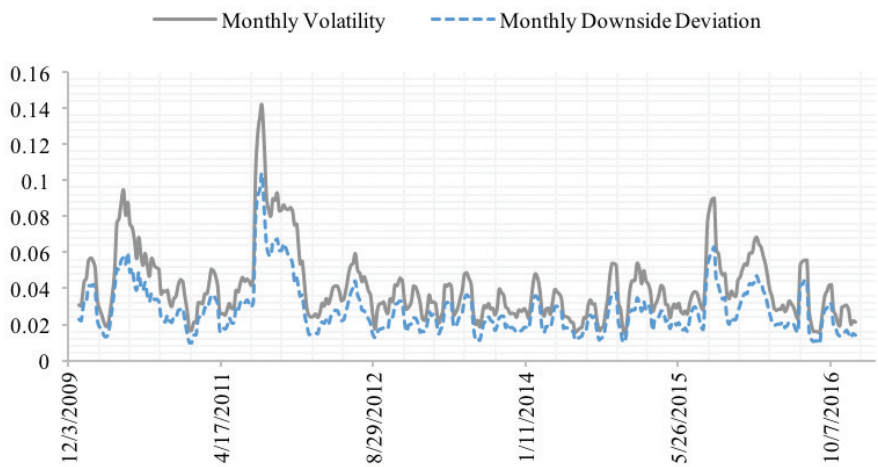
Notice, from Table 4, the following key caveat: both VT and DDT underperformed the market by a significant margin. This can actually be attributed to a particular characteristic of both strategies. Notice in Figure 9 that we have some small increases in price followed by a decrease in price of slightly less magnitude. Referring to Figure 11, we can observe how these fluctuations caused the underperformance. We observe that the very small cyclical positive to negative returns result in losses from buying when the price is high and selling when the price is low, since both metrics lower rebalances

ex post and thus are reactionary. This can be seen at 6/18/2013 and many other similar dates. This sort of loss is similar to losses from delta hedging an option.

TABLE 4
RETURNS STATISTICS FOR CASE 4

Strategies	Return (% for the whole period)	Monthly Volatility	Maximum Draw-down
VT	39.178%	2.348%	-3.179%
DDT	55.595%	3.431%	-4.905%
S&P 500	97.604%	4.486%	-6.663%

FIGURE 11
MONTHLY VOLATILITY AND DOWNSIDE DEVIATION FOR CASE 4, RECALCULATED WEEKLY



SUMMARY OF THE CASES

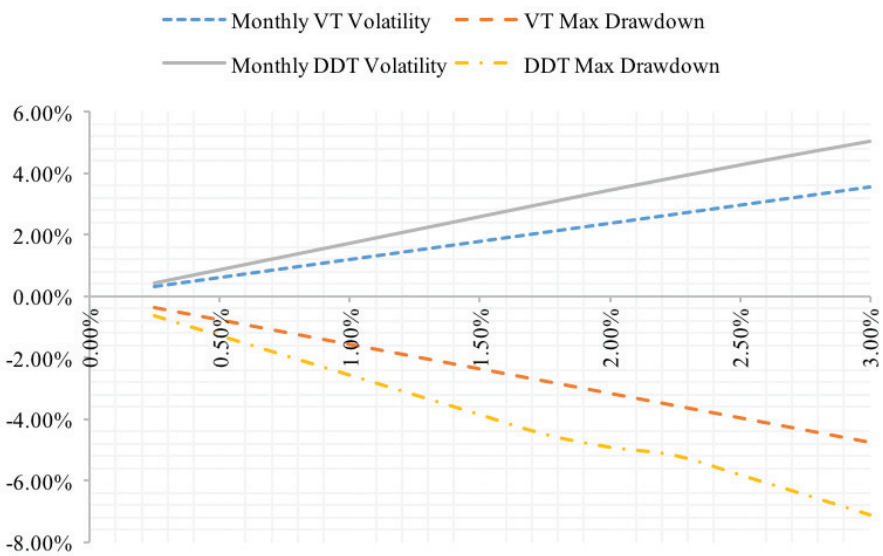
In summary, from the four cases above we can observe that DDT allocates more to the risky assets, therefore as a result performs better. Further, as a way to limit market risk while also maintaining risky exposure, it is prudent to note that the DDT has a high correlation to the marketplace (in all four cases above, correlation is at 0.95 or above; and we can also observe that by comparing the equity curves and the closing prices for the S&P 500). Since DDT's chief objective is to limit risk to a certain level while also maintaining market exposure, and in all four cases we have a maximum drawdown lower than the marketplace, from that perspective DDT is successful in controlling risk. The higher maximum drawdown when compared to VT is certainly a function of the downside deviation target, and we can change the target to lower the maximum drawdown.

THE IMPACT OF RISK TARGETS

Using case 4 as an example, if we use a downside deviation target of 0.01 monthly instead of 0.02 in all four cases, we have a standard deviation of 1.713% with a maximum drawdown of -2.563% and a return of 30.598%. Notice that these numbers offer significantly better standard deviation of the portfolio value and much smaller maximum drawdown when compared to VT. However, in the standard case 4, since we made our downside deviation target larger, our maximum drawdown, volatility, and returns all increase. Figure 13 below shows the relationship between the risk target to volatility and max drawdown for both DDT and VT.

FIGURE 12

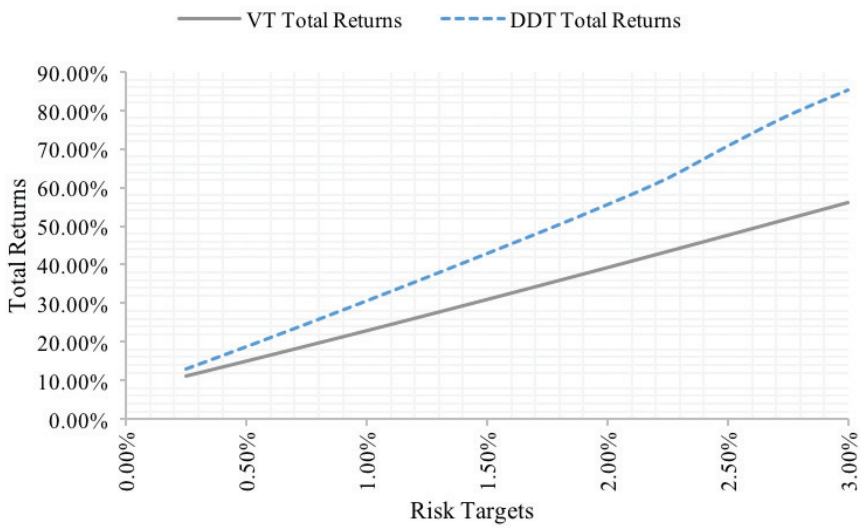
VARIOUS DOWNSIDE DEVIATION AND VOLATILITY TARGETS AND THEIR MONTHLY



VOLATILITY AND MAX DRAWDOWNS USING CASE 4′S DATA

In Figure 12 above, the topmost line is monthly DDT volatility and the bottommost line is DDT's maximum drawdown. This graph shows the various maximum drawdowns and average volatilities realized using case 4. As the risk target increases, DDT and VT volatility, as well as maximum drawdown, all increase as we take on more risky assets. At this point, it is important to note that the linear relationship is positive between the monthly volatility as the byproduct of incremental additional risk tolerance. This is also true of the relationship between max drawdown and target monthly downside deviation. Notice that DDT's reaction to a stricter risk target is stronger when compared to VT. Now let us look at the effect of different risk targets on the return of VT and DDT over the timespan of case 4.

FIGURE 13
DIFFERENT RISK TARGETS EFFECTS ON RETURNS



Referring to the two different returns curves in Figure 13, we see the same relationship between risk targets and returns. As risk target increases, the returns increase. Notice here, too, that the slope for the VT's relationship between returns and risk target is more gradual when compared to DDT. DDT's behavior in response to a changing risk metric is more drastic and sensitive to a different target when compared to VT. This is the result of how downside deviation is calculated, and can offer an investor some advantages in gaining market exposure; however, this has the cost of additional risk exposure. It is for the investor to decide if this trade-off is justifiable.

In summary, the historical cases confirmed our initial conclusions about the downside deviation strategy, made in Section III of this paper. We confirmed that, during times of volatile single directional markets, DDT tends to perform better than VT as seen in cases 2 and 3. Additionally, we observed that the dynamically rebalancing strategy makes adjustments to the portfolio ex post, hence behaving similarly to a delta hedging strategy for options. This is also true, and we can observe phenomena in case 4 where the market's micro fluctuations caused both the DDT and the VT strategies to underperform the marketplace.

SOME COMMENTS ON REAL-WORLD IMPLEMENTATION

When implementing this model in the real world, an investor would need to conduct several follow-up tests to ensure that a strategy like DDT or VT is optimized for his or her objectives. For example, if the investor is attracted by the additional controlled exposure they gain when using DDT, then to optimize it they would have to do the following tests in order to ensure that the strategy is worthy of being invested

with actual capital. Firstly, an investor needs to determine the optimal target downside deviation; to do so, the investor needs to determine his or her risk tolerance. Then, using some recent data, back the DDT strategy with the various levels of risk targets to choose the one that generates a portfolio that suits his or her needs. Then, to find a good rebalancing window, the investor needs to take into account the cost of trading as well as his or her potential market impact. A good way of doing this could be to implement a transaction cost additive into the calculation of the portfolio value. These numbers need to be the investor's own cost. For finding a good estimation window, the investor needs to determine a good period of time that he or she can use to estimate future behavior. Generally, the industry uses a monthly or 3-month window. System-generated data with stochastic equations can also represent a way to test out which metrics are best. Back tests are a simple way of seeing how different inputs can affect the output of the DDT or VT strategy in a real market environment. If the investor also wants to run with simulated data, then he or she can do so with several methods. Possible methods can range from using historical density functions to a simple Brownian motion.

Possible Areas of Improvement

We can address some of the disadvantages described above with a modification onto downside deviation into another ratio such as the Calmar ratio, which is average returns divided by downside deviation, which would address this particular disadvantage (Young 1991). This modification can address disadvantage 1 and disadvantage 3, since it will incorporate average returns, because the Calmar ratio increases when the portfolio is profitable and decreases when the portfolio is not profitable, thereby cutting down on risky allocations even when the market is in a steady downward trend. However, since the nature of these strategies involves ex post rebalancing, disadvantage number 4 will always persist unless the investor implements some predictive measurement into the risk metric.

As such, we can also improve multiple areas which this strategy can be modified to include. First, we can include liquidity risk and transactional cost by including average daily volume into the calculation of the risky and riskless percentage of the portfolio. Second, by balancing using downside deviation, we are trying to gear the portfolio toward a certain target downside deviation. However, since our goal is to maximize returns, we should incorporate some sort of average returns into the calculation of the rebalancing metric as an improvement. As seen from the above discussion, one can, theoretically, have a portfolio that has a very narrow spread of very large average downside returns. Thus, if we include a measure of average returns into the mechanism, such as the Calmar ratio, we would address the aforementioned disadvantages.

This strategy, as with any strategy that rebalances based on some criterion, will behave differently depending on different market conditions. In the next section, using historical data from the S&P 500 as our risky instrument and assuming a generic savings account as our riskless instrument, we will back-test the performance of our

downside deviation strategy in comparison to rebalancing using volatility. Before we go any further, we should also note that this dynamically rebalancing strategy works with other forms of risk metrics. One particularly interesting metric worthy of additional testing and calibration would be metrics using Extrema Value Theory. Due to the complexity of the topic, we will explore this in future research.

V. Conclusion

Downside deviation dynamically rebalancing (DDT) offers an alternative to volatility rebalancing (VT) that is more attractive to a portfolio investor in certain market conditions. The use of minimal acceptable return in calculation of downside deviation offers investors greater flexibility, as it can be not just a static coefficient but also an aggregation function over the historical estimation period—therefore, it is more adaptable to different user requirements. Downside deviation does not perform well when the market is in a steady negative trend. Like all ex-post rebalancing strategies that react to market stimuli, downside deviation does not behave well in a market where the volatility of volatility is high; this can be seen recently, where both volatility and downside deviation strategy underperformed the marketplace. DDT is far more complex to construct when compared to VT portfolios, but with accessible high-level programming languages and APIs, it is far more accessible than ever. Using real market data, throughout our historical cases we found that, until recently, this strategy outperformed VT and pure market exposure using the same risk target of monthly 2% downside deviation or volatility.[3] DDT has a steeper risk–return trade-off, which is better for investors who hypothesize a cyclical or positively trending market.

Since the goal of these strategies is to allow investors to limit their risk while also maintaining market exposure, downside deviation's customizability and adaptability to the user's intentions may offer additional value to an investor seeking a very specific set of returns. The DDT strategy offers an investor more exposure to the marketplace in exchange for a higher possible drawdown. Before concluding this paper, we should note that a more complex and different way to create a risk metric would be to use Extrema Value Theory. Extrema Value Theory can also offer a way in which we can estimate the probability of tail events. It can perhaps offer a more appealing set of portfolio features if incorporated into the VT mechanism, and warrants future research into its application.

[3] Although, 2% downside deviation target is actually a much more liberal risk target when compared to 2% volatility target.

References

Albeverio, S., Steblovskaya, V., & Wallbaum, K. (2013). "Investment instruments with volatility target mechanism." *Quantitative Finance*, 13(10), 1519-1528. doi:10.1080/14697688.2013.804943

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath (1999). "Coherent Measures of Risk." *Mathematical Finance* 9, no. 3 (1999): 203-28. doi:10.1111/1467-9965.00068.

Embrechts, Paul, Alexander J. Mcneil, and Daniel Straumann (2002). "Correlation and Dependence in Risk Management: Properties and Disadvantages." *Risk Management: Value at Risk and Beyond*, M. Dempster (Ed.), Cambridge University Press, 2002, 176-223. doi:10.1017/cbo9780511615337.008.

Embrechts, Paul, Sidney I. Resnick, and Gennady Samorodnitsky (1999). "Extreme Value Theory as a Risk Management Tool." *North American Actuarial Journal* 3, no. 2 (1999): 30-41. doi:10.1080/10920277.1999.10595797.

King, A. J. (1993). "Asymmetric risk measures and tracking models for portfolio optimization under uncertainty." *Annals of Operations Research*, 45(1), 165-177. doi:10.1007/bf02282047

Lai, T. L., & Xing, H. (2008). "Markowitz's Portfolio Theory". In *Statistical models and methods for financial markets* (pp. 67-72). New York: Springer.

Lopez, Jose, and Christian Walter (2001). "Evaluating Covariance Matrix Forecasts in a Value-at-risk Framework." *The Journal of Risk* 3, no. 3 (2001): 69-97. doi:10.21314/jor.2001.044.

Markowitz, H. (1952). "Portfolio Selection." *The Journal of Finance*, 7(1), 77-91. Retrieved November 10, 2015, from http://www.jstor.org/stable/10.2307/2975974?ref=search-gateway:e20372f3c722fc40a0d2aac8ca719fb0

Markowitz, H (1959). *Portfolio Selection; Efficient Diversification of Investments*. New York: Wiley, 1959.

Mina, Jorge, and Jerry Yi Xiao (2001). "Return to RiskMetrics: The Evolution of a Standard." Technical paper. New York: RiskMetrics, 2001.

Rachev, S. T., Stoyanov, S. V., & Fabozzi, F. J. (2008). "Risk and Uncertainty." In: *Advanced stochastic models, risk assessment, and portfolio optimization: The ideal risk, uncertainty, and performance measures* (pp. 171-206). Hoboken, NJ: Wiley.

Young, W., Terry (1991). "Calmar ratio: A smoother tool." *Managed Money*.

Physical Deviance: The Stigma Surrounding Underweight Women

By Jennifer Miller and Brittani Murphy*

Weight stigma and discrimination are prevalent issues in the United States, and numerous studies have addressed their causes and effects as they relate to overweight and obese individuals. However, very little attention has been given to the stigma surrounding underweight women. Our research aims to examine two of the possible causes of this underweight stigma: anti-thin prejudice and attribution of blame. Participants were shown images of several women with varying body weights and medical conditions, and asked questions about each one's likeability and eating habits. Our findings suggest that attribution of blame plays a large role in causing weight stigma.

Keywords: Weight, underweight, stigma, discrimination, gender.

I. Introduction

Weight stigma and discrimination are critical issues plaguing both underweight and overweight individuals in our nation. Between 1996-2006, the prevalence of weight discrimination – the unequal treatment of people because of their weight – in the United States increased by 66% (Andreyeva, Puhl, & Brownell, 2008). The purpose of our research is to examine two of the possible causes of the weight stigma – defined as prejudice or negative attitudes towards one's weight – surrounding underweight women; specifically, we studied anti-thin prejudice and attribution of blame. The stigma surrounding underweight women can have harmful social, psychological, and physical consequences. We hope that gaining insight into two of the possible causes of this stigma will allow other researchers to build off our findings and possibly find a way to reduce this stigma and its effects.

II. Literature Review

The existing literature related to weight discrimination and stigma is extensive; however, a majority of peer-reviewed journal articles focus solely on the stigma surrounding overweight and obese individuals, with very little attention given to the other end of the weight spectrum (Swami, Chan, Wong, Furnham, & Tovée, 2008). This is problematic because underweight individuals (BMI < 18.5) are the victims of weight stigma, too.

---

* Corresponding email: brittani.murphy@gmail.com. We would like to thank Professor Miriam Boeri for her continuous support and constructive feedback throughout the research processs.

Manifestation of Stigma

This underweight stigma is present in occupational settings as emaciated (BMI < 15) individuals have been shown to be the most likely to be terminated and least likely to be promoted of all the weight classes (Swami, Pietschnig, Stieger, Tovee, & Voracek, 2010). Underweight individuals also face stigma in academic settings: previous research has shown that emaciated individuals are the least likely of all weight classes to be admitted to institutions of higher education (Swami & Monk, 2013). Federal legislation prohibiting age, gender, and race discrimination have been in effect for more than half a century; however, there are not currently any laws in place that prohibit weight discrimination (Puhl, Andreyeva, & Brownell, 2008). The stigma surrounding underweight individuals is also present in the media. News sources, such as the New York Times and Newsweek, often portray those who are underweight - especially those who suffer from an eating disorder - as having a disease or psychological problem (Saguy & Gruys, 2010).

Stigma Consequences

This seemingly omnipresent weight stigma can have detrimental social, psychological, and physical effects on the underweight community (Carr & Friedman, 2005). Underweight men have been found to have an extremely negative self-image and poor social adjustments in college settings (Harmatz, Gronendyke, & Thomas, 1985). Some individuals view and describe underweight women as "walking skeletons" who are "sickening" and "disgusting," and therefore, these women sometimes have a difficult time forming close social relationships (Anderson & Bresnahan, 2013, p. 611). Weight stigma and insecurities also contribute to romantic relationship difficulties (Boyes & Latner, 2013). Body-shaming and weight-related teasing often begins in elementary school where underweight children report higher levels of teasing related to being "weak" (O'Dea & Amy, 2011). Later in life, underweight individuals often report feeling "deeply discredited" because of their "spoiled identity" (Schafer & Ferraro, 2011, p. 79). Underweight individuals often feel intense pressure to achieve society's beauty norms, which has been shown to cause psychological distress and depression (Hackman, Maupin, & Brewis, 2016).

Other research adds to this list of consequences by reporting that individuals report delaying or avoiding health care because they do not want to be weighed on the scale or be told that they need to change their weight (Drury & Aramburu, 2002). Delaying or avoiding health care may only make a person's weight problem worse, which could cause the stigma to increase, and the cycle to continue. Therefore, a change must occur to lessen weight stigma in our society.

Next Steps

In order to combat this discrimination and reduce the stigma surrounding being underweight, one must understand its underlying causes. One possible explanation for the stigma surrounding underweight individuals is that they are seen as being responsible for and in control of their weight. Previous literature has shown that subjects were less likely to blame obese individuals for their body size when their weight was the result of a "disease" rather than individual choices (Hoyt, Burnette, Auster-Gussman, Blodorn, & Major, 2017).

To test whether this holds true for underweight women, we conducted a survey similar to the one conducted by DeJong (1980). DeJong (1980) presented participants with images of four women - two at a healthy weight and two obese - with various medical conditions. He then asked the subjects multiple questions related to the women's likeability and lifestyle habits. Our questions are very similar, but our participants were presented with images of underweight women rather than obese women. The primary goal of our research is to examine two of the possible causes of the stigma surrounding underweight women that the literature identified: anti-thin prejudice and attribution of blame. Anti-thin prejudice can be defined as having a generally negative attitude towards thin individuals, and attribution of blame can be defined as assigning fault to an individual, with or without proper factual evidence to support the claim.

III. Hypotheses

Previous literature has shown that individuals report being less likely to be friends with obese women without a thyroid problem, and we believe the same will hold true for underweight women (DeJong, 1980). As such, we hypothesize that a woman's weight and the cause of her weight will affect her likeability, and that respondents will report being less likely to be friends with the underweight woman without hyperthyroidism. Hyperthyroidism is a condition that causes one's thyroid to produce an excessive amount of thyroxine, which can increase one's metabolism and cause sudden weight loss (Bahn et al., 2011). We thus propose Hypotheses 1 and 2.

Hypothesis 1: The women's weight will affect their likeability.

Hypothesis 2: The causes of the women being underweight will affect their likeability.

A previous study by DeJong (1980) found that individuals tend to assume that obese people have poor eating habits. We predict that the same will hold true for underweight individuals. We thus propose Hypotheses 3-5.

HYPOTHESIS 3: INDIVIDUALS WILL TEND TO BELIEVE THAT THE UNDERWEIGHT WOMEN EAT FEWER CALORIES THAN THE WOMEN AT A HEALTHY WEIGHT DO.

HYPOTHESIS 4: INDIVIDUALS WILL TEND TO BELIEVE THAT THE UNDERWEIGHT WOMEN EAT LESS FREQUENTLY THAN THE WOMEN AT A HEALTHY WEIGHT DO.

HYPOTHESIS 5: INDIVIDUALS WILL GIVE MORE DIET-RELATED HEALTH RECOMMENDATIONS TO THE UNDERWEIGHT WOMAN WITHOUT HYPERTHYROIDISM THAN THE UNDERWEIGHT WOMAN WITH HYPERTHYROIDISM.

IV. Methods

Data Collection

The data used for this quantitative analysis were collected using a survey. The link to the survey was posted in three university Facebook pages and sent through GroupMe to 72 students who reside in two on-campus residence halls at a small business university in the suburbs of Boston, MA. We chose to survey college students because they have diverse backgrounds with various body norms, and all students at the university have a personal computer that would allow them to take this online survey if they were to receive the link. Data collection began on March 24, 2017 and continued through March 31, 2017.

Participants

Our convenience sample consists of 225 undergraduate students, who attend a small business university in the suburbs of Boston, who accessed Facebook and/or GroupMe - a mobile messaging app owned by Microsoft - between March 24, 2017 and March 31, 2017. The link to this survey was made available to approximately 3,100 students, giving us an estimated response rate of 7.26%. 67.11% of respondents are females, and 32.89% are males.

Procedure

Participants were presented with four images of real women in bikinis and asked to answer the questions that followed to the best of their ability using the pictures shown. The women in the images varied in weight and health status: “Sarah” is at a healthy weight according to her doctor, “Taylor” is also at a healthy weight and has hyperthyroidism, “Mariah” is underweight according to her doctor, and “Kayla” is also underweight and has hyperthyroidism. The participants were made aware of these differences and told that hyperthyroidism “can increase [one’s] metabolism.” Please refer to Figure 1 to view all four women featured in our survey.

We asked all respondents a variety of questions based on the pictures shown. All participants viewed the same four pictures in the same order, as shown above. After viewing each picture, respondents were asked if they would be friends with the women shown based on looks alone. Answer choices ranged from 1 (I definitely would not be friends with her) to 10 (I definitely would be friends with her). This question was included to help determine whether a woman's weight, and the cause of her weight, affects her likeability (Research Questions 1 and 2).

Figure 1
Four Women Featured in Survey

Next, participants were asked how many calories they believe each woman consumes per day, and the question clearly stated that 2,000 calories is the standard for daily caloric intake. Answer choices ranged from "Less than 500 calories" to "3,001 calories or more," with 500-calorie intervals in between. This question was included to help determine whether individuals tend to believe that underweight women eat fewer calories than women at a healthy weight do (Research Question 3).

Respondents were also asked how many meals they believe each woman skips per week, on average. Answer choices ranged from "0" to "17-21," with four-meal intervals in between. This question will help determine whether individuals tend to believe that underweight women eat less frequently than women at a healthy weight do (Research Question 4).

Lastly, participants were presented with two open-ended questions asking them to give one health recommendation to each of the two underweight women shown. We hypothesize that respondents will give more diet-related health recommendations to the underweight woman without hyperthyroidism than the underweight woman with hyperthyroidism (Research Question 5).

V. Analysis

Dataset Preparation

While preparing our survey, we required every question to be mandatory. After closing our survey, we examined the data to confirm all responses were complete. We removed six survey responses due to incomplete open-ended answers, leaving us with 225 complete responses. We also checked for outliers and did not find any.

To better understand the answers to our two open-ended questions, we went through all the responses and coded them into one or two different categories depending on how many pieces of advice the respondent gave. Please refer to Table 4 below in the *Findings* section for a complete breakdown of the codes utilized.

Testing Our Hypotheses

A paired t-test was conducted at the 5% significance level to analyze research questions one and two. A column was calculated to find the difference between the likeability of Sarah, a woman at a healthy weight without hyperthyroidism, and Mariah, an underweight woman without hyperthyroidism. This column was utilized to find the sample standard deviation, as well as the average difference of likeability between Sarah and Mariah. The average and standard deviation were used to solve for our test statistic. The same process was performed again when comparing Taylor, a woman at a healthy weight with hyperthyroidism, and Kayla, an underweight woman with hyperthyroidism, as well as when comparing Mariah and Kayla, the two underweight women.

A z-test for the proportion in terms of the number of events of interest was used to analyze research questions three, four, and five. When comparing the proportion of respondents who chose 2,000+ calories between Sarah, a woman at a healthy weight without hyperthyroidism, and Mariah, an underweight woman without hyperthyroidism, we were able to utilize Sarah's proportion as the population proportion within our hypothesis. After finding our test statistic, we found the corresponding p-value within the z-table, and compared it to our alpha of 0.05. The same process was repeated when examining the amount of meals skipped, as well as recommendations given.

Assumptions

When using the z-test for the proportion in terms of the number of events of interest, we assumed that the proportion for a certain woman was representative of our population of all undergraduate students at this university who accessed Facebook and GroupMe between March 24, 2017 and March 31, 2017. Our other assumptions were that the observations are independent of one another and the dependent variable is approximately normally distributed.

VI. Findings

RESULT 1: *A WOMAN'S WEIGHT DOES AFFECT HER LIKEABILITY*

Our paired *t*-test indicated that there was a statistically significant difference ($T_{0.05,224}$ = 1.653, t_1=14.779 and t_2=10.279) between the likeability scores of the two women at a healthy weight and the two underweight women at the 5% significance level. The average likeability score for Sarah, a woman at a healthy weight without hyperthyroidism, was 7.960, while the average likeability score for Mariah, an underweight woman without hyperthyroidism, was 5.591. For the two women with hyperthyroidism, the one at a healthy weight received an average likeability score of 7.618, which is significantly higher than the underweight woman's average likeability score of 5.94. We were able to reject the null hypothesis and conclude that a woman's weight does affect her likeability. The survey respondents were less likely to become friends with the two underweight women than the two women at a healthy weight, which supports our hypothesis. Please refer to Table 1 to see the average likeability scores for all four women.

TABLE 1
DOES A WOMAN'S WEIGHT AFFECT HER LIKEABILITY?
DOES THE CAUSE OF A WOMAN BEING UNDERWEIGHT AFFECT HER LIKEABILITY?

Woman	Weight Type	Medical Condition	Average Likeability (Out of 10)	Standard Deviation for Average Likeability (Out of 10)
Sarah	Healthy Weight	None	7.960	2.124
Taylor	Healthy Weight	Hyperthyroidism	7.618	2.425
Mariah	Under-weight	None	5.591	2.605
Kayla	Under-weight	Hyperthyroidism	5.940	2.602

RESULT 2: *THE CAUSE OF A WOMAN BEING UNDERWEIGHT AFFECTS HER LIKEABILITY*

The results of our paired t-test showed that there was a statistically significant difference ($T_{0.05,224}$ = 1.653, t = 4.037) between the likeability scores of the two underweight women at the 5% significance level. Mariah, the underweight woman without hyperthyroidism, had an average likeability score of 5.591, while Kayla, the under-

weight woman with hyperthyroidism, had an average likeability score of 5.94. We were able to reject the null hypothesis and conclude that the cause of a woman being underweight affects her likeability. The survey respondents were less likely to become friends with the underweight woman who does not have hyperthyroidism than the underweight woman who does.

RESULT 3: *INDIVIDUALS TEND TO BELIEVE THAT UNDERWEIGHT WOMEN EAT FEWER CALORIES THAN WOMEN AT A HEALTHY WEIGHT DO*

The results of our *z*-test for the proportion in terms of the number of events of interest showed that there was a statistically significant difference (Z-stat = -17.413 and p-value < 0.0001, Z-stat = -13.430 and p-value < 0.0001) between the proportion of respondents who believe the women at a healthy weight consume 2,000 or more calories each day and the proportion who believe the underweight women consume 2,000 or more calories each day at the 5% significance level. 57.78% of respondents believe that Sarah, the woman at a healthy weight without hyperthyroidism, consumes at least 2,000 calories on a daily basis, while only 0.44% of respondents believe Mariah, the underweight woman without hyperthyroidism, consumes the same amount. When asked about the two women with hyperthyroidism, 56.00% of respondents believe the woman at a healthy weight consumes at least 2,000 calories each day, while only 11.56% of respondents answered the same for the underweight woman. We were able to reject the null hypothesis and conclude that individuals tend to believe that underweight women eat fewer calories than women at a healthy weight do. Please refer to Table 2 for a breakdown of the calorie counts selected for all four women.

TABLE 2

DO INDIVIDUALS TEND TO BELIEVE THAT UNDERWEIGHT WOMEN EAT FEWER CALORIES THAN WOMEN AT A HEALTHY WEIGHT DO?

Woman	Weight Type	Medical Condition	< 500 Cal.	500 - 1,000 Cal.	1,001 - 1,500 Cal.	1,501 - 2,000 Cal.	2,001 - 2,500 Cal.	2,501 - 3,000 Cal.	> 3,000 Cal.
Sarah	Healthy Weight	None	0.00%	0.44%	4.44%	37.33%	48.89%	6.67%	2.22%
Taylor	Healthy Weight	Hyperthyroidism	0.00%	0.44%	10.22%	33.33%	29.78%	21.78%	4.44%
Mariah	Under-weight	None	10.67%	37.78%	44.00%	7.11%	0.00%	0.00%	0.44%
Kayla	Under-weight	Hyperthyroidism	12.00%	22.22%	29.33%	24.89%	8.89%	1.33%	1.33%

RESULT 4: *INDIVIDUALS TEND TO BELIEVE THAT UNDERWEIGHT WOMEN EAT LESS FREQUENTLY THAN WOMEN AT A HEALTHY WEIGHT DO*

The results of our z-test for the proportion in terms of the number of events of interest showed that there was a statistically significant difference (Z-stat = 9.056 and p-value < 0.0001, Z-stat = 6.909 and p-value < 0.0001) between the proportion of respondents who believe the women at a healthy weight skip at least one meal per week and the proportion who believe the underweight women skip at least one meal per week at the 5% significance level. 70.67% of respondents believe Sarah, the woman at a healthy weight without hyperthyroidism, skips at least one meal per week, while 98.22% answered the same for the underweight woman without hyperthyroidism. When asked about the two women with hyperthyroidism, 63.11% of respondents believe that the woman at a healthy weight skips at least one meal per week, while 85.33% answered the same for the underweight woman. We were able to reject our null hypothesis and conclude that individuals tend to believe that underweight women eat less frequently than women at a healthy weight do. Please refer to Table 3 for a complete breakdown of the skipped meal counts selected for all four women.

Result 5: *Individuals gave more diet-related health recommendations to the underweight woman without hyperthyroidism than the underweight woman with hyperthyroidism*

The results of our z-test for the proportion in terms of the number of events of interest showed that there was a statistically significant difference (Z-stat = –3.902 and p-value < 0.0001) between the proportion of respondents who gave "diet-related" health recommendations to the underweight women with hyperthyroidism and without hyperthyroidism at the 5% significance level. Diet-related health recommendations included those that mentioned eating more, not skipping meals, eating more protein, etc. 69.33% of respondents gave diet-related health recommendations to Mariah, the underweight woman without hyperthyroidism, while only 57.33% of respondents did the same for Kayla, the underweight woman with hyperthyroidism. We rejected our null hypothesis and concluded that individuals gave more diet-related health recommendations to the underweight woman without hyperthyroidism than the underweight woman with hyperthyroidism. Please refer to Table 4 for a more detailed breakdown of the health recommendations given to both underweight women.

TABLE 3
DO INDIVIDUALS TEND TO BELIEVE THAT UNDERWEIGHT WOMEN EAT LESS FREQUENTLY THAN WOMEN AT A HEALTHY WEIGHT DO?

Woman	Weight Type	Medical Condition	0	1 - 4	5 - 8	9 - 12	13 - 16	17 - 21
Sarah	Healthy Weight	None	29.33%	62.22%	8.00%	0.44%	0.00%	0.00%
Taylor	Healthy Weight	Hyper-thyroidism	36.89%	53.78%	8.00%	1.33%	0.00%	0.00%
Mariah	Under-weight	None	1.78%	19.11%	34.22%	33.33%	8.89%	2.67%
Kayla	Under-weight	Hyper-thyroidism	14.67%	29.78%	28.89%	16.00%	6.67%	4.00%

*Number of meals skipped per week

TABLE 4
DO INDIVIDUALS GIVE MORE DIET-RELATED HEALTH RECOMMENDATIONS TO UNDERWEIGHT WOMEN WITHOUT HYPERTHYROIDISM THAN UNDERWEIGHT WOMEN WITH HYPERTHYROIDISM?

Woman	Weight Type	Medical Condition	A	B	C	D	E	F	G
Mariah	Under-weight	None	69.33%	13.78%	7.56%	11.56%	8.89%	1.33%	0.89%
Kayla	Under-weight	Hyper-thyroidism	57.33%	34.67%	4.44%	6.67%	4.44%	0.89%	0.00%

*Sums of percentages are greater than 100% since some respondents gave more than one health recommendation.

KEY: TABLE 4 CODING BREAKDOWN

Code	Meaning
A	Diet (eat more/do not skip meals)
B	See doctor/take medicine
C	Counseling/eating disorder
D	Exercise
E	Words of encouragement
F	Nothing/keep doing what you're doing
G	Other suggestions

VII. Discussion

These survey results support the claim that a woman's weight, as well as the cause of her weight, affect her likeability. Respondents seemed to blame the underweight woman without hyperthyroidism for becoming underweight; they believe that she consumes less calories and skips more meals – both actions that are under the individual's control – than the women at a healthy weight and the underweight woman with hyperthyroidism. Respondents also gave more diet-related health recommendations to the underweight woman without hyperthyroidism, implying that choosing to eat different foods will help her reach and maintain a healthy weight. This study suggests that underweight individuals are stigmatized, and those without an "excuse" for their weight, such as hyperthyroidism, are looked upon less favorably than those with a known medical condition. Previous studies have had similar findings; Hoyt et al. (2017) found that subjects were less likely to blame obese individuals for their body size when their weight was the result of a "disease" rather than individual choices.

These results have important implications for attempting to reduce this weight stigma going forward. The stigma surrounding underweight women seems to stem from individuals assuming that the women are underweight because of their individual choices, such as choosing to consume less calories and skip meals. However, when individuals recognize that a woman's weight may be out of her control, this stigma lessens. Educating the public about the many different factors that can cause a woman to be underweight – including cancer, hyperthyroidism, eating disorders, and certain medications – may help to reduce the stigma and lessen the blame attributed to these underweight individuals. In doing so, underweight women may find it easier to form and maintain strong friendships.

Limitations & Opportunities for Future Research

As mentioned above, the link to the survey was posted in three university Facebook pages and sent through GroupMe to 72 students who reside in two on-campus residence halls. It is important to note that some undergraduate students at this university were not presented with the opportunity to take this survey. All students who do not have a Facebook or GroupMe account, or did not visit either of these two sites between March 24, 2017 and March 31, 2017, did not receive the link. Therefore, the results of this study cannot be generalized to the entire undergraduate population at this university. Researchers looking to replicate our study at a different university should use a random sample of undergraduate students so the results can be generalized.

It is important to note that all participants saw all four pictures in the same order, which could have caused an order effect. To avoid this, surveys completed in the future should randomize the order in which participants see each picture.

Another limitation of this study is that the individuals pictured in our survey were only women. In the future, researchers can create a similar survey that features pictures

of men to examine whether or not underweight men experience the stigma of being underweight to the same extent women do.

In addition, all of the women had their faces showing, so their perceived attractiveness could have influenced respondents' answers, especially for likeability. Future studies should use consistent pictures, without faces, with the sole difference between them being the shape of the bodies and the associated medical conditions.

Also, we purposely made our survey relatively short in an attempt to increase our response rate. However, should researchers conduct a similar survey in the future, additional questions can be asked. For example, respondents can be asked to rate the sexual attractiveness or the perceived intelligence of the individuals in the photos. The final limitation of this study is that we only examined two of the possible causes of the stigma surrounding underweight individuals: anti-thin prejudice and attribution of blame. Future research can focus on identifying and investigating other possible causes of this stigma, such as jealousy.

Practical Application

In December 2015, France adopted a law that bans underweight models (BMI < 18), and any photos of women that are photoshopped to make them appear thinner must be labeled as a "retouched photo" (Zarya, 2015). Israel, Italy, and Spain all have similar laws. Future research can be done to examine the impact of these policies. These bans, which have been said to be a form of weight discrimination, could have actually increased the stigma surrounding underweight women, which will especially hurt the underweight women who are thin due to factors out of their control. It will be interesting to note whether or not there has been a decrease in the number of models suffering from eating disorders; it is possible that the models who previously suffered from anorexia switched to binge eating to be eligible to model again. Knowing the impact these bans have had may help the United States make policy decisions in the future.

References

Anderson, J., & Bresnahan, M. (2013). "Communicating stigma about body size." *Health Communication*, 28(6), 603-615.

Andreyeva, T., Puhl, R. M., & Brownell, K. D. (2008). "Changes in perceived weight discrimination among Americans, 1995–1996 through 2004–2006." Obesity, 16(5), 1129-1134.

Bahn, R., Burch, H., Cooper, D., Garber, J., Greenlee, M., Klein, I., & Ross, D. (2011). "Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists." *Endocrine Practice*, 17(3), 456-520.

Boyes, A. D., & Latner, J. D. (2009). "Weight stigma in existing romantic relationships." *Journal of Sex & Marital Therapy*, 35(4), 282-293.

Carr, D., & Friedman, M. A. (2005). "Is Obesity Stigmatizing? Body Weight, Perceived Discrimination, and Psychological Well-Being in the United States." *Journal of Health and Social Behavior*, 46(3), 244-259.

DeJong, W. (1980). "The stigma of obesity: The consequences of naive assumptions concerning the causes of physical deviance." *Journal of Health and Social Behavior*, 21(1), 75-87.

Drury, A., Aramburu, C., & Louis, M. (2002). "Exploring the association between body weight, stigma of obesity, and health care avoidance." *Journal of the American Academy of Nurse Practitioners*, 14(12), 554-562.

Hackman, J., Maupin, J., & Brewis, A. A. (2016). "Weight-related stigma is a significant psychosocial stressor in developing countries: Evidence from Guatemala." *Social Science & Medicine*, 161, 55-60.

Harmatz, M. G., Gronendyke, J., & Thomas, T. (1985). "The underweight male: The unrecognized problem group of body image research." *Journal of Obesity & Weight Regulation*, 4(4), 258-267.

Hoyt, C. L., Burnette, J. L., Auster-Gussman, L., Blodorn, A., & Major, B. (2017). "The obesity stigma asymmetry model: The indirect and divergent effects of blame and changeability beliefs on antifat prejudice." *Stigma and Health*, 2(1), 53-65.

O'Dea, J. A., & Amy, N. K. (2011). "Perceived and desired weight, weight related eating and exercising behaviours, and advice received from parents among thin, overweight, obese or normal weight Australian children and adolescents." *International Journal of Behavioral Nutrition and Physical Activity*, 8(1), 68.

Puhl, R. M., Andreyeva, T., & Brownell, K. D. (2008). "Perceptions of weight discrimination: Prevalence and comparison to race and gender discrimination in America." *International Journal of Obesity,* 32(6), 992-1000.

Saguy, A. C., & Gruys, K. (2010). "Morality and health: News media constructions of overweight and eating disorders." *Social Problems*, 57(2), 231-250.

Schafer, M. H., & Ferraro, K. F. (2011). "The stigma of obesity: Does perceived weight discrimination affect identity and physical health?" *Social Psychology Quarterly*, 74(1), 76-97.

Swami, V., Chan, F., Wong, V., Furnham, A., & Tovée, M. J. (2008). "Weight-Based Discrimination in Occupational Hiring and Helping Behavior." *Journal of Applied Social Psychology*, 38(4), 968-981.

Swami, V., & Monk, R. (2013). "Weight bias against women in a university acceptance scenario." *The Journal of General Psychology*, 140(1), 45-56.

Swami, V., Pietschnig, J., Stieger, S., Tovee, M. J., & Voracek, M. (2010). "An investigation of weight bias against women and its association with individual difference factors." *Body Image*, 7(3), 194-199.

Zarya, V. (2015, Dec. 18)." It's now illegal to hire ultra skinny models in France." *Fortune*. Retrieved from http://fortune.com/2015/12/18/france-model-law/

Appendix – Survey

Body Types Survey

* Required

Informed Consent

You must read and consent to the information below before beginning the survey

The purpose of this research project is to examine college students' first impressions of women. This is a research project being conducted by Jennifer Miller and Brittani Murphy for a class project.

You are invited to participate in this research project because you are an undergraduate student. Your participation in this research study is voluntary.

The procedure involves completing an online survey. It will take approximately five minutes to complete. Your responses will be confidential and anonymous. No identifying information is collected. Findings are reported only for a class paper and aggregate findings might be published in a journal.

If you have any questions about the study please contact Brittani Murphy at Brittani.Murphy24@aol.com.

ELECTRONIC CONSENT:
Clicking on the "I consent" button below indicates that:
- You have read the above information.
- You voluntarily agree to participate.
- You are at least 18 years of age.

1. *

Mark only one oval.

- I consent

Untitled Section

2. **Which gender do you most closely identify with?** *

Mark only one oval.

- Male
- Female
- Other
- Prefer not to say

Sarah

Please use the picture and description below to answer the following questions to the best of your ability.

Sarah is at a healthy weight according to her doctor.

Picture taken from https://s-media-cache-ak0.pinimg.com/736x/ef/cb/92/efcb9245012326c0ac82d4e549e6a49c.jpg

3. **Based on looks alone, do you think you would be friends with her?** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
I definitely would not be friends with Sarah	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	I definitely would be friends with Sarah

4. **On average, how many calories do you think Sarah consumes each day? (2000 calories is standard for daily caloric intake)** *

Mark only one oval.

- ◯ Less than 500 calories
- ◯ 500 - 1000 calories
- ◯ 1001 - 1500 calories
- ◯ 1501 - 2000 calories
- ◯ 2001 - 2500 calories
- ◯ 2501 - 3000 calories
- ◯ 3001 or more calories

5. On average, how many meals do you think Sarah skips each week? (assuming 3 meals a day) *

Mark only one oval.

- 0
- 1 - 4
- 5 - 8
- 9 - 12
- 13 - 16
- 17 - 21

Taylor

Please use the picture and description below to answer the following questions to the best of your ability.

Taylor is at a healthy weight according to her doctor. However, she has hyperthyroidism, which can increase metabolism.

Picture taken from http://www.dhoste.com/product/unique-vintage-charlotte-red-black-white/229418853.html

6. **Based on looks alone, do you think you would be friends with her?** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
I definitely would not be friends with Taylor	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	I definitely would be friends with Taylor

7. **On average, how many calories do you think Taylor consumes each day? (2000 calories is standard for daily caloric intake)** *

Mark only one oval.

- ◯ Less than 500 calories
- ◯ 500 - 1000 calories
- ◯ 1001 - 1500 calories
- ◯ 1501 - 2000 calories
- ◯ 2001 - 2500 calories
- ◯ 2501 - 3000 calories
- ◯ 3001 or more calories

8. **On average, how many meals do you think Taylor skips each week? (assuming 3 meals a day)** *

Mark only one oval.

- ◯ 0
- ◯ 1 - 4
- ◯ 5 - 8
- ◯ 9 - 12
- ◯ 13 - 16
- ◯ 17 - 21

Mariah

Please use the picture and description below to answer the following questions to the best of your ability.

Mariah is underweight according to her doctor.

Picture taken from http://www.thefashionpolice.net/2011/11/is-it-offensive-drop-dead-clothings-anorexic-model.html

9. Based on looks alone, do you think you would be friends with her? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
I definitely would not be friends with Mariah	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	I definitely would be friends with Mariah

10. On average, how many calories do you think Mariah consumes each day? (2000 calories is standard for daily caloric intake) *

Mark only one oval.

- ◯ Less than 500 calories
- ◯ 500 - 1000 calories
- ◯ 1001 - 1500 calories
- ◯ 1501 - 2000 calories
- ◯ 2001 - 2500 calories
- ◯ 2501 - 3000 calories
- ◯ 3001 or more calories

11. On average, how many meals do you think Mariah skips each week? (assuming 3 meals a day) *

Mark only one oval.

- 0
- 1 - 4
- 5 - 8
- 9 - 12
- 13 - 16
- 17 - 21

12. Please give Mariah one piece of advice to improve her health. *

Kayla

Please use the picture and description below to answer the following questions to the best of your ability.

Kayla is underweight according to her doctor. She has hyperthyroidism, which can increase metabolism.

Picture taken from https://www.thesun.co.uk/living/1331565/anorexia-survivor-overcomes-crippling-eating-disorder-after-instagram-photo-of-healthy-woman-saves-her-life/

13. **Based on looks alone, do you think you would be friends with her?** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
I definitely would not be friends with Kayla	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	I definitely would be friends with Kayla

14. **On average, how many calories do you think Kayla consumes each day? (2000 calories is standard for daily caloric intake)** *

Mark only one oval.

- ◯ Less than 500 calories
- ◯ 500 - 1000 calories
- ◯ 1001 - 1500 calories
- ◯ 1501 - 2000 calories
- ◯ 2001 - 2500 calories
- ◯ 2501 - 3000 calories
- ◯ 3001 or more calories

15. **On average, how many meals do you think Kayla skips each week? (assuming 3 meals a day)** *

Mark only one oval.

- ◯ 0
- ◯ 1 - 4
- ◯ 5 - 8
- ◯ 9 - 12
- ◯ 13 - 16
- ◯ 17 - 21

16. **Please give Kayla one piece of advice to improve her health.** *

_____

Perceptions of E-Cigarettes Among College Students

By Haley Morgan*

E-cigarettes are a relatively new technology and much uncertainty remains about their impact on human health. While studies are forthcoming, few address college students and their perceptions of e-cigarettes. I used survey methodology to gauge undergraduate students' perceptions of e-cigarettes, regarding both users and bystanders. The survey asked a sample of undergraduate students at Bentley University a variety of questions and was designed to serve as input for data analysis of the findings. It was successfully completed by approximately 15% of students. The survey addressed the use of both e-cigarettes and traditional cigarettes.

The survey revealed that students largely felt traditional cigarettes were harmful to both smokers and non-smokers. Participants seemed more uncertain about the potential effects of e-cigarettes, yet most expressed the belief that the devices do impact the health of users. No trend in agreement was evident regarding the impact of e-cigarettes on bystanders.

This study has several implications. It suggests that students may be unaware or uninformed of the potential effects e-cigarettes can have, particularly for bystanders. As younger people seem to be the target market for e-cigarette companies, this may become concerning. Consumers cannot make educated decisions if the effects of the devices are unclear.

Keywords: E-Cigarettes, perceptions, technology, health.

I. Introduction

Cigarette smoking has been a large part of society's culture for many years. Studies in recent decades have exposed some major health concerns associated with this activity, yet many smokers are not deterred. There is a continued demand for cigarettes in society today. However, some consumers have taken the cigarette warnings into account and have looked for ways to limit the negative impact traditional smoking can have on their health. Innovation in the industry has led to the production of electronic cigarettes, or e-cigarettes, with which users inhale vapor instead of smoke. In recent years, the number of smokers of traditional cigarettes has decreased while the number of e-cigarette users has increased substantially (Abadi, Couch, Chaffee & Walsh, 2017).

---

*Email: haley.morgan17@gmail.com. I would like to thank my advisor, Professor Johannes Eijmberts, for his support and guidance throughout the capstone process. His insights were incredibly valuable and appreciated.

Many of the concerns associated with traditional cigarettes center around smoke inhalation. Smoke in human lungs has been linked to cancer and other diseases (National Institute on Drug Abuse [NIDA], 2016). E-cigarettes take this factor out of the equation by creating vapor instead of smoke, but are they actually risk-free?

Health Effects of Traditional Cigarettes

Traditional cigarette smoking contributes to the deaths of more than 480,000 Americans each year (Centers for Disease Control and Prevention [CDC], 2016b). This is no small number, yet as of 2015 approximately 36.5 million adults in the United States were still considered current smokers (CDC, 2016b). The implications of this practice are concerning. Not only is the health of active smokers at risk, but the lives of non-smokers, too. Of the 480,000 Americans that die each year from smoking-related causes, more than 41,000 die due to secondhand smoke exposure (CDC, 2016b). Secondhand smoke refers to the smoke produced by burning tobacco products as well as the smoke exhaled by the person using such products (CDC, 2017). This is a major issue, as even people who choose not to participate in the smoking of traditional cigarettes can still be negatively affected. As a way of acknowledging this issue, people have started to look for alternatives that may be safer for all.

Smoking can harm virtually every part of the human body. It affects not only the mouth and lungs, but also the brain, the heart and the digestive system (CDC, 2016a). The combustion of tobacco and other carcinogens included in traditional cigarettes has many negative health effects. Tobacco use puts people at a higher risk for certain autoimmune diseases, diabetes, osteoporosis, cardiovascular disease, coronary heart disease, and stroke, among other things (U.S. Department of Health and Human Services [HHS], n.d.-a). The act of lighting a cigarette serves to burn the tobacco, thus producing tar and other chemicals that largely contribute to these health issues (NIDA, 2016).

Alternatives to Traditional Cigarettes

Researchers and industry specialists have looked for safer alternatives to the chemicals included in traditional cigarettes. E-cigarettes have emerged as an option which eliminates the burning of tobacco in the smoking process. Instead, these devices heat flavored liquid containing nicotine and other chemicals (NIDA, 2016).

An article published by Medical Daily discusses the level of safety of e-cigarettes as compared to traditional cigarettes (Bushak, 2013). The article raises many interesting points, maintaining that while e-cigarettes are likely safer in terms of the lack of smoke exposure they can still cause significant damage (Bushak, 2013). More so than other chemicals included in traditional cigarettes, nicotine may be responsible for damage to arteries in the body. Nicotine is one of the main components found in e-cigarettes, leading some researchers to express concern about their relative safety (Bushak, 2013).

Historically, scientists have looked at other chemicals and carcinogens in regular cigarettes as the possible causes of negative health effects on the human body (Bushak,

2013). Based on this train of thought and research, e-cigarettes appear to be a safer alternative. However, research has shown that e-cigarettes do have health consequences of their own, illustrating the need for further study on the matter. Much research still has to be done, especially concerning the health effects of long-term e-cigarette use. The lack of definitive scientific conclusions is concerning as e-cigarettes are being marketed to consumers as a safe option, yet the research to either confirm or deny this claim has not been thoroughly completed.

The points brought forth by Bushak (2013) are largely validated by scientific research, but it is important to note that the topic has also gained the attention of various media sources and trade journals. These sources draw from scientific studies, but present the information in a way designed to target a different group of readers than those of purely scientific articles. As more and more people are taking note of e-cigarettes and their effects, it is important to address not only scientific publications but also sources that may pique the interest of a wider group of readers.

Possible Implications of E-Cigarettes

E-cigarettes are a relatively new technology, but stand to gain an even larger consumer base through advertising and other sources of media. Marketing these devices as safe alternatives to traditional cigarettes may actually reveal an untapped market of consumers who would not consider using traditional cigarettes but who may think they have nothing to lose by trying e-cigarettes. Many people have either never heard of the devices or know very little about how they work. This unfamiliarity together with product marketing could lead consumers to make misinformed decisions that could ultimately have serious negative impacts on their health. In this same sense, e-cigarettes could serve as a gateway to traditional cigarettes, or even other drugs, for younger consumers.

Careful analysis of data could show which age groups are the most frequent users of e-cigarettes compared to traditional cigarettes. Further analysis of the results would address questions such as: Why is a certain age group more likely to use e-cigarettes? What factors influence a person's decision to choose e-cigarettes as opposed to regular cigarettes? How many people who first used e-cigarettes have now switched over to using traditional cigarettes? Such information could be helpful to researchers in gaining an understanding of just how widespread the use of e-cigarettes may be, as well as an understanding of the types of people who may be affected by the technology.

The effect of e-cigarettes on human health is largely unknown. Significantly more research must be done to isolate the causes of certain health issues that could arise from this technology before it can be marketed as a safe alternative to traditional cigarette smoking. Additionally, further research should be conducted to analyze the both the addiction risk of e-cigarettes as well as the risk they pose to others.

Introduction to the Study

This study explores people's perceptions of e-cigarettes and their effects on human health. Many people have heard through either media sources or word of mouth that these devices are safer to use than regular cigarettes. Perceptions are a key component to understanding this practice and its potential effects on human health. Thus, in the research it is important to consider not only the tangible health effects that can result from the use of e-cigarettes but also the perceptions that surround the activity.

The two-part research question to be addressed is as follows: Are e-cigarettes less detrimental to the health of both smokers and non-smokers than traditional cigarettes? Are e-cigarette users and people around them aware of the potential health effects these devices can have?

Due to substantial research in recent years, most people are aware of the effects that traditional cigarettes can have on their health, whether or not they are smoking firsthand. E-cigarettes have been subject to less research, likely leaving consumers less aware of the potential health effects such devices can have. It seems reasonable that without this knowledge, consumers would be more likely to assume that e-cigarettes are a safer alternative to regular cigarettes, based largely on the lack of a tobacco burning process. Research will have to be done to come to a better understanding of people's perceptions on the matter.

This research will be done through a combination of different methods. First, a thorough examination of existing literature will be completed to provide a basis for the argument. Once completed, there should be enough of a comprehensive understanding of the topic to proceed with original research.

Data collection will come in the form of a survey delivered through an online medium called Qualtrics. The survey will focus on the perceptions of students at Bentley University with regard to smoking. Survey questions about both e-cigarettes and traditional cigarettes will be included in an attempt to gauge perceptions about the practice. Additionally, there will be questions that address Bentley University's "smoke-free campus" policy to see whether that has affected perceptions of smoking in any way.

The focus on perception, specifically that of students, will be the factor that makes this research noteworthy. Much of the research that has been completed to date focuses on other aspects of the practice, such as its relative safety and health effects. While these topics will also be touched upon in this paper, the main focus will be on perception of e-cigarettes and their effects.

As with any research, there are issues that could arise. The main problems that could be faced in the course of this research project relate to the amount of existing research on the topic. E-cigarettes are a relatively recent technology and have yet to undergo much of the testing needed to form definitive conclusions regarding their toxicity and potential effects they may have on both the human body and the environment around the user.

Although it seems unlikely that the survey will garner a lack of response, this too could be a potential issue that would need to be addressed. It is believed that a substantial number of students will respond to the survey, especially if it is simple to complete. The survey will be thorough in nature but will also be something that the students find interesting.

Definition of Terms and Concepts

In order to avoid comprehension problems, one aspect of the project that will need special attention centers around defining the research concepts. Some of the terms used in this research and analysis will need to be explained so that participants are well aware of what they are being asked. An understanding of these concepts will allow readers to better grasp the main points of the research and to put the knowledge to good use.

There are three main concepts found in this analysis: risk, health effects and perception. Each will be further defined and explained in the following section, but they are all necessary to the research in their own ways.

Risk is one of the main concepts that will be addressed in the research. Risk can be defined as the likelihood that smoking will affect either the user or people around him or her (CDC, 2016a). Risk is important in talking about the impact that smoking has on people's lives, and should largely be kept to a minimum. In this research, risk will be used to refer to the ideas of risk to the smoker, risk to others, and risk of addiction.

Health effects are another concept addressed in the work. They can really be defined as any health problem that occurs as a result of exposure to cigarettes and e-cigarettes. Individual health effects vary in severity and location within the body, but all have a negative impact on the body's health as a whole.

Perception is a very important concept in this analysis. Much of the research done for this analysis is focused on the perception of students at Bentley University with regard to e-cigarettes. Perception can be defined as the way an individual views and item or a concept. Individuals often perceive things differently than their peers, so it is useful to examine a larger sample size.

Risk is a relatively tough concept to measure, especially as it is used in three different contexts. The term risk could be used to describe risk to the smoker, risk to others and risk of addiction. These could likely be measured through data analysis and statistics. Knowledge of the impact the practice of smoking has on both smokers and non-smokers could prove beneficial in measuring this concept. Risk of addiction is something that can likely be measured through knowledge of how much nicotine is ingested and statistics from surveys and datasets where people either report addiction or report smoking every day.

Health effects would be defined in various ways depending on what part of the human body is affected. For instance, health effects of smoking on the lungs would be measured differently than health effects of smoking on the brain. In the simplest of

terms, health effects are measured by damage to the body. Damage is something that can be seen through medical testing and can often be portrayed in data sets. This damage would cause parts of the human body to stop functioning in the correct way or at all.

Perception will be measured through the survey portion of the research. Questions will be asked to gauge how people feel about e-cigarettes and whether or not they think the devices are actually safer than traditional cigarettes.

Initial Hypotheses

It appears reasonable to assume that e-cigarettes are significantly less damaging to the health of both smokers and non-smokers than traditional cigarettes. It seems likely that if not exposed to the burning of tobacco in regular cigarettes, lungs would be less damaged by tar formation and damage from associated chemicals. With regard to non-smokers, secondhand smoke would no longer be an issue. Smoke is not released from the burning interaction, and thus is not a factor that could affect the health of non-smokers.

Perception of e-cigarettes is expected to be in accordance with these hypotheses. The absence of smoke leads users to believe that e-cigarettes are much less harmful and damaging to the smoker and are not harmful for those around the user. The perceived risk will likely be lower with regard to all three areas previously discussed.

II. Literature Review

Health Effects of E-Cigarettes

It is common knowledge that the smoking of traditional cigarettes has negative health effects on the human body (CDC, 2016a). The combustion of tobacco produces smoke that impacts not only the smoker, but people around him as well. When discussing traditional cigarettes it is important to address potential health effects to both of these groups of people.

As times have changed, technology has continuously improved and has provided consumers with new devices and methods of inhaling tobacco products. Use of electronic cigarettes in particular is on the rise, creating some concern about the practice's potential health effects. Abadi et al. (2017) noted that not only has the availability of e-cigarettes risen in recent years, but affordability has also improved. Trumbo and Harper (2013) agreed that the popularity of e-cigarettes has increased significantly since they were first brought to market. Additionally, Ayers, Ribisl and Brownstein found that internet searches related to e-cigarettes have skyrocketed, surpassing the number of searches for other alternative smoking devices in several of the world's largest countries (as cited in Trumbo & Harper, 2013).

E-cigarettes have not yet undergone thorough testing as to their effects on human health. In their scientific review of e-cigarettes, Grana, Benowitz and Glantz (2014) argued that many questions still exist about their safety for both smokers and non-smokers. An article published by the US Department of Health and Human Services also highlighted the many unknowns associated with these electronic devices. Without the proper research and testing, there is really no way to know if e-cigarettes are safe or not (HHS, n.d.-b).

Smith, Brar, Srinivasan, Enja and Lippmann (2016) studied the composition of the vapor being inhaled by e-cigarette users and found that this practice exposes people to lead, cadmium and nickel, among other things. Smith et al. (2016) found that these metals in particular are related to various health issues such as neurologic damage, organ failure and an inflammatory pulmonary reaction. Additionally, the vapor was found to contain formaldehyde, acrolein and hydrocarbons, all three of which are known to be toxic to the human body. Hiemstra and Bals (2016) also touched on the topic, confirming that when heated at high temperatures through use of high voltage, e-cigarettes can produce formaldehyde.

A recent study conducted by professionals at the Johns Hopkins Bloomberg School of Public Health confirmed the presence of similar metals in e-cigarette liquid: The study by Hess et al. (2017) differed from that of Smith et al. (2016) as it looked only at the liquid that forms the e-cigarette vapor, not the vapor itself. Hess et al. (2017) found cadmium, chromium, lead, manganese and nickel in the liquids of various brands of e-cigarettes. These metals are toxic to the human body and have been found to be carcinogenic. While e-cigarette users do not ingest the liquid directly, researchers worry that the metals it contains could be inhaled by users when converted into vapor (Hess et al., 2017).

Users of e-cigarettes are not the only people with the potential to be affected by the practice. With regard to secondhand exposure, Grana et al. (2014) found that e-cigarettes still pose a risk to bystanders even though they do not inhale secondhand smoke. Exhalation by a person using e-cigarettes can expose bystanders to a mixture of vapor and other dangerous chemicals. On another note, the study found that secondhand vapor only exposes non-smokers to approximately 1/10th the amount of nicotine that they would inhale via traditional cigarette secondhand smoke. Czogala et al. (2013) supported this claim through their own research, stating that nicotine is 10 times more concentrated in exhaled smoke from traditional cigarettes than in exhaled vapor from e-cigarettes. However, these two studies differ as Czogala et al. (2013) specifically tested e-cigarette vapor for combustion toxins, ultimately maintaining that the devices are not a source of secondhand exposure to such particles.

Smith et al. (2016) also discussed the issue of secondhand exposure but focused more on the impact of chemicals and other toxic particles. They argued that bystanders are very much exposed to chemicals and toxins produced through use of e-cigarettes. Smith et al. (2016) noted that formaldehyde, acetone, propanol and propylene glycol, among other substances, were found in the air after vapor had been exhaled. However,

they noted that their studies remained inconclusive as to whether these particles affected the body in the same way as traditional cigarette smoke.

Some of these chemicals were also analyzed by Wang et al. (2017) in a study examining how propylene glycol and glycerol react to heat. These substances are two of the more prominent solvents found in e-cigarettes and were discovered to be major sources of toxic compounds. Wang et al. (2017) noted the presence of formaldehyde and acetaldehyde with the heating of each solvent but found that glycerol produced substantially greater amounts of formaldehyde than did propylene glycol. The levels of these toxins that were found in the study exceeded the amounts deemed acceptable by the Environmental Protection Agency (EPA), indicating that further study is warranted to determine potential health risks.

In a review of available data on e-cigarettes, Callahan-Lyon (2014) wrote of the known health effects of these specific compounds. Propylene glycol and glycerol have been known to cause irritation of the mouth and throat as well as a dry cough. Callahan-Lyon (2014) noted that not enough information existed to come to definitive conclusions about neither short nor long term effects of these chemicals on the human body. This is an opportunity for important future research as use of e-cigarettes becomes more widespread.

Rouabhia et al. (2017) also found e-cigarette vapor to have negative health effects on cells in the mouth. The vapor actually altered the shape and composition of the cells and increased levels of L-lactate dehydrogenase (LDH). Additionally, the vapor was found to have an apoptotic effect on gingival epithelial cells. In simpler terms, apoptosis is a process in the body that equates to programmed cell death. Chemicals in the e-cigarette vapor contributed to the death of many epithelial cells in the human mouth. Hiemstra and Bals (2016) also observed negative effects on cultured cells, but looked into effects on live animals as well. Such effects included inflammatory responses, changes in behavior and the suppression of the pulmonary host defense (Hiemstra & Bals, 2016).

Shahab et al. (2017) did not focus on potential negative health effects as did other researchers, but instead sought to compare the health effects resulting from use of e-cigarettes and traditional cigarettes. They found evidence supporting the idea that in the long run, use of e-cigarettes as opposed to traditional cigarettes can lead to lower levels of known carcinogens and toxins in the body. However, levels of nicotine did not vary between long term users of e-cigarettes and traditional cigarettes. Shahab et al. (2017) defined long term as a period greater than or equal to six months, so more research would have to be done to address health effects over longer periods of time. Such research will ultimately be critical in determining the relative safety of e-cigarettes.

Largely due to the uncertainty surrounding the health effects of e-cigarettes, many people have made arguments concerning the health and regulatory policies related to the devices. Meernik, Baker, Lee and Goldstein (2017) made the argument that additional policies need to be implemented due to the increased prevalence of e-cigarette

use. However, the policies would have to be carefully considered in order to be effective in regulating the use of such electronic devices. Grana et al. (2014) touched on the policy aspect more in regard to smoke-free laws. They maintained that e-cigarettes are marketed as solution for getting around such laws and that they can be used anywhere since they do not involve actual smoke. The argument was made that e-cigarettes are still tobacco products although not combustible ones. Grana et al. (2014) supported the introduction of the same marketing restrictions for e-cigarettes that traditional tobacco products face.

In terms of regulation, the Food and Drug Administration (FDA) has decided to include e-cigarettes in their jurisdiction in the following way:

> In 2016, FDA finalized a rule extending our regulatory authority to cover all tobacco products, including vaporizers, vape pens, hookah pens, electronic cigarettes (E-Cigarettes), e-pipes, and all other ENDS. FDA now regulates the manufacture, import, packaging, labeling, advertising, promotion, sale, and distribution of ENDS. (U.S. Food and Drug Administration [FDA], 2017)

Although e-cigarettes do not involve the combustion of tobacco they do still contain nicotine in varying amounts. Nicotine is derived from tobacco plants and is thus included in the definition of a tobacco product. Many people seem to be under the impression that e-cigarettes do not have any risk attached, but in reality all tobacco products are associated with some level of health and addiction risk. Hiemstra and Bals (2016) discussed the effects of nicotine, stating that the substance is addictive and can contribute to cancer growth. However, the study also showed that nicotine is not the only factor contributing to the toxicity of e-cigarettes (Hiemstra & Bals, 2016). Smith et al. (2016) noted that e-cigarette packaging will now be required to list all ingredients, which will allow people to gain a better understanding of the contents of the vapor. As of now, it is hard to know what the e-cigarette liquid contains and thus people are left largely unaware of what they are ingesting.

This leads into a separate discussion about perception and beliefs regarding potential harm and risk. The perception of e-cigarettes needs to be addressed as use of the devices becomes more widespread. People are making uninformed decisions based largely on inaccurate or misleading marketing. Marketing campaigns target highly varied groups of people in an effort to promote e-cigarettes to a wider audience and expand their use (Hiemstra & Bals, 2016).

In addition to the influence of marketing, the lack of research concerning long-term consequences of e-cigarette use may also lead to uninformed consumer decisions. The technology is relatively new and has not been around long enough for many lengthy studies to have taken place. Potential health effects need to be assessed over a longer period of time. Researchers largely agree that further scientific testing is necessary to isolate the specific health effects that can be caused by e-cigarettes.

Perceptions of E-Cigarettes

E-cigarettes do not involve the combustion of tobacco and do not produce smoke as do traditional cigarettes. Thus, many people are led to believe that the risk of using such devices is slim to none, especially compared to the significant risk of using traditional cigarettes. Risk can be thought of as both health risk and addiction risk, and can be applied to both smokers and non-smoking bystanders.

This belief seems to be prominent among younger populations, as the reported use of e-cigarettes among middle school and high school students has increased considerably in recent years (Simon, 2016). However, different studies have looked at e-cigarette use among various age groups, thus indicating that the trend is affecting older people as well.

Ahern and Mechling (2014) studied the trend of e-cigarette use among adolescents. They found that use of the devices is rapidly increasing and voiced concerns about possible implications. The study noted that the way in which e-cigarettes are marketed may be contributing to favorable perceptions of the devices among users. Carr (2014) noted that e-cigarettes are marketed as healthy alternatives to traditional cigarettes. The industry seems to be trying to capitalize on this position by pushing this type of marketing message.

Smokers seem to believe that e-cigarettes are safer to use than traditional cigarettes. Pepper, Emery, Ribisl, Rini and Brewer (2015) found that smokers of many different demographics felt e-cigarettes were less likely to cause cancer and other diseases than traditional cigarettes. The study focused on smokers' perceptions of e-cigarettes contributing to lung cancer, oral cancer and heart disease, all of which have been found to be associated with smoking traditional cigarettes. Pepper et al. (2015) noted that this favorable perception of e-cigarettes as compared to traditional cigarettes may be contributing to the rapid increase in reported use of the devices.

Berg et al. (2015) agreed with this idea based on their own findings. E-cigarettes were found to be more positively perceived among participants than traditional cigarettes. Participants in the study believed that e-cigarettes represented a lower overall harm to health than did regular cigarettes, and also perceived the associated risk of addiction to be lower. This more positive perception of e-cigarettes may stem from marketing efforts depicting the products as safe alternatives to traditional cigarettes (Berg et al., 2015).

Choi, Fabian, Mottey, Corbett and Forster (2012) also found perceptions of e-cigarettes to be highly favorable, largely because they came in different flavors. Participants reported willingness to experiment with the products in general as well as with different flavors of the same product. Abadi et al. (2017) noted that flavors such as fruit, mint and candy were preferred when selecting an e-cigarette liquid for use. Carr (2014) also discussed the variety of flavors and noted that e-cigarettes come in different colors as well. These marketing strategies are thought to be directed toward younger consumers.

Choi et al. (2012) noted that this willingness and intent to try the devices is of concern as consumers may use e-cigarettes as a gateway to traditional cigarettes. Favorable perceptions of e-cigarettes may contribute to their increased use, and therefore expose users to the dangers of nicotine. Addiction here may lead users to eventually become dual users of both e-cigarettes and regular cigarettes, or even switch entirely to use of traditional cigarettes. Hiemstra and Bals (2016) expressed concern that use of e-cigarettes by young individuals who do not smoke traditional cigarettes may foster a dependence on nicotine. McMillen, Maduka and Winickoff were in agreement with this notion, arguing that the potential of both non-smokers becoming nicotine-dependent and existing smokers furthering their nicotine dependence is highly troubling (as cited in Trumbo & Harper, 2013).

Martinez-Sanchez et al. (2015) took a slightly different approach and looked at perceptions of the general population with regard to the effects of e-cigarettes both on users and on bystanders. Approximately 40% of those studied thought e-cigarettes were harmful to users, but only about 27% thought that e-cigarettes were harmful to bystanders. Martinez-Sanchez et al. (2015) found that more people perceive e-cigarettes to be useful than harmful. Respondents thought that e-cigarettes were actually helpful for reducing tobacco use. Carr (2014) agreed with this finding and noted that e-cigarettes are widely perceived as tools for smoking cessation.

In terms of social acceptability, Trumbo and Harper (2013) found that, in general, people tend to think use of e-cigarettes in public is more acceptable than use of traditional tobacco products. Abadi et al. (2017) were in agreement with this claim, indicating that the majority of respondents in their study shared the same beliefs.

Much of the existing literature on perceptions of e-cigarettes looks at the effect on users but not on bystanders. This is something that should be addressed as there is a potential health concern for bystanders. People may be either misinformed about the danger their smoking activities pose to others or may even be uninformed entirely.

Information and the way it is delivered is a key component in influencing consumers' perceptions. Mays, Smith, Johnson, Tercyak and Niaura (2016) found that people exposed solely to warnings about e-cigarettes were more likely to view the devices as harmful and addictive. The warning seemed to deter people from wanting to try the products. When an advertisement was included, Mays et al. (2016) found that people paid little to no attention to the warning and felt that e-cigarettes were not significantly harmful or addictive.

Sanders-Jackson, Schleicher, Fortmann and Henriksen (2015) obtained similar results through their study, finding that warnings about e-cigarettes lowered smokers' cravings for the devices and also lowered both smokers' and non-smokers' intentions to buy the products. Industry-themed warnings seemed to generate more perceptions of e-cigarettes as harmful devices than did ingredient-themed warnings (Sanders-Jackson et al., 2015).

As the use of e-cigarettes has grown, researchers have become increasingly interested in learning about the devices and how they fit into society. Health concerns regarding the technology have continued to surface in recent years, yet e-cigarette use remains high. Societal perceptions of e-cigarettes seem to vary based on a lot of different factors. Abadi et al. (2017) noted that there are both social and physical perceived benefits and risks associated with the use of e-cigarettes. Studies have tried to gain a general understanding about how the technology is perceived, yet have found responses to be different based on age, smoking status and other demographic qualities.

The goal of this research is to examine college students' perceptions of e-cigarettes. Participants will be pursuing a degree in higher education and will differ in ethnicity and gender. The research is being conducted after the FDA's announcement to regulate e-cigarettes as tobacco products, and will take place on a smoke-free campus. Participants will be asked a series of questions to try to gain a better understanding of how students perceive these devices with regard to the health of both smokers and non-smokers.

III. Analysis

Methodology

To answer the research questions at hand, original data needed to be collected and analyzed. The goal of the research was to gauge students' perceptions of e-cigarettes and their use. Specifically, research was gathered to better understand whether students felt there were health and or addiction risks associated with use of the technology.

Data for this analysis was collected in the form of an online survey. The survey was created through the use of Qualtrics and was distributed via email to all Bentley University undergraduate students, 4,021 in total. The Qualtrics survey ran from Friday, March, 17, 2017 to Wednesday, March, 22, 2017 and gathered both complete and partial responses. The survey was started by 654 students and completed by 594. Of the 594 completed responses, 591 were considered valid. The response ratio for the survey was approximately 14.70%. The three that were not considered valid included irrational responses and were not included in the analysis. The survey was initially distributed on Friday, March 17, 2017, and a reminder email was sent on Monday, March 20, 2017. All data in this analysis, unless otherwise stated, came from the aforementioned survey.

Bentley University's Institutional Review Board approved the project under the conditions that the data collected remain entirely anonymous and that participation in the survey was entirely voluntary. There was no penalty for choosing to stop the survey at any point. Prior to beginning the survey, participants were asked to give their consent in order to proceed. Formal documents were submitted to the Institutional Review Board prior to the project's approval, outlining all steps that would be taken and all information that would be obtained.

An online survey was chosen as the data-collection tool as it enabled more people to voice their opinions. This method was efficient and allowed for a substantial amount of data to be compiled in a relatively short period of time. The survey was simple to complete and addressed three main concepts related to traditional cigarettes and e-cigarettes, namely risk, health effects, and perception. Each concept was extremely important to the analysis. For the purposes of this study, risk was thought of in terms of perceived risk and was fairly generalized. This concept was broken down into both health risk and addiction risk and participants were asked to evaluate both. The concept concerning health effects was evaluated in a similar fashion and participants were asked to select responses that best matched their opinion as to both health effects and relative harm. The study in general was designed to measure students' perceptions, and as a result all questions except those concerned with demographics and smoking habits contained an element of opinion. Each individual's opinion forms a general idea of his or her perception of e-cigarettes and their place in society. Collectively, the perceptions of all participants were used to gain a better understanding of the general perceptions of the student body as a whole.

The survey itself consisted of 13 multiple choice questions and was designed to take no more than two minutes to complete. Participants were first asked basic demographic questions such as year in school, gender and ethnicity. Questions were then asked about the participants' use of both traditional cigarettes and e-cigarettes as well as whether use of each affected health. The survey was designed to gauge how undergraduate students felt about traditional cigarettes and e-cigarettes separately, as well as in comparison to each other. Lastly, participants were asked whether or not they felt e-cigarettes should be included under "smoke-free" laws. Appendix A shows the survey questions.

The data collection process was fairly smooth. However, with regard to the data itself there were a few issues that had to be addressed. In an effort to be inclusive, the survey offered three options for gender: male, female and other. A few participants chose the 'other' option, and reported their gender in irrational terms. Likewise, the following question addressed ethnicity and also included an 'other' option. While most of the responses indicated a valid ethnicity, some noted irrational responses.

Data Analysis

Analysis was done using the tools provided by Qualtrics. The data gathered from the survey was used to create graphs, charts and statistics which were all used for research purposes. The analysis included only complete, valid responses, and thus the three that were considered invalid were removed from the dataset in order to ensure the accuracy of corresponding reports. These three responses were saved outside of the dataset.

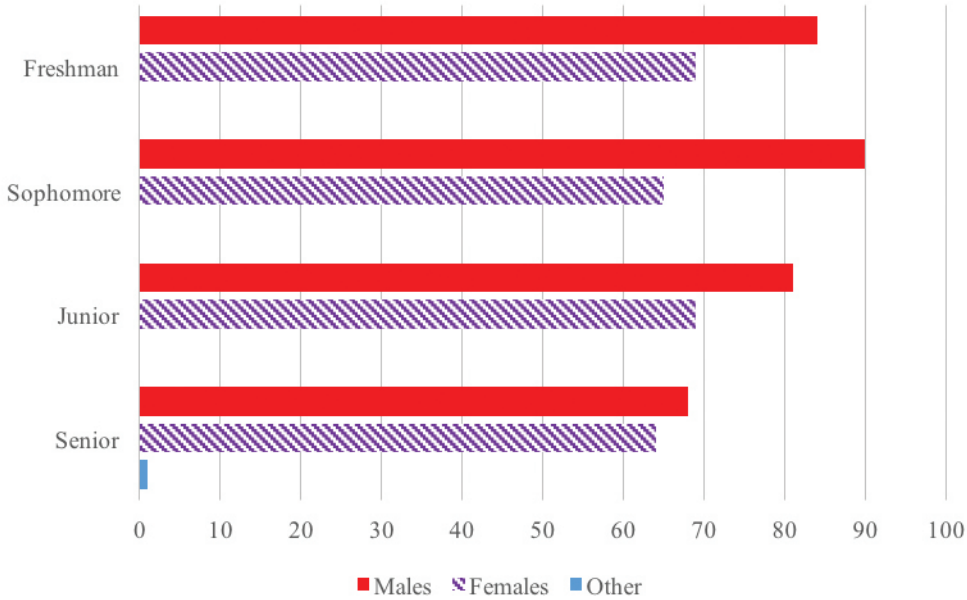
A. Demographics

The first question of the survey asked participants for their grade level. The four options, freshman, sophomore, junior and senior, gathered fairly equal numbers of responses. Of the 591 valid and completed responses, 153 were freshmen, 155 were sophomores, 150 were juniors and 133 were seniors. Sophomores represented the largest category with 26.23% of the response and seniors represented the smallest group with only 22.50% of the response.

The breakdown for gender showed that more males completed the survey than did females or people of 'other' gender. Responses were received from 323 males and 267 females. Only one participant selected 'other'. In percentage terms, this broke down into 54.65% males, 45.18% females and 0.17% other.

A look at both of these characteristics together showed a fairly even distribution of grade levels among female respondents, but not so uniform among males, as illustrated in Figure 1. Of the 267 females that completed the survey, 69 were freshmen, 65 were sophomores, 69 were juniors and 64 were seniors. Of the 323 male respondents, 84 were freshmen, 90 were sophomores, 81 were juniors and 68 were seniors. The person of 'other' gender was a senior. The largest population represented in the study was sophomore males, making up 15.23% of the response.

Figure 1
Breakdown of Gender Along Grade Level



The survey then asked participants about their ethnicity. They were provided with six choices: White, Hispanic or Latino, Black or African American, Native American or American Indian, Asian or Pacific Islander and other. Of the 591 respondents, 457 were White, 63 were Asian or Pacific Islanders, 46 were Hispanic or Latino, 18 were of 'other' nationalities, 7 were Black or African American and none were Native American or American Indian. When looking at this question in conjunction with the responses on gender, the largest response came from White males, totaling 260 responses.

B. Usage Patterns

The demographic questions concluded with ethnicity, and the survey switched focus to participants' smoking activities. Of the 591 responses, 539 people, or 91.20% reported that they did not currently smoke traditional cigarettes. Only 52 respondents, or 8.8%, reported that they were current smokers. A following question asked participants whether or not they had ever smoked traditional cigarettes to which 192, or 32.49%, responded that they had. A total of 399 respondents, or 67.51%, reported that they had never smoked traditional cigarettes. Analysis of these two questions together showed that 140 people who were not current smokers had smoked traditional cigarettes in the past, amounting to 25.97% of the response.

As for e-cigarette use, 47 people, or 7.95%, reported that they were current users while 544, or 92.05%, reported that they did not currently use the technology. However, when asked if they had ever used e-cigarettes, the numbers changed significantly. A total of 206 respondents, or 34.86%, reported that they had used e-cigarettes in the past while 385 people, or 65.14%, said that they had never used the devices. Analysis of the two questions together showed that 159 of the 544 people that did not currently use e-cigarettes had used the technology in the past. This amounted to 29.23% of the response, slightly higher than the equivalent comparison with regard to traditional cigarettes.

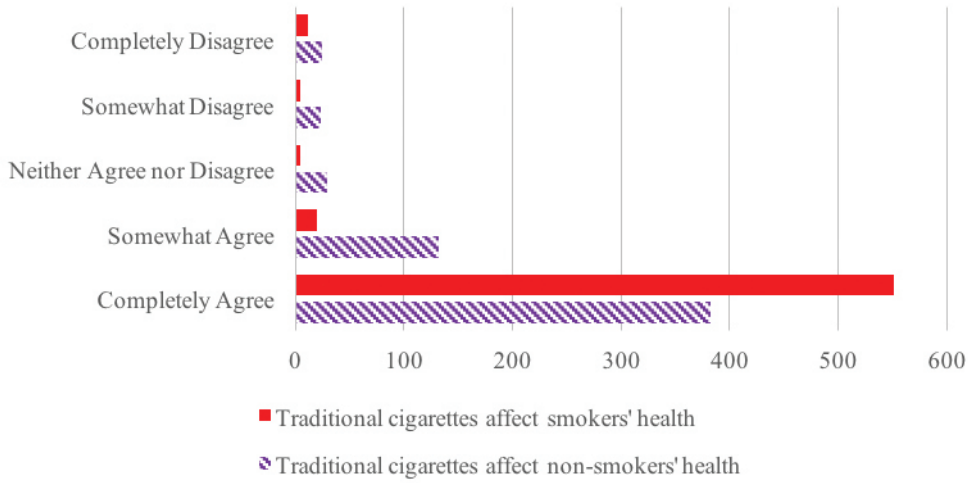
C. Perceptions

The survey then switched course to focus more about respondents' perceptions of both traditional cigarettes and e-cigarettes and how they affect human health. The questions used a five-point scale with the following options: 'completely disagree', 'somewhat disagree', 'neither agree nor disagree', 'somewhat agree' and 'completely agree'.

When asked what they thought about the statement "traditional cigarettes affect smokers' health," the vast majority of respondents thoroughly agreed, as shown in Figure 2. A total of 551 people noted that the 'completely agree' option best matched their opinion on the matter, a number which amounted to 93.23%. Looking at the complete opposite side of the spectrum, only 11 respondents selected 'completely disagree'

as the response that best matched their opinion. This amounted to only 1.86% of the response.

Figure 2
Respondents' Perceptions of the Effects of Traditional Cigarettes on the Health of Both Smokers and Non-Smokers

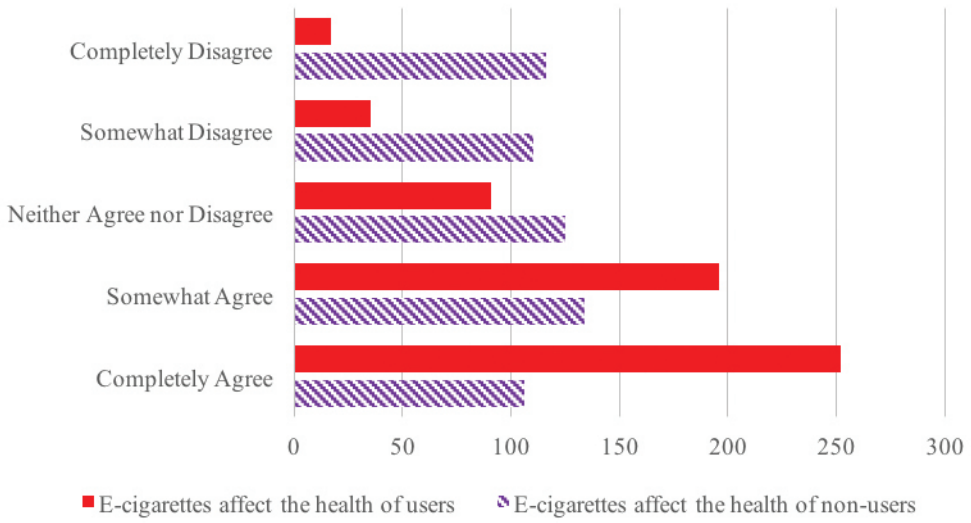


As for the statement "traditional cigarettes affect non-smokers' health," the response was slightly more varied but most people still noted their agreement. A total of 382 students indicated that 'completely agree' best matched their opinion. This was a significantly lower number of respondents than the 551 that selected 'completely agree' with regard to the previous statement, however it still accounted for 64.64% of the response. On the other end of the scale, 25 people, or 4.23% selected 'completely disagree'.

When asked about the phrase "e-cigarettes affect the health of users," most respondents agreed on some level. A total of 252 students, or 42.64%, noted that 'completely agree' best matched their opinion on the matter while 196, or 33.16%, selected 'somewhat agree'. On the other end, only 17 respondents, or 2.88%, indicated that 'completely disagree' best matched their opinion.

Responses were extremely varied when participants were asked to evaluate the statement "e-cigarettes affect the health of non-users". As shown in Figure 3, each of the five options was selected with similar frequency. A total of 116 students, or 19.63%, chose 'completely disagree' while only 106 students, or 17.94%, selected 'completely agree'. The 'somewhat disagree' category was selected by 110 students, making up 18.61% or the response, while the 'somewhat agree' option was selected by 134 people, accounting for 22.67% of the response. The 'neither agree nor disagree' option was the second most frequently selected, chosen by 125 respondents, or 21.15%.

FIGURE 3
RESPONDENTS' PERCEPTIONS OF THE EFFECTS OF E-CIGARETTES ON THE HEALTH OF BOTH SMOKERS AND NON-SMOKERS



The breakdown of responses to the statement "e-cigarettes affect the health of non-users" was interesting in that there was no clear leading answer. Each category was so close in the number of responses received so the distribution was fairly even. This seemed to indicate that students were uneducated as to exactly how e-cigarettes impact non-users, thus leading to such a spread of responses.

When asked about the relative harm of traditional cigarettes and e-cigarettes, the vast majority of respondents replied that traditional cigarettes were more harmful to both the health of smokers and non-smokers than e-cigarettes. A total of 451 participants, or 76.31%, indicated that they felt traditional cigarettes were more harmful to the health of smokers while only 9 people, or 1.52% felt that e-cigarettes were more harmful. A fair amount of people, however, felt that traditional cigarettes and e-cigarettes were both equally harmful. This group amounted to 122 people, or 20.64% of the response. Only 9 respondents, or 1.52%, felt that neither traditional cigarettes nor e-cigarettes were harmful to the health of smokers.

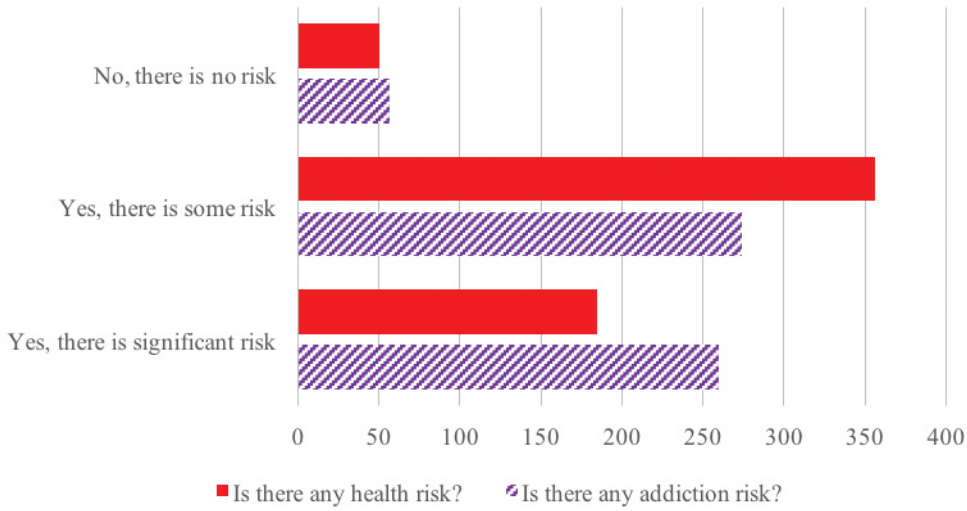
As for harm to non-smokers, 503 students, or 85.11%, felt that traditional cigarettes were more harmful. Only 5 people, or 0.85%, felt that e-cigarettes were more harmful. This represented the smallest part of the response. A total of 65 students, or 11.00%, felt that both traditional cigarettes and e-cigarettes were equally harmful while only 18 people, or 3.05%, felt that neither were harmful to the health of non-smokers.

Based on these responses, it seemed that students viewed e-cigarettes more favorably than traditional cigarettes with regard to health effects. While it seemed as though some students were aware of the potential impact e-cigarettes can have on human

health, it appeared that the majority was relatively unaware.

When asked whether there was any risk attached with the use of e-cigarettes most respondents indicated that there was at least some risk associated with the activity, as can be seen in Figure 4. Participants were presented with three possible answers to choose from: 'no, there is no risk,' 'yes, there is some risk' and 'yes, there is significant risk'. With regard to health risk, the majority of the response fell under the 'yes, there is some risk' category. A total of 356 respondents selected this option, thereby making up 60.24% of the response. Only 50 people, or 8.46%, felt that there was no risk involved while 185 people, or 31.30%, felt that there was significant health risk.

Figure 4
Respondents' Perceptions of Health and Addiction Risk With Regard to Use of E-Cigarettes



As for addiction risk, the responses were slightly more varied. Again, the majority of the response fell under the 'yes, there is some risk' category. A total of 274 people, or 46.36%, selected this response, while 260 people, or 43.99%, felt that there was significant addiction risk. Only 57 respondents, or 9.64%, felt that there was no addiction risk associated with use of e-cigarettes.

D. Smoke-Free Laws and Regulation

The last question participants were asked pertained to "smoke-free" laws and regulation. When asked whether or not e-cigarettes should be regulated under "smoke-free" laws, 314 of the 591 respondents felt that the devices should be included while 277 felt that they should not. This amounted to 53.13% and 46.87% respectively.

When this data was looked at in conjunction with data asking about the use of e-cigarettes, either currently or in the past, the breakdown was mostly as expected. Of the 206 people who had used e-cigarettes in the past, 41, or 19.90%, felt that they should be included under "smoke-free" laws while 165 people, or 80.10%, felt that they should not be. Of the 385 people that had never used e-cigarettes, 273, or 70.90%, felt that they should be included under "smoke-free" laws while 112, or 29.10%, did not. It seemed reasonable that more people that had used the technology would feel it should not be included under such legislation as compared to those who have not.

FINDINGS

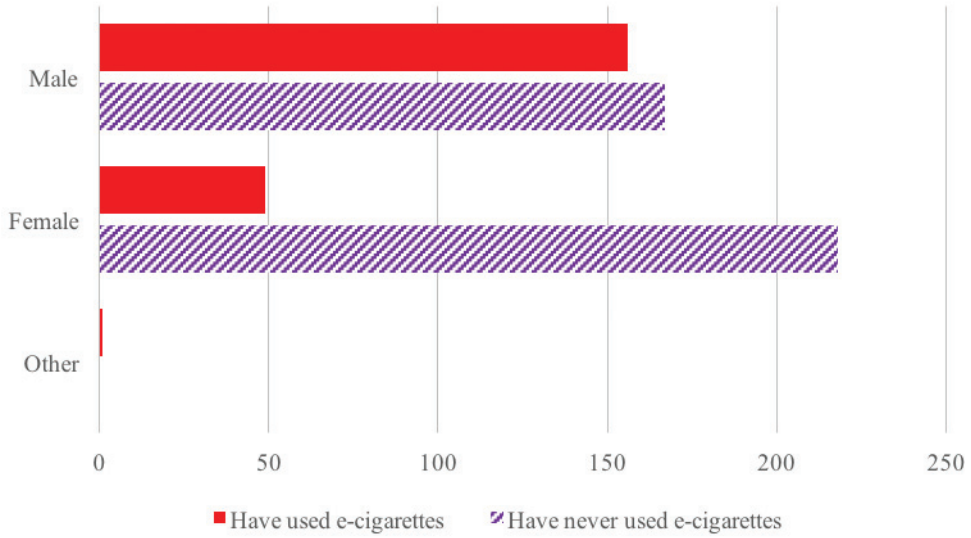
Analysis of the survey yielded some interesting results. These results were gathered by using the question breakout option on Qualtrics, allowing for the study of multiple questions at once. Ultimately, this allowed for a better understanding of how certain factors were related to each other with regard to the use of e-cigarettes.

As seen in Figure 5, the gender breakdown of those who had used e-cigarettes in the past showed that more males than females had used the technology. A total of 48.30% of males reported that they had used e-cigarettes at some point in their lives, while only 18.35% of females had participated in the activity. As there was only one person of 'other' gender, the findings could not truly be applied to a general population as it accounted for only one individual's response.

As far as school experience was concerned, fewer seniors had used e-cigarettes in the past than any other grade. The freshman and sophomore groups had very similar numbers of students who had used e-cigarettes in the past, however the junior category had the highest percentage of all four grades. This was interesting as there was no apparent trend in the data due to the high level of use among the junior respondents. Without this value, it would seem as though the lower the grade level, the higher the likelihood that a student would have used e-cigarettes at some point in their lives. Grade level is typically highly correlated with age, so this could potentially speak to the marketing strategies of e-cigarette companies and their efforts to market the technology to a younger crowd.

It seemed that there could also be a small cultural component to address. The Hispanic or Latino category had the highest rate of use among all ethnicities included in the study. However, the percentage of those that had used the technology at some point in life was only 54.35%, so this did not seem to suggest a strong relationship between ethnicity and use of e-cigarettes. That being said, a relationship could still exist that would likely have to be examined along with other cultural factors.

Figure 5
E-Cigarette Use by Gender



The data showed that among the respondents, current use of traditional cigarettes and e-cigarettes was fairly equivalent, yet traditional cigarette use was slightly higher. A total of 8.80% of respondents were current smokers of traditional cigarettes, while 7.95% of respondents were current users of e-cigarettes. Interestingly, the data also showed a similar equivalence of use in terms of having tried traditional cigarettes or e-cigarettes in the past. In percentage terms, more respondents had previously used e-cigarettes then traditional cigarettes. The percentages were fairly close, but 34.86% of respondents had previously used e-cigarettes while only 32.49% had previously smoked traditional cigarettes.

Another interesting finding concerned relationships between respondents' perceptions of both e-cigarettes and traditional cigarettes. People who said that they had used e-cigarettes at some point in their lives were significantly less likely to select 'completely agree' with regard to the statements "traditional cigarettes affect smokers' health" and "traditional cigarettes affect non-smokers' health" as illustrated in Tables 1 and 2. While 88.83% of those who had previously used e-cigarettes completely agreed with the statement regarding traditional cigarettes' impact on smokers' health, 95.58% of those who had never used e-cigarettes selected 'completely agree' as a response. With regard to traditional cigarettes' effect on non-smokers, the results were even more varied. Only 46.60% of people who had previously used e-cigarettes completely agreed that traditional cigarettes affect the health of non-smokers, as compared to the 74.29% of people who had never used e-cigarettes.

TABLE 1
OPINIONS OF PEOPLE WHO HAVE PREVIOUSLY USED E-CIGARETTES ON THE IMPACT TRADITIONAL CIGARETTES HAVE ON THE HEALTH OF BOTH SMOKERS AND NON-SMOKERS

	Completely Disagree		Somewhat Disagree		Neither Agree nor Disagree		Somewhat Agree		Completely Agree		Total
	#	%	#	%	#	%	#	%	#	%	#
Traditional cigarettes affect smokers' health	4	1.94%	5	2.43%	3	1.46%	11	5.34%	183	88.83%	206
Traditional cigarettes affect non-smokers' health	12	5.83%	10	4.85%	20	9.71%	68	33.01%	96	46.60%	206

TABLE 2
OPINIONS OF PEOPLE WHO HAVE NEVER USED E-CIGARETTES ON THE IMPACT TRADITIONAL CIGARETTES HAVE ON THE HEALTH OF BOTH SMOKERS AND NON-SMOKERS

	Completely Disagree		Somewhat Disagree		Neither Agree nor Disagree		Somewhat Agree		Completely Agree		Total
	#	%	#	%	#	%	#	%	#	%	#
Traditional cigarettes affect smokers' health	7	1.82%	0	0.00%	1	0.26%	9	2.34%	368	95.58%	385
Traditional cigarettes affect non-smokers' health	13	3.38%	13	3.38%	9	2.34%	64	16.62%	286	74.29%	385

People who had used e-cigarettes in the past were also less likely to completely agree with the statement "e-cigarettes affect the health of users" than were people who had never used the technology. As shown in Tables 3 and 4, only 23.30% of people who had used e-cigarettes at some point in their lives selected 'completely agree' as compared to the 52.99% of respondents who had never used e-cigarettes. When asked about how e-cigarettes affect non-smokers, only 4.85% of people who had used e-cigarettes completely agreed that "e-cigarettes affect the health of non-users". However, people who had never used e-cigarettes were more likely to select 'completely agree' as 24.94% noted that this option best matched their opinion on the matter.

Table 3
Opinions of People Who Have Previously Used E-Cigarettes on the Impact E-Cigarettes Have on the Health of Both Users and Non-Users

	Completely Disagree		Somewhat Disagree		Neither Agree nor Disagree		Somewhat Agree		Completely Agree		Total
	#	%	#	%	#	%	#	%	#	%	#
E-cigarettes affect the health of users	10	4.85%	24	11.65%	41	19.90%	83	40.29%	48	23.30%	206
E-cigarettes affect the health of non-users	75	36.41%	58	28.16%	40	19.42%	23	11.17%	10	4.85%	206

Table 4
Opinions of People Who Have Never Used E-Cigarettes on the Impact E-Cigarettes Have on the Health of Both Users and Non-Users

	Completely Disagree		Somewhat Disagree		Neither Agree nor Disagree		Somewhat Agree		Completely Agree		Total
	#	%	#	%	#	%	#	%	#	%	#
E-cigarettes affect the health of users	7	1.82%	11	2.86%	50	12.99%	113	29.35%	204	52.99%	385
E-cigarettes affect the health of non-users	41	10.65%	52	13.51%	85	22.08%	111	28.83%	96	24.94%	385

The data showed that people who had previously used e-cigarettes were more likely to believe that there was no health risk and no addiction risk associated with use of e-cigarettes than people who had never used the devices. A total of 16.50% of people who had previously used e-cigarettes felt that there was no health risk associated with use of the devices, as compared to the 4.16% of people who had never used e-cigarettes. With regard to addiction risk, 18.45% of those who had previously used e-cigarettes felt that that there was no risk, as compared to the 4.94% of people who had never used the technology. On the other hand, people who had never used e-cigarettes were much more likely to believe that there was both significant health risk and addiction risk than people who had used the technology. A total of 41.82% of those respondents who had never used e-cigarettes felt that there was significant health risk involved, while only 11.65% of people who had previously used the devices felt the same way. Similarly, 49.61% of people who had never used e-cigarettes felt that there

was significant addiction risk involved, as compared to the 33.50% of people who had previously used e-cigarettes.

This leads into a discussion about the perception of smoking as a whole. From the data, it seemed that people who had smoked e-cigarettes had a more favorable perception of traditional cigarettes than did people who had never used the technology. Research has shown that e-cigarettes and traditional cigarettes are largely different products with very different effects on human health. A possible explanation of the support of traditional cigarettes by people who had, at some point in time, used e-cigarettes could relate to the support of smoking as a general practice. These people may not differentiate as much between traditional cigarettes and e-cigarettes as do people who have never participated in use of either item.

However, the data also seemed to indicate that people really did not know the exact effects of e-cigarettes, both for smokers and non-smokers. The responses were largely varied across the five options presented in the survey. Most participants responded to the statement "e-cigarettes affect the health of users" with either 'somewhat agree' or 'completely agree', however 15.40% still responded that they neither agreed nor disagreed with the statement. Only 8.8% of respondents selected either 'somewhat disagree' or 'completely disagree'.

The variation was even more apparent with regard to the statement "e-cigarettes affect the health of non-users". Each of the five categories received a relatively equal response. The percentages of the response for each category ranged from 17.94% to 22.67%. This data alone was cause for concern. The closeness of the response for each of the five categories indicated that people, students especially, were not well-informed of the potential effects that e-cigarettes can have, especially with regard to non-smokers.

This same analysis was done with regard to traditional cigarettes as well. Here, there was a very clear trend with the vast majority of respondents agreeing on some level with both statements presented. A total of 96.61% of respondents selected either 'somewhat agree' or 'completely agree' in response to the statement "traditional cigarettes affect smokers' health" and 86.98% selected one of these categories in response to the statement "traditional cigarettes affect non-smokers' health". These numbers were quite different from those collected with regard to e-cigarettes.

In recent history, many efforts have been made to educate the public on the effects of traditional cigarettes on both smokers and bystanders. There have been campaigns dedicated to reducing the amount of incorrect information about the products that would ultimately reach consumers and expanding the public's knowledge on both health and addiction risks. E-cigarettes are a relatively new technology and have not regularly been the focus of such campaigns. This may have been a contributing factor to the discrepancy between the perceptions of traditional cigarettes and e-cigarettes among respondents.

While the findings obtained from this survey offered new insights, it must be acknowledged that, as with any study, this research had its limitations. The 591 valid

responses received could be considered a good response rate for a survey of this type, yet it must be noted that the survey was distributed to 4,021 undergraduate students. This was a response rate of approximately 14.70%.

As far as health risks were concerned, it seemed that participants in the study perceived traditional cigarettes as more harmful to the health of both smokers and non-smokers than e-cigarettes. However, many participants did express the belief that e-cigarettes affect the health of users to some extent. The varied responses received when participants were asked about the impact e-cigarettes have on the health of non-users suggested that there was a high level of uncertainty among the student body. It seemed as though people were unsure as to the true effects of e-cigarettes on bystanders and thus produced a wide range of opinions. Perceived risk did appear to be lower when discussing e-cigarettes as opposed to traditional cigarettes. As for general perceptions, it seemed as though participants largely perceived e-cigarettes to be a safer and lower-risk alternative than traditional cigarettes.

IV. Conclusions

Analysis of the survey responses yielded some notable results. However, the survey was designed to address only one of the two research questions discussed in the introduction. The analysis was only concerned with the question regarding students' perceptions of e-cigarettes, as the research question regarding health effects was addressed in the literature review.

Using the data obtained from the survey, I was able to see some prominent trends in perceptions among students. Students largely felt that traditional cigarettes were harmful to both smokers and non-smokers. When asked about e-cigarettes, participants seemed slightly more uncertain as to their potential effects, but most did express the belief that e-cigarettes have an impact on the health of users. The interesting piece gathered through analysis of the survey concerned the impact of e-cigarettes on the health of non-users. Unlike with the previous questions, no clear trend was evident and the responses ranged from complete agreement to complete disagreement.

This study has several implications. It suggests that students may be unaware or uninformed of the potential effects e-cigarettes can have, particularly with regard to bystanders. As younger people seem to be the target market for e-cigarette companies, this has the potential to be an enormous issue. Consumers cannot make educated decisions if the effects of the devices are unclear.

The research question addressed through analysis of the survey is as follows: Are e-cigarette users and people around them aware of the potential health effects these devices can have? The hypothesis developed in response to this question was that the absence of smoke will lead users to believe that e-cigarettes are less harmful and damaging to the smoker than traditional cigarettes and that they are not harmful to bystanders. While it appears that respondents believe e-cigarettes are less harmful than traditional cigarettes, the research model did not allow for in-depth analysis as to why

participants felt that way. To do this, the survey would have had to be much longer and much more specific. It seemed that students would be more likely to complete a brief, simple survey than a long, complex one. Therefore, the survey was designed accordingly and simply addressed what students' perceptions were, not why they came about.

This points to an opportunity for further research about why students perceive the products in specific ways. Understanding of this component could provide valuable insight as to what factors are most likely to shape consumers' opinions. This would likely also affect marketing practices and product development, as new information could allow for further improvements to be made.

Additionally, the survey and corresponding analysis did not extensively address the presence of "smoke-free" laws. The survey included one question asking participants whether they believed e-cigarettes should be included under smoke free laws, but again did not ask why they selected a specific answer. This is another opportunity for further research, especially as more and more areas and institutions are implementing "smoke-free" laws.

The research provided interesting insights into the perceptions of the Bentley University undergraduate student population, and can potentially be generalized to represent the views of a current population of college-age individuals and beyond. However, it must be noted that there are other demographic factors that could influence results that were not directly addressed in this study. Given that these factors could have a significant impact on both results and subsequent conclusions, it would be reasonable to conduct additional research before further generalizations are made. It would be interesting to do this research on people both younger and older than college-age students to see if there is any sort of usage trend associated with age.

The increased use of new technologies, such as e-cigarettes, brings forth new perceptions of risks associated with their use. Further research will allow for a better understanding of such risks.

References

Abadi, S., Couch, E. T., Chaffee, B. W., & Walsh, M. M. (2017). "Perceptions related to use of electronic cigarettes among California college students." *Journal of Dental Hygiene*, 91(1), 35-43.

Ahern, N. R., & Mechling, B. (2014). "E-cigarettes: A rising trend among youth." *Journal of Psychosocial Nursing & Mental Health Services*, 52(6), 27-31. http://dx.doi.org/10.3928/02793695-20140506-01

Berg, C. J., Stratton, E., Schauer, G. L., Lewis, M., Wang, Y., Windle, M., & Kegler, M. (2015). "Perceived harm, addictiveness, and social acceptability of tobacco products and marijuana among young adults: Marijuana, hookah, and electronic cigarettes win." *Substance Use & Misuse*, 50(1), 79-89. http://dx.doi.org/10.3109/10826084.2014.958857

Bushak, L. (2013, December 19). "It's not the smoke, it's the nicotine: E-cigarettes may damage arteries." *Medical Daily*. Retrieved from http://www.medicaldaily.com/its-not-smoke-its-nicotine-e-cigarettes-may-damage-arteries-265498

Callahan-Lyon, P. (2014). "Electronic cigarettes: Human health effects." *Tobacco Control*, 23(Suppl. 2), ii36-ii40. http://dx.doi.org/10.1136/tobaccocontrol-2013-051470

Carr, E. R. (2014). "E-cigarettes: Facts, perceptions, and marketing messages." *Clinical Journal of Oncology Nursing*, 18(1), 112-116. http://dx.doi.org/10.1188/14.CJON.112-116

Centers for Disease Control and Prevention. (2016a, December 01). "Health effects of cigarette smoking." Retrieved from https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/

Centers for Disease Control and Prevention. (2016b, December 28). "Burden of tobacco use in the U.S." Retrieved from https://www.cdc.gov/tobacco/campaign/tips/resources/data/cigarette-smoking-in-united-states.html

Centers for Disease Control and Prevention. (2017, February 21). "Secondhand smoke (SHS) facts." Retrieved from https://www.cdc.gov/tobacco/data_statistics/fact_sheets/secondhand_smoke/general_facts/

Choi, K., Fabian, L., Mottey, N., Corbett, A., & Forster, J. (2012). "Young adults' favorable perceptions of snus, dissolvable tobacco products, and electronic cigarettes: Findings from a focus group study." *American Journal of Public Health*, 102(11), 2088–2093. http://doi.org/10.2105/AJPH.2011.300525

Czogala, J., Goniewicz, M. L., Fidelus, B., Zielinska- Danch, W., Travers, M. J., & Sobczak, A. (2014). "Secondhand exposure to vapors from electronic cigarettes." *Nicotine & Tobacco Research*, 16(6), 655-662. http://dx.doi.org/10.1093/ntr/ntt203

Grana, R., Benowitz, N., & Glantz, S.A. (2014). "E-cigarettes: A scientific review." *Circulation, 129*(19), 1972-1986. https://doi.org/10.1161/CIRCULATIONAHA.114.007667

Hess, C. A., Olmedo, P., Navas-Acien, A., Goessler, W., Cohen, J. E., & Rule, A. M. (2017). "E-cigarettes as a source of toxic and potentially carcinogenic metals." *Environmental Research*, 152, 221-225. http://dx.doi.org/10.1016/j.envres.2016.09.026

Hiemstra, P. S., & Bals, R. (2016). "Basic science of electronic cigarettes: Assessment in cell culture and in vivo models." *Respiratory Research*, 17, 1-5. doi:10.1186/s12931-016-0447-z

Martinez-Sanchez, J., Fu, M., Martin-Sanchez, J., Ballbè, M., Saltó, E., & Fernandez, E. (2015). "Perception of electronic cigarettes in the general population: Does their usefulness outweigh their risks?" *BMJ Open*, 5(11), 1-6. http://dx.doi.org/10.1136/bmjopen-2015-009218

Mays, D., Smith, C., Johnson, A. C., Tercyak, K. P., & Niaura, R. S. (2016). "An experimental study of the effects of electronic cigarette warnings on young adult nonsmokers' perceptions and behavioral intentions." *Tobacco Induced Diseases*, 14, 1-10. http://dx.doi.org/10.1186/s12971-016-0083-x

Meernik, C., Baker, H. M., Lee, J. G. L., & Goldstein, A. O. (2017). "The tobacco 21 movement and electronic nicotine delivery system use among youth." *Pediatrics, 139*(1), 1-2. http://dx.doi.org/10.1542/peds.2016-2216

National Institute on Drug Abuse. (2016, May). "Electronic cigarettes (e-cigarettes)." Retrieved from https://www.drugabuse.gov/publications/drugfacts/electronic-cigarettes-e-cigarettes

Pepper, J. K., Emery, S. L., Ribisl, K. M., Rini, C. M., & Brewer, N. T. (2015). "How risky is it to use e-cigarettes? Smokers' beliefs about their health risks from using novel and traditional tobacco products." *Journal of Behavioral Medicine, 38*(2), 318-326. http://dx.doi.org/10.1007/s10865-014-9605-2

Rouabhia, M., Park, H. J., Semlali, A., Zakrzewski, A., Chmielewski, W. & Chakir, J. (2017). "E-cigarette vapor induces an apoptotic response in human gingival epithelial cells through the caspase-3 pathway." *Journal of Cellular Physiology*, 232(6), 1539–1547. http://dx.doi.org/10.1002/jcp.25677

Sanders-Jackson, A., Schleicher, N. C., Fortmann, S. P., & Henriksen, L. (2015). "Effect of warning statements in e-cigarette advertisements: An experiment with young adults in the United States." *Addiction*, 110(12), 2015-2024. http://dx.doi.org/10.1111/add.12838

Shahab, L., Goniewicz, M.L., Blount, B.C., Brown, J., McNeill, A., Alwis, K.U., & Feng, J. (2017). "Nicotine, carcinogen, and toxin exposure in long-term e-cigarette and nicotine replacement therapy users: A cross-sectional study." *Annals of Internal Medicine*, 166(6), 390-400. Retrieved from http://annals.org/aim/article/2599869/nicotine-carcinogen-toxin-expo

Simon, S. (2016, April 19). "Use of e-cigarettes rising among middle and high school students." Retrieved from https://www.cancer.org/latest-news/use-of-e-cigarettes-rising-among-middle-and-high-school-students.html

Smith, L., Brar, K., Srinivasan, K., Enja, M., & Lippmann, S. (2016). "E-cigarettes: How "safe" are they?" *Journal of Family Practice*, 65(6), 380-385.

Trumbo, C. W., & Harper, R. (2013). "Use and perception of electronic cigarettes among college students." *Journal of American College Health*, 61(3), 149-155. doi:10.1080/07448481.2013.776052

U.S. Department of Health and Human Services. (n.d.-a). "Effects of smoking on your health." Retrieved from https://betobaccofree.hhs.gov/health-effects/smoking-health/

U.S. Department of Health and Human Services. (n.d.-b). "Electronic cigarettes." Retrieved from https://betobaccofree.hhs.gov/about-tobacco/Electronic-Cigarettes/index.html

U.S. Food and Drug Administration. (2017, February 13). "Vaporizers, e-cigarettes, and other electronic nicotine delivery systems (ENDS)." Retrieved from https://www.fda.gov/TobaccoProducts/Labeling/ProductsIngredientsComponents/ucm456610.htm

Wang, P., Chen, W., Liao, J., Matsuo, T., Ito, K., Fowles, J., Kumagai, K. (2017). "A device-independent evaluation of carbonyl emissions from heated electronic cigarette solvents." *PLoS One*, 12(1), 1-12. http://dx.doi.org/10.1371/journal.pone.0169811

Appendix – Survey

Grade level

- ◯ Freshman
- ◯ Sophomore
- ◯ Junior
- ◯ Senior

Gender

- ◯ Male
- ◯ Female
- ◯ Other

Ethnicity

- ◯ White
- ◯ Hispanic or Latino
- ◯ Black or African American
- ◯ Native American or American Indian
- ◯ Asian / Pacific Islander
- ◯ Other

Do you currently smoke traditional cigarettes?

- ◯ Yes
- ◯ No

Have you ever smoked traditional cigarettes?

- ◯ Yes
- ◯ No

Do you currently use e-cigarettes?

- ◯ Yes
- ◯ No

Have you ever used e-cigarettes?

- ◯ Yes

Select the choice that best matches your opinion:

	Completely Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Completely Agree
Traditional cigarettes affect smokers' health	○	○	○	○	○
Traditional cigarettes affect non-smokers' health	○	○	○	○	○

Select the choice that best matches your opinion:

	Completely Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Completely Agree
E-cigarettes affect the health of users	○	○	○	○	○
E-cigarettes affect the health of non-users	○	○	○	○	○

In your opinion, which is more harmful to the health of a smoker: traditional cigarettes or e-cigarettes?

- ○ Traditional cigarettes are more harmful
- ○ E-cigarettes are more harmful
- ○ Both are equally harmful
- ○ Neither are harmful

In your opinion, which is more harmful to the health of a non-smoker: traditional cigarettes or e-cigarettes?

- ○ Traditional cigarettes are more harmful
- ○ E-cigarettes are more harmful
- ○ Both are equally harmful
- ○ Neither are harmful

Select the option that best matches your opinion with regard to the use of e-cigarettes:

	No, there is no risk	Yes, there is some risk	Yes, there is significant risk
Is there any health risk?	○	○	○
Is there any addiction risk?	○	○	○

Should e-cigarettes be included under "smoke-free" laws?

- ○ Yes
- ○ No

ARTICLES

Bentley University Honors Program
Waltham, MA, USA
fusio@bentley.edu

BENTLEY
UNIVERSITY