How does searching for health information on the Internet affect individuals' demand for health care services?

Agne Suziedelyte*

School of Economics, University of New South Wales, Level 4, Australian School of Business Building, UNSW Sydney NSW 2052, Australia,
email: agne@unsw.edu.au, tel: +61 2 9385 6965, fax: +61 2 9313 6337.

Manuscript Number: SSM-D-11-01784R2

*I am grateful to Denise Doiron, Denzil Fiebig, Shiko Maruyama, Mark Rosenzweig, Kevin Lang, Hong Il Yoo, Adeline Tubb, two anonymous reviewers, and the participants of the 7th Summer Workshop in Health Economics, the 1st Annual Workshop on Health IT and Economics, and the 8th World Congress on Health Economics for their helpful comments and advice. All opinions and any mistakes are my own.

Abstract

The emergence of the Internet made health information, which previously was almost exclusively available to health professionals, accessible to the general public. Access to health information on the Internet is likely to affect individuals' health care related decisions. The aim of this analysis is to determine how health information that people obtain from the Internet affects their demand for health care. I use a novel data set, the U.S. Health Information National Trends Survey (2003-07), to answer this question. The causal variable of interest is a binary variable that indicates whether or not an individual has recently searched for health information on the Internet. Health care utilization is measured by an individual's number of visits to a health professional in the past 12 months. An individual's decision to use the Internet to search for health information is likely to be correlated to other variables that can also affect his/her demand for health care. To separate the effect of Internet health information from other confounding variables, I control for a number of individual characteristics and use the instrumental variable estimation method. As an instrument for Internet health information, I use U.S. state telecommunication regulations that are shown to affect the supply of Internet services. I find that searching for health information on the Internet has a positive, relatively large, and statistically significant effect on an individual's demand for health care. This effect is larger for the individuals who search for health information online more frequently and people who have health care coverage. Among cancer patients, the effect of Internet health information seeking on health professional visits varies by how long ago they were diagnosed with cancer. Thus, the Internet is found to be a complement to formal health care rather than a substitute for health professional services.

1. Introduction

This research is motivated by the observation that a large and increasing proportion of the population in developed countries use the Internet as a health information source. An

estimated 27.5 percent of the U.S. adult population looked online for information about a health or medical issue in 2000 (Rice, 2006). This figure increased to 40 percent in 2002 and to 61 percent in 2008 (Rice, 2006; Fox & Jones, 2009). It is likely that health information that people obtain from the Internet influences their health related decisions, including their demand for health care services. Indeed, most of the individuals who look for health information online report that this information had a major or minor impact on their own health care or the way they cared for someone else (Rice, 2006; Fox & Jones, 2009). Additionally, some of the health information seekers say that the information obtained from the Internet led them to ask a doctor new questions or to get a second opinion from another doctor or affected their decision about whether to see a doctor or not (Rice, 2006).

There is no consensus in the literature on whether Internet health information is a substitute for or complement to health care. For example, the medical sociology literature has two opposing hypotheses related to this topic (Lee, 2008). The first hypothesis is that by diffusing health knowledge, which was previously available only to health professionals, the Internet may reduce people's dependence on health professionals as a source of health information and, consequently, lower the frequency of their visits to health professionals. Patients may elect to use online health information to diagnose and treat themselves rather than visit a doctor to reduce monetary and time costs. The second hypothesis states that, despite people's access to health knowledge, the knowledge gap between the general public and health professionals remains, since new information constantly emerges and is first available to health professionals. Furthermore, health and medical information involves uncertainty and error; therefore, people rely on health professionals for the interpretation of Internet health information. Online health information may also make individuals more concerned about their health and well-being. As a result, increasing access to health information on the Internet may, in fact, increase the frequency of people's visits to health professionals.

Existing empirical research investigating the effect of Internet health information on the demand for health care reaches inconclusive and inconsistent conclusions. For example, Khechine et al. (2007) find a positive association between individuals' intensity of Internet use for health purposes and their utilization of health care services. These results are, however, based on a small and non-random sample of individuals who are Internet users and have a long-term medical condition. The authors do not control for unobserved individual characteristics that are likely to be correlated with both health information seeking intensity and health care utilization. Therefore, the positive correlation between Internet use for health purposes and health care utilization cannot be interpreted as causal. There are two studies that do address the endogeneity of Internet health information seeking. The first study uses panel data to estimate a model that controls for an individual's health care utilization in the base period and finds that the intensity of Internet health information exposure has a positive effect on two health care utilization measures, contacting a physician for information and visiting a physician for treatment (Lee, 2008). To the contrary, the second study finds a small, negative, and insignificant effect of computerized health information use on an individual's number of doctor visits, using instrumental variable and fixed effect methods (Wagner & Jimison, 2003). Additionally, it is investigated whether computerized health information affects the probability of visiting a doctor and a positive, but statistically insignificant effect is found (Wagner & Jimison, 2003). Thus, this paper contributes to the literature by providing additional evidence on the effect of Internet health information seeking on health care utilization.

2. Methodology

In this analysis, an individual's demand for health care HC_i is modeled as a function of a binary variable indicating whether or not an individual is an Internet health information seeker eHi_i, observed demographic characteristics X_i (age, sex, education, race, and marital

status), and unobserved variables c_i:

$$HC_i = \beta_0 + \beta_1 eHi_i + X_i'\beta_2 + c_i. \quad (1)$$

The variable of interest in this analysis is eHi_i. A positive coefficient on eHi_i would suggest that Internet health information is a complement to health care. A negative coefficient on eHi_i would imply that it is a substitute for health care. The above mentioned demographic characteristics are included in the model, as there is evidence that not only health care utilization, but also the demand for Internet health information varies by age, sex, education, and race (Rice, 2006). Additionally, the data used for this analysis shows that there are differences in online health information seeking behavior by marital status (see Table 1). I do not include other variables that may affect an individual's demand for health care, such as health care coverage, in the baseline model, because these variables may be endogenous, which would bias the estimate of β_1.

Internet health information seekers may be different from non-seekers in their unobserved characteristics. Therefore, estimating equation (1) by ordinary least squares (OLS) may produce biased estimates of the effect of Internet health information on the demand for health care. I use two strategies to identify the effect of Internet health information seeking on the demand for health care that take into account the unobserved variables. First, I include the following additional variables in equation (1): household income, an individual's employment status, a binary variable indicating whether or not an individual has health care coverage, and variables describing an individual's smoking behavior, exercising, and nutrition. Second, I use the instrumental variable (IV) method. This approach requires data on at least one variable that is correlated with Internet health information seeking behavior (the relevance assumption), but not with the unobserved variables (the exogeneity assumption).

3. Data and variables

3.1. Data

For the empirical analysis, I use the data from the U.S. Health Information National Trends Survey (HINTS) (National Cancer Institute, 2006-2009). Given that this data is de-identified and publicly available, no ethical approval was needed for this study. The HINTS is a repeated cross-sectional survey of the U.S. civilian non-institutionalized adult population. The National Cancer Institute (NCI) manages and funds the survey. The NCI is part of the National Institutes of Health, a medical research agency of the U.S. Department of Health and Human Services. Although the main purpose of the survey is to collect data about the public's use of cancer-related information, the HINTS contains questions about the exposure to, and search for, general health information in different media, including the Internet. The survey additionally asks respondents about their health care utilization. The availability of these key variables makes this data set suitable for analyzing the relationship between Internet health information and the demand for health care.

The HINTS data is collected mainly via telephone interviews. The sample is drawn from all telephone exchanges in the U.S. One adult (18 years or older) is randomly selected from all household members to answer the survey. In the last survey, mail questionnaires have supplemented telephone interviews to reach people who do not use a landline telephone. The mail sample is drawn from the national listing of addresses. All adult household members are asked to fill in the questionnaire in this sample. To produce reliable estimates for minority groups, stratified (non-clustered) random sampling is used. Households from the stratum with a higher proportion of black and Hispanic population are over-sampled. The response rates of HINTS are not high (21-33 percent), but comparable to the response rates of other surveys on health-related Internet use (Fox & Jones, 2009). To investigate whether the low response rates affect the representativeness of the HINTS, I compare selected socio-demographic characteristics between the HINTS and the American Community

Survey (ACS). The ACS has high response rates and is considered to be nationally representative. There are no substantial differences in race, education, and marital status between the HINTS and ACS, but males and younger individuals are under-represented in the HINTS sample. These differences are taken into account by including gender and age in the regressions. Additionally, sampling weights are used to estimate the descriptive statistics of the variables used in the analysis. The data is currently available for the years 2003, 2005, and 2008 (The last survey is called HINTS 2007, although the data was collected in 2008). To increase the precision of the parameter estimates, I pool the data over all available years. Excluding observations with missing values, the size of the analysis sample is 16,677 observations - 5,618 from the year 2003, 4,924 from the year 2005, and 6,135 from the year 2008. Observations with missing income values are included and indicated by a dummy variable.

3.2. Key variables

Next, I describe the key variables used in this analysis. Health care utilization is measured by an individual's number of visits to a health professional within a 12 month period. In the survey, respondents are asked: "During the past 12 months, not counting times when you went to an emergency room, how many times did you go to a doctor, nurse, or other health professional to get care for yourself?" Thus, the definition of a health professional is broad and includes not only doctors, but also nurses and other health professionals. Based on the answers to other survey questions, it appears that respondents may include chiropractors and other alternative therapists, dentists, midwives, physical and occupational therapists, psychiatrists, and psychologists in their understanding of "other health professionals". Possible answers to the survey question on the number of health professional visits are "None", "1 time", "2 times", "3 times", "4 times", "5-9 times", and "10 or more times". I recode the health care utilization variable by assigning numerical values to the last two categories. In the baseline specification, the dependent variable takes the value 7, a midpoint of the

interval, if a respondent visited a health professional 5 to 9 times. The dependent variable takes the value 12 if a respondent visited a health professional 10 or more times, which corresponds to visiting a doctor once a month on average. The results are robust to the changes in the coding of this variable. The estimate of the mean number of health professional visits in the population is close to four visits in the past 12 months.

As a measure of Internet health information, I use a binary variable indicating whether a respondent has searched for information about a health or medical topic, such as the symptoms, diagnosis, causes, treatment, or prevention of a disease, illness, or health condition, on the Internet for *him/herself* in the past 12 months. I refer to the individuals who look for health information on the Internet for themselves as "e-health information seekers" and to those who do not use the Internet to look for health information as "non-seekers". Non-seekers include individuals who do not look for health information at all and those who do not use the Internet to search for such information. In 2003, a third of the population is estimated to have used the Internet as a health information source. This figure increases to 37.4 percent in 2005 and further rises to 40.0 percent in 2008. Looking at the Internet users only, the percentage of Internet health information seekers in this group varies from 51.2 percent in 2003 to 54.5 percent in 2008. In 2003, the survey also included a question on how often a respondent had searched the Internet for health information. The answers to this question show that 37.8 percent of the e-health seekers looked for health information on the Internet regularly (once a week or once a month). The rest searched for such information every few months (34.9 percent) or less often (27.4 percent).

According to the HINTS, the Internet is the most widely used source of health information. In 2008, 61 percent of the people who looked for health information searched for this information on the Internet first. Only 14 percent first contacted their doctor or health care provider, and only 10 percent used books as the primary source of information. Other sources (such as brochures, magazines, newspapers, family, friends, and co-workers) were

used even less often. Moreover, more than a quarter of the people who first looked for health information elsewhere used the Internet as the secondary information source. Most people chose to look for health information on the Internet, rather than somewhere else, because there was a lot of information available on the Internet, they could get information immediately and it was convenient. Some individuals chose the Internet, because they could not get access to other sources, such as a doctor or books. As reasons for health information search, respondents indicated having a health issue or concern; being curious and wanting to have knowledge; looking for a second opinion; checking, and supplementing information from other sources, including a doctor; and being referred to the Internet by a doctor or somebody else. The last two reasons indicate the need to account for possible reverse causality between Internet health information and the demand for health care.

Table 1 shows that the individuals who search for health information on the Internet are systematically different in their characteristics from the individuals who do not use the Internet for this purpose. E-health information seekers are relatively younger and have a higher educational attainment. Female, white, and married (or living with a partner as a married couple, but not legally married) individuals are more likely to search for health information on the Internet. A larger proportion of e-health information seekers are employed and have health care coverage relative to non-seekers. Individuals who search for health information on the Internet are also more likely to live in higher income households. These two groups are different in their health-related behaviors. E-health seekers are more likely be non-smokers, engage in moderate intensity physical activity at least once a week, and consume the recommended quantity of vegetables. All these differences are statistically significant at the one percent level.

Figure 1 about here

Figure 1 presents the distribution of the number of visits to a health professional separately for the e-health information seekers and non-seekers. The proportion of individuals with no visits is substantially lower among e-health seekers (9.4 percent compared to 19.5 percent). E-health seekers also have a higher proportion of frequent users with 5-9 and 10 or more health professional visits. The difference in the mean number of visits between e-health information seekers and non-seekers was 0.9 (p-value < 0.01) in 2003. This difference increased slightly to 1.1 (p-value < 0.01) in 2005 and 2008. Thus, the raw data shows that e-health information seekers have higher health care utilization, as measured by the number of visits to a health professional, than non-seekers. This relationship cannot be interpreted as causal, however, as the confounding variables are not taken into account.

3.3. Instrumental variable

I use information on U.S. states' right-of-way regulations to construct an instrument for Internet health information. A review of the telecommunications literature reveals that access to public rights-of-way is one of the most relevant issues for Internet penetration (National Telecommunications and Information Administration, 2003; Study Committee on Public Rights-of-Way, 2002). Right-of-way is a privilege of someone to pass over a property belonging to someone else. Internet providers require access to public lands and infrastructure in order to build their own infrastructure. If a state's rights-of-way regulations make it difficult or costly for Internet providers to access public lands and infrastructure, it may reduce their supply of Internet services in that state. This hypothesis is supported by the empirical findings that the states that specifically grant telecommunication firms access to public rights-of-way have higher high-speed Internet penetration (Wallsten, 2005). An increase in the supply of Internet services and a decrease in their price is expected to increase the probability of an individual using the Internet to look for health information, but not affect his/her health care utilization directly.

The instrument is a binary variable that takes the value 1 if an individual resides in a state where right-of-way regulations are relatively favorable to telecommunication providers and the value 0 otherwise. I define that a state has relatively favorable right-of-way regulations if it explicitly grants telecommunication firms access to local or state public rights-of-way or restricts local governments' authority to prevent access or both. Information about right-of-way regulations is obtained from a survey of U.S. state laws conducted by NTIA (National Telecommunications and Information Administration, 2010). I verify and supplement this information using the official statutes of US states (FindLaw, 2012). Based on my definition, 39 states have right-of-way policies that are relatively favorable to telecommunication providers. Appendix 1 lists these states.

4. Results

The first stage results, presented in Table 2, show that the instrument is indeed strongly correlated to the variable of interest. The probability of being an Internet health information seeker is 4 percentage points higher for individuals living in states with more favorable right-of-way policies compared to individuals living in states with less favorable right-of-way policies, holding other factors fixed. The estimated coefficient on the instrument is statistically significant at the one percent level. The probability of looking for health information on the Internet is also positively associated with education, younger age, being female, white, and married (or living with a partner). In this and other estimations, the standard errors are adjusted (clustered) to account for the fact that a state-level variable is used as the instrument for Internet health information seeking (Wooldridge, 2003). More specifically, the standard errors are computed allowing for correlation across observations within a state (using the *cluster(state)* option of `regress` and `ivregress` commands in Stata 11).

The main results, reported in Table 3, show that the positive relationship between Internet health information seeking and health care utilization persists when the endogeneity

of Internet health information seeking is addressed. Holding observed characteristics age, sex, race, education, and marital status fixed, e-health information seekers are estimated to have approximately 1.2 health professional visits per year more than non-seekers, which is a 27.7 percent increase from the sample mean of 4.2 visits (column 1). The estimated coefficient on the e-health information seeking variable practically does not change when the other variables (an individual's health care coverage, employment status, household income, and health-related behaviors) are included in the model (column 2). Column 3 presents the results of the IV model, which also support the hypothesis that Internet health information seeking has a positive effect on the demand for health care. The IV estimate is larger in magnitude than the OLS estimates, which implies that Internet health information is negatively correlated with the unobserved determinants of the demand for health care. In other words, the individuals who use the Internet to look for health information are different from the individuals who do not use the Internet for this purpose, and these differences cannot be captured solely by the observed individual characteristics. Failing to account for this leads to the under-estimation of the effect of Internet health information seeking on the demand for doctor visits.

The other variables have the expected effects on health care utilization. The number of visits to a health professional increases with age and is higher for females. Holding other individual characteristics fixed, race is not significantly related to health care utilization. Being married (or living with a partner) has a negative effect on the number of health professional visits, but it becomes insignificant once the other observed variables are included in the model. Education appears to be negatively related to health care utilization. There is a strong positive association between health care coverage and health care utilization. Employed individuals have a lower number of health professional visits than people who are unemployed or out of the labor force (including retired individuals). Household income is also negatively associated with health care utilization, once health care coverage and other factors

are held fixed. The correlation between smoking and health care utilization is insignificant. An individual's number of health professional visits is positively associated with a healthier diet, but it is negatively related to exercising. The OLS estimate of the effect of Internet health information seeking on health care utilization is similar in magnitude to the estimated effect of employment. The effect of Internet health information seeking is smaller than the effect of health care coverage, but larger than the gender effect. The IV estimate of the effect of being an e-health information seeker is almost three times as large as the effect of having college education (relative to less than high school education).

Next, I investigate whether the effect of Internet health information seeking on health care utilization varies across individuals (OLS model is used for this purpose due to the lack of instruments for additional variables included in these models). The coefficient on the e-health information seeker variable measures the total effect of all searches made by an individual within the past 12 months on his/her visits to a health professional in the same period (rather than the effect of a single search occasion). As mentioned in the data section, a substantial proportion of the population look for health information on the Internet more than once a year. This may explain the large magnitude of the coefficient on the e-health information seeker variable. Furthermore, if the frequency of search were taken into account, individuals who search for health information on the Internet more often should make more visits to a health professional compared to those who look for such information less often. I can test this hypothesis using data on the frequency of health information searches on the Internet, which is available for one year of the survey (2003). Table 4 presents estimates obtained from regressing the number of health professional visits on the dummy variables indicating whether an individual has looked for health information on the Internet for him/herself once a week, once a month, or less frequently (the reference group consists of the individuals who have not looked for health information for themselves). These estimates are as expected. The effect of Internet health information seeking monotonically

declines as the frequency of search decreases. Looking for Internet health information once a week increases an individual's number of health professional visits by 1.8 visits (relative to non-seekers), whereas the difference in the number of health professional visits between infrequent e-health information seekers and non-seekers is 0.5 visits. All coefficients are statistically significant at the one percent level.

Whether Internet health information is a substitute for or complement to health care may depend on the stage of a disease or medical condition, especially in the context of chronic conditions, such as diabetes, asthma, arthritis, osteoporosis, heart diseases, and cancer. According to one hypothesis, information obtained from the Internet is most likely to complement health care services at the onset of a disease (Smith, 2005). At this stage, patients may be not familiar with their disease and, therefore, need their doctor's assistance in interpreting health information. As patients become familiar with their condition, they may become less reliant on their doctors for the interpretation and verification of health information. At this stage, patients may start substituting Internet health information for their doctor's advice, since on the Internet, they can find the most up-to-date information that is specifically applicable to their personal circumstances, whereas a doctor possesses more general medical knowledge and may not have the newest information on all medical conditions. On the other hand, patients whose condition worsens may look for health information on the Internet more often over time and, therefore, have more doctor visits. If the first hypothesis were true, the effect of Internet health information seeking would decrease with the time since the diagnosis. If the second hypothesis were true, the effect of Internet health information seeking would increase with the time since the diagnosis. To test these hypotheses, I restrict the sample to the individuals who have ever had cancer and interact the e-health information seeker variable with the dummy variables indicating the time since an individual's cancer diagnosis (3-5 years, 6-8 years, or 9 or more years, the omitted category is 2 years or less). The estimated effect of Internet health information for the patients

who were diagnosed with cancer recently (up to 2 years ago) is 1.2 (p-value = 0.01), which is larger compared to this effect for the patients who were diagnosed 3-5 years ago ($\hat{\beta}_{eHi} = 0.5$, p-value=0.33) or 6-8 years ago ($\hat{\beta}_{eHi} = 0.9$, p-value=0.05). So, there appears to be some support for the first hypothesis. On the other hand, the effect of Internet health information seeking is higher for patients who were diagnosed with cancer 6 or more years ago comparing to the effect for patients who were diagnosed 3-5 years ago, suggesting that the second hypothesis may be true as well.

I also investigate if the effect of Internet health information seeking on the demand for health care varies by education and health care coverage status. I expect the effect of Internet health information on the demand for health care to be lower for people with higher levels of education, since education may be correlated with information-processing skills. On the other hand, I expect this effect to be higher for individuals with health care coverage, as they face a lower price of health care. There is little support for the first hypothesis. The effect of Internet health information on health professional visits does not vary significantly with an individual's educational attainment. On the other hand, the data provides support for the second hypothesis. The effect of Internet health information on health professional visits is higher for individuals with health care coverage ($\hat{\beta}_{eHi} = 1.1$, p-value<0.01), who face a lower price of health care. For individuals without health care coverage, this effect is still positive and significant ($\hat{\beta}_{eHi} = 0.8$, p-value<0.01). Thus, Internet health information seeking could have a larger effect on health care utilization if more people in the U.S. had health care coverage. The data also shows that the effect of Internet health information seeking on the demand for health care is higher for females and individuals of white race. This effect does not vary significantly by marital status or age.

4.1. Sensitivity analysis

In this section, I check the sensitivity of the results to the assumptions made in this analysis. First, I investigate how the results are affected by the changes in how the HINTS phrases the question regarding an individual's online health information seeking behavior. In 2003 and 2005, respondents were asked whether they had searched online for health information in the past 12 months, whereas in 2008, respondents were asked whether they had used the Internet in their most recent search for health information. Some of the respondents in 2008 may have last searched for health information on the Internet more than 12 months ago. These individuals would be incorrectly coded as e-health information seekers in the data. On the other hand, some of the respondents in 2008 may have used the Internet to look for health information within the past 12 months, but not in their most recent search. These individuals would be incorrectly coded as non-seekers in the data. In both cases, the positive effect of online health information on an individual's number of doctor visits in the past 12 months would be under-estimated in the 2008 sample. Additionally, mail surveys were introduced in 2008. To account for these changes to the survey, I estimate the model separately for the 2003-05 and 2008 samples. In the 2003-05 sample, the IV estimate of the coefficient on the Internet health information seeking variable is 4.4 (p-value = 0.06). In the 2008 sample, the estimated coefficient decreases to 3.4 (p-value = 0.15). The difference between these two coefficient estimates may reflect the above mentioned changes to the questionnaire and/or sampling methodology in 2008. On the other hand, the effect of Internet health information seeking on the demand for health professional visits may have indeed decreased over time.

Next, I verify that the coefficient on the e-health information seeker dummy indeed measures the effect of health information seeking on the Internet and not the effects of other Internet related activities. It is likely that individuals who look for information about diseases or medical conditions on the Internet also use it for other health-related purposes, which

could affect their demand for health care. This would alter the interpretation of my results. To investigate this possibility, I estimate a model that controls for other health-related uses of the Internet. In particular, this model includes dummy variables that indicate whether or not an individual has participated in an online support group for people with similar health issues (*E-support group*), communicated with a doctor or doctor's office via e-mail or the Internet (*Email doctor*), or bought medicines or vitamins online (*E-pharmacy*). The OLS estimates of this model are presented in Table 5. Holding an individual's Internet use for other health-related purposes fixed, the coefficient on the e-health information seeker dummy decreases, but only slightly, and remains positive and statistically significant. In 2005, the HINTS respondents were also asked whether they had looked for information about exercise, diet, sun protection, or quitting smoking on the Internet. In 2007, the survey included a question of whether a respondent used the Internet to look for a health care provider. I estimate models that additionally control for these uses of the Internet (for 2005 and 2007 samples separately). I observe the same pattern as that in Table 5. The coefficient on the e-health information seeker variable decreases only slightly and remains positive and significant, once Internet use for other purposes is held constant. These observations support the assumption that the coefficient on the variable of interest captures the effect of health information seeking, not the effects of other uses of the Internet, on an individual's demand for health care.

5. Discussion and conclusions

The results of the empirical analysis presented in the previous section suggest that Internet health information has a positive effect on the demand for health care. Other factors being equal, e-health information seekers demand more health care than non-seekers. One possible explanation for this finding is that the information that e-health information seekers obtain from the Internet makes them more concerned about their health compared to

otherwise similar non-seekers. Most of the individuals who look for health information on the Internet report that they learn something new and that this information affects the way they care for their health (Rice, 2006). Greater health awareness, in turn, may drive e-health information seekers to visit a health professional. Moreover, surveys of doctors and patients suggest that most doctors view the information that patients bring to their offices positively (Rice & Katz, 2006; Rice, 2006), which may encourage patients to discuss online health information with their doctors. If the results of this analysis are indeed driven by Internet health information making people more aware of their health and therefore increasing their utilization of health care, their long-term health care utilization and costs may decrease. The earlier a patient sees a doctor regarding a potential health problem, the earlier it can be diagnosed and treated, which usually increases the chance that treatment is successful and lowers the patient's long-term health care utilization and/or costs (Clark et al., 2000; Leifer, 2003; Etzioni et al., 2003). Further analysis is needed to test this hypothesis.

The estimated positive effect of Internet health information on health care utilization is also consistent with the interpretation that patients lack the ability to understand medical information, as some surveys of Internet health information seekers suggest (McMullan, 2006; Tang & Lee, 2006; Ahmad et al., 2006). Therefore, even if patients have access to health information, they may not be able to interpret this information and make health-related decisions independently and thus seek health care. Finally, it is also possible that incorrect online health information leads some patients to a wrong self-diagnosis, which in turn increases their number of doctor visits. The survey data suggests that health-related information on the Internet is largely correct. For example, more than 80 percent of the e-health information seekers who discussed the information obtained from the Internet with a health professional said that the health professional agreed with this information (Fox & Rainie, 2002). Nonetheless, patients may interpret online health information incorrectly. Therefore, this interpretation of the results cannot be ruled out and deserves further investigation.

To summarize, the aim of this paper is to determine whether Internet health information seeking has an effect on individuals' health care utilization. More specifically, I examine if using the Internet to search for health or medical information affects an individual's number of visits to a health professional. The results of this analysis show that the effect of Internet health information seeking on health care utilization is positive, relatively large, and statistically significant. Additionally, this effect is found to vary across individuals. This analysis could be extended in at least three ways. The first is exploring whether the relationship between Internet health information and health care utilization varies by the type of health information and by the type of health care. Another possible extension of this analysis is an investigation of the effects of Internet health information on the demand for medicines, time spent improving health, and risky behaviors. Finally, it is important to understand how these different channels combine to affect patient health outcomes and health care costs in the long term.

Ahmad, F., Hudak, P. L., Bercovitz, K., Hollenberg, E., & Levinson, W. (2006). Are physicians ready for patients with Internet-based health information? *Journal of Medical Internet Research*, *8*.

Clark, C. M., Fradkin, J. E., Hiss, R. G., Lorenz, R. A., Vinicor, F., & Warren-Boulton, E. (2000). Promoting early diagnosis and treatment of type 2 diabetes: The national diabetes education program. *The Journal of the American Medical Association*, *284*, 363–365.

Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., Radich, J., Anderson, G., & Hartwell, L. (2003). The case for early detection. *Nature Reviews Cancer*, *3*, 243–252.

FindLaw (2012). Official state codes. FindLaw, a Thomson Reuters business, <http://law.findlaw.com/state-laws/state-codes.html>, accessed on 24 March 2012.

Fox, S., & Jones, S. (2009). The social life of health information. Pew Internet & Americal Life Project, <http://pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf>, accessed on 10 June 2010.

Fox, S., & Rainie, L. (2002). Vital decisions. Pew Internet & American Life Project, <http://pewinternet.org/~/media//Files/Reports/2002/PIP_Vital_Decisions_May2002.pdf.pdf>, accessed on 23 March 2012.

Khechine, H., Pascot, D., & Prémont, P. (2007). Relationship betwen the use of Internet information for health purposes and medical resource consumption for an English-speaking sample. *The Journal of American Academy of Business, Cambridge*, *12*, 31–37.

Lee, C.-J. (2008). Does the Internet displace health professionals? *Journal of Health Communication: International Perspectives*, *13*, 450 – 464.

Leifer, B. P. (2003). Early diagnosis of Alzheimer's disease: Clinical and economic benefits. *Journal of the American Geriatrics Society*, *51*, S281–S288.

McMullan, M. (2006). Patients using the Internet to obtain health information: How this affects the patient-health professional relationship. *Patient Education and Counseling*, *63*, 24–28.

National Cancer Institute (2006-2009). Health information national trends survey public use dataset, 2003-2007. U.S. National Institutes of Health, <http://hints.cancer.gov/dataset.jsp>, accessed on 1 April 2010.

National Telecommunications and Information Administration (2003). State and local rights of way success stories. National Telecommunications and Information Administration, U.S. Department of Commerce, <http://www.ntia.doc.gov/ntiahome/staterow/ROWstatestories.htm>, accessed on 23 August 2010.

National Telecommunications and Information Administration (2010). State and local rights of way. National Telecommunications and Information Administration, <http://www.ntia.doc.gov/ntiahome/staterow/statelocalrow.html>, accessed on 5 June 2010.

Rice, R. E. (2006). Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys. *International Journal of Medical Informatics*, *75*, 8–28.

Rice, R. E., & Katz, J. E. (2006). Internet use in physician practice and patient interaction. In M. Murero, & R. E. Rice (Eds.), *The Internet and health care: Theory, research, and practice* (pp. 149–176). New Jersey: Lawrence Erlbaum Associates, Inc.

Smith, R. B. (2005). An alternative perspective on information asymmetry implications for consumer authority in physician services markets. *Journal of Economics & Management Strategy*, *14*, 665–699.

Study Committee on Public Rights-of-Way (2002). Promoting broadband access through public rights-of-way and public lands. The National Association of Regulatory Utility Commissioners, <http://www.naruc.org/Publications/row_summer02.pdf>, accessed on 23 August 2010.

Tang, E., & Lee, W. (2006). Singapore Internet user's health information search: Motivation, perception of information sources, and self-efficacy. In M. Murero, & R. E. Rice (Eds.), *The Internet and health care: Theory, research, and practice* (pp. 107–126). New Jersey: Lawrence Erlbaum Associates, Inc.

Wagner, T. H., & Jimison, H. B. (2003). Computerized health information and the demand for medical care. *Value in Health*, *6*, 29–39.

Wallsten, S. (2005). Broadband penetration: An empirical analysis of state and federal policies. AEI-Brookings Joint Center Working Paper 05-12.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *The American Economic Review*, *93*, 133–138.

Appendix 1. Classification of the U.S. states by their right-of-way policies

States with more favorable right-of-way policies		States with less favorable right-of-way policies
Arizona	Missouri	Alabama
Arkansas	Montana	Alaska
California	Nebraska	Hawaii
Colorado	Nevada	Illinois
Connecticut	New Jersey	Indiana
Delaware	New York	New Hampshire
District of Columbia	North Carolina	New Mexico
Florida	Ohio	North Dakota
Georgia	Oklahoma	Utah
Idaho	Oregon	West Virginia
Iowa	Pennsylvania	Wisconsin
Kansas	Rhode Island	Wyoming
Kentucky	South Carolina	
Louisiana	South Dakota	
Maine	Tennessee	
Maryland	Texas	
Massachusetts	Vermont	
Michigan	Virginia	
Minnesota	Washington	
Mississippi		

Figure 1: Distribution of the number of visits to a health professional for e-health information seekers and non-seekers. *Notes*: Author's estimations using the analysis sample and sampling weights. The size of the e-health information seeker sample is 6,164, and the size of the non-seeker sample size 10,513.

Table 1: Means of control variables

	E-health info seekers	Non-seekers	
Age	41.05	48.54	(20.37)***
Male	0.43	0.50	(6.70)***
White	0.84	0.78	(6.45)***
Married/living with a partner	0.64	0.61	(2.77)***
Education			
less than high school	0.04	0.19	(20.27)***
high school	0.20	0.35	(15.83)***
some college or technical school (1-3 years)	0.38	0.28	(9.15)***
college (4 or more years)	0.38	0.18	(22.23)***
Employed (for wages or self-employed)	0.66	0.56	(9.93)***
Has health care coverage	0.90	0.84	(7.01)***
Household income			
< \$20,000	0.09	0.19	(13.32)***
\$20,000 to \$50,000	0.25	0.33	(7.62)***
\$50,000 to \$75,000	0.21	0.15	(6.73)***
> \$75,000	0.36	0.19	(16.98)***
missing	0.09	0.14	(6.19)***
Smokes (everyday or some days)	0.19	0.24	(5.41)***
Engages in physical activity once a week or more	0.76	0.65	(11.64)***
Consumes recommended quantity of fruits	0.30	0.30	(0.08)
Consumes recommended quantity of vegetables	0.12	0.07	(7.66)***
Observations	6,164	10,513	

Note: The estimates are calculated using the analysis sample and sampling weights. Absolute t-statistics for mean differences are in parenthesis. Recommended quantities of vegetable and fruits are adjusted to age and sex and are taken from the brochure "How many fruits and vegetables do you need?" provided to the public by Centers for Disease Control and Prevention (CDC) and available on <http://www.fruitsandveggiesmatter.gov/downloads/General_Audience_Brochure.pdf>
*** denotes that means are significantly different at the 1% level.

Table 2: First stage results, OLS estimates (dependent variable: e-health information seeker)

	Coefficient	Standard error
State has favorable right-of-way laws (IV)	0.04^{***}	0.01
Age	-0.01^{***}	0.00
Male	-0.06^{***}	0.01
White	0.06^{***}	0.01
Married/living with a partner	0.03^{***}	0.01
High school	0.11^{***}	0.01
Some college	0.29^{***}	0.01
College graduate	0.40^{***}	0.01
Adjusted R^2	0.15	
F-stat(1,50) (IV)	33.29	
Observations	16,677	

Notes: The instrument is a binary variable that takes the value 1 if an individual resides in a state where right-of-way regulations are relatively favorable to telecommunication providers and the value 0 otherwise. Standard errors are clustered by state. A constant and the year effects are included.
*** denotes statistical significance at the 1% level.

Table 3: OLS and IV estimates (dependent variable: the number of visits to a health professional in the past 12 months)

	(1) OLS	(2) OLS	(3) IV
E-health info seeker	1.17***	1.17***	4.08**
	(0.06)	(0.06)	(1.93)
Age	0.04***	0.02***	0.06***
	(0.00)	(0.00)	(0.01)
Male	−0.99***	−0.74***	−0.82***
	(0.07)	(0.06)	(0.14)
White	−0.04	−0.02	−0.21
	(0.07)	(0.07)	(0.13)
Married/living with a partner	−0.12**	0.00	−0.20***
	(0.05)	(0.05)	(0.08)
High school	−0.24*	−0.14	−0.57**
	(0.12)	(0.10)	(0.27)
Some college	−0.05	0.09	−0.88
	(0.14)	(0.12)	(0.57)
College graduate	−0.25*	0.01	−1.40*
	(0.13)	(0.12)	(0.77)
Employment	No	Yes	No
Health care coverage	No	Yes	No
Household income	No	Yes	No
Health-related behaviors	No	Yes	No
Adjusted R^2	0.07	0.11	-
Parameters	11	21	11
Observations	16,677	16,677	16,677

Notes: Standard errors (clustered by state) are reported in parentheses. All regressions include a constant and the year effects.

* denotes statistical significance at the 10% level.

** denotes statistical significance at the 5% level.

*** denotes statistical significance at the 1% level.

Table 4: Effect of Internet health information seeking on an individual's number of health professional visits by frequency of search, OLS estimates

	Coefficient	Standard error
Search for e-health info:		
Once a week	1.77***	0.21
Once a month	1.48***	0.22
Every few months	1.00***	0.12
Less often	0.53***	0.12
Adjusted R^2	0.08	
Observations	5,618	

Notes: Reference group consists of the individuals who did not look for health information on the Internet for themselves. Estimates are based on the 2003 sample. Standard errors are clustered by state. A constant, age, sex, race, marital status, education variables, and the year effects are included.
*** denotes statistical significance at the 1% level.

Table 5: OLS estimates controlling for other uses of the Internet (dependent variable: the number of visits to a health professional in the past 12 months)

	(1)		(2)	
	Coefficient	Standard error	Coefficient	Standard error
E-health info seeker	1.17***	0.06	1.07***	0.06
E-support group			1.11***	0.23
Email doctor			0.70***	0.15
E-pharmacy			0.03	0.09
Adjusted R^2	0.07		0.07	
Observations	16,636		16,636	

Notes: Standard errors are clustered by state. All regressions include a constant, age, sex, race, marital status, education variables, and the year effects.
*** denotes statistical significance at the 1% level.