5-13-2016

Race, Online Health Resources and their Contribution to Health Disparities

Elizabeth Keck

Recommended Citation

Keck, Elizabeth, "Race, Online Health Resources and their Contribution to Health Disparities." Thesis, Georgia State University, 2016.
doi: https://doi.org/10.57709/7981282

RACE, ONLINE HEALTH RESOURCES AND THEIR CONTRIBUTION TO HEALTH DISPARITIES

by

ELIZABETH KECK FREEMAN

Under the Direction of Erin Ruel, PhD

ABSTRACT

With increasing usage of the Internet for a variety of activities, including health information seeking, there is opportunity for the technology to have a pivotal impact on reducing health disparities. Using a Fundamental Causes framework (Link and Phelan 1997), this thesis explores whether or not Internet usage for health information seeking reduces racial health disparities. Using data from the Pew Research Center this study examined active and passive health information seeking and the impact they have on health outcomes among blacks, whites, and Hispanics. The health conditions included self-rated health, high blood pressure, diabetes, lung disease, and heart disease. The results indicate participating in active or passive health information seeking had little or no impact on decreasing the chronic health conditions. Additionally, the results show participating in active or passive health seeking activities has a mixed impact on decreasing chronic health conditions across racial groups.

INDEX WORDS: Race, Health, Chronic Health Conditions, Fundamental Causes

RACE, ONLINE HEALTH RESOURCES AND THEIR CONTRIBUTION TO HEALTH DISPARITIES

by

ELIZABETH KECK FREEMAN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Arts

in Sociology

in the College of Arts and Sciences

Georgia State University

2015

RACE, ONLINE HEALTH RESOURCES AND THEIR CONTRIBUTION TO HEALTH DISPARITIES

by

ELIZABETH KECK FREEMAN

Committee Chair: Erin Ruel

Committee: Dawn Baunach

Deirdre Oakley

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2015

DEDICATION

This thesis is dedicated to my husband, Ian Freeman. It would never be finished if it weren’t for your constant encouragement and patience. Thank you and I love you.

ACKNOWLEDGEMENTS

Thank you to my chair, Erin, and the rest of my committee, Dawn and Deirdre. This was a long process for me- thank you for sticking with me and believing in me! I’ll never forget Dawn’s words in my first graduate course at Georgia State — “A good thesis is a done thesis!” Amen!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

1 INTRODUCTION 1

2 LITERATURE REVIEW 3

1.1 Fundamental Cause Theory 3

1.2 Online Health Information Seekers 8

1.3 Reasons for Online Health Information Seeking 10

1.4 Online Health Information Seeking and Health Outcomes 12

1.5 The Hispanic Paradox 13

3 HYPOTHESES 17

4 RESEARCH METHODS 17

4.1 Data 18

4.2 Constructs 21

4.2.1 Dependent Variables *21*

4.2.2 Primary Independent Variables *23*

4.2.3 Secondary Independent Variables *25*

4.2.4 Control Variables *26*

4.3 Analytical Plan 28

5 RESULTS 33

5.1 Univariate Analysis 33

5.2 Bivariate Analysis 35

5.3 Binary Logistic Regression 38

5.3.1 Poor Health 38

5.3.2 Diabetes 41

5.3.3 Heart Disease 44

5.3.4 Lung Disease 47

5.3.5 High Blood Pressure 50

5.3.6 High Level Analysis of Odds Ratios 52

6 DISCUSSION 54

7 CONCLUSION 63

8 RECOMMENDATIONS 64

9 REFERENCES 66

LIST OF TABLES

Table 4.1.1: Demographics by Data Set 20

Table 4.2.1: Dependent Variables.. 23

Table 4.2.2: Primary Independent Variables 25

Table 4.2.3: Secondary Independent Variables: Race 26

Table 4.2.4: Control Variables... 27

Table 5.1.1: Univariate Analysis 33

Table 5.2.1: Bivariate Analysis: Race and Internet Health Seeking Activities by Health Conditions.. 35

Table 5.3.1: Regressing Self-Reported Poor Health on Passive and Active Online Resources 38

Table 5.3.2: Regressing Diabetes on Passive and Active Online Resources.................... 41

Table 5.3.3: Regressing Heart Disease on Passive and Active Online Resources 44

Table 5.3.4: Regressing Lung Disease on Passive and Active Online Resources............ 47

Table 5.3.5: Regressing High Blood Pressure on Passive and Active Online Resources 50

Table 5.3.6: Significant Odds Ratios by Chronic Condition 53

1 INTRODUCTION

Despite the increase in positive health interventions research indicates health disparities continue to exist for blacks and Hispanics. The occurrence of diabetes in blacks is somewhere between 1.4 and 2.2 times that of whites, and Hispanics also have a higher incidence than non-Hispanics for the disease. Additionally, the mortality rate from diabetes is higher for all minorities when compared to whites (UDHHS 2006). Furthermore, a 2003 study by the Centers for Disease Control found African Americans reported higher prevalence of two or more risk factors for heart disease than other racial and ethnic groups. Heart disease and diabetes are two of the top causes of mortality, and race is a large factor in a person's likelihood to experience the disease.

Health disparities likely exist because of the negative differences marginalized people experience because of certain characteristics or life circumstances. For example, perhaps because of lower socioeconomic status some people are unable to afford or understand treatments that could improve their health; they may not even be able to physically get to a treatment location. Or, perhaps because of their education or income levels they aren't able to eat healthy, cannot afford more expensive fruits and vegetables, or have little time to cook healthy meals. All of these individual experiences could potentially have a huge impact on a person's health. These characteristics or life circumstances are called Fundamental Causes (Link and Phelan 1995) or key factors of inequality continue to sustain a person's position in society. Fundamental Causes essentially perpetuate one's position of inferiority despite positive interventions or actions that could otherwise help improve his or her health outcomes.

Use of the Internet for health information seeking could act as a positive intervention and help improve the health of minorities, mitigating Fundamental Causes. A 2013 Pew Internet and American life study reports a total of 86% of American adults use the Internet, which is an increase of 14% since 1995. People have transitioned from utilizing the Internet every so often to integrating its use with their daily lives. This increase in use also plays into the ways people use the Internet for health information seeking activities. Previously people used the Internet to search for a doctor near them. However, now they are not only finding health care providers but also using the Internet to learn more about their diagnosis, research their medications, and find and connect with others who have similar health issues. These connections can potentially lead to a greater understanding of their illness, better treatment, more positive well-being, and overall better health outcomes.

In the same way Fundamental Causes (Link and Phelan 1995) impact a person’s life in the offline space the Internet could also be subject to the perpetuation of inequality. People may not be able to afford a computer or Internet access or may not have time to visit their local library to utilize community options. Inequality has always played a role in the way people receive health care, and Fundamental Causes continue to perpetuate inequality through access and treatment to health care. Understanding the impact Fundamental Causes of inequality have on the utilization and effectiveness of online health resources will help the public health and medical communities create ways to overcome these barriers and help improve the health of disparate populations.

This research explores how Link and Phelan’s Fundamental Causes impact the utilization and effectiveness of online health resources on health outcomes and specifically the incidence of cancer, heart disease, and diabetes as well as overall self-rated health. This research works to answer the following question: Does the use of Internet health seeking activities work to reduce the disparity of poor self-rated health and chronic conditions among blacks and Hispanics when compared to whites? To answer this question I will first review existing literature on fundamental causes, online health information seeking, and health outcomes. Then, I will introduce the data and methods followed by the findings, discussions, and conclusions for a primary research analysis utilizing data from the 2012 Pew Internet and American Life study.

2 LITERATURE REVIEW

1.1 Fundamental Cause Theory

The most appropriate theory to look to when questioning why racial health disparities exist, despite positive interventions, comes from Link and Phelan (1995). Link and Phelan's theory of Fundamental Causes describes the factors that perpetuate health inequality. These are factors that involve a person's ascribed characteristics such as race and social class, socioeconomic status, and gender. The reason these factors are fundamental is they are the foundations on which one's experiences are based, and they are most often unchangeable. Fundamental Causes do not just impact whether or not someone is likely to seek or act on an intervention; their mere existence means a person will face other barriers to positive change. Those who aren't part of the minority or disadvantaged group — or in other words, are white, have a high socioeconomic status, or are male — are more likely to have the characteristics associated with resources such as knowledge, money, power, prestige, social connections, and other resources, which allows them an increased ability to navigate or prevent disease.

Link and Phelan (1995) explain that when it comes to health Fundamental Causes and their association to the incidence of disease continues to exist, because even if there is an intervening mechanism introduced or if the disease completely eradicated, the disease or another will re-emerge. They argue that one of the reasons these associations persist is because the focus is often on reducing or eliminating one cause or element of an individual's specific disease rather than focusing on eliminating the social factors related to Fundamental Causes. The reasons Fundamental Causes of disease are so complex is because they are difficult or non-changing, and the focus on interventions occurs at a micro-individual level rather than a higher level focused on eliminating fundamental inequalities.

In Fundamental Cause Theory, Link and Phelan (1995) present the concept of efforts or interventions that attempt to neutralize Fundamental Causes and put all people on the same playing field. They use the example of AIDS to illustrate the importance of conceptualizing a person's experience to understand what factors they face that may impact their ability to avoid health risk. They provide the example of an intervention which includes educating people on how to reduce their risk of contracting HIV. Some people are better able to utilize this education than others, so it is important to understand why some are less able to avoid putting themselves at risk. For example, it is important to consider some poor women may need to engage in prostitution for money and are therefore unable to reduce their risk of acquiring HIV (Link and Phelan 1995).

Conceptualizing barriers to engaging in positive health or avoiding health risk is one way Link and Phelan (1995) present as to get to the root cause of health inequality. They then further explore how understanding the reasons for the inequality lead back to the Fundamental Causes of disease. Again, using the example of AIDS, they connect poor health outcomes from drug use by those with low Socioeconomic Status (SES) in the 1980s, which had an even larger impact with the emergence of AIDS. Furthermore, they point out that this SES link will continue as AIDS contributes to SES differences in the future. The researchers illustrate that understanding the context (why someone puts themselves at risk of a disease) along with how it is connected to a Fundamental Cause (in this case SES) leads to the perpetuation of health inequality.

Chang and Lauderdale (2009) specifically test Link and Phelan’s theory of Fundamental Causes by looking at two waves of the National Health and Nutrition Examination survey to see if total cholesterol levels of different racial groups changed once a new treatment option entered the market. Historically, people with higher SES had higher cholesterol levels than those with lower SES. However, after the introduction of the new treatments incidence of high cholesterol completely flipped; those with high SES are now more likely to have healthy cholesterol levels than those with low SES. The effect also presents itself across racial lines; whites are more likely than blacks and Hispanics to have healthy cholesterol levels. Those with more power, in the form of race and SES, are able to take advantage of the treatment. Those who experience a factor considered a Fundamental Cause face both challenges in learning about the treatment and access to the new treatment. Therefore, the results support Link and Phelan’s theory in that those who experienced race or low SES as a Fundamental Cause are unlikely to be benefitted by health advances.

A study by Suziedelyte (2012), which examined the relationship between health information seeking and fundamental causes utilizing Health Information National Trends Survey (HINTS) data, found online health information seekers are more likely to seek out health care than non-online health seekers. This could be evidence that Link and Phelan’s theory of Fundamental Causes will be supported in that those who already have health seeking behaviors are only boosted by interventions whereas those who typically do not already seek help see very little improvement with interventions. However, the study does not explore whether or not people seek online health information, and subsequently health care, because they are unhealthy or whether or not these are actions of healthier people who then avoid further risk.

Link, Phelan, Miech, and Westin (2008) look at the impact intelligence, rather than just education, has on health disparities. Link, Phelan, Miech, and Westin report intelligence could be emerging as another determinant of health much like SES; so, they study it to determine if it indeed follows the pattern of Fundamental Cause Theory and can explain any health differences. They utilize the Wisconsin Longitudinal Study of high school graduates and find intelligence has little impact after controlling for education and income. They also find the effect of SES remained significant and changed very little when controlling for intelligence. These findings indicate the Fundamental Causes of inequality, such as education or socioeconomic status, continue to explain inequality, and intelligence, as another potential determinant of health is ruled out. This provides further support for the impact of Fundamental Causes on health outcomes.

Lutfey and Freese (2005) use a year's worth of ethnographic data from two clinics with vastly different populations to explore how differences in SES impact a patient's ability to follow diabetes treatment regimens. They evaluate clinic visits at Park Clinic — a mostly white, upper-middle class population, and County Clinic, which serves a large working class, minority, and underinsured population. The purpose of their research is to provide support for Link and Phelan's Fundamental Cause Theory by exploring how SES effects the following: the providers assessment of how the patient is able to manage diabetes treatments; how provides obtain the information about the patient's problems; how the provider identifies solutions; and an assessment of how likely the treatment plan is to be implemented successfully. Their findings show even within one specific disease there can be a multitude of factors influenced by SES that can prevent those with lower SES from improving their health status. They find that the Park Clinic population is provided with more experienced doctors and new, cutting-edge treatment technology in their clinics. However, those at the County clinic are not provided the same quality of resources. These structural inequalities lead to further perpetuation of poor health conditions of those at County Clinic. Lutfey and Freese's (2005) findings continue to show that, despite the existence of interventions, the distribution of these interventions is inherently unequal; thus, health disparities are perpetuated. They hypothesize these results stem from one of four reasons: biological differences; structural or interpersonal discrimination; differences in early life experience and life course experiences with disease; differences in health risk behaviors or the exposure; and effect of cumulative stress processes. They call for more research around race and gender as fundamental causes to further explore these relationships.

In a new study by Masters, Link, and Phelan (2015), the researchers work to incorporate race (blacks and whites only) and gender into their analysis of educational gradients and their impact on adult mortality from non-heart disease related preventable causes. They note that at this point there has been a great deal of research supporting Fundamental Cause Theory, next they decide to include race and gender to see if the theory can be expanded to these two demographic factors. They use 19 waves of the National Health Interview Survey and link them to official death records from the 2006 National Death Index. Their research finds evidence that there are strong gender differences in mortality trends; the data indicates no reductions in women's mortality from non-heart disease preventable causes. When it came to race they find no reduction among US black mortality from preventable causes, but they find significant reductions for both white men and women.

These studies have all explored Link and Phelan’s Fundamental Cause Theory by examining the social factors that could have an impact on health disparities. However, despite other positive interventions, researchers continue to indicate Fundamental Cause Theory holds true. Factors such as income, education, and socioeconomic status allow people to obtain and maintain a position of power within society and, in turn, perpetuate health disparities. Newer research also calls for further investigation into gender and race as fundamental causes.

1.2 Online Health Information Seekers

The majority of previous research related to online health information seekers explores differences in those who seek health information online versus those who do not. The following studies provide evidence of the types of people who are found to seek online health information. Following that, there is some research indicating what impact health information seeking actually has on health outcomes. However, there is very little of said research, so the primary research outlined in this thesis will further expand on the exploration into the impact health information seeking has on leveling health disparities.

Chen and Lee (2014) conducted a study with 594 college students in an introductory course at a large public university. They evaluated the students’ online health information seeking behaviors and found there were no significant race differences in online health information seeking behaviors. However, Ono and Zavodny (2000) who surveyed 1,009 U.S. residents between the ages of 15 and 59, found blacks and Hispanics were less likely than whites to own and use a computer at home. Although these results cannot be generalized to seeking health information online, it does indicate there could be differences in online health seeking based on race due to computer ownership and usage. Additionally, Yabarra and Suman (2006) found blacks were 60% less likely than whites to be health information seekers when controlling for other variables. However, they did not specify whether or not the health information seeking was online or offline.

Lorence, Park, and Fox (2006) found a relationship between income and online health information seeking. They utilized the 2000 and 2002 versions of the Pew Internet and American Life data set, which includes a respondent group of 1,509 online health information seekers. They found race had little impact on the use of computers, the Internet, and online health information seeking when controlling for income. Although this could indicate online health information seeking activities among blacks and Hispanics may not vary compared to whites, it does not show whether or not the usage of these resources by different groups leads to worse or improved health outcomes.

Another study by Lorence and Park (2007) utilizing 2000 and 2002 Pew Internet and American Life data found there were large differences in the use of the computer and Internet for health information across education levels; those with less education were significantly less likely to use online health information. Additionally, a study by Suziedelyte (2012) utilizing HINTS data further supports the finding that online health information seeking is associated with higher education and a younger age. Ybarra and Suman (2006) found, among their sample of 1,454 telephone interviews with Spanish or English speaking Americans older than 12 years, that for each increase in education level the odds someone reported health information seeking increased 10%. Since race and education levels are often linked, with minorities often having less education than whites, these studies indicate that being a member of a minority group could also lead to a decrease in likelihood to utilize online health information.

1.3 Reasons for Online Health Information Seeking

In addition to exploring demographic characteristics on online health information seekers, Ybarra and Suman (2006) examined some interesting relationships between a person's contact with a health professional and their likelihood to seek support from others online. They found that those who had an interaction with a health professional were 2.2 times more likely to also report seeking help from others online. This finding has important implications to this study as it could indicate that getting a diagnosis leads to online health information seeking. This could indicate those who are already sick are more likely to participate in online health information seeking activities rather than people participating in these activities as preventative measures.

Fox and Duggan (2013) utilize the 2012 Pew Internet and American Life study data to examine how Americans living with chronic conditions utilize the Internet for information. They find Internet users living with one or more chronic health conditions are significantly more likely to participate in one of the following online health activities: gathering information online about medical problems, treatments, or drugs; consulting online reviews of drugs or other treatments; and reading or watching something online about someone else's personal health experience. The relationship remains significant even after they control for age, income, education, race, and overall health. However, their research also indicates a gap in access or usage of the Internet by those who have a chronic condition with only 72% saying they have access to the Internet compared to 98% access by those who do not have any chronic health problems. These findings indicate that 1) there is a gap in access by those with chronic health conditions and 2) those who do have access are more likely to use the Internet to seek health information. This further highlights the importance of understanding how these health-seeking activities do or do not perpetuate Fundamental Causes, because if one doesn't have access, it means he or she cannot reap the rewards, and health inequalities will continue to be perpetuated despite this potentially positive intervention.

Amante, Hogan, Pagoto, English, and Lapane (2015) examine whether or not people who experience challenges in their attempt to access health care are more likely to turn to the Internet for answers to their health questions. They utilize the National Health Interview Survey of 32,129 adults. When controlling for health insurance related issues they find almost 55% of people experienced at least one barrier in their attempt to access health care, and 9% experienced three or more barriers to care. The most common reasons for reporting a barrier to care were delay in getting care because they could not get an appointment soon enough or because they had to wait too long to see a doctor. Those who experienced issues with access to care were twice as likely to report using the Internet to search for health information when controlling for sex, age, race, education, marital status, presence of disease, and insurance coverage. Additionally, they found those who had one or more chronic medical conditions had greater odds of using the Internet for online health information. These results provide further evidence that prior diagnosis, or perhaps even the knowledge one is sick, may lead to increased usage of the Internet for health information.

Rice (2006) did an analysis of Pew Internet and American life data from a variety of years. He specifically analyzed the June 2001 data set of 500 Internet users who used the Internet to find health information. He conducted a linear regression analysis to examine the reasons why respondents reported going online for health information. He finds more frequent online health seekers stated their primary motivations to seek health information as the following: current health problems; problems with physician access; to diagnose or treat a condition on their own; or to look for information about a sensitive topic. This source provides further support that those who seek health information online are doing so because they have already become aware or suspect they have a health condition. Hence, whether or not people go online for health information may have quite a bit to do with their overall feeling of well-being (or suspicion they might have a health problem) along with whether or not they have recently seen a health care provider or have been diagnosed with a disease or condition.

1.4 Online Health Information Seeking and Health Outcomes

Cotton and Gupta (2004) utilized the 2000 General Social Survey data and found the more an individual sought health information online, the better self-rated health and happiness they reported. Their findings also support the previous studies that indicate health information seekers have higher incomes and more education. Cotton and Gupta (2002) do not examine race as a factor in their analysis, so although their study provides evidence that health information seeking through the Internet could in fact have a positive impact on self-rated health, it does not do enough to examine if online health information seeking has enough impact to overcome Fundamental Causes of inequality.

Wang, Clouston, Rubin, Colen and Link (2012) do not focus their research on online information specifically; however, their findings have direct implications regarding health information and colorectal cancer mortality. They utilized 1968 to 2008 data from the National Center for Health Statistics to better understand how areas where information on behaviors related to disease prevention and treatment were more likely to spread and how that impacted colorectal cancer mortality rates. They developed an innovative score that identified states that had a higher and lower propensity for diffusion of prevention information. Their results showed states with higher propensity for diffusion have higher SES. However, some states, such as New Hampshire, had high SES, but low diffusion. They found colorectal cancer mortality rates increased steadily in the lowest SES counties and declined in the highest. When controlling for sex, race, and age, they found SES and diffusion were still significantly associated with mortality. Their research provides evidence that SES and the diffusion of health information are related and they impact colorectal cancer rates.

Very few studies have focused on the multivariate relationship between online health information seeking, race (or other factors like socioeconomic status or education), and actual health outcomes. Liszka, Steyer, and Hueston (2006) conducted a survey at a family medical practice among English speaking patients, over 18 years of age. The survey asked about their online health information seeking habits. Their sample of 300 people found that 77% accessed the Internet at least once, but those who were more likely to have experience on the Internet were non-Hispanic white patients, less than 50 years old, with both higher education and income. Those with a higher level of experience reported higher self-rated health and had no high risk health factors. However, this is a sample of people who have already sought out health care, and the analysis doesn't focus specifically on race or the impact of Internet usage on actual health outcomes.

1.5 The Hispanic Paradox

It is important to discuss the commonly studied Hispanic Paradox as a primary focus of this research is the Hispanic population. The Hispanic Paradox is seen when Hispanics live longer than non-Hispanic whites despite higher risk factors for certain diseases and an overall lower socioeconomic status (Inosa-Medina, Jean, Cortes-Bergoderi, and Lopez-Jimeneze 2014). Jose Inosa-Medina, et al. (2014) do a content analysis of nine influential studies that focused specifically on health differences of Hispanics and their incidence of heart disease and rates of mortality compared to whites and non-Hispanic blacks. They found that of these nine studies, eight of them found evidence in support of the Hispanic Paradox. Although the exact reason for this paradox still remains largely unknown, there is speculation and some evidence that it is likely because of nutritional differences in diet, geographic differences and/or genetic or psychosocial differences from non-Hispanic whites and blacks.

The Hispanic Paradox, however, becomes increasingly complex. Hummer, Rogers, Amir, Forbes, and Frisbie (2000) conduct an analysis to uncover the differences among sub-categories of Hispanics. They look at how mortality from a variety of factors including circulatory diseases, cancers, external causes, and other causes, varies by the following sub-groups of Hispanics: Puerto Ricans, Mexicans, Cubans, Central/South Americans, and other Hispanics. They also examine if mortality for these groups depends on whether or not someone is born in the United States or abroad. They find the paradox varies across each sub-category when compared to non-Hispanic whites; Puerto Ricans have a clear disadvantage, Central/South Americans have a clear advantage, and Mexican, Cubans, and other Hispanics have mortality rates about equal to whites. These findings are particularly interesting when looking at the Mexican Hispanics as they typically experience the most disadvantage socioeconomically yet still experience the same mortality risk as non-Hispanic whites. This analysis indicates that treating Hispanics as a homogeneous category may in fact mask the heterogeneity which shows true differences among sub-categories of Hispanics.

Borrell and Dallo (2008) examine the differences in Hispanic sub-groups in slightly alternative ways in the NHIS data set of 127,596 adults 18 and up from years 2000, 2001, 2002, and 2003. They look at how self-rated health varies among non-Hispanic blacks and whites as well as Hispanic blacks and Hispanic whites. They find interesting differences among these four groups. Specifically, they find black Hispanics are more likely to rate their health as fair or poor than non-Hispanic whites. However, black Hispanics and non-Hispanic blacks do not have much difference in the way they rate their health, but Hispanic whites are less likely to rate their self-rated health as poor when compared to Hispanic blacks and blacks. The authors conclude that perhaps lumping black and white Hispanics into one group ignores important differences in how these groups experience disadvantage and discrimination. Subsequently, this is important to keep in mind with this analysis as these nuances are not considered in the analysis perhaps leading to misleading results regarding Hispanics.

This literature review has examined whether or not Internet access and usage has proven strong enough to overcome the Fundamental Causes of inequality and have a positive impact on health outcomes. The literature shows online health information seeking is already closely linked to higher education (Lorence and Park 2007, Suziedelyte 2012, Ybarra and Suman 2006, Liszka, Steyer, and Hueston 2006) and higher income (Lorence, Park, and Fox 2006; Liszka, Steyer, and Hueston 2006; Link, Phelan, Miech, and Weston 2008; Wang, Clouston, Rubin, Colen and Link 2012). The majority of prior research (Ono and Zavodny 2000; Ybarra and Suman 2006; Masters, Link and Phelan 2015) also indicates racial differences do indeed exist when it comes to online health information seeking behaviors. However, many of these studies were conducted almost 10 years ago. One of the more recent studies conducted by Chen and Lee (2014), found there were no significant differences in online health information seeking behaviors and race. However, the newest study by Masters, Link and Phelan (2015) specifically includes race in the analysis to start to understand whether or not race can be included as a Fundamental Cause; the researchers find evidence supporting that it is indeed a Fundamental Cause. There has not been much research related to how utilizing online health information impacts health outcomes; a couple of studies have shown a relationship between online health information seeking and positive self-reported health (Chen and Lee 2014; Liszka, Steyer, and Hueston 2006), while some have shown greater likelihood to seek online health information if you have a known medical condition (Fox and Duggan 2013, Ybarra and Suman 2006, Rice 2006) or issues accessing care (Amante, Hogan, Pagoto, English, and Lapane 2015, Rice 2006). Therefore, perhaps more recent data and a focus specifically on online health information seeking, race, and health outcomes will contribute new insights into health disparities.

Additionally, this literature review highlights important considerations when it comes to the Hispanic population. It will be important when analyzing the results to consider the impact of the Hispanic Paradox on Hispanic health outcomes; as studies show despite disadvantages, Hispanics often experience lower rates of mortality than non-Hispanic whites (Inosa-Medina, Jean, Cortes-Bergoderi, and Lopez-Jimeneze 2014). However, it is also important to keep in mind the limitations of this research, namely that the Hispanic category is not examined by sub-group, consequently important variations may be hidden by treating this group as a homogeneous population (Borrell and Dallo 2008, Hummer, Rogers, Amir, Forbes, and Frisbie 2000).

There is support indicating that the Internet can provide access to more information on managing chronic conditions or overall healthier living, which could have an impact on health outcomes. However, Link and Phelan’s Fundamental Cause Theory argues that regardless of positive interventions Fundamental Causes of inequality always prove stronger. For example, despite the potential positive interventions or information the Internet could offer, marginalized people will face barriers to access and use as well as potentially interpret the information they find. If the consumer does not know how to evaluate information coming across the Internet it could increase health disparities. It is important to understand if online health information seeking has had any positive impact on health outcomes for those who have to face Fundamental

Causes of inequality. There is a clear gap in the current literature exploring the relationship of online health information seeking, race, and health outcomes; therefore, the following sections outline the primary research that was conducted, and the results that followed, to further explore these relationships.

3 HYPOTHESES

Based on the literature review the following hypotheses have been formed and were tested with the data analysis:

H1- Blacks and Hispanics will have a higher incidence of diabetes, heart disease, lung disease, high blood pressure and worse self-rated health than whites.

H2- As online health information seeking increases the incidence of diabetes, heart disease, lung disease, high blood pressure, and negative self-rated health of all races will improve.

H3- When controlling for online health information seeking the relationship between race and health will get smaller but will not be eliminated.

H4- Online health information seeking will interact with race such that whites who use online health information seeking will be associated with a smaller incidence of diabetes, heart disease, lung disease, high blood pressure, and negative self-rated health than blacks or Hispanics who use online health information seeking.

H5- Controlling for age, gender, marital status, and income, online health information seeking use will remain significantly associated with race and diabetes, heart disease, lung disease, high blood pressure, and negative self-rated health.

4 RESEARCH METHODS

4.1 Data

This research utilizes secondary data from the Pew Internet & American Life survey, which is a longitudinal survey asking Americans questions about their Internet use and a variety of other topics. The data was found on the Pew Research website (pewInternet .org) and was obtained through the public use data set download option. Information on the usage of the data was provided to the Pew Internet organization in return for the data set download.

The data was collected from a sample of American adults 18 and older, which is derived through random digit dialing of a random sample of landline and cell phone numbers. A variety of sampling procedures were used in order to obtain the best possible random sample of the general population. Interviewers attempted to reach respondents on a variety of days at various times throughout the day; each number was tried seven times. I used the 2012 data set, which consists of 3,014 respondents prior to weighting.

The data set was weighted in a few ways. The data was first weighted based on household size as those who are part of a larger household are less likely to be selected. Then, the sample was weighted based on U.S. Census population parameters, which are: gender by age; gender by education; age by education, region, race and Hispanic origin that includes a break for Hispanics based on whether or not they were born in the U.S. or not; population density and among non-Hispanic whites – age, education and region. Although these processes cannot account for all non-response bias, the weighting does ensure that the sample represents the total U.S. population and controls for any error in sampling as much as possible. This weight will need to be used during analysis to correct for these issues. The final sample size after the weights are applied is n=18,322.

For the purpose of this research the questions from the Internet use and the health information section from the 2012 survey have been used. The data set from 2012 was used as it is the last year that both data on health and Internet health-seeking behaviors were asked of the same group of respondents. Health status questions were asked of all respondents, while health information seeking questions were asked of those respondents who say they use the Internet at least occasionally. Data for those who have indicated they use the Internet at least occasionally was utilized. While a longitudinal analysis would provide additional insight into how behaviors have changed over time, due to time and other limitations, this research focused specifically on understanding the link between online health information seeking and health outcomes at a single point in time.

Additionally, Asians and Native Americans were excluded from the analysis. While it would be beneficial and interesting to look at the relationship between all racial/ethnic groups, health information seeking, and health outcomes, it was determined for this analysis it was best to limit the scope. Masters, Link, and Phelan’s (2015) research is one of the first attempts at expanding Fundamental Cause Theory to include race, and it focuses only on white and black non-Hispanic people. Until the research is further developed around the majority of racial/ethnic groups, it has been decided this analysis will be limited to non-Hispanic whites and blacks and Hispanics.

Before starting the analysis it was important to explore the data to understand who was represented in the data and how it changed once those who do not use the Internet were removed. Additionally, since list-wise deletion occurred if there were any missing variables, it was also important to understand how this impacted the sample. First, all missing data was deleted from the analysis data set list-wise. Then, anyone who does not use the Internet at least occasionally was excluded from the data set as well as Asians and Native Americans. The final analysis data set included 10,919 cases. The demographic composition before and after removing the missing cases is shown in Table 4.1.1 below. The results show very few differences, which shows there is nothing to indicate the reduced analysis data set greatly differs from the original.

Table 4.1.1: Demographics by Data Set

	Demographics	Original Data Set, Internet Users n=13,016	Missing, Asians, and Native Americans Excluded, Internet Users n=10,919
Education	Less than High School	1.7%	1.6%
	Some High School Education	4.2%	4.1%
	High School Graduate (Reference)	27.8%	25.8%
	Some College	21.4%	21.2%
	Two Year Associates Degree	10.8%	11.7%
	Four Year Degree	19.3%	20.1%
	Some Post Graduate Education	.8%	.9%
	Post Graduate Degree	13.7%	14.4%
Sex	Male (Reference)	48.2%	48.8%
	Female	51.8%	51.2%

Race	White, non-Hispanic (Reference)	75.7%	75.9%
	Black, non-Hispanic	11.5%	12%
	Hispanic	12.8%	9.4%
Marital Status	Currently Married or Living with Partner (Reference)	61%	62.3%
	Formerly Married	13.1%	12.6%
	Never Married	25.9%	25.2%
Annual Income	Less than $30,000	30.1%	29.7%
	$30,000 to less than $75,000 (Reference)	39.6%	39.1%
	$75,000 to more than $150,000	30.3%	31.3%
Age	Average Age	43.64	42.95

4.2 Constructs

4.2.1 Dependent Variables

The focus of this research was to examine how utilizing online health information impacts health outcomes. In order to measure health, both a respondents’ self-rated health, along with his or her reported disease incidence, was taken into account.

Self-rated health has been found to be an important indicator of an individual's overall well-being (CDC 2000). There is some research that finds this type of health measurement is less valid at actually predicting disease risk or mortality, but it still provides important information on the quality of life and overall well-being a person feels (CDC 2000). Self-rated health was measured by the respondents' answer to a scale variable that asks, "In general, how would you rate your own health — excellent, good, only fair, or poor?" The options, "Do not know" or "Refused" were recoded and excluded from analysis. This variable was transformed into a dichotomous variable that included "Excellent" or "Good" as a "0" for having good self-rated health and only fair and poor as "1" for the presence of poor self-rated health.

Chronic health problems were measured by the responses to the nominal and dichotomous variable asking, "Are you now living with any of the following health problems or conditions?" The selection options include the diseases that most often lead to an increased risk of mortality (CDC 2013); Diabetes or sugar diabetes, high blood pressure, asthma, emphysema or other lung conditions, heart disease, heart failure or attack, cancer, or any other chronic health condition not mentioned. Cancer was excluded because of the very small sample sizes. Each chronic condition was measured as a dummy variable with "1" indicating presence of the condition. This allowed a comparison of differences in health information utilization not only across race but also across disease type.

Table 4.2.1: Dependent Variables

Variable	Question	Choice Categories	Recoded Categories
Poor Health	Switching topics…In general, how would you rate your own health — excellent, good, only fair or poor?	Excellent Good Only fair Poor Do not Know Refused	0= excellent and good (Reference Category) 1= only fair and poor Excluded= do not know/refused
Chronic Diseases	Are you now living with any of the following health problems or conditions? a. Diabetes or sugar diabetes b. High blood pressure c. Asthma, bronchitis, emphysema, or other lung conditions d. Heart disease, heart failure or heart attack e. Cancer (excluded)	Yes No Do not know Refused	0= No (Reference Category) 1=Presence of disease Excluded = Do not know/refused

4.2.2 Primary Independent Variables

There are two questions that asked about using online health resources for health information, both of which measure a variety of online health related actions respondents could take. They are both nominal, categorical variables. The first asks, "Apart from looking for information online, there are many different activities related to health and medical issues a person might do on the Internet. I'm going to read a list of online health-related activities you may or may not have done in the last 12 months. Just tell me if you happened to do each one or not." The respondents select all options that apply, including: "Signed up to receive email updates or alerts about health or medical issues;" "Read or watched someone else's commentary or personal experience about health or medical issues online:" and/or "Gone online to find others who might have health concerns similar to yours, download forms online, or applied for health insurance online, including private insurance, Medicare, or Medicaid;" Respondents who completed these activities simply went online looking for more health information, but did not engage actively with others. Therefore, this group of activities will be called "Passive Online

Health Seeking Activities." First, each of the activities had to be recoded into a dichotomous dummy variable; a response indicating they had completed the activity was coded as a "1" while a lack of the activity was coded as a "0" and "Do not know" and "Refused" were excluded. Then, a new variable was computed to count the total number of each incidence and then was treated as a continuous variable for the analysis. The more online health seeking behaviors respondents indicate they have participated in the higher their score. The maximum health information seeking score for this variable is four and the minimum is zero. The reference category for the regression analysis was those who had not participated in any health information seeking activities.

The second variable included to measure online health information seeking used the following question: "Thinking again about health-related activities you may or may not do online have you…" and provides the following answer options: "Consulted online rankings or reviews of doctors or other providers;" "Consulted online rankings or reviews of hospitals or other medical facilities;" "Consulted online reviews of particular drugs or medical treatments;" "Posted a review online of a doctor;" "Posted a review online of a hospital;" and/or "Posted your experiences with a particular drug or medical treatment." These activities require more effort of the respondent such as actually sharing their own person experience or digging deeper into a medical treatment type, doctor, or drug, thus they are labeled as "Active Health Seeing Activities." Each response category was transformed into a dichotomous dummy variable where "1" indicates the response of "Yes, they had completed the activity" and "0" if they had not. "Do not know" and "Refused" were recoded to "missing." Then, a count was computed to measure the total number of activities the respondent completed with zero being the minimum and four or more activities being the maximum. The reference category for the logistic regression analysis included people who did not participate in health information seeking activities.

Table 4.2.2: Primary Independent Variables

Variable	Question	Choice Categories	Recoded Categories
Passive Health Information Seeking Activities	Apart from looking for information online, there are many different activities related to health and medical issues a person might do on the Internet. I'm going to read a list of online health-related activities you may or may not have done in the last 12 months. Just tell me if you happened to do each one or not. First, in the last 12 months, have you... [INSERT ITEM; RANDOMIZE]? In the last 12 months, have you...[INSERT ITEM]? a. Signed up to receive email updates or alerts about health or medical issues b. Read or watched someone else's commentary or personal experience about health or medical issues online. c. Gone online to find others who might have health concerns similar to yours d. Download forms online or applied for health insurance online, including private insurance, Medicare, or Medicaid	Yes No Do not Know Refused	Excluded= Do not know/refused New Variable with scale from 0-4 Reference category are those who do not participate
Active Health Information Seeking Activities	Thinking again about health-related activities you may or may not do online, have you… a. Consulted online rankings or reviews of doctors or other providers b. Consulted online rankings or reviews of hospitals or other medical facilities c. Consulted online reviews of particular drugs or medical treatments d. Posted a review online of a doctor e. Posted a review online of a hospital f. Posted your experiences with a particular drug or medical treatment	Yes No Do not know Refused	Excluded= do not know/refused New Variable with scale from 0-4 Reference category are those who do not participate

4.2.3 Secondary Independent Variables

In addition to measuring online health information seeking race is also an important factor in understanding the relationship between online health information seeking and health outcomes. Race was measured using the questions: "What is your race? Are you black, white, Asian, or some other race?" As well as the question, "Are you, yourself, of Hispanic or Latino origin or descent such as Mexican, Puerto Rican, Cuban, or some other Latin American background?"

These variables were then manipulated into the following dichotomous categories: non-Hispanic white, black and non-Hispanic, and Hispanic. All of those reporting “Do not Know,” “Refused” or “Other” were excluded from the analysis. White, non-Hispanics were utilized as the reference category for the logistic regression analysis.

Table 4.2.3: Secondary Independent Variables- Race

Variable	Question	Choice Categories	Recoded Categories
Hispanic	Are you, yourself, of Hispanic or Latino origin or descent such as Mexican, Puerto Rican, Cuban, or some other Latin American background?	1 Yes 2 No 8 Do not know 9 Refused	1= Hispanic 2= non-Hispanic Do not know and Refused excluded
Race	What is your race? Are you white, black, Asian, or some other race?	1 White 2 Black or African-American 3 Asian or Pacific Islander 4 Mixed race 5 Native American/American Indian 6 Other 8 Do not know 9 Refused	White, non-Hispanic (Reference Category) Black, non-Hispanic Hispanic Other, Do not know, Refused excluded

4.2.4 Control Variables

For the research it was important to control for other demographic variables so an understanding of the true relationship of race, online health information seeking, and health outcomes could be acquired. This allowed me to indicate how much of the equation could be explained by these variables and not just the independent and dependent variables. Age was measured using the scale variable for birth year. While gender was indicated by the dichotomous variable indicating self-reported male or female, both of these excluded responses that were “Do not know” or “Refused.” During regression analysis, female was utilized as the reference category. Socioeconomic status was measured by income by utilizing reported household income, which was recoded into three categories: under $10,000 to under $30,000; $30,000 to under $75,000; and $75,000 to over $150,000. Education level was also used as an additional measure to represent socioeconomic status; each level of education was transformed into a dummy variable with the following categories: “Less than High School;” “Some High School;” “High School Diploma;” “Two Year Degree;” “Some College;” “Four Year Degree;” “Some Post Grad;” and “Post Grad Degree.” Those having a high school degree were utilized as the reference category for the regression analysis. Finally, marital status was included and measured by the question, “Are you currently married, living with a partner, divorced, separated, widowed, or have you never been married?” This variable was transformed into a dichotomous dummy variable with one group being those who are currently married (includes “married,” or “living with partner”), one group formerly married (includes “divorced,” “separated,” or “widowed”), and a group for those who never have been married (“never married” or “single”). “Do not know” and “Refused” answers were excluded from the analysis for all variables and “Currently Married” was utilized as the reference category.

Table 4.2.4: Control Variables

Variable	Question	Choice Categories	Recoded Categories
Income	Last year, that is in 2011, what was your total family income from all sources before taxes? Just stop me when I get to the right category...	1 Less than $10,000 2 $10,000 to under $20,000 3 $20,000 to under $30,000 4 $30,000 to under $40,000 5 $40,000 to under $50,000 6 $50,000 to under $60,000 7 $60,000 to under $75,000 8 $75,000 to under $100,000 9 $100,000 to over $150,000 98 Do not know 99 Refused	Less than $30,000 a year Exactly $30,000 to less than $75,000 a year (Reference Category) $75,000 to more than $150,000 a year Excluded 98 and 99
Education	What is the last grade or class you completed in school?	1 Less than High School 2 Some High School Education 3 High School Graduate 4 Some College 5 Two Year Associates Degree 6 Four Year Degree 7 Some Post Graduate Education 8 Post Graduate Degree	Less than High School Some High School Education High School Graduate (Reference Category) Some College Two Year Associates Degree Four Year Degree Some Post Graduate Education Post Graduate Degree

Age	What is your age?	Age in years	Scale of Age
Gender	Are you?	Male Female	0=Female (Reference Category) 1=Male
Marital Status	Are you currently married, living with a partner, divorced, separated, widowed, or have you never been married?	1 Married 2 Living with a partner 3 Divorced 4 Separated 5 Widowed 6 Never been married 7 **(VOL.)** Single 8 **(DO NOT READ)** Do not know 9 **(DO NOT READ)** Refused	Never married Formerly Married Currently Married (Reference Category) Excluded 8 and 9

4.3 Analytical Plan

In order to gain a better understanding of the relationships between the dependent and independent variables Binary Logistic Regression analysis was conducted. With logistic regression we can predict the odds of the dependent variable occurring, or overall heath, based on the presence of the independent variable, which in this study is online health information seeking behaviors. Binary logistic regression is the most appropriate type because the primary dependent variable, in this case health outcomes, is dichotomous. One either has one of the chronic diseases, or they do not; there is not scale associated with incidence. It is also important to note, that while self-reported health was a liner variable, the majority of the other health variables were dichotomous, leading me to determine that utilizing binary logistic regression in this thesis was most appropriate. Self-rated health is only one indicator of health, while four others were also utilized as dependent variables and they are all binary variables. There are certain assumptions made in OLS regression that cannot be made with a dichotomous dependent variable; namely, the values have a normal distribution and there is homoscedasticity within the sample. Utilizing Binary Logistic Regression works around these assumptions to provide a better model for binary data by unbounding the probability so there is no upper or lower limit (Allison 2012).

It is important to note there are some assumptions that must be considered when choosing to use logistic regression; the dependent variables are all binary. In order to interpret how the independent variable affects the dependent we will look at the adjusted odds ratios to explain the likelihood of the health event occurring with the presence of the independent variable with all other variables held constant. The strength of each model will be measured by the Nagelkerke R Square, which is interpreted by evaluating how close the statistic is too "1:" the closer to one, the stronger the model. Additionally, each odds ratio will be determined to be significant by its p-value.

The analysis will be presented in four models; the first model will evaluate the relationship between race and chronic and self-rated health outcomes.

$$log(Chronic\ Health\ Conditions) = a + b_1(Race - Black) + b_2(Race - Hispanic)$$

$$log(Self - rated\ Health) = a + b_1(Race - Black) + b_2(Race - Hispanic)$$

Where "a" is logit for the average health of whites, "b_1" is the deviation of blacks' average health compared to whites' and where "b_2" is the deviation of Hispanics' average health compared to whites'.

The second model added the passive and active health information seeking activities.

$$log(Chronic\ Health\ Conditions) = a + b_1(Race - Black) + b_2(Race - Hispanic) + b_3(passive\ online\ health\ information\ seeking) + b_4(active\ online\ health\ information\ seeking)$$

$$log(Self - rated\ Health) = a + b_1(Race - Black) + b_2(Race - Hispanic) + b_3(passive\ online\ health\ information\ seeking) + b_4(active\ online\ health\ information\ seeking)$$

Where “a” is logit for the average health of whites who do not participate in any health information seeking activities, “b_1” is logit for the main effect of blacks’ health compared to whites’ and where “b_2” is the main effect of Hispanics’ average health compared to whites’. This model also includes “b_3,” which is the logit for passive health seeking activities when race is held constant and “b_4” is the logit for active online health seeking activities.

The third model looked at the interaction between race and online health information seeking. The equations are as follows:

$$\log(Chronic\ Health\ Conditions) = \alpha + b\_1\ (Race - Black) + b\_2\ (Race - Hispanic) + b\_3\ (passive\ online\ health\ information\ seeking) + b\_4\ (active\ online\ health\ information\ seeking) + b\_5\ (Black * passive\ online\ health\ information\ seeking) + b\_6\ (Black * active\ online\ health\ seeking\ activities) + b\_7\ (Hispanic * passive\ online\ health\ information\ seeking) + b\_8\ (Hispanic * active\ online\ health\ information\ seeking)$$

$$log(Self - rated\ health) = a + b\_1\ (Race - Black) + b\_2\ (Race - Hispanic) + b\_3\ (passive\ online\ health\ information\ seeking) + b\_4\ (active\ online\ health\ information\ seeking) + b\_5\ (Black * passive\ online\ health\ information\ seeking) + b\_6\ (Black * active\ online\ health\ seeking\ activities) + b\_7\ (Hispanic * passive\ online\ health\ information\ seeking) + b\_8\ (Hispanic * active\ online\ health\ information\ seeking)$$

Where “a” is logit for the average health of whites who do not participate in any health seeking activities, “b_1” is the main effect of blacks’ average health compared to whites’ and where “b_2” is the main effect of Hispanics’ average health compared to whites’. The outcomes for race in this model is when the online health information seeking is valued at zero or there is no online health info seeking. “b_3” is the main effect of passive online health seeking behavior for whites and “b_4” is the main effect of active online health seeking for whites. “b_5” and “b_6” are the interaction effects of being black and engaging in passive or active online health seeking activities, respectively. Finally, “b_7” and “b_8” are the interaction effects of being Hispanic and engaging in passive or active online health seeking activities, respectively.

Finally, the fourth model added in all of the controls:

$$log(Chronic\ Health\ Conditions) = a + b_1(Race - Black) + b_2\ (Race - Hispanic) + b_3(passive\ online\ health\ information\ seeking) + b_4(active\ online\ health\ information\ seeking) + b_5(Black * passive\ online\ health\ information\ seeking) + b_6(Black * active\ online\ health\ seeking\ activities) + b_7(Hispanic * passive\ online\ health\ information\ seeking) + b_8(Hispanic * active\ online\ health\ information\ seeking) + b_9(income) + b_{10}(age) + b_{11}(education) + b_{12}(marital\ status)$$

$$log(Self - rated\ health) = a + b_1(Race - Black) + b_2\ (Race - Hispanic) + b_3(passive\ online\ health\ information\ seeking) + b_4(active\ online\ health\ information\ seeking) + b_5(Black * passive\ online\ health\ information\ seeking) + b_6(Black *$$

$$active\ online\ health\ seeking\ activities) + b_7(Hispanic *$$

$$passive\ online\ health\ information\ seeking) + b_8(Hispanic *$$

$$active\ online\ health\ information\ seeking) + b_9(income) + b_{10}(age) +$$

$$b_{11}(education) + b_{12}(marital\ status)$$

Where “a” is logit for the average health of whites who do not participate in any health information seeking activities, “b_1” is the main effect of blacks’ average health compared to whites’ and where “b_2” is the main effect of Hispanics’ average health compared to whites’. The outcomes for race in this model are when the online health information seeking is valued at zero or there is no online health info seeking. “b_3” is the main effect of passive online health seeking behavior for whites and “b_4” is the effect of active online health seeking for whites. “b_5” and “b_6” are the interaction effects of being black and engaging in passive or active online health seeking activities, respectively. Finally, “b_7” and “b_8” are the interaction effects of being Hispanic and engaging in passive or active online health seeking activities, respectively. “b_9” is the added effect of income, “b_{10}” the added effect of age, “b_{11}” the added effect of education, and “b_{12}” the added effect of marital status to the equation.

Binary Logistic Regression takes these equations, assumes there is non-linear relationship between the independent and dependent variables, and then estimates the coefficients by using maximum likelihood estimation. This equation gives us the odds ratios, or the likelihood of the dependent variable equaling one with the presence of the independent variable and all other variables in the model held constant. Adding control variables in steps will allow us to understand how the addition of each variable increases or decreases the likelihood of the dependent variable occurring with the presence of the control or independent variables with all other variables held constant (Allison 2012).

The strength of each model will be measured by observing Nagelkere's R squared, which, when interpreted, tells us the closer the R squared to 1, the stronger the model. Finally, we will evaluate the "P" value of each model, which will describe how confident we can be the dependent and independent variables have a relationship with each other. After completing the analysis and interpreting the statistics the hypotheses will be either supported or not supported based on the data on the strength of the associations and our determined confidence in the model (Allison 2012).

5 RESULTS

5.1 Univariate Analysis

Before the more advanced binary regression it was important to gain a better understanding of the make-up of the sample. The table below indicates the percentage of the total sample that participates in passive or active health seeking activities, the racial mix of the respondents, and the incidence of each health condition.

Table 5.1.1: Univariate Analysis

Race, Health Information Seeking, and Health Conditions		n count	Percentage of total sample
Passive Health Seeking Activities	0 passive activities	6456	59.2%
	1 passive activity	2429	22.2%
	2 passive activities	1363	12.5%
	3 passive activities	529	4.8%
	4 or more passive activities	142	1.3%
Active Health Seeking Activities	0 passive activities	7270	66.5%
	1 active activity	1839	16.8%

	2 active activities	1069	9.8%
	3 active activities	539	4.9%
	4 or more active activities	202	1.8%
Race	White	8290	75.9%
	Black, non-Hispanic	1309	12.0%
	Hispanic	1023	9.4%
Health Conditions	Poor Self-rated Health	1542	14.1%
	Heart Disease	529	4.8%
	Diabetes	1309	12.0%
	Lung Disease	1251	11.5%
	High Blood Pressure	2228	20.4%
Education	Less than High School	177	1.6%
	Some High School	446	4.1%
	High School Graduate	2822	25.8%
	Some College	2315	21.2%
	Two Year Degree	1272	11.7%
	Four Year Degree	2191	20.1%
	Some Post Graduate Education	103	.9%
	Post Graduate Degree	1574	14.4%
Gender	Male	5333	48.8%
	Female	5586	51.2%
Marital Status	Currently Married	6797	62.3%
	Never Married	2747	25.2%
	Formerly Married	1374	12.6%
Income	Less than $30,000	3242	29.7%
	Exactly $30,000, but less	4264	39.1%

	than $75,000		
	Exactly $75,000 to more than $150,000	3413	31.3%
Age	Average Age	10919	42.95

Table 5.1.1 shows the frequencies of the key dependent and independent variables of the entire sample. The table indicates the most common health condition in the sample is high blood pressure with 20.4% of the sample reporting this condition. Cancer is the least common health condition with only 3.1% of the sample reporting this disease. The sample is majority white (75.9%), followed by 12% black, non-Hispanic, and 9.4% Hispanic. The majority of the sample does not participate in any health information seeking activities. 22.2% of respondents said they participate in one passive health seeking activity and 18.6% in more than one. 16.8% said they participate in at least one active activity and 1.5% participates in more than one active activity.

5.2 Bivariate Analysis

Before conducting the Binary Logistic Regression it was also important to understand the relationship between the independent and dependent variables in a bivariate analysis. This allowed us to identify relationships between health information seeking, race, and health outcomes before completing the regression. Table 5.2.1 contains the Bivariate Analysis results.

Table 5.2.1: Bivariate Analysis: Race and Internet Health Seeking Activities by Health Conditions

Demographics	**Poor Health**	**Heart Disease**	**Diabetes**	**Lung Disease**	**High Blood Pressure**

Passive Health Seeking Activities	0 activities	16.8%	82.6%	74%	53.1%	36.5%
	1 activity	13.5%	4.7%	6.0%	11.4%	17.3%
	2 activities	11.6%	4.5%	9.0%	10.4%	16.6%
	3 activities	18.7%	4.7%	9.6%	12.3%	16.1%
	4 or more activities	39.4%	3.5%	1.4%	12.8%	13.5%
Active Health Seeking Activities	0 activities	32.6%	74.5%	67.3%	55.2%	20.8%
	1 activity	15.1%	5.1%	5.7%	10.4%	21.8%
	2 activities	10.8%	4.1%	10.7%	8.1%	19.2%
	3 activities	18.7%	6.9%	8.4%	11.9%	15.4%

	4 or more activities	22.8%	9.4%	7.9%	14.4%	22.8%
Race	White, non-Hispanic	13.5%	4.7%	6.3%	10.9%	21.9%
	Black, non-Hispanic	16.3%	4.2%	13.1%	13.4%	22.8%
	Hispanic	14.1%	6.9%	9.0%	9.2%	7.6%

The bivariate table comparing race and health seeking activities with health outcomes indicates the occurrence of health information seeking on reporting a health condition varies by condition. The table also indicates whites are the least likely to report having all diseases when compared to black, non-Hispanics. Whites are also less likely to report most of the diseases, except poor health, and high blood pressure, when compared to Hispanics. Hispanics only report having heart disease more often than black, non-Hispanics. Additionally, there are interesting trends in the data in regards to the number of health seeking activities by condition. Some of these differences could be attributed to the cross-sectional nature of the data, however, there is little in the data to explain why these trends are occurring, which indicates another limitation of this study and data.

5.3 Binary Logistic Regression

Next, the Binary Logistic Regression was completed with a set of models for each dependent variable: poor health, heart disease, diabetes, and lung disease. The results are shown in tables 5.3.1-5.3.5.

5.3.1 Poor Health

Table 5.3.1: Regressing Self-Reported Poor Health on Passive and Active Online Resources
Raw Logits are presented with standard errors in parentheses and odds ratios in brackets.

	Model 1	Model 2	Model 3	Model 4+
Black, non-Hispanic	.20 (.08)* [1.221]	.21 (.08)* [1.234]	.394 (.10)*** [1.483]	.208 (.10)* [1.232]
Hispanic	.02 (.10) [1.020]	.006 (.10) [1.006]	-.268 (.13)* [.765]	-.316 (.14)* [.729]
Passive Internet Health Info Seeking		.094 (.03)* [1.099]	.045 (.04)* [1.089]	.216 (.04)*** [1.241]
Active Internet Health Info Seeking		.023 (.03) [1.023]	.045 (.04) [1.046]	.170 (.04)*** [1.185]
Black Passive			-.498 (.13)*** [.607]	-.571 (.13)*** [.565]
Black Active			.143 (.10) [1.154]	.172 (.11) [1.188]
Hispanic Passive			.483 (.10)*** [1.620]	.321 (.11)* [1.378]
Hispanic Active			-.369 (.12)* [.691]	-.313 (.12)* [.731]
Nagelkerke R Square/Model Strength	.001*	.003***	.011***	.126***

* p<.05 **p<.001 ***p<.0001

Note: reference categories are those that do not engage in passive online activities or active online activities, White, Currently Married, Income $30,000-Less than $75,000/year, High School Diploma, and Female.

+Model four includes all control variables for Age, Gender, Education, Income, and Marital Status.

The first model shown in Table 5.3.1 shows the regression of poor self-rated health on race. Being black, non-Hispanic does have a significant association with having poor health. The odds a person reports poor health are 1.221, or 22% higher, if a person is black versus if he or she is white. The results indicate being Hispanic is not statistically associated with the presence of poor health. The model's Nagelkerke R square indicates this model is not very strong with only .1% of the variability in the dependent variable explained, however, the model is significant at the p<.05 level.

The second model adds in active health seeking activities. Here, the model indicates that, holding active and passive health seeking activities constant, blacks are still 23% more likely to experience poor health than those who are not black, which the "p" value also indicates is significant. The relationship between being Hispanic and reporting poor health remains non-significant. When holding the variables black and Hispanic constant, and controlling for all active health seeking activities, we see a significant association between passive health seeking activities and reporting poor health with a 9.9% increase in likelihood to report poor health for each additional passive health seeking activity. There is no significant association between active health seeking activities and reporting poor health. The Nagelkerke R square for this model is slightly higher, indicating the model explains .3% of variability of the dependent variable. This model is significant at the p<.0001 level.

The third model includes interaction variables for non-Hispanic blacks and Hispanics who conduct health information seeking activities. In this model, non-Hispanic blacks who do not participate in any health seeking activities are significantly associated with reporting poor health with a 48% increased likelihood to do so over whites. Hispanics who do not participate in health seeking activities are 23.5% less likely than whites to report poor health. The main effect of passive health seeking activities for whites is an 8.9% increase in reporting poor health for each additional activity. The main effects of online health seeking activities are not significantly associated with reporting poor health for whites. For non-Hispanic Blacks, each additional passive health seeking activity means they are 39.3% less likely to report poor health than whites, while each additional active health seeking activity is not significantly associated with health for blacks. For each additional passive health seeking activity Hispanics conduct they are 62% more likely to report poor health compared to whites. For each additional active health seeking activity Hispanics conduct they have a 30.9% decreased likelihood in reporting poor health when compared to whites. Once again, this model increases the Nagelkerke R square, which indicates the model explains 1.1% of the variability of poor health. The model is significant at the $p<.0001$ level.

The fourth model includes all Internet health seeking variables as well as the control variables. The control variables, income, marital status, education, and gender are included in this model. Non-Hispanic blacks who do not participate in any health seeking activities are significantly associated with reporting poor health with a 23% increased likelihood to do so than whites. Hispanics who do not participate in health seeking activities are 27% less likely than whites to report poor health. For each additional passive health seeking activity a non-Hispanic black conducts they are 43.5% less likely to report poor health than whites. Blacks who participate in active health seeking activities are not significantly associated with poor health. Hispanics who conduct passive health seeking activities are 37.8% more likely to report poor health, while conducting active health seeking activities leads to a 26.9% decreased likelihood in reporting poor health. The model is significant at the p<.0001 level, with a Nagelkerk R square indicating the model explains 12.6% of the variability of poor health.

5.3.2 Diabetes

Table 5.3.2: Regressing Diabetes on Passive and Active Online Resources
Raw Logits are presented with standard errors in parentheses and odds ratios in brackets.

	Model 1	Model 2	Model 3	Model 4+
Black, non-Hispanic	.781 (.09)*** [2.184]	.783 (.09)*** [2.189]	.517 (.12)*** [1.676]	.708 (.13)*** [2.029]
Hispanic	.361 (.12)* [1.434]	.370 (.12)* [1.448]	.162 (.15) [1.175]	.611 (.17)*** [1.843]
Passive Internet Health Info Seeking		-.036 (.05) [.964]	-.030 (.06) [.970]	.208 (.06)*** [1.231]
Active Internet Health Info Seeking		.088 (.04)* [1.092]	-.056 (.05) [.946]	.088 (.06) [1.092]
Black Passive			-.099 (.12) [.906]	-.248 (.13)* [.780]
Black Active			.489 (.10)*** [1.630]	.526 (.11)*** [1.693]
Hispanic Passive			.049 (.13) [1.050]	-.270 (.14)* [.763]
Hispanic Active			.262 (.14)* [1.299]	.326 (.14)* [1.385]
Nagelkerke R Square/Model Strength	.015***	.016***	.022***	.182***

* p<.05 **p<.001 ***p<.0001

Note: reference categories are those that do not engage in passive online activities or active online activities, White, Currently Married, Income $30,000-Less than $75,000/year, High School Diploma, and Female.

+Model four includes all control variables for Age, Gender, Education, Income, and Marital Status.

Table 5.3.2 shows the regression of diabetes on race/ethnicity. The results indicate being black, non-Hispanic has a highly significant association with reporting diabetes while being Hispanic is has a slight significant association. Being black, non-Hispanic, means someone is 1.18 times more likely to report having diabetes than whites while Hispanics are 43.4% more likely to report having diabetes than whites. The Nagelkerke R square of this model indicates 1.5% of the dependent variable is explained by the dependent variables. The model is significant at the p<.0001 level.

The second model adds in active and passive health seeking activities. Here, the model indicates that holding active and passive health seeking activities constant, blacks are still 1.189 times more likely to report experiencing diabetes than those who are white with a highly significant association. The relationship between being Hispanic and reporting diabetes is slightly less significant but still indicates Hispanics are 44.8% more likely to report experiencing diabetes. Participating in passive health seeking activities is not significantly associated with reporting diabetes. However, when holding race constant, participating in active health seeking activities is significantly associated with reporting the disease with a 9.2% increased likelihood for each additional activity. This model only explains slightly more of the variability of the dependent variable, “Diabetes”, with a Nagelkerke R square of 1.6%. The model is significant at the p<.0001 level.

The third model includes the interaction variables. Non-Hispanic blacks, when holding health seeking activities constant, are 67.6% more likely to report diabetes than whites. When holding health seeking activities constant, being Hispanic is not significantly associated with reporting diabetes. Blacks' participation in passive health seeking activities is not significantly associated with the likelihood to report diabetes. However, for each additional active health seeking activity blacks participate in leads to a 63% increased likelihood of reporting diabetes when compared to whites. Being Hispanic and participating in active health information seeking activities has a slight significant association with each additional active activity Hispanics are 29.9% more likely to report diabetes than whites. The Nagelkerke R square of this model indicates the independent variables explain 2.2% of the dependent variable's variability. The model is significant at the p<.0001 level.

The fourth model includes all Internet health seeking variables as well as the control variables. This model indicates that, holding all other variables constant, the significant associations between being black or Hispanic and reporting diabetes remains significant for blacks and becomes highly significant for Hispanics. The model shows black, non-Hispanics are still 2.029 times more likely to report having diabetes and Hispanics are 1.843 times more likely than whites when all other variables are held constant. For every additional passive health seeking activity whites are 23.1% more likely to report diabetes, while whites' participation in active health seeking activities is not significantly associated with reporting diabetes. Blacks who participate in passive health seeking activities are 22% less likely to report diabetes, for each additional activity, compared to whites. For each additional active health seeking activity blacks participate in they have a 69.3% increased likelihood to report diabetes than whites. Hispanics who participate in passive health seeking activities are 23.7% less likely for each additional activity to report diabetes than whites. Conversely, Hispanics are 38.5% more likely for each active activity to report diabetes than whites. This group of independent and control variables explain 18.2% of diabetes' variability with an overall model significance at the p<.0001 level.

5.3.3 Heart Disease

Table 5.3.3: Regressing Heart Disease on Passive and Active Online Resources
Raw Logits are presented with standard errors in parentheses and odds ratios in brackets.

	Model 1	Model 2	Model 3	Model 4+
Black, non-Hispanic	-.107 (.15) [.899]	-.108 (.15) [.897]	-.191 (.18) [.827]	.021 (.19) [1.022]
Hispanic	.417 (.13)* [1.517]	.450 (.13)** [1.568]	-.216 (.20) [.805]	.274 (.21) [1.316]
Passive Internet Health Info Seeking		-.157 (.10)* [.855]	-.284 (.07)*** [.753]	-.086 (.07) [.918]
Active Internet Health Info Seeking		.189 (.05)*** [1.209]	.158 (.06)* [1.171]	.349 (.06)*** [1.418]
Black Passive			-.194 (.23) [.824]	-.320 (.23) [.726]
Black Active			.250 (.16) [1.284]	.231 (.17) [1.259]
Hispanic Passive			.631 (.14)*** [1.880]	.243 (.17) [1.275]
Hispanic Active			.091 (.14) [1.095]	.198 (.16) [1.219]
Nagelkerke R Square/Model Strength	.003*	.007***	.018***	.172***

* p<.05 **p<.001 ***p<.0001

Note: reference categories are those that do not engage in passive online activities or active online activities, White, Currently Married, Income $30,000-Less than $75,000/year, High School Diploma, and Female.

+Model four includes all control variables for Age, Gender, Education, Income, and Marital Status.

Table 5.3.3 show the regression of heart disease on race/ethnicity. The results show being black, non-Hispanic does not have a significant association with reporting heart disease. Being Hispanic has a slightly significant impact on incidence with Hispanics being 51.7% more likely to report heart disease. This model only explains .3% of the dependent variability, according to the Nagelkerke R square. However, the model is significant at the p<.05 level.

The second model adds in active and passive health seeking activities. The relationship between being black and having heart disease remains non-significant. The model indicates that, holding active and passive health seeking activities constant, being Hispanic becomes more significantly associated with heart disease than those who are not Hispanic, with 20.9% higher likelihood to report the disease. Holding race and active health seeking activities constant, those who participate in passive Internet health seeking activities, with a slightly significant association, are 14.5% less likely to report the disease for each additional activity. Participating in active Internet health seeking activities has a highly significant 20.9% increased likelihood to report heart disease for each additional health seeking activity. The second models Nagelkerke R square indicates there is only a slight increase in the explanation of the variability of the dependent variable to .7%, however, the model is significant at the p<.0001 level.

The third model includes the interaction variables. Being a non-Hispanic black when holding health seeking activities constant, is not significantly associated with reporting heart disease more or less often than whites. When holding health seeking activities constant being Hispanic is not significantly associated with reporting heart disease. The model also indicates significant associations between being white and participating in active and passive health seeking activities; whites who participate in passive health seeking activities are 24.7% less likely to report heart disease, for each additional activity, and those who participate in active activities are 17.1% more likely to report heart disease for each additional activity. Being black and participating in active or passive health information seeking activities is not significantly associated with a difference in likelihood to report heart disease. Being Hispanic and participating in passive health information seeking activities is significant, and this interaction leads to an 88% increase in likelihood to report heart disease with each additional activity. The interaction between being Hispanic and participating in active online health information seeking activities does not lead to a significant association. The Nagelkerke R square for this model indicates the independent variables explain 1.8% of the dependent variables variability. The model is significant at the p<.0001 level.

The fourth model includes all Internet health seeking variables as well as the control variables. This model indicates that, holding all other variables constant, there is no significant relationship between being black, non-Hispanic, or Hispanic and reporting heart disease. There is also no significant association among the interaction of being black or Hispanic and participating in any type of health seeking activity. The model does indicate that, holding all other variables constant, whites who participate in active health seeking activities have a 41.8% increased likelihood of reporting heart disease for each additional activity. The model itself is highly significant, and the addition of the control variables helps to explain more of the model for heart disease, which has a larger impact on the disease incidence than race or health seeking activities. This final model explains much more of heart disease's variability with a Nagelkerke R square of 17.2% and a significance level of p<.0001.

5.3.4 Lung Disease

Table 5.3.4: Regressing Lung Disease on Passive and Active Online Resources
Raw Logits are presented with standard errors in parentheses and odds ratios in brackets.

	Model 1	Model 2	Model 3	Model 4+
Black, non-Hispanic	.180 (.09)* [1.197]	.179 (.09)* [1.196]	.314 (.10) * [1.369]	.136 (.10) [1.146]
Hispanic	-.239 (.11)* [.787]	-.248 (.11)* [.780]	-.007 (.138) [.993]	-.228 (.15) [.796]
Passive Internet Health Info Seeking		.038 (.04) [1.038]	.122 (.04)* [1.130]	.180 (.04)*** [1.197]
Active Internet Health Info Seeking		-.076 (.04)* [.927]	-.111 (.04)* [.895]	-.042 (.04) [.959]
Black Passive			-.478 (.14)*** [.620]	-.538 (.14)*** [.584]
Black Active			.232 (.11)* [1.262]	.307 (.31)* [1.360]
Hispanic Passive			-.468 (.15)** [.626]	-.478 (.14)** [.620]
Hispanic Active			.155 (.15) [1.168]	.197 (.16) [1.217]
Nagelkerke R Square/Model Strength	.002*	.003*	.007***	.082***

Note: reference categories are those that do not engage in passive online activities or active online activities, White, Currently Married, Income $30,000-Less than $75,000/year, High School Diploma, and Female.

+Model four includes all control variables for Age, Gender, Education, Income, and Marital Status.

Table 5.3.4 shows the first model regressing lung disease incidence on race/ethnicity. Being black, non-Hispanic, does not have a significant association with lung disease incidence.

The results indicate being Hispanic means one is 22.3% less likely to report lung disease than whites; The Nagelkerke R square indicates this model only explains .3% of the variability of the dependent variable. The model is significant at the p<.05 level.

The second model adds in active and passive health seeking activities. When holding health information seeking activities constant, the relationship between being black and having lung disease is slightly significant with blacks being 19.6% more likely to report having lung disease. Holding active and passive health seeking activities constant being Hispanic remains slightly significantly associated with lung disease with Hispanics 22% less likely to report having the disease. Participating in passive health information seeking activities is not significantly associated with reporting lung disease. However, participating in active health information seeking activities is slightly significant, when holding race constant for each active activity, there is a 7.3% decrease in likelihood of reporting lung disease. This model only explains slightly more of the dependent variability based on a Nagelkerke R square of .3%. The model is significant at the p<.0001 level.

The third model includes the interaction variables. Being a Non-Hispanic black, when holding health seeking activities constant, results in a 36.9% increased likelihood of reporting lung disease. When holding health seeking activities constant being Hispanic is not significantly associated with reporting lung disease. Being white and participating in passive health seeking activities has a significant association with a 13% increase in likelihood for each additional activity. “Whites who participate in active health seeking activities” is also significant and results in a 10.5% decrease for each additional activity. For each additional active or passive health seeking activity conducted by blacks there is a 38% decrease and 26.2% increase in reporting lung disease, respectively. Being Hispanic and participating in passive health information seeking activities is significant, and this interaction leads to a 37.4% decrease in likelihood to report lung disease for each additional activity. The interaction between being Hispanic and participating in active online health information seeking activities does not lead to a significant association. Again, this model only explains .4% more than the previous with a Nagelkerke R square of .7% at a significance level of p<.0001.

The fourth model includes all Internet health seeking variables as well as the control variables. This model indicates that, holding all other variables constant, there is no longer a significant association between being black, non-Hispanic or Hispanic and reporting lung Disease. However, there is a significant association between the interaction of being black and participating in passive and active health seeking activities. For each additional passive health seeking activity blacks are 41.6% less likely to report lung disease, and for each additional active activity they are 36% more likely to report the disease. Hispanics who participate in passive health seeking activities are 38% less likely, for each additional activity, to report lung disease. There is no significant association between the interactions of being Hispanic, participating in active health seeking activities, and reporting lung disease. This model explains 8.2% of the variability of lung disease as a dependent variable with a significance of p<.0001.

5.3.5 High Blood Pressure

Table 5.3.5: Regressing High Blood Pressure on Passive and Active Online Resources
Raw Logits are presented with standard errors in parentheses and odds ratios in brackets.

	Model 1	Model 2	Model 3	Model 4+
Black, non-Hispanic	.072 (.07) [1.075]	.058 (.07) [1.059]	-.232 (.09)* [.793]	.046 (.10) [1.047]
Hispanic	-1.203 (.12)*** [.300]	-1.177 (.12)*** [.308]	-1.500 (.12)*** [.223]	-1.150 (.17)*** [.317]
Passive Internet Health Info Seeking		-.212 (.03)*** [.809]	-.299 (.04)*** [.742]	-.084 (.04)* [.920]
Active Internet Health Info Seeking		.059 (.03)* [1.061]	.049 (.03) [1.050]	.118 (.04)** [1.125]
Black Passive			.366 (.09)*** [1.422]	.234 (.10)* [1.253]
Black Active			.140 (.08) [1.150]	.124 (.09) [1.132]
Hispanic Passive			.649 (.13)*** [1.913]	.445 (.13)** [1.561]
Hispanic Active			-.385 (.16)* [.681]	-.428 (.17)* [.652]
Nagelkerke R Square/Model Strength	.020***	.027***	.035***	.235***

* p<.05 **p<.001 ***p<.0001

Note: reference categories are those that do not engage in passive online activities or active online activities, White, Currently Married, Income $30,000-Less than $75,000/year, High School Diploma, and Female.

+Model four includes all control variables for Age, Gender, Education, Income, and Marital Status.

Table 5.3.5 show the regression of high blood pressure incidence on race/ethnicity. Being black non-Hispanic does not have a significant association with high blood pressure. The results indicate being Hispanic means one is 68.5% less likely to report High Blood Pressure than whites with a highly significant association. The Nagelkerke R square indicates this model explains 2% of the variability of the dependent variable, “High Blood Pressure,” at a significance level of p<.0001.

The second model adds in active and passive health seeking activities. Here, the model indicates that, holding active and passive health seeking activities constant, being Hispanic remains slightly significantly associated with high blood pressure with Hispanics 69.2% less likely to report having the disease than whites. The relationship between being black and having high blood pressure remains non-significant. This model increases the explanation of the high blood pressure to a Nagelkerke R square of 2.7% at a significance of p<.0001.

The third model includes the interaction variables. In this model, being a Non-Hispanic black and reporting high blood pressure becomes significant with a 20.7% decreased likelihood of reporting high blood pressure compared to whites when holding health information seeking activities constant. Being black and participating in passive health seeking activities is significantly associated when compared to whites with a 42.2% increase in likelihood of reporting high blood pressure with each additional activity. Being black and participating in active health seeking activities is not significantly associated with high blood pressure. When holding health seeking activities constant being Hispanic means a 77.7% decreased likelihood in reporting high blood pressure compared to whites. Additionally, being Hispanic and participating in passive health information seeking activities is significant, and this interaction leads to a 91.3% increase in likelihood to report high blood pressure for each additional activity. The interaction between being Hispanic and participating in active online health information seeking activities has the opposite effect with a 31.9% decrease in likelihood of reporting high blood pressure for each additional active health seeking activity. This model explains 3.5% of the variability of the dependent variable as indicated by the Nagelkerke R square. The model is significant at the p<.0001 level.

The fourth model includes all Internet health seeking variables as well as the control variables. The model shows being black non-Hispanics is no longer significant when adding in the control variables. Holding all other variables constant the association between being Hispanic and reporting high blood pressure remains highly significant with Hispanics having 68.3% decreased likelihood to report the disease compared to whites. However, the interaction between being black and participating in passive health seeking activities is significant with a 23.5% increased likelihood in reporting high blood pressure with each additional passive activity. Being Hispanic and participating in active or passive health information seeking is also significantly associated with reporting high blood pressure. Hispanics who participate in passive activities are 56.1% more likely to report high blood pressure for each additional activity participating in active activities leads to a 34.8% decrease for each additional activity. This model explains 23.5% of the variability of the dependent variable at a significance level of p<.0001.

5.3.6 High Level Analysis of Odds Ratios

Table 5.3.6 shows each chronic condition along with the odds ratios from models one through three of each. The purpose of this table is to provide a snapshot look of the impact online health information seeking has on health outcomes across racial groups. It indicates the directional differences outcomes for Hispanics and blacks first when passive and active health information seeking was added and then when all control variables and interactions were added.

Table 5.3.6: Significant Odds Ratios by Chronic Condition

	Self-Rated Health			Diabetes			Heart Disease			Lung Disease			High Blood Pressure		
Model	1	2	4	1	2	4	1	2	4	1	2	4	1	2	4
Black	1.22	1.23	1.23	2.18	2.18	2.02	NS	NS	NS	1.19	1.19	NS	NS	NS	NS
Hispanic	NS	NS	.729	1.43	1.44	1.84	1.51	1.56	NS	.787	.780	NS	.300	.308	.317
Black Passive			.565			.780			NS			.584			1.25
Black Active			NS			1.69			NS			1.36			NS
Hispanic Passive			1.37			.763			NS			.620			1.56
Hispanic Active			.731			1.38			NS			NS			.652

Note: NS means Not Significant

This table shows blacks who participate in passive online health information seeking have a decreased likelihood of reporting poor health, but active activity participation has no significant impact on reporting poor health. The effect on Hispanics is an increase in likelihood to report poor health when participating in passive activities and a decrease when participating in active activities compared to whites. When it comes to diabetes, blacks and Hispanics who participate in passive health seeking activities experience a decreased likelihood to report having the disease when compared to whites. For active health information seeking both blacks and Hispanics experience an increased likelihood to report diabetes when compared to whites. When examining heart disease we see there is little association between the disease, race, and online health information seeking. Blacks and Hispanics who conduct passive health seeking activities have a decreased likelihood to report lung disease when compared to whites. For blacks who participate in active health seeking activities there is an increased likelihood of reporting lung disease compared to whites. Hispanics who participate in active activities are not significantly associated with a decreased or increased likelihood to report lung disease. Participating in passive health information seeking activities results in an increased likelihood to report high blood pressure for both blacks and Hispanics, when compared to whites. However, Hispanics who participate in active health seeking activities are less likely to report high blood pressure when compared to whites.

Based on these results it seems as though health information seeking activities have little impact on blacks and Hispanics likelihood to report a condition in the first and second models. In the fourth model, after controlling for other factors, we do see significant associations between the interactions of race and health information seeking, however, the direction of these is somewhat inconsistent across disease type and race. It seems there is a slight trend towards passive activities leading to a decreased likelihood to report a condition and active activities leading to an increased likelihood. Therefore, while it looks like health information seeking does have an impact on health outcomes the type of impact has variations that are not easily explained by the data. Subsequently, we cannot conclude health information seeking always leads to the attenuation of diseases for blacks and Hispanics.

6 DISCUSSION

The purpose of this research has been to explore the relationship between race, online health information seeking, and health outcomes and, more specifically, to answer the following research question: Does the use of Internet health seeking activities work to reduce the disparity of poor self-rated health and chronic conditions among blacks and Hispanics when compared to whites? A binary logistical regression analysis was completed utilizing the 2012 Pew Internet and American Life data set to determine what effect the independent variables “race” and “health information seeking” had on the dependent variable, “health outcomes.” Specifically, the analysis addressed the following chronic health conditions: self-rated health, heart disease, lung disease, high blood pressure, and diabetes. This discussion will address the findings related to each of the following hypotheses proposed at the beginning of this thesis:

H1- Blacks and Hispanics will have a higher incidence of diabetes, heart disease, cancer, and worse self-rated health than whites.

H2- As online health information seeking increases the incidence of diabetes, heart disease, cancer, and negative self-rated health of all races will improve.

H3- When controlling for online health information seeking the relationship between race and health will get smaller but will not be eliminated.

H4- Online health information seeking will interact with race such that whites who use online health information seeking will be associated with a smaller incidence of diabetes, heart disease, cancer, and negative self-rated health than blacks or Hispanics who use online health information seeking.

H5- Controlling for age, gender, marital status, and income, online health information seeking use will remain significantly associated with race and diabetes, heart disease, cancer, and negative self-rated health.

Before analyzing how the data supports or does not support each hypothesis it is important to keep in mind the limitations of this study. This data only provides a cross sectional look at the relationship between race, health information seeking, and health outcomes. A longitudinal analysis could potentially provide more detailed insight as far as how health information seeking has varying impacts at different stages of health (i.e. prevention through diagnosis and treatment of a specific condition). The research also focuses specifically on those who indicate they use the Internet at least occasionally, meaning there are already differences in the demographic characteristics of the sample from the total population. These differences may be very important as we already know health information seeking is conducted by those with a higher socioeconomic status (Chang and Lauderdale 2009, Luftey and Freese 2005, Link and Phelan 1995, Lorence, Park, and Fox 2006, Lorence and Park 2007, Suziedelyte 2012, and Ybarra and Suman 2006). This could mean we are leaving out a group of people who are more likely to suffer from these conditions but do not use the Internet. Finally, the research does not take into account every possible chronic disease but focuses on those that are more common. The data relies on participants to self-report their chronic conditions and does not account for those who misreport their diagnosis or who have not yet been diagnosed. The data also does not take into account the severity of a condition nor does it link a condition to actual mortality.

The first hypothesis stated blacks and Hispanics would have a higher incidence of diabetes, heart disease, lung disease, high blood pressure, and worse self-rated health than whites. The results of the analysis indicate black, non-Hispanics and Hispanics indeed do have a higher likelihood of reporting some diseases than whites; however, they are also sometimes less likely to report other diseases. Blacks are more likely to report poor health, diabetes, and lung disease than whites, but they are neither more nor less likely to report heart disease or high blood pressure. Hispanics are more likely to report diabetes and heart disease than whites but are less likely to report lung disease or high blood pressure. Additionally, Hispanics are neither more nor less likely to report poor health than whites.

These results do not match the majority of the literature that indicates blacks and Hispanics report higher mortality from these top diseases (UDHHS 2006, CDC 2013). However, there is some support for this Hispanic Paradox in that Hispanics should experience more poor health than whites but do not appear to across many conditions as seen in the data (Inosa-Median, Jean, Cortes-Bergoleri, Lopez-Jimeneze 2014). Additionally, the study eliminated anyone who does not use the Internet, which often includes those in a lower socioeconomic status. The literature indicates that those who have a higher income (Lorence, Park, and Fox 2006) and education (Lorence and Park 2007, Suziedelyte 2012, Ybarra and Suman 2006) are more likely to utilize the Internet for health information. Knowing that those with lower education and income are often black and Hispanic means the health of these groups in our sample may be underestimated because of the population was excluded from the sample. Additionally, some of the Hispanic paradox literature indicates health disparities could be hidden based on key differences between sub-groups in the population (Hummer, Rogers, Amir, Forbes, and Frisbie 2000 and Borrell and Dallo 2008). We have no information on the composition of this Hispanic sample, so it is unknown as to whether or not it represents various sub-groups appropriately.

Although not backed by the literature it is important to note this analysis also did not take into consideration if these respondents have recently (or ever) visited the doctor or been screened for any of these conditions, which may have an impact on their current likelihood to report a disease. They may not be reporting one of these diseases because they are currently going undiagnosed. However, as raised in the limitations section, this could be because the data includes people who self-report their disease incidence, which includes only those who know about their condition and understand their diagnosis. Looking more specifically at time-series mortality data may provide support to prior research as the data would utilize actual cause of death data. These speculations on the reasons for these departures from the literature may provide evidence for why health disparities in mortality from these diseases may continue to exist; there could be a lack of diagnosis and early intervention to prevent these conditions.

Hypothesis two stated that as online health information seeking increased the incidence of diabetes, heart disease, lung disease, high blood pressure, and poor self-rated health of all races would improve. Accordingly, we would expect that for each online health seeking activity there would be a decrease in the odds a person would report their experience with a specific condition. The results show people are more likely to report poor self-rated health if they participate in passive health seeking activities. People who participate in passive health seeking activities are less likely to report heart disease and high blood pressure. Participation in active activities leads to a higher likelihood of reporting diabetes, heart disease, and high blood pressure. Lung disease is the only chronic condition that results in a decreased likelihood associated with each additional passive and active activity. We start to see a pattern indicating passive health information seeking activities are more often associated with a decreased likelihood to report a disease, while active activities more often result in an increased likelihood. Therefore, the hypothesis is mostly supported when it comes to passive health information seeking activities. However, the hypothesis is not supported when active health seeking activities are introduced.

The findings suggest health information seeking isn’t a significant factor for health outcomes, meaning just because someone seeks health information it doesn’t mean they have more positive or negative health outcomes. Rather, the results show there must be some type of variability in when and why one decides to participate in a health seeking activity. We know from the literature this is likely true in that those who are sick or have an issue accessing care are more likely to seek online health information (Yabarra and Suman 2006, Fox and Duggan 2013, and Rice 2006). There is some evidence to suggest passive activities may function as more of a preventative measure, which is logical considering these are the activities included in the passive measure: signing up to receive email updates or alerts about health or medical issues; reading or watching someone else's commentary or person experience about health or medical issues online; going online to find others who might have health concerns similar to yours; and downloading forms online or applying to health insurance online (including private insurance, Medicare, or Medicaid). These are activities that may be likely to happen before being diagnosed with a condition, or at very early stages, when a positive resolution is more likely. Conversely, active activities seem to potentially be more of a reaction to a negative diagnosis, which, once again, makes sense considering the activities included in this measure are: consulting online rankings or reviews of doctors or other providers; consulting online rankings or reviews of hospitals or other medical facilities; consulting online reviews or particular drugs or medical treatments; posting a review of a doctor online; posting a review of a hospital online; or posting your experience with a particular drug or medical treatment. An unknown related to these findings is whether or not the active activities have a positive impact on mortality, which this data is unable to answer. The question becomes whether or not participating in these activities results in a decline in mortality from these diseases, which cannot be answered with this data and analysis. More research would need to be conducted with longitudinal mortality data to truly understand whether or not active health seeking activities lead to reduced mortality.

Hypothesis three stated that when controlling for online health information seeking the relationship between race and health would get smaller, but would not be eliminated. The second model across all chronic condition shows being black, non-Hispanic and Hispanic means people are more likely to report having one of the chronic health conditions. Additionally, after holding health information seeking constant, the significant relationship between being black, non-Hispanic or Hispanic and reporting a chronic health condition remains. The strength of the odds ratios either decreases slightly or in some cases even increases slightly across all conditions. The only exception to this pattern is a Hispanic’s likelihood to report heart disease, which increases in significance when holding health information seeking constant. The hypothesis is not supported and health information seeking explains little, if any, of the relationship between race and health outcomes. The lack of support for this hypothesis provides support for Fundamental Cause Theory in that, despite positive interventions, the relationship between race and poor health continues to exist. Race as a Fundamental Cause has begun to be examined and there is already some existing support this relationship exists (Masters, Link, and Phelan 2015).

Hypothesis four said online health information seeking would interact with race such that whites who use online health information seeking will be associated with a smaller incidence of diabetes, heart disease, and negative self-rated health than blacks or Hispanics who participate in online health information seeking. Without controlling for age, gender, income, or marital status, we do not see any consistent trends in the likelihood of one racial group experiencing a disease over another. Differences do exist across conditions, but they vary in an inconsistent way. Whites experience an increased likelihood when participating in passive activities for poor health, heart disease and active activities for lung disease. They experience a decreased likelihood when participating in passive activities for heart disease and high blood pressure or active activities for lung disease. Blacks experience a decreased likelihood of poor health and lung disease when participating in passive activities. When participating in active activities blacks see an increase likelihood of diabetes, lung disease, and high blood pressure. Hispanics experience an increased likelihood of poor health, heart disease, and high blood pressure when they participate in passive activities and an increased likelihood of diabetes when they participate in active activities. Hispanics see a decreased likelihood of poor health and high blood pressure when they participate in active activities. These inconsistencies may exist because the model does not yet control for other variables, and there may be some explanation related to health information seeking and outcomes that is explained by these variables.

The final hypothesis, number five, said that when controlling for age, gender, marital status, and income, online health information seeking would remain significantly associated with race and diabetes, heart disease, lung disease, high blood pressure, and negative self-rated health. The results indicate some common trends in direction of the significant odds ratios. When it comes to black, non-Hispanics and Hispanics and the interactions between race and health information seeking the odds ratios follow the pattern seen in the second regression models and discussed previously. Participation in passive health seeking activities leads to a decreased likelihood to report a chronic condition, while participation in active health seeking activities lead to an increased likelihood to report a chronic condition. These generalizations are not true for the following: Hispanics participating in passive activities and their likelihood to report high blood pressure; Hispanics participating in active activities and reporting high blood pressure; and blacks participating in passive activities and reporting high blood pressure. Consequently, there must be some opposite relationship in terms of likelihood to report high blood pressure when it comes to health information seeking and race. The inclusion of the control variables do not explain away any of the interactions that were significant in the fourth model and some significant associations actually appear after controlling for other factors. This indicates that where there is a significant interaction between race, health information seeking, and health outcomes, these associations remain even when controlling for education, income, gender, marital status, and age.

Therefore, based on these results, the hypothesis is supported in some instances while in others it is not. The data does not provide much clarity to why these differences and inconsistences exist. Some of the differences among Hispanics can be support by the Hispanic Paradox literature (Insoa-Medina, Jean, Cortes-Bergoderi, and Lopez-Jimeneze 2014); however, some of the differences are unclear. The trends show active activities lead to an increased likelihood to report a chronic condition, which support existing research that indicates participation in these activities occurs when one is already sick and diagnosed versus pre-diagnosis (Yabarra and Suman 2006, Fox and Duggan 2013, Rice 2006). However, there is still room for additional research on the impact specific activities have on health outcomes perhaps with a closer focus on preventative activities versus post-diagnosis activities.

There are some limitations in the design and analysis of this research. First, the study utilizes secondary data, which means the questions could not be designed to specifically focus on only health and online health information seeking behaviors. Additionally, the study only provides a cross sectional analysis, while a longitudinal analysis could provide rich insight into how online health information seeking impacts health at different levels and stages throughout various stages of chronic conditions. The research also focuses specifically on those who indicate they use the Internet at least occasionally, meaning there are already differences in the demographic characteristics of the sample from the total population. Additionally, the data relies on participants to self-report their chronic conditions. This doesn’t account for those who misreport their diagnosis or who have not yet been diagnosed. Utilizing data that reports on mortality would be more accurate, as then the actual cause of death would be known rather than self-reported health conditions. Finally, the research does not take into account every possible chronic disease but rather focuses on those that are more common. It also does not take into account the severity of a disease but rather just whether or not a person has been diagnosed or not, thus we are not able to understand the link between severity and online health information seeking behaviors.

7 CONCLUSION

The purpose of this quantitative, multivariate analysis was to examine the relationship between race, health information seeking and health outcomes. The results provide evidence to support some of the hypotheses, and at a broader level, they provide support for Fundamental Cause Theory. The results show there is weak support for online health information seeking to act as an intervening variable and close the gap in health disparities. Subsequently, race as a Fundamental Cause is supported as many disparities continue to exist even after controlling for health information seeking and other factors. These results provide opportunities for policy recommendations and imply a need for additional research.

This study provides support for Fundamental Cause Theory because even in the final regression model for each health outcome, which includes control variables and variables to account for health information seeking, racial disparities still exist in most health outcomes. Based on the results, it is hypothesized that some people are going undiagnosed. Additionally, those excluded from the sample because they do not use the Internet may be the most sick, meaning even larger disparities likely exist. Now, in future research, it will be important to compare mortality rates for these diseases. That way, associations between race, use of the Internet for health information, and mortality will be clearer. This research, however, does lead me to believe, despite positive interventions, race often remains significantly associated with poor health outcomes meaning it can be considered a Fundamental Cause of disease.

The results show that for blacks and Hispanics passive health information seeking activities lead to better health while active health information seeking activities are associated with worse health. For whites, however, any health information seeking activities are negatively associated with health conditions. This result may be evidence disparities exist in terms of diagnosis. Blacks and Hispanics may remain undiagnosed, providing a possible reason their participation in passive activities leads to a decreased likelihood to report a chronic condition, whiles whites experience the opposite effect. Additionally, along with these differences, the health outcome incidence results provide some indirect evidence a gap likely exists in access to health care and resources. It remains unknown how many of the non-Internet users are actually the most sick and what impact online health information may have on them if they had access and knowledge to use the Internet as a resource. Consequently, the true impact of health information seeking on health outcomes remains unclear.

8 RECOMMENDATIONS

Based on the results of this study, it is important for public policies to be created encouraging access and usage of doctors and preventative health measures. Ensuring people have access and utilize health care will help to reduce preventable diseases and keep people from going undiagnosed and dying from treatable conditions. Additionally, if further research indicates health information seeking does indeed decrease mortality from disease; expanding Internet access would be helpful in encouraging people to take advantage of this positive intervention. The research and analysis presented in this thesis also leads to additional questions that, if answered, would have a significant impact on the understanding of health disparities and, in turn, health policy. These questions include the following: How does health information seeking impact mortality? How does this impact vary by race, and does race continue to show significant association with mortality even when controlling for health information seeking and other variables? How does the timing and type of health information seeking impact its ability to act as a positive intervention from mortality?

9 REFERENCES

Allison, Paul D. 2012. “Logistic Regression Using SAS: Theory and Application, Second Edition.” Cary, NC: SAS Institute, Inc.

Amante DJ, Hogan TP, Pagoto SL, English TM, and Lapane KL. 2015. “Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey.” Journal of Medical Internet Research. Vol. 17.4.

Borrell, Luisa N. and Florence J. Dallo. 2008. “Self-Rated Health and Race Among Hispanic and Non-Hispanic Adults.” Journal of Immigrant and Minority Health.10:229-238.

Centers for Disease Control. 2003. “Racial/Ethnic and Socioeconomic Disparities in Multiple Risk Factors for Heart Disease and Stroke.” Retrieved October 10, 2015 (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5405a1.htm)

Centers for Disease Control. 2013. “Health Related Quality of Life” All pages, retrieved October 10, 2015 (http://www.cdc.gov/hrqol/about.htm).

Chang, Virginia W. and Diane S. Lauderdale. 2009. “Fundamental Cause Theory, Technological Innovation, and Health Disparities: The Case of Cholesterol in the Era of Statins.” Journal of Health and Social Behavior. Vol. 50: 245-260

Chen, Wenhong and Kye-Hyoung Lee. 2014. “More than search? Informational and participatory eHealth behaviors.” Computers in Human Behavior. Vol. 30: 103-109.

Cotton, Shelia R. and Sipi S. Gupta. 2004. “Characteristics of online and offline health information seekers and factors that discriminate between them.” Social Science and Medicine. Vol. 59: 1795-1806

Department of Health and Human Services. 2006. “Fact Sheet: Protecting the Health of Minority Communities.” Retrieved October 10, 2015 (http://www.hhs.gov/news/factsheet/minorityhealth.html).

Fox, Susannah and Maeve Duggan. 2013. “The Diagnosis Difference: A portrait of the 45% of U.S. adults living with chronic health conditions.” Pew Research Center. Retrieved October 10, 2015 (http://www.pewinternet .org/files/old-media//Files/Reports/2013/PewResearch_DiagnosisDifference.pdf.)

Hummer, Robert A., Richard G. Rogers, Sarit H. Amir, Douglas Forbes, and W. Parker Frisbie. 2000. “Adult Mortality Differentials Among Hispanic Subgroups and Non-Hispanic Whites.” Social Science Quarterly. Vol 81.1: 459-476.

Inoiosa-Medina, Jose, Nathlie Jean, Mery Cortes-Bergoderi, and Francisco Lopez-Jimenez. 2014. “The Hispanic Paradox in Cardiovascular Disease and Total Mortality.” Progress in Cardiovascular Diseases. 57: 286-292.

Lorence, Daniel and Heeyoung Park. 2007. “Study of Education Disparities and Health Information Seeking Behavior.” CyberPsychology and Behavior. Vol. 10.1 :149-151.

Lutfey, Karen and Jeremy Freese. 2005. “Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in Routine Clinic Visits for Diabetes.” American Journal of Sociology. Vol. 110.5: 1326-1372

Link, Bruce G. and Jo Phelan. 1995. “Social Conditions as Fundamental Causes of Disease.” Journal of Health and Social Behavior. Extra Issue: 80-94

Link, Bruce, Jo C. Phelan, Richard Miech and Emily Leckman Westin. 2008. “The Resources That Matter: Fundamental Social Causes of Health Disparities and the Challenge of Intelligence.” Journal of Health and Social Behavior. Vol. 49, no.1: 72-91.

Liszka, Heather A., Terrence Steyer, and William J. Hueston. 2006. “Virtual Medical Care: How are our patients using online health information?” Journal of Community Health. Vol. 31.5: 368-378.Social Science & Medicine. Vol. 127: 19-28.

Lorence, Daniel, Heeyoung Park, and Susannah Fox. 2006. “Racial Disparities in Health Information Access: Resilience of the Digital Divide.” Journal of Medical Systems. 30:241-249.

Masters, Ryan K., Bruce G. Link, and Jo C. Phelan. 2015. “Trends in education gradients of ‘preventable’ mortality: A test of fundamental cause theory.”

Ono, Hiroshi and Madeline Zavodny. 2003. “Race, Internet Usage, and E-Commerce.” The Review of Black Political Economy. Winter.

Rice, Ronald E. 2006. “Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys.” International Journal of Medical Informatics. Vol 75: 8-28.

Suziedelyte, Agne. 2012. “How does searching for health information on the Internet affect individuals demand for health care services?” Social Science and Medicine. 75: 1828-1835.

Wang, Andrew, Sean A.P. Clouston, Marcie S. Rubin, Cynthia G. Colen, and Bruce G. Link. 2012. “Fundamental Causes of Colorectal Cancer Mortality: The Implications of Informational Diffusion.” Milbank Quarterly. Vol. 10.3: 592-618.

Ybarra, Michele and Michael Suman. 2006. “Help seeking behavior and the Internet: A national survey.” International Journal of Medical Informatics. 75: 29-41.