The Impact of Online Health Searches on Medical Outcomes: Some Evidence from Google Trends

Completed Research

Tamuchin McCreless
Arizona State University
tam.mccreless@asu.edu

Abstract

The search for online health information has received some attention for many years now. Patients like to search for health information online because they are able to get information that helps them to potentially determine what may be causing them discomfort and some trips on how to treat it. Physicians are often critical of this information because they feel that patients often use it as a substitute for seeking professional help, thus potentially exacerbating an existing problem by misdiagnosing and not getting proper attention for an ailment. There have been few studies which have measured the impact of these online tools on medical outcomes. This study provides some evidence to suggest that online health searches have a positive impact on patient care by comparing search trends from Google to medical outcomes by geographic region.

Keywords (Required)

Online health education, health outcomes

Introduction

The search for health information online engines has become quite popular in recent years. According to a 2013 Pew Research Survey, about 35% of U.S. adults have said that they have searched for information online about their own or someone else’s medical condition (Fox & Duggan, 2013), and 80% of internet users have searched for a health related topic online (Weaver, 2013). Patients have reported that using a search engine to find health-related information online has helped them to deal with illnesses (Fox & Duggan, 2013). The popularity of these searches has raised questions as to whether or not they are effective in helping patients to deal with illnesses. The research question raised by these searches is whether or not they help patients achieve better outcomes. This question has not been well explored thus far. Although studies have measured the effects of the use of these searches on the doctor patient relationship (Murray et al., 2003; Anderson et al., 2003; McMullan, 2006), and the impact of the use of tailored web-based interventions on patient outcomes has been explored (Eysenbach, 2003; Rains, 2009; Ryhänen et al., 2010; Bessel et al., 2002; White & Horvitz, 2013; Nguyen et al., 2004), there have been few if any studies that have measured the impact of patient web use on medical outcomes. This study measures the effect of the use of internet-based health information searches by exploring search results from Google Trends by region and comparing the results of web-based searches in certain regions to various health-related outcomes in those regions. The study provides evidence that regions with higher rates of use of internet-based health search engines have better health outcomes.

Background and Literature Review

The process of seeking help for a health related symptom begins with a patient. Patients begin to feel discomfort or pain and will either seek help from a medical professional or wait to see if the symptoms go away. If symptoms persist, patients will either go to a professional or seek information to determine whether or not seeing a professional is necessary. As visits to a physician outside of annual physicals are often costly,
many patients seek to determine the cause of their symptoms without seeking professional help ( Fox & Duggan, 2013). Historically this could have been done by talking with friends or relatives. Today many patients seek health information on the web.

The use of the internet for health related information can generally fall into one of four categories: communication, content, community, and e-commerce (Eysenbach, 2003). Communication involves the use of email, instant messaging, teleconferencing or other modes of internet based communication. Content involves the use of online content such as information available on health-focused search engines, health provider websites, and other sources of information. Community involves the use of online forums, support groups, and other community based sites. E-commerce involves the purchase of health related products through the internet (Eysenbach, 2003). This study focuses on the use of content and its impact on health outcomes. Patients often search for information simply by using a search engine such as Google, but will also look for information directly through a health-based internet search engine such as WebMD (Fox & Duggan, 2013).

A recent survey found that about 77% of health information seekers begin their search at a search engine such as Google, Bing, or Yahoo and about 13% begin their searches at health specific sites such as WebMD (Fox & Duggan, 2013). By far, the most frequently visited health specific website is WebMD with an estimated 80 million unique monthly visitors (eBizMBA, 2017). This study focuses on the question of whether the use of these search engines and health specific websites improves the overall health of a population.

A literature review found few articles that measured the impact of internet based information on health outcomes. A number of studies measured the impact of internet health seeking on the patient-physician relationship, and a number of studies explored the way in which patients use the internet and the factors affecting this use. Few, however, explored the impact of internet use on health related outcomes.

The only studies found which measured the impact of the internet on actual patient outcomes were related to cancer patients. A systematic review of the effect of the internet on cancer patients found 15 randomized trials to evaluate the impact of information based interventions on patients with cancer and determined that patients self-reported positive impact of the internet on a number of outcomes (Eysenbach, 2003). However, these trials focused on the impact of personalized intervention systems rather than the use of general non-directed search (Eysenbach, 2003). Two more recent systematic reviews found studies that measured the impact of web based information on cancer outcomes, but focused on patient reported outcomes such as improved emotions, increased hope, etc. (Ryhänen et al, 2010; Erfani & Abedin, 2014). It is important to note that most of the studies in these two reviews also did not focus on general web use but rather the use of tailored web based interventions. A search of the literature for studies which measured the impact of the use of general internet-based health seeking behavior did not yield any studies which fall into this area. This study, therefore, addresses a major gap in the literature related to whether or not internet-based health information has any impact on patient health.

### Research Questions and Hypotheses

The research question being addressed in this study is the extent to which internet-based health searches affect the health of the overall population. Although there are many ways to measure the health of the overall population, the availability of data on the health of the population are limited. This study will look at healthcare costs as well as utilization of healthcare services. Thus, the two hypotheses that will be tested are:

**H1:** Increased use of internet-based health searches has a negative impact on overall health expenditures.

**H2:** Increased use of internet-based health searches has a negative impact on the utilization of health services.

### Methods

The unit of analysis for this study is a metropolitan region, or specifically a Designated Market Area (DMA). This is an area used by the Nielsen research company to track measures related to market based research. This is also the unit used by Google to track search term data related to metropolitan areas. To test the hypotheses of this study, a regression analysis was conducted using the equation below:
Outcome = $\beta_0 + \beta_1 \times \text{SearchRate} + \beta_2 \times \text{HealthConsciousRank} + \beta_3 \times \text{PCPRatio}$

In this equation, outcome represents the medical outcome of interest. There were more than one outcomes used in this study, with further detail provided in the next section. SearchVolume represents the volume of searches in any given DMA, as determined by Google. The PCPRatio represents the ratio of people to primary care physicians in any given county. Details related to how each of these variables was quantified is given in the following sections.

**Medical Outcomes**

There were three outcomes used in this study: standardized risk-adjusted per capita medical costs, inpatient covered days per 1000 people, and outpatient stays per 1000 people. Cost data were used to measure the impact of the independent variables on health expenditures. A population that is more educated about its health will presumably incur fewer medical costs. Inpatient days were used because they are an indication of a population’s health. Patients that are educated about health will be able to recognize issues more quickly and thus spend fewer days in the hospital. This is also true of outpatient visits.

Because these statistics are not available for the general public, the data used for this study come from the Centers for Medicare and Medicaid Services and are limited only to Medicare beneficiaries. Because Medicare utilization is often the major driver of healthcare costs in a given area, these data are being used as a proxy for the healthcare costs of the general population. The dataset used in this study comes from the State and County public use file available from Medicare. This file contains data as recent as 2015 with cost and utilization data for all Medicare beneficiaries. The file contains a Federal Information Processing Standard (FIPS) code which was mapped to the DMA using a file from the Harvard Dataverse (Sood, 2016). Because there was generally more than one FIPS per DMA, the results were aggregated using the sum of the raw numbers for each FIPS and dividing by the total number of beneficiaries in that DMA.

**Health Conscious Rank**

The health conscious rank was used as a proxy for how health conscious a population is. This measure was used to control for the fact that regions with more health conscious people will naturally be more healthy than those with less health conscious people. It is important to include a measure such as this because it is likely that an underlying factor such as the level of health consciousness will drive people to conduct more health related searches on the web. The results will show that the search rate and the health conscious rank are orthogonal and that both impact medical outcomes.

These ranks were obtained from the company Value Penguin. This organization ranked cities in the United States on how health conscious they were based on health amenities, availability of healthy food, the overall health of residents, and environmental factors (Value Penguin, 2018). Cities with smaller ranking numbers are more health conscious than those with larger ranking numbers. These cities were mapped to DMAs manually using the names of each city and mapping them to the names of the DMA.

**PCP Ratio**

The PCP ratio is a measure of the ratio of people to primary care physicians in a given county. PCP Ratio is included in this analysis as a covariate to serve as a proxy for access to health care, especially primary care. Although the health conscious rank included health amenities as a factor in its rankings, these amenities were not related to healthcare providers but rather to health clubs and other wellness related facilities.

This measure comes from the organization Forum One which provided health related statistics at the county level in collaboration with University of Wisconsin Public Health Institute and the Robert Wood Johnson Foundation. Counties were mapped to DMAs manually. Since most DMAs lie in a single county, this process was simple. Many DMAs spread across counties. These DMAs were mapped to the county for which they were the county seat.

**Search Rate**

The search rate comes from a measure used on the Google Trends website. The measure is, as its name suggests, not a measure of the total number of raw searches of any given term, but rather a measure of the
total number of searches for any given search term as a proportion of the total number of all searches for
the given unit of analysis. The unit of analysis in this study is the DMA so the search rate is a reflection of
the proportion of searches for a health related search term to the total number of searches in that area. The
proportions for each region are then divided by the proportion of the region with highest proportion of
searches for that given term so that the measure for any given region is a number from zero to one hundred
(Stephens-Davidowitz, 2014). For example, one search term used in this study is the search term "WebMD".
The measure of the volume of searches for this search term can be calculated using the formula below
(Stephens-Davidowitz, 2014):

\[
WebMDSearchRate_j = 100 \times \frac{\left( \frac{\text{Searches containing the term "webmd"}}{\text{Total Number of all searches}} \right)_j}{\left( \frac{\text{Searches containing the term "webmd"}}{\text{Total Number of all searches}} \right)_{\text{max}}}
\]

Where j represents any DMA region and max represents the DMA region with the highest proportion of
searches of WebMD to total searches.

Google Trends allows one to measure these rates by region for certain terms, or for search terms falling into
a specific category. Since two of the most common reasons why people search for health information online
is to search for information related to a specific symptom or to search for information related to a specific
disease, data for searches in these two categories were both used as search rates in this study and their
impact on outcomes were measured separately. The rates for the search term WebMD were also used as
independent variables for search rate and their impact on health outcomes was measured separately.

**Regression Analysis**

All datasets were combined into a single dataset, with one record for each DMA. The R software was used
to combine datasets as well as to run the regression analysis. Correlations were tested for all independent
variables to ensure there was no multicollinearity which would violate the assumptions of the general linear
model. To test the first hypothesis, the per capita medical costs were used. The second hypothesis was tested
using two different dependent variables: inpatient days per 1000 members and outpatient visits per 1000
members.

**Results**

This analysis included search data from Google for 210 DMA regions. After matching to the CMS data and
the data from Value Penguin, there were 65 regions which contained values that could be used for the study.
The correlation analysis of the dependent variables is shown in Table 1. Although there is a correlation
between PCP_Ratio and HC_Rank, it is not high enough to be of concern in this analysis.

<table>
<thead>
<tr>
<th></th>
<th>Search_Rate</th>
<th>HC_Rank</th>
<th>PCP_Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search_Rate</td>
<td>1</td>
<td>0.145</td>
<td>-0.134</td>
</tr>
<tr>
<td>HC_Rank</td>
<td>0.145</td>
<td>1</td>
<td>0.269</td>
</tr>
<tr>
<td>PCP_Ratio</td>
<td>-0.134</td>
<td>0.269</td>
<td>1</td>
</tr>
</tbody>
</table>

The distribution of each dependent variable was analyzed to determine if it fit a normal distribution. These
results are shown in Figure 1. The results show that costs per 1000 beneficiaries and inpatient days per
1000 beneficiaries both fit a normal distribution well. The outpatient visits per 1000 beneficiaries
distribution does not fit a normal distribution as neatly, but is deemed closed enough for this analysis.
Figure 1. Distribution of Dependent Variables

The results of the regression analysis using search terms falling into the disease category are shown in Table 2. Results when using the other two search terms were not significant.

Table 2. Regression Results

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Per Capita Costs</th>
<th>IP Days Per 1000</th>
<th>OP Visits per 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Search_Rate</td>
<td>-26.380*</td>
<td>-7.897***</td>
<td>24.530**</td>
</tr>
<tr>
<td>HC_Rank</td>
<td>17.767***</td>
<td>3.479***</td>
<td>3.729</td>
</tr>
<tr>
<td>PCP_Ratio</td>
<td>0.638*</td>
<td>-0.025</td>
<td>-1.262***</td>
</tr>
<tr>
<td>Constant</td>
<td>8,551.450***</td>
<td>1,583.050***</td>
<td>4,539.735***</td>
</tr>
<tr>
<td>Observations</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>R²</td>
<td>0.328</td>
<td>0.305</td>
<td>0.243</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.294</td>
<td>0.271</td>
<td>0.205</td>
</tr>
<tr>
<td>Residual Std. Error (df = 61)</td>
<td>1,063.895</td>
<td>212.195</td>
<td>1,144.258</td>
</tr>
<tr>
<td>F Statistic (df = 3; 61)</td>
<td>9.903***</td>
<td>8.943***</td>
<td>6.511***</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

These results provide support for hypothesis 1 and partial support for hypothesis 2. The first hypothesis is supported by looking at the impact of the search rate on per capita costs. The coefficient of -26 suggests that
DMA regions with 1 unit of higher search rate had approximately $26 fewer in medical costs per capita. This number may not seem large in comparison to the total costs per capital of approximately $8,500. However, this number can be meaningful when one takes into account the number of beneficiaries that may live in any given region. A region with one million members could potentially have tens of millions of dollars in lower expenses due to obtaining information regarding health information on the web. This number can also be meaningful when one considers that in this dataset there is a difference of 58 units from the DMA region with the lowest search rate to the region with the highest. This could mean a difference of more than $1,000 in per capita medical expenses.

The second hypothesis is supported by looking at the coefficient of the search rate for both inpatient days per 1000 beneficiaries. This result is more meaningful than the results for per capita costs. There are a number of factors that could influence cost aside from both inpatient and outpatient visits. These results provide support that regions which had higher search rates for disease related terms, had members which spent fewer days in the hospital. Again, when considering the range of values from highest to lowest, this could mean a difference of hundreds of inpatient days per thousand members.

The impact of search rate on outpatient visits shows the opposite result than expected. Rather than patients having fewer outpatient visits in regions with higher search rates, patients had more visits. This could mean that patients who are more likely to search for health data online, are also more likely to visit a physician. It is possible that higher rates of visits to physician offices could also be leading to fewer days in the hospital.

**Discussion**

The results of this study provide some support that online health search activity may have an impact on costs as well as the utilization of health services. While it could be that another underlying factor is affecting both items, an attempt has been made to account for this by including the regional health conscious ranking as a factor in this study. While it could also be that access to healthcare resources is perhaps impacting both variables, an attempt has been made to account for this by including the PCP ratio as a factor in the analysis.

The results also provide some explanation for why increased use of online health searches may lead to lower costs as well as less inpatient utilization. It is possible that patients who search for online health information may become more informed about when to see a physician. For much of the population, this may mean that they avoid seeking physician help. However, for patients who have more acute conditions, this may actually increase the frequency with which they visit the physician. This increased use of physician services may lead to increased provision of preventive care, thus curbing the need for inpatient services at a later time.

The implications of this research is important for both patients, providers and payers. The results suggest that patients should make an effort to seek out health information on the internet to determine what the best course of action is for any given ailment, especially to determine whether or not to seek help from a professional. Providers often invest resources in online health information to help patients in determining potential causes of ailments. These results should help providers see that the investments could be having a positive effect on patient outcomes. Finally, payers should be encouraged to contribute to the growth of online health information resources to aid in curbing medical costs.

**Limitations**

There are many factors that impact medical costs and the utilization of healthcare resources. While this study shows a relationship between search rates for medical information and these outcomes, there are other factors that may impact these outcomes as well. Specifically, it is difficult to separate one’s intents from actions associated with those intents. That is, if a person intends to stay healthy, there are many actions they may take, some of which may impact their health, some of which may not. One of these actions may be to search for health information online. It could therefore be that the underlying factor impacting these outcomes is actually a person’s intent to improve their health, and that search frequency is an indicator of this intent. This could potentially be teased out by conducting a study using the individual patient as the unit of analysis and gathering information related to intent to improve one’s health.

While health searches can be conducted by anyone, and while the Google search health data is for all members, the medical outcomes were only measured for Medicare patients, as this is the only publicly available data that is easily accessible. While a large percentage of medical costs are driven by Medicare
patients, and while Medicare costs could be a proxy for overall costs, it is likely that total overall costs differ from Medicare costs, especially where the demographic makeup of a region is largely composed of a younger population.

The unit of analysis is slightly different for more than one of these measures. The CMS data is at the FIPS level, which does not match perfectly to the DMA level. The PCP ratio is at the county level which is not exactly the same as the DMA level either.

**Future Research**

This study measures the impact of one type of search term, disease related search terms, on the impact of health outcomes. There could be many other types of terms that could have impacts on different types of outcomes. A more in-depth study which could determine which search terms have a bigger impact on outcomes, and which outcomes are affected by which search terms, may reveal more about the nature of online health searches.

While this study has been done at the regional level, a study at the individual level could also reveal more about the nature of online health searches, and overcome some of the limitations of conducting a study at the regional level. A study at the individual level could allow for a controlled environment and help measure a more direct effect of search terms on health outcomes.

**REFERENCES**


