The European Journal of Health Economics (2022) 23:337–356
https://doi.org/10.1007/s10198-021-01352-7

ORIGINAL PAPER

Online and offline health information seeking and the demand for physician services

Hiroaki Suenaga[1] · **Maria Rosalía Vicente**[2]

Received: 19 November 2020 / Accepted: 8 July 2021 / Published online: 7 September 2021

Abstract

We examine the relationship between Internet-based health information seeking and the demand for physician services, using data collected from the 28 European Union states in 2014. Unlike previous research, our analysis distinguishes seekers of health information into those who use only non-Internet sources and those who use the Internet and possibly non-Internet sources. By comparing the frequencies of physician visits among the two groups of health information seekers and non-seekers, we infer the net association between online health information and the demand for physician services while partially controlling for the effects of concurrent seeking of offline health information. The following are the two key findings: (1) individuals' health status and sociodemographic factors shape online and offline health information seeking patterns in similar ways; and (2) the demand for physician services is positively associated with offline health information seeking and not with online health information. The net association with online health information would be even smaller after controlling for the effect of concurrent offline health information seeking. These results suggest that extending the availability of online health information would potentially reinforce the unequal access to health information, which could create greater variation in individuals' health management skills and benefits from health care in the long term. However, it would be associated with little or no increase in the demand for physician services, unlike the implications of previous research.

Keywords Health information · Health care demand · Internet · Physician visits · Sample selection

JEL codes I11 · D83 · C35

Introduction

Since Arrow's [1] seminal work, uncertainty and information asymmetry have been recognized as two main features of the health care market. Uncertainty arises mainly in relation to the incidence of illness and the effectiveness of treatments provided by physicians. Patients generally possess limited information, relative to physicians, about their own health conditions, treatments provided by physicians and their likely benefits, physicians' quality, and other treatment options [2, 3]. This information asymmetry allows physicians to act as agents for their patients and creates economic incentives for physicians to induce demand for their medical services so long as it brings extra benefits to them even when the costs outweigh patients' benefits [4, 5].

While uncertainty and information asymmetry still characterize current health care markets,[1] a rapid diffusion of the Internet has allowed individuals to access a large variety of health information with practically no monetary and low opportunity (time) costs [6], which has potentially brought diverse impacts on the demand for health care. On the one hand, the extended access to health information can help patients to improve their position in the relationship with physicians and to evade physician-induced demand. It can also act as a cost-effective substitute for medical information from physicians. On the other hand, it reduces uncertainties concerning patients' health investment decisions and thus increases health care demand for risk-averse patients. Health information can also raise the demand for health care if it enhances patients' appreciation of the benefits from health

✉ Hiroaki Suenaga
hiroaki.suenaga@cbs.curtin.edu.au

[1] School of Accounting, Economics and Finance, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

[2] Department of Applied Economics, University of Oviedo, Campus del Cristo, 33006 Oviedo, Asturias, Spain

[1] Specially, in the current times of the COVID-19 pandemic.

care or raises their concerns about their health status. Due to these opposing impacts, the net effects of increased availability of online health information on the demand for health care are ambiguous.

Despite the importance of information in health care market, empirical evidence on the effects of online health information on the demand for health care is scarce and mixed, with recent studies examining the effects of Internet-based health information reporting either positive or non-significant impacts on health care demand [7–9]. This paper empirically examines the relationship between Internet-based health information and the demand for health care, specifically the demand for physician services, and contributes to this line of the literature in two important ways. First, while previous empirical studies have mainly focused on the United States and on periods before the extensive diffusion of the Internet, this study examines the data collected from the 28 European Union states in 2014, a period after the extensive diffusion of the Internet with ubiquitous access through smart phones. Second, while the media complementarity hypothesis [10] suggests that individuals seeking health information online tend to seek comparable information on other, conventional (non-Internet) media, previous studies have not considered the potential impacts of concurrent seeking of offline health information when examining the effect of online health information on health care demand. Thus, it is not clear how much of the reported impacts are attributable to information obtained from the Internet and that from other, conventional media. To address this issue, we explore potential differences between online and offline health information in their association with the demand for physician services through distinguishing individuals seeking health information exclusively on offline sources from those who seek it on online and potentially on offline sources too.

Our results indicate that individuals' health status and sociodemographic factors affect online and offline health information seeking patterns in a similar manner. Furthermore, while the demand for physician services is significantly and positively associated with offline health information seeking, it has no significant association with online health information seeking, indicating that the net association with online health information seeking, after controlling for potential concurrent offline health information seeking, would be even weaker. This result provides a stark contrast to previous studies reporting significantly positive effect of online health information. Our results suggest that fostering the availability of health information on the Internet will promote online health information seeking to wider population, yet this will be associated with little or no change in the demand for physician services in the short term. However, it can reinforce unequal distribution of health information and create even greater variation in individuals' health management skills and their health care benefits in the long term.

The rest of the paper is organized as follows: The next section discusses the roles of health information in health care investment decision and reviews previous empirical findings on the effects of health information seeking on the demand for health care, with particular emphasis on those focusing on health information from the Internet. Third and fourth section describe the data and introduce three econometric models that measure the differential effects of online and offline health information seeking on health care demand. Fifth section presents the results of the empirical analysis. Sixth section concludes the paper with a synopsis of the results.

Effects of health information on health care demand

Conceptual framework

Theoretically, a rational individual will determine the optimal level of health care investment by equating the marginal benefits and marginal costs of health care. However, in the absence of perfect information, individuals do not fully appreciate the true benefits of health care. Health information helps them to make better health care investment decisions by reducing the uncertainty of their returns. It can also alter an individual's preference and enhance his/her appreciation of good health condition, relative to other goods. Acquiring new health information, however, incurs monetary and non-monetary (time) cost. Thus, an individual determines the optimal level of health information seeking by equating the marginal benefits and costs of seeking health information from various sources. This decision is made simultaneously with the health care investment decisions, as the benefits of health information emanates at least partly through enhancing benefits of health care.

A rapid diffusion of the Internet has facilitated fast access to health information at reduced cost and multiplied the volume of available information [11, 12]. This cost reduction can promote health information seeking among individuals with low expected benefits from the acquired information, which potentially creates two opposing impacts on the demand for health care. First, health information increases the demand for health care as it reduces uncertainties of health care benefit and enhances consumers' appreciation of health and wellbeing. For example, acquiring health information prior to visiting physicians can enhance patients' benefits from their visits [13, 14]. Ultimately, this could lead patients to create excess

demand for medical services from their physicians [15].[2] Health information can also act as a complement to the services provided by medical professionals. For example, patients with insufficient medical knowledge might seek physicians' support to understand information obtained from the Internet or other media. In another case, patients might use the Internet to gain further insights about the diagnoses and treatments recommended by physicians [6, 16]. In both cases, the cost reduction in health information will increase the demand for both health information and physician services.

Second, additional health information can decrease the demand for health care if it improves patients' position in their relationship with physicians and thus allows them to evade physician-induced demand. Health information can also act as a substitute for medical services provided by physicians when patients face high monetary or opportunity costs of gathering information from physicians [6]. Such would be the case for uninsured individuals and those with high travel costs to their usual health center. With these opposing impacts, the net effects of increased availability of online health information on the demand for health care are ambiguous and vary among individuals due to the heterogeneity in the costs and benefits of the acquired information.

Empirical work on health information's impacts on health care demand

Health economists have long tried to empirically estimate the effects of health information on the demand for health care and determine whether positive or negative effects prevail. Empirical research has shown mixed evidence. For example, using direct measures of health information based on individuals' health knowledge, Kenkel [17] reports that consumer health information increases the probability of physician visits while it has no significant effect on the number of visits. Hsieh and Lin [2] confirmed the positive effect on the demand for preventive care among elders. Dwyer and Liu [18] also found a positive effect, but they identified weak yet negative effects on the use of physician services and emergency rooms among individuals with low trust of doctors. Contrary to these results, Schmid [3] reports a significant negative effect.

More recently, a few studies have investigated on how health information from the Internet impacts the demand for health care. Lee [8] reports that online health information seeking had positive and unidirectional impacts on the demand for two types of medical services: medical information and physical treatment. Suziedelyte [9] confirms the positive effect of online health information seeking on the number of physician visits. In contrast, Beck et al. [7] report survey results in which 88.6% of young online health information seekers self-report that the use of the Internet for health matters has not changed the frequency of their medical consultations.[3]

Inconsistency of the results previously reported on the effect of health information on the demand for health care can be attributable to several factors, yet two are of strong relevance for our analysis. Firstly, previous studies analyzed data from early stages of the diffusion of Internet services, when collecting intended information was still costly due to high cost of Internet connection and limited availability of reliable health information. This high cost possibly limited online health information seeking among individuals expecting high benefits from the intended information, leading previous studies to report positive impacts of online information on the demand for health care. Secondly, the media complementarity hypothesis [10] suggests that individuals seeking information on the Internet also seek comparable information on other, conventional media. The hypothesis has received strong empirical support for the case of health information [21–25].[4] However, no studies have explicitly considered the effects of offline health information seeking on the demand for health care when examining those of online information seeking. The subsequent sections develop models to address these limitations, analyzing data from a period after the extensive diffusion of the Internet, in particular of broadband services.

Data

Data analyzed in this study are sourced from the Flash Eurobarometer 404 "European Citizens' Digital Health Literacy" coordinated by the European Commission [28].[5] The survey collected data from 26,566 individuals aged 15 years and older about their use of Internet resources to manage their health. It was conducted through phone interview in

[2] Shih and Tai-Seale [15] define the term "demand-induced supply" as additional medical services provided by physicians at the request of patients that otherwise would not have been offered. We refer to this phenomenon as patient-induced supply.

[3] The self-reported behavior can differ from revealed behavior, thus its validity is often discussed in the health literature. However, previous studies commonly report that the difference is small (see for example, [19]) except that self-reported ambulatory physician visits tend to be downward biased among male aged 40–79 years [20].

[4] Other studies also commonly report that sociodemographic factors shape online and offline health information seeking in a similar manner, which indirectly supports the media complementarity hypothesis [26, 27].

[5] Details of the survey are available at European Commissions' Eurobarometer survey's website: http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/General/index.

the 28 Member States of the European Union over September 18–20, 2014.[6] Data include detailed information about respondents' frequencies of health information search and physician visits, their health conditions, and sociodemographic features.

Table 1 lists the variables used in the analysis and provides their descriptive statistics. The frequency of health information seeking in the last 12 months is originally measured in the questionnaire on an ordinal scale from 1 (never) to 6 (once a week or more often). Importantly, the survey distinguishes between Internet-based and offline health information resources.[7] However, while the frequency of online health information search (*HIi*) was asked to all sampled individuals, that of offline search (*HIni*) was only asked to individuals who never sought health information online. The empirical models developed in the subsequent section address potential sample selection bias resulting from this survey design. As an alternative measure, we also construct a multinomial variable to classify respondents into three groups according to their health information seeking: (i) non-seekers, individuals who do not seek health information either offline or online ($HIS = 0$); (ii) online health information seekers, who seek health information online and do not reveal their offline health information seeking (they might or might not seek health information offline) ($HIS = 1$); and (iii) offline health information seekers, who only seek health information offline and not online ($HIS = 2$). For the second key variable, the demand for health care is proxied by the frequency of physician visits in the last 12 months (*PV*), which is originally measured in the questionnaire on an ordinal scale from 1 (never) to 4 (6 times or more). This is a conventional measure in the literature.

For factors influencing the demand for health care and health information, three variables relating to individuals' health conditions are considered: respondents' self-assessed health status (*Health*), coded in four ordered categories from 1 (very bad) to 4 (very good); a binary variable indicating whether the respondent suffers from a long-term health issue (*LONGILL*); and the frequency of physical activity (*Sport*), coded in six ordered categories from 1 (never) to 6 (5 times a week or more). Respondents' self-assessed health measures might not accurately reflect respondents' true health condition. However, they are considered to be relevant for individuals' health information seeking and health investment decisions, and thus are commonly used in the literature [2, 3, 17]. For all three variables, better health is expected to relate inversely to the frequency of physician visits and health information seeking, as a priori, individuals in poor health are expected to receive higher benefits from health care services and health information than those in better health [6]. Nonetheless, the inverse health information law posits that the use of health information relates inversely to the needs of the population [7, 29, 30]. Hone et al. [31] and Vicente and Madden [32] provide empirical support to this law, reporting that individuals in poorer health condition seek health information online less frequently than those in better condition, due to their lower abilities to search online.

Respondents' sociodemographic factors are included into both health information search and physician visitation equations, including age (*AGE*), gender (*FEMALE*), employment status (*Employment*) in three categories (employed, unemployed, inactive), household size (*ONEADULT*), type of locality (*Area*) in three categories (rural, small, large), and country of residence (*Country*). These variables are commonly considered in the literature as determinants of health information seeking and health care demand. In particular, individuals who are younger, married, and female tend to seek health information more frequently on the Internet (see for example, [6, 13, 33–35]).

As for the factors determining the demand for health care, previous evidence suggests a positive relationship between availability of health care service and health care demand. In particular, high physician density implies more competition which could strengthen incentives for physician-induced demand and lower health care costs [3, 18, 36]. Thus, data on the density of medical doctors per resident at NUTS2 level (*DENSITY*) are collected from Eurostat [37].

In the literature, the price of medical services is often approximated by patients' health insurance coverage [17]. Unfortunately, our dataset provides no information about respondents' insurance status. The exclusion of the price variable might raise a concern about potential omitted variable bias. However, such bias would be negligible for the case of the European Union where public health coverage

[6] The sample is distributed across 28 Member States with about 1000 respondents (3.8% of sample size) per country for Austria, Belgium, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, the Netherlands, Poland, Portugal, Spain, Sweden, Romania, Slovakia, Slovenia, and the United Kingdom; and 500 respondents (1.9% of sample size) for Cyprus, Luxembourg, and Malta.

[7] Specifically, questions on health information seeking are stated in the survey as: for Internet sources, "Within the last 12 months, have you used the Internet to search for health-related information?"; and for non-Internet sources, "Within the last 12 months, how often have you looked for health-related information, whatever the means used?" Offline sources include friends, relatives, books/encyclopedias, newspapers/magazines, and medical programmes/documentaries on radio and TV. For each of the two questions, respondents were asked to choose one from six ordered categories: (1) never, (2) a few times within the past year, (3) approximately once every two months, (4) approximately once a month, (5) several times a month, and (6) once a week or more often.

Table 1 Description of variables

Name	Description	Mean (SD)
HIi	Frequency of health information search on the Internet in the last 12 months	
	= 1 if the respondent did not seek health information online	
	= 2 if the respondent sought health information online a few times a year	0.17
	= 3 if the respondent sought health information online every two months	0.04
	= 4 if the respondent sought health information online once a month	0.12
	= 5 if the respondent sought health information online several times a month	0.08
	= 6 if the respondent sought health information online once a week or more often	0.09
HIni	Frequency of health information search on non-Internet sources in the last 12 months	
	= 1 if the respondent did not seek health information offline	
	= 2 if the respondent sought health information offline a few times a year	0.16
	= 3 if the respondent sought health information offline every two months	0.03
	= 4 if the respondent sought health information offline once a month	0.08
	= 5 if the respondent sought health information offline several times a month	0.05
	= 6 if the respondent sought health information offline once a week or more often	0.08
HIS	Types of health information seekers	
	= 0 if the respondent reports not seeking health information	
	= 1 if the respondent reports using the Internet for health information purposes	0.51
	= 2 if the respondent reports seeking health information offline and not online	0.19
PV	Frequency of physician visitation in past 12 months	
	= 1 if the respondent did not visit physician	
	= 2 if the respondent visited physician 1 or 2 times	0.37
	= 3 if the respondent visited physician 3 to 5 times	0.24
	= 4 if the respondent visited physician 6 times or more	0.23
DENSITY	Number of medical doctors per hundred thousand inhabitants at NUTS2 level	325.83 (114.73)
INTERNET	Internet penetration rate (in percentage) at NUTS2 level	77.92 (11.10)
Health-related variables		
Health	Self-assessed health status (reference category: very bad)	
FBAD	= 1 if the respondent reports health status as fairly bad	0.11
FGOOD	= 1 if the respondent reports health status as fairly good	0.61
VGOOD	= 1 if the respondent reports health status as very good	0.25
LONGILL	= 1 if respondent reports suffering from a long-term health issue	0.46
Sport	Respondent's frequency of sport or exercise practice (reference category: none)	
MONTH	= 1 if the respondent plays sport or exercises less than once a month	0.08
1–3 MONTH	= 1 if the respondent plays sport or exercises 1 to 3 times a month	0.05
1–2 WEEK	= 1 if the respondent plays sport or exercises 1 to 2 times a week	0.22
3–4 WEEK	= 1 if the respondent plays sport or exercises 3 to 4 times a week	0.17
5 WEEK	= 1 if the respondent plays sport or exercises 5 times a week or more	0.23
Socio-demographic variables		
FEMALE	= 1 if respondent is female	0.61
ONEADULT	= 1 if household consists of one adult	0.25
Employment	Respondent's employment status (reference category: inactive)	
EMPLOYED	= 1 if the respondent is employed	0.46
UNEMPLOY	= 1 if the respondent is unemployed	0.04
Area	Size of the residence area (reference category: rural area or village)	
SMALL	= 1 if the respondent lives in a small medium-sized town	0.37
LARGE	= 1 if the respondent lives in a large town/city	0.31
Education	Respondent's age when stopped full-time education (reference category: no full-time education)	
EDU15	= 1 if the respondent stopped full-time education at 15 years old or less	0.13
EDU19	= 1 if the respondent stopped full-time education between 16 and 19 years old	0.37
EDU20	= 1 if the respondent stopped full-time education at 20 years old or higher	0.43
STUDY	= 1 if the respondent is still studying	0.05
AGE	Respondent's age (in years)	53.25 (18.04)
Country	28 dummies, in which = 1 for the country where the respondent lives	

Note: For the four variables on health information seeking and physician visitation, the reported numbers represent the share of respondents reporting the corresponding ordered choice. SD stands for standard deviation

is practically universal.[8] Furthermore, our regression equations include country-specific effects as well as other variables such as employment status, density of medical doctors, and size of residence area, which together will capture the remaining cross-country variation in health care coverage and differences in travel and time cost of visiting a physician.

As for the demand for health information, unequivocal evidence exists that education is among the most important determinants of individuals' health knowledge and health information demand [2, 3, 17, 18, 39, 40]. Health information on the Internet in particular varies substantially in quality and reliability; thus, its benefits depend largely on individuals' previous experience and ability to assess information quality [6, 13, 32]. Education improves such ability and lowers learning and searching costs. Thus, we include respondents' educational attainment (*Education*), measured on a five-ordinal scale, in our models of online and offline health information seeking.[9]

Both online and offline health information seeking can be also affected by Internet accessibility. Availability of a high-speed Internet connection reduces the cost of acquiring health information online relative to other sources and hence encourages online seeking while potentially discouraging offline seeking. For example, Costa-Font et al. [13] report a positive correlation between the frequency of online health information search and Internet penetration. Accordingly, our empirical models of online and offline health information seeking include regional (NUTS2) Internet penetration rate (*INTERNET*), measured as the percentage of households with Internet access.[10]

Empirical model

The extended sample selection model

Empirical models to examine the effect of health information seeking on the demand for health care need to address two issues. The first is sample selection: the Eurobarometer survey collected data on offline health information search only for individuals who never sought health information online. This survey design creates a setting similar to Heckman's [42] sample selection model; yet the dependent variables in the sample selection and activity equations (respondents' online and offline health information search, respectively) are both measured on ordinal scales in our setting. An ordinal dependent variable in the activity equation makes the model non-linear, for which the standard two-step approach proposed by Heckman [42] yields biased and inconsistent estimates. Greene and Hensher [43] suggest estimating selection and activity equations jointly by the method of maximum likelihood when the two equations are specified as binary and ordered probit models, respectively. Their approach is extended directly to our case where the two equations form a bivariate ordered probit model.

The second issue, as commonly considered by previous empirical studies, is the endogeneity of health information variables in the health care demand equation [17, 18]. This issue arises for two reasons. First, online and offline health information seeking and physician visitation are considered as simultaneously determined in our models. This treatment is consistent with the way the data were recorded: the survey collected data on the frequencies of health information seeking and physician visits over the same twelve-month period, and not on the detailed sequence of how these activities were conducted over that period. Second, health information seeking and physician visitation are potentially affected by unobserved common factors such as individuals' preferences for health, health care, time as well as risk [6, 17, 18]. For example, an individual appreciating good health is likely to exhibit high demand for both health information and services. Also, an individual with high time cost might consider searching health information to substitute the comparable services provided by physicians. Risk-averse individuals might consume more health services and information than risk lovers to reduce their health risk; yet, it is also possible that they perceive physician provided services more (or less) reliable than self-collected health information.

To address the endogeneity issue, we estimate each of the online and offline health information search equations simultaneously with the physician visitation equation. Thus, for the selection equation, a system of two ordered probit equations is specified for the frequencies of online health information search (*HIi*) and physician visitation (*PV*):[11]

[8] Public coverage reaches 100% in 18 Member States, while it ranges between 91 and 98% in five countries and between 83 and 89% in the other five nations [38].

[9] In the survey, education is measured in five ordered categories of the age at which the respondent completed his/her full-time education, which differs from the standard measure of education (typically, the number of years in education) for respondents experiencing some breaks before completing their final year of education.

[10] Data are sourced from Eurostat [41].

[11] Demand for physician services is commonly modelled by the hurdle model, particularly when it is measured by the number of physician visits (PVs) over a specified period. Such model allows the covariates to affect differently the probability of PV and the number of PVs conditional on the patient visiting physicians. In our case, the ordered probit model is the natural choice given that online and offline health information seeking and PV are all measured in ordinal scale. Besides, the threshold parameters, c_{3,j_3} and c_{4,j_4}, are estimated endogenously in Eqs. (1) and (2), which accommodates potential non-linear relationship between PV and the covariates that are commonly modelled by the hurdle model in the literature.

$$\begin{aligned} HIi_i^* &= \mathbf{X}_i\beta_{11} + \mathbf{Z}_{1,i}\beta_{12} + \varepsilon_{1,i} \\ PV_i^* &= \mathbf{X}_i\beta_{31} + \mathbf{Z}_{3,i}\beta_{32} + \beta_{33}HIi_i^* + \varepsilon_{3,i} \end{aligned} \tag{1}$$

where the observed ordered categorical variable HIi_i for individual i takes a value j_1 ($j_1 = 1, ..., 6$) if the associated latent variable HIi_i^* takes a value within the thresholds ($c_{1,j_1-1} < HIi_i^* < c_{1,j_1}$) with $c_{1,0} = -\infty$ and $c_{1,6} = +\infty$. Similarly, $PV_i = j_3$ ($j_3 = 1, ..., 4$) if $c_{3,j_3-1} < PV_i^* < c_{3,j_3}$ with $c_{3,0} = -\infty$ and $c_{3,4} = +\infty$. The vector $\mathbf{X}_i$ represents the set of covariates that affect the two latent variables, HIi_i^* and PV_i^*, whereas $\mathbf{Z}_{1,i}$ and $\mathbf{Z}_{3,i}$ represent the set of covariates that uniquely affect online health information search and physician visitation, respectively. The inclusion of the vector $\mathbf{Z}_{1,i}$ assures the identification of the model.[12]

For individuals not seeking health information online ($HIi = 1$), the outcome equations are formed by two ordered probit equations, corresponding to the frequencies of offline health information search ($HIni$) and physician visitation as follows:

$$\begin{aligned} HIni_i^*\big|_{HIi_i=1} &= \mathbf{X}_i\beta_{21} + \mathbf{Z}_{2,i}\beta_{22} + \varepsilon_{2,i}, \\ PV_i^*\big|_{HIi_i=1} &= \mathbf{X}_i\beta_{41} + \mathbf{Z}_{3,i}\beta_{42} + \beta_{43}HIni_i^* + \varepsilon_{4,i}, \end{aligned} \tag{2}$$

where $HIni_i = j_2$ ($j_2 = 1, ..., 6$) if $c_{2,j_2-1} < HIni_i^* < c_{2,j_2}$ and $PV_i = j_4$ ($j_4 = 1, ..., 4$) if $c_{4,j_4-1} < PV_i^* < c_{4,j_4}$ with $c_{2,0} = c_{4,0} = -\infty$ and $c_{2,6} = c_{4,4} = +\infty$. The vector $\mathbf{Z}_{2,i}$ contains the set of covariates that are unique to the offline health information search equation.

In Eqs. (1) and (2), the matrix of common exogenous variables includes respondents' sociodemographic features and health status $\mathbf{X} = \{AGE, FEMALE, ONEADULT, Employment, Area, Country, Health, Sport, LONGILL\}$. It does not include individuals' educational attainment (*Education*). This is a standard practice in the literature [2, 3, 17, 18] and presumes that education may impact the health care demand indirectly through affecting respondent's propensity to health-oriented practices and health information seeking. These indirect effects are controlled for by including on the right-hand side of the physician visitation equation the covariates such as health conditions, sport and physical activity (as measures of respondent's past and current health-oriented practices), employment status, and health information seeking.[13] The equation specific factors include $\mathbf{Z}_1 = \{Education, INTERNET\}$, $\mathbf{Z}_2 = \{Education\}$, and $\mathbf{Z}_3 = \{DENSITY\}$.

The vector of error terms is assumed to be distributed as follows:[14]

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & 0 & \rho_{24} \\ \rho_{13} & 0 & 1 & 0 \\ \rho_{14} & \rho_{24} & 0 & 1 \end{pmatrix}\right) = N(0, \mathbf{\Sigma}) \tag{3}$$

Under this normality assumption, the coefficients in the two bivariate ordered probit models are estimated jointly by the method of maximum likelihood, using the likelihood function derived in the Appendix.

The specification in (1) and (2) allows the coefficients for $\mathbf{X}$ and $\mathbf{Z}_3$ in the physician visitation equations to differ between seekers and non-seekers of online health information. Alternatively, a more restrictive specification sets these coefficients to be identical between the two groups, $\beta_{31} = \beta_{41}$ and $\beta_{32} = \beta_{42}$, which simplifies the physician visitation equations in (1) and (2) as follows:

$$\begin{aligned} PV_i^* = \; & \mathbf{X}_i\beta_{31} + \mathbf{Z}_{3,i}\beta_{32} + I(HIi_i > 1)\left(\beta_{33}HIi_i^* + \varepsilon_{3,i}\right) \\ & + I(HIi_i = 1)\left(\beta_{43}HIni_i^* + \varepsilon_{4,i}\right), \end{aligned} \tag{4}$$

where $I(\cdot)$ is the indicator variable that takes a value 1 if its argument is true and 0 otherwise.

Alternative models for robustness checks

We also consider two alternative models to assess the robustness of the estimates of the extended sample selection model. The first alternative specifies a bivariate ordered probit model for the frequencies of online health information search and physician visitation, without modelling offline health information search as follows:

$$\begin{aligned} HIi_i^* &= \mathbf{X}_i\alpha_{11} + \mathbf{Z}_{1,i}\alpha_{12} + v_{1,i} \\ PV_i^* &= \mathbf{X}_i\alpha_{21} + \mathbf{Z}_{3,i}\alpha_{22} + \alpha_{23}HIi_i^* + v_{2,i} \end{aligned} \tag{5}$$

Unlike the extended sample selection models specified in (1) through (4), model (5) does not address the potential effects of concurrent seeking of offline health information on physician visitation. This is the approach commonly used by previous studies [8, 9]. The two equations in (5) are jointly estimated by the method of maximum likelihood.

[12] Specifically, the identification requires the dimension of $\mathbf{Z}_{1,i}$ to be no less than the number of endogenous variables on the right-hand side of the physician visitation (PV) equation. PV equations in (1) and (2) are specified as a function of the latent variable of online and offline health information seeking (HIi^* and $HIni^*$) rather than the observed categorical variables (HIi and $HIni$), which reduces the dimensions of $\mathbf{Z}_{1,i}$ and $\mathbf{Z}_{2,i}$ required for identification.

[13] All three models presented in Sect. 4 were also estimated with the education variables on the right-hand side of the physician visitation equation, yet their coefficients were not statistically significant. Lee [8] and Suziedelyte [9] similarly report that education variables have no direct impacts on the demand for health care.

[14] In (3), $\rho_{23} = \rho_{34} = 0$ is imposed because the error term ε_3 in the physician visitation equation for online health information seekers is not observed jointly with ε_2 and ε_4 from information seeking and physician visitation equations, respectively, for non-online seekers.

For the second alternative, we construct a two-equation model, using a multinomial measure of respondents' health information seeking activity (*HIS*):

$$\begin{aligned} HIS_i &= \mathbf{X}_i\gamma_{11} + \mathbf{Z}_{2,i}\gamma_{12} + u_{1,i}, \\ PV_i^* &= \mathbf{X}_i\gamma_{21} + \mathbf{Z}_{1,i}\gamma_{22} + \gamma_{23}I(HIS_i = 1) + \gamma_{24}I(HIS_i = 2) + u_{2,i}, \end{aligned} \tag{6}$$

where the health information search and physician visitation equations are modelled as a multinomial probit and an ordered probit model, respectively.

Model (6) is estimated by the two-stage residual inclusion method which yields biased yet consistent estimates of the coefficients [44]. Under this method, the health information equation is estimated first, by the method of maximum likelihood; then, the physician visitation equation is estimated by including, on the right-hand side of the equation, the observed two indicator variables ($I(HIS_i = j)$, $j = 1$ and 2) and the residuals from the estimated health information equation in the first stage.

Results

Extended sample selection model—health information search equations

Table 2 shows the estimates of the extended sample selection model specified in (1) through (4).[15] As to the use of the Internet to search health information (column 1 in Table 2), practically all the explanatory variables are statistically significant at the 1% level. Specifically, the estimated coefficients are significantly negative for age and living alone. These following results are as expected: older people are less likely to seek health information online, as they generally show low rates of Internet use due to technology anxiety and limited ability to deal with new technologies. Living alone means less need to seek health information for other household members. Results also suggest that being active in the labor market, having more education, and living in regions with high Internet penetration rates are all positively related to online health information seeking. Women also seek health information online more frequently than men. This is consistent with a common finding in the literature that women tend to take charge of their households' health matters [7, 17, 46].

Concerning health status, individuals with a long-term illness tend to seek health information online more frequently than those without a long-term illness. On the contrary, self-assessed health status is positively related to the frequency of online health information seeking, i.e., respondents reporting a better health status search health information online more frequently than those describing their health status as very bad. These contradictory results are not surprising given that previous research also reports mixed or weak evidence on the link between individuals' health condition and information demand (e.g., [2, 3, 17, 18]). A plausible interpretation for this counterintuitive result can be sought in the heterogeneity of individuals' health interests. Specifically, those in better health condition tend to look for information on general health and wellbeing, whereas those in worse condition tend to seek information specific to some particular illness or treatment [31]. Accordingly, the estimated positive coefficient of self-assessed health status might reflect a major use of the Internet by healthy individuals to gather general health information. This interpretation also aligns with positive coefficients estimated for the frequency of sports practice (*Sport*).

In column 2 of Table 2, the estimates of the offline health information search equation among non-online seekers indicate similar patterns to online search behaviour. Although age is expected to affect offline search less negatively than online search, its coefficient estimate keeps a negative sign. Previous studies also recognize a complex relationship between age and health information [17] and report inconclusive empirical evidence [2, 3, 18]. A non-monotonic relationship between health condition and information seeking is also observed in the offline setting, with positive coefficients estimated for both long-term illness and individuals' self-assessed health status.

Overall, the similarity of the results obtained for the sociodemographic determinants of online and offline health information seeking implies that fostering Internet usage or increasing availability of online health information potentially exacerbates inequalities in health information access, rather than alleviating them. It also supports indirectly the media complementarity hypothesis, i.e., individuals seeking health information on the Internet also seek comparable information on conventional media. Finally, the relationship between respondents' health status and information seeking is non-monotonic and complicated, due possibly to the heterogeneity in individuals' health interests. This result suggests that empirical support for the inverse information law is sensitive to the measures of health status and types of health information.

[15] The extended sample selection models (1)–(4) and bivariate ordered probit model (5) were estimated through Stata's cmp routine [45]. For all three models, the standard errors are obtained based on the variance–covariance matrix that accounts for correlation among individuals residing in the same region. This correlation results from the use of region-specific instrumental variables (*DENSITY* and *INTERNET*).

Table 2 Maximum likelihood estimates of extended sample selection model

Variables	(i) Unconstrained				(ii) Constrained			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	HIi	*HIni*	*PV* for *HIi*	*PV* for *HIni*	*HIi*	*HIni*	*PV* for *HIi*	*PV* for *HIni*
*HIi**			−0.034				0.042	
			(0.045)				(0.060)	
*HIni**				0.187**				0.247***
				(0.084)				(0.040)
INTERNET	0.006***				0.006***			
	(0.002)				(0.002)			
DENSITY			0.312**	0.320**			0.309***	0.309***
			(0.126)	(0.163)			(0.103)	(0.103)
AGE	−0.023***	−0.010***	−0.003***	0.002	−0.023***	−0.010***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
FEMALE	0.220***	0.345***	0.222***	0.112***	0.222***	0.345***	0.156***	0.156***
	(0.021)	(0.027)	(0.022)	(0.030)	(0.021)	(0.025)	(0.019)	(0.019)
ONEADULT	−0.190***	−0.140***	0.010	−0.056**	−0.188***	−0.142***	−0.009	−0.009
	(0.021)	(0.021)	(0.026)	(0.023)	(0.021)	(0.021)	(0.018)	(0.018)
Employment (reference is "inactive")								
EMPLOYED	0.227***	0.095***	−0.203***	−0.187***	0.228***	0.101***	−0.225***	−0.225***
	(0.020)	(0.027)	(0.027)	(0.030)	(0.020)	(0.026)	(0.023)	(0.023)
UNEMPLOY	0.105**	0.090**	−0.131**	−0.102*	0.107**	0.097**	−0.139***	−0.139***
	(0.044)	(0.042)	(0.052)	(0.061)	(0.045)	(0.042)	(0.043)	(0.043)
Area (reference is "rural/village")								
SMALL	0.101***	0.074***	0.019	0.017	0.102***	0.075***	0.011	0.011
	(0.019)	(0.023)	(0.024)	(0.024)	(0.019)	(0.023)	(0.018)	(0.018)
LARGE	0.152***	0.092***	0.052*	0.098***	0.151***	0.093***	0.065***	0.065***
	(0.021)	(0.026)	(0.028)	(0.031)	(0.021)	(0.026)	(0.022)	(0.022)
Health (reference is "very bad")								
FBAD	0.203***	0.177***	−0.363***	−0.212***	0.206***	0.188***	−0.256***	−0.256***
	(0.063)	(0.058)	(0.123)	(0.058)	(0.063)	(0.057)	(0.056)	(0.056)
FGOOD	0.329***	0.208***	−0.986***	−0.736***	0.331***	0.219***	−0.829***	−0.829***
	(0.060)	(0.058)	(0.116)	(0.061)	(0.060)	(0.057)	(0.060)	(0.060)
VGOOD	0.230***	0.068	−1.327***	−1.041***	0.232***	0.081	−1.146***	−1.146***
	(0.059)	(0.061)	(0.119)	(0.064)	(0.059)	(0.060)	(0.060)	(0.060)
LONGILL	0.235***	0.378***	0.800***	0.803***	0.234***	0.381***	0.782***	0.782***
	(0.016)	(0.023)	(0.023)	(0.039)	(0.017)	(0.023)	(0.021)	(0.021)
Sport (reference is "none")								
MONTH	0.247***	0.230***	0.056	0.094*	0.247***	0.232***	0.058*	0.058*
	(0.030)	(0.037)	(0.044)	(0.049)	(0.030)	(0.037)	(0.032)	(0.032)
1 – 3MONTH	0.247***	0.285***	−0.007	0.041	0.247***	0.288***	−0.001	−0.001
	(0.034)	(0.042)	(0.054)	(0.053)	(0.034)	(0.042)	(0.036)	(0.036)
1 – 2WEEK	0.322***	0.304***	0.013	0.082*	0.321***	0.307***	0.018	0.018
	(0.025)	(0.039)	(0.035)	(0.046)	(0.024)	(0.039)	(0.029)	(0.029)
3 – 4WEEK	0.371***	0.396***	0.037	0.099**	0.369***	0.399***	0.035	0.035
	(0.027)	(0.037)	(0.040)	(0.043)	(0.027)	(0.037)	(0.029)	(0.029)
5 WEEK	0.259***	0.316***	0.004	0.031	0.259***	0.317***	−0.001	−0.001
	(0.029)	(0.030)	(0.041)	(0.045)	(0.028)	(0.030)	(0.032)	(0.032)
Education (reference is "no full-time education")								
EDU15	0.209*	0.173*			0.222*	0.215**		
	(0.126)	(0.099)			(0.128)	(0.092)		

Table 2 (continued)

Variables	(i) Unconstrained				(ii) Constrained			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	HIi	*HIni*	*PV* for *HIi*	*PV* for *HIni*	*HIi*	*HIni*	*PV* for *HIi*	*PV* for *HIni*
EDU19	0.709***	0.518***			0.724***	0.561***		
	(0.125)	(0.104)			(0.126)	(0.097)		
EDU20	1.050***	0.800***			1.066***	0.841***		
	(0.125)	(0.107)			(0.126)	(0.098)		
STUDY	0.769***	0.710***			0.789***	0.765***		
	(0.135)	(0.119)			(0.135)	(0.110)		
Thresholds								
$C_{k,1}$	0.788***	0.433***	−1.939***	−1.299***	0.802***	0.527***	−1.546***	−1.546***
	(0.184)	(0.155)	(0.138)	(0.147)	(0.188)	(0.118)	(0.112)	(0.112)
$C_{k,2}$	1.344***	0.705***	−0.552***	−0.056	1.358***	0.799***	−0.223**	−0.223**
	(0.183)	(0.161)	(0.140)	(0.149)	(0.188)	(0.122)	(0.114)	(0.114)
$C_{k,3}$	1.491***	0.775***	0.338**	0.735***	1.505***	0.869***	0.621***	0.621***
	(0.183)	(0.164)	(0.142)	(0.155)	(0.188)	(0.124)	(0.117)	(0.117)
$C_{k,4}$	1.944***	0.975***			1.958***	1.069***		
	(0.183)	(0.173)			(0.188)	(0.129)		
$C_{k,5}$	2.369***	1.147***			2.383***	1.242***		
	(0.184)	(0.185)			(0.189)	(0.140)		
Correlations								
		ρ_{12}	ρ_{13}	ρ_{14}		ρ_{12}	ρ_{13}	ρ_{14}
		0.885***	0.095**	0.222***		0.883***	0.009	0.184***
		(0.035)	(0.042)	(0.061)		(0.029)	(0.063)	(0.045)
				ρ_{24}				ρ_{24}
				0.134**				0.090***
				(0.052)				(0.031)
Observations	24,971				24,971			

Note: The table reports the results from estimating the extended sample selection model defined in Eqs. (1)-(4) by the method of maximum likelihood. Columns (1)–(4) show the estimates of health information and physician visitation equations for seekers and non-seekers of online health information. *HIi* (*HIni*) refers to the equation for online (offline) health information seeking, and *PV* for *HIi* (*PV* for *HIni*) refers to the equation for physician visitation for online health information seekers (non-seekers). Columns (5)–(8) report the estimates of these equations, constraining the coefficients to be identical for the common variables between the two physician visitation equations. Robust standard errors are in parentheses. Country dummies are included in the estimations and their coefficient estimates are reported in Table 2A in the Appendix

***$p<0.01$, **$p<0.05$, *$p<0.1$

Extended sample selection model—physician visitation equations

Columns 3 and 4 of Table 2 show the results for the physician visitation equations. The first thing to notice is that online and offline health information search relate to physician visits differently: the estimated coefficient is negative and not significant for the former, whereas it is positive and significant for the latter. As previously explained, the reference populations differ between the two physician visitation equations as follows: they are, for offline health information search, individuals who did not seek health information on the Internet and, for online search, those who sought health information online and possibly on non-Internet sources as well. Thus, the two coefficients can be interpreted, respectively, as the *net* association of offline information search with the demand for physician services (conditional on not seeking health information online), and the *gross* association of online search. The significant positive coefficient for offline health information search suggests that individuals seeking health information offline visit physicians more frequently than those not seeking health information offline. Several explanations are possible for this complementary relationship. First, seeking health information offline before or after visiting physicians could enhance the benefits from physician services by increasing patients'

appreciation of these services or by reducing their uncertainty. These increased (expected) benefits would in turn lead the better-informed patients to demand more physician services. Second, health information could increase patients' anxiety about their health condition and hence make them resort more to support and advice from physicians. In relation to the long-standing hypotheses in the health economics literature, the positive relationship is in line with the patient-induced supply hypothesis, while disagreeing with the physician-induced demand hypothesis.

By contrast, the estimated coefficient for online health information search is not statistically significant. This coefficient represents the sum of the direct (or *net*) association between online health information search and the demand for physician services, and the possible indirect link through concurrent offline searches. This indirect link would be positive (negative) if online and offline health information seeking were complements (substitutes), coherently with the significantly positive coefficient of offline search. Our dataset does not allow us to test for the complementarity of online and offline searches explicitly, yet previous studies commonly support the complementarity of health information from various media [21–25]. If this was the case, the indirect link through concurrent offline search would be positive and the *net* association of online search would be strongly negative, to make the *gross* association non-significant yet negative. Were online and offline health information substitutes, the indirect link would be negative, and the *net* association of online search would be positive yet weaker than the *net* effect of offline search. The magnitude of the indirect link and the net association of online search fall between the above two extreme cases if online and offline health information are imperfect substitutes or complements. In either case, our result provides no support for either the patient-induced supply or physician-induced demand hypotheses for online health information search.

Scarce previous evidence has shown mixed evidence on the effect of health information on the demand for health care, with two recent studies reporting significantly positive estimates for the online case [8, 9]. While various factors can contribute to the difference between our results and previous evidence, the most plausible one will be a substantial reduction in the cost of information seeking on the Internet. Both Lee [8] and Suziedelyte [9] analyzed data from 2008 and earlier, the periods before extensive diffusion of the Internet. In these periods, online health information seeking was still costly due to high Internet access cost, low connection speed, and limited availability of reliable health information, and hence it was limited to individuals expecting high benefits from the acquired information. An extensive diffusion of the Internet, in particular broadband services, has reduced the cost of information seeking substantially and promoted online health information seeking among those with low expected benefits from the available information.[16] These benefits emanate at least partly from better appreciation of health care services including physician visits. Thus, extending online health information seeking to those with low expected benefits weakens a positive or complementary relationship between online health information and physician visits. Our analysis also indicates that individuals with good health status seek health information more frequently and visit physicians less frequently than those with worse health status. This suggests a possibility that the reduction in online search cost has promoted online seeking of general health information by healthy individuals more than that of specific information about illnesses and treatments by unhealthy individuals. These changes weaken the link between online health information seeking and health care utilization as measured by the frequency of physician visits.

For the remaining variables, the estimated coefficients mostly confirm previous findings in the literature [3, 17, 18]. Women are found to visit physicians more frequently than men, while individuals with a long-term illness or bad health status visit physicians more frequently than those in a better health condition. Residing in a large town implies more visits, not surprisingly because its higher population density means closer proximity and hence better access to medical services than in small towns and rural areas. High physician density means better availability of medical services and also more competition among them, which can lower service costs and strengthen incentives for physician-induced demand. As for employment status, both the employed and unemployed visit physicians less frequently than those inactive in the labor market, which possibly reflects the higher opportunity (time) costs of the former groups than the latter. The coefficient for living alone is negative, suggesting that they visit physicians less frequently, possibly because they do not need to accompany other household members, in particular children. However, the coefficient is significant only for non-seekers of online health information. The coefficient for age is negative and significant only for online health information seekers. This counterintuitive result is common in the literature [3, 17, 18, 47] and possibly attributable to a high correlation between age and health status measures: elders tend to be in worse health condition, and after controlling for the effects of health condition, age itself has a weak,

[16] Suziedelyte [9] reports that only one third, 37%, and 40% of observations sought health information online, in 2003, 2005, and 2008, respectively. The corresponding number for our data is 51%.

Table 3 Maximum likelihood estimates of the bivariate ordered probit model

Variables	*HIi*	*PV*	VARIABLES	*HIi*	*PV*
*HIi**		−0.018	Sport (reference is "none")		
		(0.029)	*MONTH*	0.234***	0.096***
INTERNET	0.007***			(0.030)	(0.031)
	(0.002)		*1–3 MONTH*	0.242***	0.044
DENSITY		0.293***		(0.033)	(0.033)
		(0.101)	*1–2 WEEK*	0.317***	0.065**
AGE	−0.023***	−0.001		(0.025)	(0.027)
	(0.001)	(0.001)	*3–4 WEEK*	0.365***	0.090***
FEMALE	0.224***	0.195***		(0.026)	(0.025)
	(0.021)	(0.016)	*5 WEEK*	0.260***	0.038
ONEADULT	−0.193***	−0.021		(0.028)	(0.029)
	(0.021)	(0.017)			
			Education (reference is "no full-time education")		
Employment (reference is "inactive")			*EDU15*	0.140	
EMPLOYED	0.228***	−0.216***		(0.119)	
	(0.020)	(0.021)	*EDU19*	0.661***	
UNEMPLOY	0.106**	−0.137***		(0.118)	
	(0.044)	(0.042)	*EDU20*	1.003***	
Area (reference is "rural/village")				(0.118)	
SMALL	0.105***	0.025	*STUDY*	0.716***	
	(0.019)	(0.017)		(0.126)	
LARGE	0.153***	0.082***			
	(0.021)	(0.021)	Thresholds		
			Cut 1	0.776***	−1.516***
Health (reference is "very bad")				(0.196)	(0.091)
FBAD	0.182***	−0.247***	Cut 2	1.333***	−0.210**
	(0.064)	(0.052)		(0.194)	(0.091)
FGOOD	0.301***	−0.809***	Cut 3	1.480***	0.626***
	(0.061)	(0.055)		(0.194)	(0.091)
VGOOD	0.203***	−1.138***	Cut 4	1.933***	
	(0.060)	(0.056)		(0.194)	
			Cut 5	2.357***	
				(0.194)	
			ρ	0.115***	
				(0.029)	
			Observations	24,971	

Note: The table reports the results from estimating a bivariate ordered probit model for the frequencies of online health information seeking and physician visitation as defined in Eq. (5) by the method of maximum likelihood. The model was estimated with country dummies and their estimated coefficients are reported in Table 3A in the Appendix. Robust standard errors are in parentheses

***$p<0.01$, **$p<0.05$, *$p<0.1$

negative effect on physician visits. It might also reflect the fact that older people tend to utilize other types of medical services, such as health services at retirement houses, and hence visit physicians less often than younger people [3, 47]. Finally, the frequency of sport practice is positively related to physician visits, but the estimated coefficients are not significant for most categories. Two competing effects might be in play: while physical exercise improves health condition (hence reducing physician visits), it also increases the probability of an injury (thus increasing visits).

Columns 5–8 of Table 2 summarize the estimation results of the restricted model, in which the coefficients of socio-demographic and health status variables (**X**) and variables

Table 4 Two-stage residual inclusion estimates of the two-equation model with multinomial health information seeking (*HIS*)

Variables	*HIS*		*PV*		*HIS*		*PV*
	HIS = 1 (Online HI search)	*HIS* = 2 (Offline HI search)			*HIS* = 1 (Online HI search)	*HIS* = 2 (Offline HI search)	
Observed binary value				Sport (reference is "none")			
I(*HIS* = 1)			0.043	*MONTH*	0.485***	0.230***	0.083***
(Online HI search)			(0.105)		(0.055)	(0.064)	(0.032)
I(*HIS* = 2)			0.616**	*1–3 MONTH*	0.571***	0.340***	0.025
(Offline HI search)			(0.312)		(0.065)	(0.074)	(0.037)
				1–2 WEEK	0.621***	0.289***	0.052*
INTERNET	0.010***	−0.006			(0.041)	(0.055)	(0.029)
	(0.003)	(0.004)		*3–4 WEEK*	0.695***	0.408***	0.069**
DENSITY			0.343***		(0.047)	(0.053)	(0.029)
			(0.106)	*5 WEEK*	0.400***	0.315***	0.014
AGE	−0.040***	0.003***	−0.002*		(0.044)	(0.042)	(0.031)
	(0.001)	(0.001)	(0.001)				
FEMALE	0.429***	0.439***	0.170***	Education (reference is "no full-time education")			
	(0.036)	(0.036)	(0.022)	*EDU15*	0.249	0.262	
ONEADULT	−0.336***	−0.101***	−0.023		(0.181)	(0.173)	
	(0.035)	(0.034)	(0.018)	*EDU19*	1.000***	0.428**	
					(0.180)	(0.177)	
Employment (reference is "inactive")				*EDU20*	1.645***	0.591***	
EMPLOYED	0.309***	−0.067*	−0.191***		(0.179)	(0.180)	
	(0.034)	(0.038)	(0.020)	*STUDY*	1.092***	0.653***	
UNEMPLOY	0.130*	0.035	−0.124***		(0.200)	(0.201)	
	(0.078)	(0.077)	(0.042)				
				Constant	−0.969***	−1.623***	
Area (reference is "rural/village")					(0.290)	(0.362)	
SMALL	0.201***	0.013	0.026				
	(0.037)	(0.038)	(0.017)	Residuals			
LARGE	0.296***	0.008	0.089***	I(*HIS* = 1)			0.256**
	(0.040)	(0.045)	(0.022)	(Online HI search)			(0.102)
				I(*HIS* = 2)			−0.356
Health (reference is "very bad")				(Offline HI search)			(0.310)
FBAD	0.347***	0.286***	-0.261***				
	(0.092)	(0.096)	(0.055)	Thresholds			
FGOOD	0.580***	0.269***	−0.817***	C_1			1.476***
	(0.087)	(0.095)	(0.057)				(0.120)
VGOOD	0.339***	0.052	−1.138***	C_2			−0.156
	(0.085)	(0.105)	(0.057)				(0.118)
				C_3			0.687***
LONGILL	0.517***	0.478***	0.794***				(0.120)
	(0.030)	(0.037)	(0.024)				
				Observations	24,624	24,624	24,522

Note: The table reports the results from estimating a bivariate model with a multinomial measure of health information seeking as defined in Eq. (6) by the method of two stage residual inclusion. The model was estimated with country dummies and their estimated coefficients are reported in Table 4A in the Appendix

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

specific to the physician visitation equation ($\mathbf{Z_3}$) are constrained to be identical in the two physician visitation equations. This restriction does not qualitatively alter the estimates. Most importantly, the estimated coefficient remains positive and significant for offline health information search and not significant for online search.

Robustness based on alternative model specifications

Tables 3 and 4 present the results of the two alternative models; the bivariate ordered probit model specified in Eq. (5), which evaluates the association of online health information seeking with physician visitation without addressing that of offline seeking; and the two-equation model specified in (6), which utilizes a multinomial measure of health information seeking. For both models, the estimation results are generally consistent with those obtained for the extended sample selection model. Most importantly, the coefficient is significantly positive for offline health information search, while it is not significant for online search in the estimated physician visitation equations. Accordingly, results are robust against changes in model specification, estimation method, and construction of the variables for health information seeking. The coefficient estimates for other control variables also confirm the results of the extended sample selection model, and are consistent with previous findings in the literature.

Conclusions

The present analysis has examined the link between health information and the demand for physician services. Specifically, the analysis has distinguished individuals seeking health information exclusively on offline sources from those seeking online and possibly offline too. Using an extended sample selection model that addresses both the sample selection issue created by the survey design and the endogeneity of the health information variables, our analysis has elucidated the determinants of online and offline health information seeking and how they relate to the demand for physician services.

The empirical analysis for the 28 European Union Member countries has revealed that sociodemographic factors shape online and offline health information seeking in similar ways. Specifically, those who are female, younger, better educated, in the labor market, living in urban areas, and with a long-term illness, are more likely to seek health information both online and offline. On the contrary, individuals' self-assessed health status is positively related to health information seeking. These results support the media complementarity hypothesis and suggest that enhancing the availability of online health information potentially exacerbates, rather than alleviates, the unequal distribution of health information traditionally observed in the offline environment.

Concerning the demand for health care, the extended sample selection model shows different results for online and offline health information seeking. Offline seeking has a significantly positive *net* association with the frequency of physician visits, that is, offline health information seekers visit physicians more frequently than non-seekers. Several explanations are possible for this complementary relationship of health information and physician services. First, seeking health information offline before or after visiting physicians could enhance the benefits of physician services, which in turn could lead into an increased demand for physician services. Second, additional information could increase individuals' anxiety about their health condition and hence make them resort more to support and advice from physicians. In relation to the long-standing hypotheses in the literature, the positive link is in line with the patient induced-supply, while it disagrees with the physician induced-demand hypothesis.

For online health information, its *gross* association with the physician visitation is not significant. Our result contrasts with previous findings by Lee [8] and Suziedelyte [9] who report significantly positive impact of online health information on health care demand. A plausible explanation for the difference in the reported results is sought in an extensive diffusion of the Internet, which substantially reduced the cost of online information seeking. This cost reduction might have extended online health information seeking to individuals with low expected benefits from the acquired information, and hence attenuated the link between online health information seeking and the demand for physician services. Our analysis has also revealed that healthy individuals seek health information more often and visit physicians less often than unhealthy individuals. Reduction in the cost of online information seeking potentially promoted health information among healthy individuals, which weakens the average link between online health information and the demand for physician services.

Some directions for further research can be derived from our analysis. First, the supporting evidence for the media complementarity hypothesis raises a concern that the

Internet could reinforce the traditional unequal access to health information. In particular, the Internet possibly helps individuals in a better health condition to acquire more information and consequently make better health care investment decisions than those in a worse condition. Types of intended health information would also vary by health condition, with individuals on a good health status seeking information on general health and those on a bad health status seeking information specific to some illnesses and treatments. Hence, strategies need to be designed to assist individuals secluded from health information to access appropriate health information online. To help design these strategies, future research should further unveil the complex association among individuals' health status, their access to health information and types of desired health information, and how these factors affect their health care investment decisions.

Second, our analysis suggests that further reduction in the cost of health information acquisition (e.g., via improved availability of reliable health information) would extend online health information seeking into wider groups within populations. However, health information seeking is little linked to health care demand for existing online seekers and this link would be even weaker for the new online seekers. Thus, further promotion of online health information would be associated with little or no change in the cost of the health care system in the short term, unlike the implications of previous studies reporting significantly positive effects of online health information. Furthermore, online health information can improve general health knowledge of the broader population and their ability to self-manage health issues, which could improve their health condition and hence relieve the cost pressure on the health care system in the long term. Future research should explore these relationships and quantify their relative magnitudes.

Finally, while our analysis has found no significant association between online health information seeking and demand for physician services, it has considered only one type of health service. It is possible that health information affects differently other types of health care services, such as those specific to mental health, and other aspects such as duration and time efficiency of physician visits [48, 49]. Of particular importance is whether health information helps patients to choose appropriate types and levels of health care. For instance, increased physician visits by better informed patients can mean more efficient use of medical service if they reduce unnecessary use of emergency services. The literature on the effect of health information on the efficient use of health care service is premature and scarce empirical evidence reports mixed results.[17] Future research should extend the scope of analysis to potentially heterogeneous impacts of online (and offline) health information on broader types of health care services and examine if it improves patient's choice of appropriate health care services.

Appendix

Likelihood function of the extended sample selection model

The contribution to the likelihood function is, for individual i who seeks health information on the Internet (i.e., $HIi_i = j_1$ $(1 < j_1 \le J_1)$ and $DV_i = j_3$ $(1 \le j_3 \le J_3)$),

$$\begin{aligned} li_i(j_1, j_3) = {} & \Phi_2\left(HIi_i^* < c_{1,j_1}, DV_i^* < c_{3,j_3}\right) \\ & - \Phi_2\left(HIi_i^* < c_{1,j_1-1}, DV_i^* < c_{3,j_3}\right) \\ & - \Phi_2\left(HIi_i^* < c_{1,j_1}, DV_i^* < c_{3,j_3-1}\right) \\ & + \Phi_2\left(HIi_i^* < c_{1,j_1+1}, DV_i^* < c_{3,j_3+1}\right) \end{aligned}$$

and for individual i who does no seek health information on the Internet (i.e., $HIi_i = 1$, $HIni_i = j_2$ $(1 \le j_2 \le J_2)$, and $DV_i = j_4$ $(1 \le j_4 \le J_4)$),

$$\begin{aligned} lni_i(1, j_2, j_4) = {} & \Phi_3\left(HIi_i^* < c_{1,1}, HIni_i^* < c_{2,j_2}, DV_i^* < c_{4,j_4}\right) \\ & - \Phi_3\left(HIi_i^* < c_{1,1}, HIni_i^* < c_{2,j_2-1}, DV_i^* < c_{4,j_4}\right) \\ & - \Phi_3\left(HIi_i^* < c_{1,1}, HIni_i^* < c_{2,j_2}, DV_i^* < c_{4,j_4-1}\right) \\ & + \Phi_3\left(HIi_i^* < c_{1,1}, HIni_i^* < c_{2,j_2+1}, DV_i^* < c_{4,j_4+1}\right), \end{aligned}$$

where Φ_k is the k-variate standard normal cumulative distribution function. The maximum likelihood estimator of the coefficient vector $\Theta = \{\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_{41}, \beta_{42}, \beta_{43}, c_{1,1}, ..., c_{1,J_1-1}, c_{2,1}, ..., c_{2,J_2-1}, c_{3,1}, ..., c_{3,J_3-1}, c_{4,1}, ..., c_{4,J_4-1}, \rho_{12}, \rho_{13}, \rho_{24}, \rho_{14}\}$ is expressed as follows:

$$\hat{\Theta}_{MLE} = \arg\max_{\Theta} \left(\sum_i \left\{ \sum_{j_1 \neq 1, j_3} li_i(j_1, j_3) + i(HIi_i = 1) \sum_{j_2, j_4} \ln i_i(1, j_2, j_4) \right\} \right).$$

[17] For example, Johnson and Rehavi [50] report that physician mothers were less prone to caesarean sections than the non-physician. Frakes et al. [51] find that physicians made only slightly better medical care choices than non-physician patients.

Table 2A Maximum likelihood estimates of country fixed effects in the extended sample selection model

Variables	(i) Unconstrained				(ii) Constrained			
	HIi	*HIni*	*PV* for *HIi*	*PV* for *HIni*	*HIi*	*HIni*	*PV* for *HIi*	*PV* for *HIni*
France	−0.052	−0.029	0.255***	0.336***	−0.050	−0.018	0.279***	0.279***
	(0.041)	(0.042)	(0.048)	(0.077)	(0.042)	(0.043)	(0.052)	(0.052)
Belgium	−0.187***	−0.211***	0.306***	0.506***	−0.186***	−0.195***	0.406***	0.406***
	(0.049)	(0.038)	(0.063)	(0.068)	(0.049)	(0.037)	(0.058)	(0.058)
The Netherlands	0.112*	−0.077	−0.097*	−0.109	0.117**	−0.069	−0.121*	−0.121*
	(0.058)	(0.053)	(0.052)	(0.092)	(0.059)	(0.049)	(0.065)	(0.065)
Germany	−0.198***	−0.107***	0.276***	0.241***	−0.193***	−0.099***	0.267***	0.267***
	(0.048)	(0.035)	(0.050)	(0.084)	(0.049)	(0.034)	(0.057)	(0.057)
Italy	0.072*	0.119***	0.233***	0.275***	0.074*	0.129***	0.245***	0.245***
	(0.040)	(0.045)	(0.044)	(0.089)	(0.040)	(0.046)	(0.054)	(0.054)
Luxembourg	−0.136***	0.044**	0.362***	0.398***	−0.131***	0.055***	0.368***	0.368***
	(0.049)	(0.018)	(0.030)	(0.061)	(0.051)	(0.018)	(0.045)	(0.045)
Denmark	0.045	−0.332***	−0.054	−0.020	0.051	−0.314***	−0.056	−0.056
	(0.053)	(0.048)	(0.078)	(0.083)	(0.055)	(0.042)	(0.060)	(0.060)
Ireland	−0.067	−0.074*	0.057	0.232***	−0.067	−0.063	0.136**	0.136**
	(0.046)	(0.044)	(0.059)	(0.068)	(0.047)	(0.043)	(0.060)	(0.060)
United Kingdom	−0.169***	−0.256***	0.050	0.036	−0.164***	−0.249***	0.055	0.055
	(0.054)	(0.051)	(0.039)	(0.072)	(0.056)	(0.052)	(0.046)	(0.046)
Greece	−0.208***	−0.330***	−0.255***	−0.343***	−0.204***	−0.325***	−0.280***	−0.280***
	(0.038)	(0.074)	(0.079)	(0.076)	(0.037)	(0.076)	(0.058)	(0.058)
Spain	−0.234***	−0.411***	0.145**	0.178**	−0.231***	−0.399***	0.177***	0.177***
	(0.046)	(0.067)	(0.069)	(0.078)	(0.045)	(0.066)	(0.064)	(0.064)
Portugal	−0.203***	−0.023	0.063	0.027	−0.199***	−0.013	0.059	0.059
	(0.053)	(0.036)	(0.046)	(0.078)	(0.054)	(0.033)	(0.060)	(0.060)
Finland	−0.148***	−0.053	−0.260***	−0.690***	−0.135***	−0.054	−0.443***	−0.443***
	(0.047)	(0.037)	(0.036)	(0.092)	(0.050)	(0.036)	(0.059)	(0.059)
Sweden	−0.089	−0.281***	−0.342***	−0.417***	−0.083	−0.276***	−0.381***	−0.381***
	(0.060)	(0.062)	(0.080)	(0.090)	(0.062)	(0.061)	(0.078)	(0.078)
Austria	−0.141***	0.041	0.338***	0.309***	−0.136***	0.048	0.329***	0.329***
	(0.045)	(0.040)	(0.057)	(0.085)	(0.045)	(0.040)	(0.062)	(0.062)
Cyprus (Republic)	−0.273***	−0.265***	−0.017	0.257***	−0.274***	−0.251***	0.153***	0.153***
	(0.021)	(0.015)	(0.039)	(0.071)	(0.021)	(0.016)	(0.049)	(0.049)
Czech Republic	−0.197***	−0.102*	0.209***	0.064	−0.190***	−0.096*	0.139**	0.139**
	(0.072)	(0.055)	(0.054)	(0.076)	(0.073)	(0.057)	(0.056)	(0.056)
Estonia	−0.322***	−0.267***	−0.229***	−0.423***	−0.314***	−0.264***	−0.313***	−0.313***
	(0.032)	(0.017)	(0.032)	(0.060)	(0.033)	(0.018)	(0.044)	(0.044)
Hungary	0.052	0.195***	0.034	0.170**	0.053	0.205***	0.088	0.088
	(0.043)	(0.039)	(0.052)	(0.071)	(0.042)	(0.039)	(0.055)	(0.055)
Latvia	−0.361***	−0.285***	0.085**	−0.171***	−0.353***	−0.282***	−0.023	−0.023
	(0.024)	(0.021)	(0.035)	(0.063)	(0.025)	(0.020)	(0.045)	(0.045)
Lithuania	−0.329***	−0.106***	−0.025	−0.129**	−0.324***	−0.098***	−0.070	−0.070
	(0.021)	(0.027)	(0.030)	(0.059)	(0.022)	(0.024)	(0.043)	(0.043)
Malta	−0.216***	−0.002	0.275***	0.130**	−0.208***	0.005	0.201***	0.201***
	(0.030)	(0.019)	(0.030)	(0.061)	(0.031)	(0.018)	(0.044)	(0.044)
Poland	−0.218***	−0.181***	0.292***	0.195**	−0.214***	−0.174***	0.247***	0.247***
	(0.050)	(0.037)	(0.036)	(0.083)	(0.050)	(0.038)	(0.054)	(0.054)
Slovakia	0.022	0.122**	0.086	0.089	0.026	0.131***	0.077	0.077
	(0.056)	(0.049)	(0.067)	(0.069)	(0.056)	(0.050)	(0.064)	(0.064)

Table 2A (continued)

Variables	(i) Unconstrained				(ii) Constrained			
	HIi	*HIni*	*PV* for *HIi*	*PV* for *HIni*	*HIi*	*HIni*	*PV* for *HIi*	*PV* for *HIni*
Slovenia	−0.261***	−0.031	−0.411***	−0.206***	−0.261***	−0.022	−0.275***	−0.275***
	(0.043)	(0.045)	(0.074)	(0.062)	(0.042)	(0.045)	(0.051)	(0.051)
Bulgaria	−0.108**	−0.067	0.115**	−0.015	−0.105**	−0.061	0.053	0.053
	(0.047)	(0.054)	(0.058)	(0.071)	(0.047)	(0.052)	(0.051)	(0.051)
Romania	−0.104***	0.044	−0.009	−0.008	−0.101***	0.055	−0.017	−0.017
	(0.039)	(0.043)	(0.047)	(0.080)	(0.038)	(0.042)	(0.056)	(0.056)

Note: The table reports the estimated coefficients for the country-specific dummy variables in the extended sample selection model specified in Eqs. (1)-(4). See notes below Table 2

Table 3A Maximum likelihood estimates of country fixed effects in the bivariate ordered probit model

Country	*HIi*	*PV*	Country	*HIi*	*PV*
France	−0.075*	0.288***			
	(0.043)	(0.050)			
Belgium	−0.198***	0.380***	Austria	−0.147***	0.340***
	(0.050)	(0.056)		(0.045)	(0.057)
The Netherlands	0.086	−0.110*	Cyprus (Republic)	−0.267***	0.128***
	(0.061)	(0.060)		(0.017)	(0.047)
Germany	−0.214***	0.259***	Czech Republic	−0.218***	0.132**
	(0.050)	(0.055)		(0.074)	(0.053)
Italy	0.065*	0.255***	Estonia	−0.360***	−0.315***
	(0.040)	(0.053)		(0.032)	(0.041)
Luxembourg	−0.177***	0.377***	Hungary	0.051	0.105**
	(0.053)	(0.042)		(0.046)	(0.053)
Denmark	0.019	−0.074	Latvia	−0.373***	−0.048
	(0.058)	(0.059)		(0.020)	(0.043)
Ireland	−0.073	0.132**	Lithuania	−0.322***	−0.084**
	(0.045)	(0.061)		(0.018)	(0.040)
United Kingdom	−0.194***	0.022	Malta	−0.208***	0.196***
	(0.058)	(0.045)		(0.030)	(0.041)
Greece	−0.209***	−0.297***	Poland	−0.226***	0.236***
	(0.039)	(0.051)		(0.050)	(0.051)
Spain	−0.237***	0.136**	Slovakia	0.019	0.092
	(0.044)	(0.061)		(0.054)	(0.058)
Portugal	−0.199***	0.057	Slovenia	−0.258***	−0.262***
	(0.059)	(0.057)		(0.040)	(0.047)
Finland	−0.196***	−0.421***	Bulgaria	−0.103**	0.061
	(0.049)	(0.058)		(0.050)	(0.049)
Sweden	−0.113*	−0.388***	Romania	−0.082**	−0.013
	(0.062)	(0.076)		(0.036)	(0.053)

Note: The table reports the coefficient estimates for the country-specific dummy variables in the bivariate ordered probit model specified in Eq. (5). See notes below Table 3

Table 4A Two-stage residual inclusion estimates of country fixed effects in the two-equation model with multinomial health information seeking (*HIS*)

Country	*HIS*		*PV*	Country	*HIS*		*PV*
	HIS = 1 (Online HI search)	*HIS* = 2 (Offline HI search)			*HIS* = 1 (Online HI search)	*HIS* = 2 (Offline HI search)	
France	0.243**	0.018	0.300***				
	(0.095)	(0.098)	(0.055)				
Belgium	−0.121	−0.178*	0.413***	Austria	0.007	0.195*	0.318***
	(0.088)	(0.101)	(0.061)		(0.077)	(0.115)	(0.061)
The Netherlands	0.420***	−0.245*	−0.069	Cyprus (Republic)	−0.496***	−0.322***	0.151***
	(0.120)	(0.139)	(0.066)		(0.035)	(0.035)	(0.053)
Germany	−0.158*	−0.027	0.266***	Czech Republic	−0.283**	0.009	0.124**
	(0.094)	(0.109)	(0.059)		(0.112)	(0.069)	(0.056)
Italy	0.319***	0.181**	0.249***	Estonia	−0.341***	−0.172**	−0.306***
	(0.075)	(0.086)	(0.056)		(0.059)	(0.068)	(0.047)
Luxembourg	−0.092	0.248**	0.380***	Hungary	0.097	0.170***	0.092
	(0.096)	(0.120)	(0.046)		(0.077)	(0.062)	(0.057)
Denmark	0.085	−0.645***	−0.011	Latvia	−0.388***	−0.096**	−0.036
	(0.095)	(0.119)	(0.066)		(0.039)	(0.041)	(0.047)
Ireland	0.308***	0.059	0.149**	Lithuania	−0.410***	−0.069*	−0.082*
	(0.107)	(0.105)	(0.063)		(0.036)	(0.038)	(0.044)
United Kingdom	0.026	−0.189*	0.064	Malta	−0.382***	−0.097	0.200***
	(0.107)	(0.114)	(0.052)		(0.057)	(0.062)	(0.046)
Greece	−0.298***	−0.358***	−0.291***	Poland	−0.278***	−0.152*	0.250***
	(0.084)	(0.124)	(0.061)		(0.066)	(0.088)	(0.056)
Spain	−0.305***	−0.551***	0.171**	Slovakia	−0.080	0.037	0.083
	(0.082)	(0.105)	(0.069)		(0.109)	(0.071)	(0.066)
Portugal	−0.196**	0.215***	0.014	Slovenia	−0.386***	0.079	−0.288***
	(0.083)	(0.077)	(0.063)		(0.058)	(0.095)	(0.051)
Finland	0.159*	0.229**	−0.432***	Bulgaria	−0.110	0.047	0.025
	(0.092)	(0.097)	(0.061)		(0.105)	(0.102)	(0.054)
Sweden	0.174	−0.343***	−0.345***	Romania	−0.261***	0.043	−0.035
	(0.130)	(0.124)	(0.082)		(0.068)	(0.091)	(0.057)

Note: The table reports the coefficient estimates for the country-specific dummy variables in a bivariate model with a multinomial measure of health information seeking as specified in Eq. (6). See notes below Table 4

Acknowledgements We would like to express our gratitude to Gary Madden for his guidance while working on earlier version of this paper. We also thank Jeffrey Wooldridge for his suggestions on the empirical approach, and Mark Harris and seminar participants at the Bankwest Curtin Economics Centre and Centre for Research in Applied Economics, Curtin University for helpful comments.

Funding The analysis was conducted while the authors were employed at public universities. No financial support was received from a third party.

Availability of data The study analyzed publicly available data from GESIS, Leibniz Institute for the Social Sciences. Individuals' identities are anonymized in this dataset. No other ethical considerations apply.

Code availability All the econometric models presented in this manuscript were estimated using Stata Ver. 16. Estimation codes are available from the authors upon request.

Declarations

Conflict of interest The authors have no conflict of interests to declare, financial or otherwise.

References

1. Arrow, K.J.: Uncertainty and the welfare economics of medical care. Am. Econ. Rev. **53**(5), 941–973 (1963)
2. Hsieh, C.-R., Lin, S.-J.: Health information and the demand for preventive care among the elderly in Taiwan. J. Hum. Resour. **32**(2), 308–333 (1997). https://doi.org/10.2307/146217
3. Schmid, C.: Consumer health information and the demand for physician visits. Health Econ. **24**(12), 1619–1631 (2015). https://doi.org/10.1002/hec.3117

4. Dranove, D.: Demand inducement and the physician/patient relationship. Econ. Inq. **26**(2), 281–298 (1988). https://doi.org/10.1111/j.1465-7295.1988.tb01494.x
5. Evans, R.G.: Supplier-induced demand: Some empirical evidence and implications. In: Perlman, M. (ed.) The Economics of Health and Medical Care, pp. 162–173. Palgrave Macmillan, London (1974)
6. Bundorf, M.K., Wagner, T.H., Singer, S.J., Baker, L.C.: Who searches the Internet for health information? Health Serv. Res. **41**(3), 819–836 (2006). https://doi.org/10.1111/j.1475-6773.2006.00510.x
7. Beck, F., Richard, J.-B., Nguyen-Thanh, V., Montagni, I., Parizot, I., Renahy, E.: Use of the Internet as a health information resource among French young adults: results from a nationally representative survey. J. Med. Internet Res. **16**(5), e128 (2014). https://doi.org/10.2196/jmir.2934
8. Lee, C.-J.: Does the Internet displace health professionals? J. Health Commun. **13**(5), 450–464 (2008). https://doi.org/10.1080/10810730802198839
9. Suziedelyte, A.: How does searching for health information on the Internet affect individuals' demand for health care services? Soc. Sci. Med. **75**(10), 1828–1835 (2012). https://doi.org/10.1016/j.socscimed.2012.07.022
10. Dutta-Bergman, M.J.: Complementarity in consumption of news types across traditional and new media. J. Broadcast. Electron. Media **48**(1), 41–60 (2004). https://doi.org/10.1207/s15506878jobem4801_3
11. Cline, R.J.W., Haynes, K.M.: Consumer health information seeking on the Internet: the state of the art. Health Educ. Res. **16**(6), 671–692 (2001). https://doi.org/10.1093/her/16.6.671
12. World Health Organization. The Health Internet. Retrieved from http://www.who.int/ehealth/programmes/governance/en/ (2018)
13. Costa-Font, J., Mossialos, E., Rudisill, C.: When is the Internet a valued communication device for health information in Europe? Econ. Innov. New Technol. **18**(5), 429–445 (2009). https://doi.org/10.1080/10438590802547159
14. Xie, B., Dilts, D.M., Shor, M.: The physician-patient relationship: the impact of patient-obtained medical information. Health Econ. **15**(8), 813–833 (2006). https://doi.org/10.1002/hec.1098
15. Shih, Y.-C.T., Tai-Seale, M.: Physicians' perception of demand-induced supply in the information age: a latent class model analysis. Health Econ. **21**(3), 252–269 (2012). https://doi.org/10.1002/hec.1710
16. Haas-Wilson, D.: Arrow and the information market failure in health care: the changing content and sources of health care information. J. Health Polit. Policy Law **26**(5), 1031–1044 (2001). https://doi.org/10.1215/03616878-26-5-1031
17. Kenkel, D.: Consumer health information and the demand for medical care. Rev. Econ. Stat. **72**(4), 587–595 (1990). https://doi.org/10.2307/2109598
18. Dwyer, D.S., Liu, H.: The impact of consumer health information on the demand for health services. Q. Rev. Econ. Finance **53**(1), 1–11 (2013). https://doi.org/10.1016/j.qref.2012.11.002
19. Cleary, P.D., Jette, A.M.: The validity of self-reported physician utilization measures. Med. Care **22**(9), 796–803 (1984). https://doi.org/10.1097/00005650-198409000-00003
20. Roberts, R.O., Bergstralh, E.J., Schmidt, L., Jacobsen, S.J.: Comparison of self-reported and medical record health care utilization measures. J. Clin. Epidemiol. **49**(9), 989–995 (1996). https://doi.org/10.1016/0895-4356(96)00143-6
21. Tian, Y., Robinson, J.D.: Incidental health information use and media complementarity: a comparison of senior and non-senior cancer patients. Patient Educ. Couns. **71**(3), 340–344 (2008). https://doi.org/10.1016/j.pec.2008.02.006
22. Tian, Y., Robinson, J.D.: Media use and health information seeking: An empirical test of complementarity theory. Health Commun. **23**(2), 184–190 (2008). https://doi.org/10.1080/10410230801968260
23. Rains, S.A., Ruppel, E.K.: Channel complementarity theory and the health information-seeking process. Commun. Res. **43**(2), 232–252 (2016). https://doi.org/10.1177/0093650213510939
24. Ruppel, E.K., Rains, S.A.: Information sources and the health information-seeking process: an application and extension of channel complementarity theory. Commun. Monogr. **79**(3), 385–405 (2012). https://doi.org/10.1080/03637751.2012.697627
25. Yang, Q., Chen, Y., WendorfMuhamad, J.: Social support, trust in health information, and health information-seeking behaviors (HISBs): a study using the 2012 Annenberg National Health Communication Survey (ANHCS). Health Commun. **32**(9), 1142–1150 (2017). https://doi.org/10.1080/10410236.2016.1214220
26. Jacobs, W., Amuta, A.O., Jeon, K.C.: Health information seeking in the digital age: analysis of health information seeking behavior among US adults. Cogent Soc. Sci. **3**(1) (2017). https://doi.org/10.1080/23311886.2017.1302785
27 Medlock, S., Eslami, S., Askari, M., Arts, D.L., Sent, D., de Rooij, S.E., Abu-Hanna, A.: Health information-seeking behaviour of seniors who use the Internet: a survey. J. Med. Internet Res. **17**(1), e10 (2015). https://doi.org/10.2196/jmir.3749
28. European Commission. Flash Eurobarometer 404 (European Citizens´ Digital Health Literacy). GESIS Data Archive, Cologne (2015). https://doi.org/10.4232/1.12194
29 Eysenbach, G.: Consumer health informatics. Br Med J **320**(7251), 1713–1716 (2000). https://doi.org/10.1136/bmj.320.7251.1713
30. Rowlands, G., Nutbeam, D.: Health literacy and the "inverse information law." Br. J. Gen. Pract. **63**(608), 120–121 (2013). https://doi.org/10.3399/bjgp13X664081
31. Hone, T., Palladino, R., Filippidis, F.T.: Association of searching for health-related information online with self-rated health in the European Union. Eur. J. Pub. Health **26**(5), 748–753 (2016). https://doi.org/10.1093/eurpub/ckw022
32. Vicente, M.R., Madden, G.: Assessing eHealth skills across Europeans. Health Policy Technol. **6**(2), 161–168 (2017). https://doi.org/10.1016/j.hlpt.2017.04.001
33. Amante, D.J., Hogar, T.P., Pagoto, S.L., English, T.M., Lapane, K.L.: Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J. Med. Internet Res. **17**(4), e106 (2015). https://doi.org/10.2196/jmir.426
34. Haluza, D., Naszay, M., Stockinger, A., Jungwirth, D.: Digital natives versus digital immigrants: influence of online health information seeking on the doctor-patient relationship. Health Commun. **32**(11), 1342–1349 (2017). https://doi.org/10.1080/10410236.2016.1220044
35. Koch-Weser, S., Bradshaw, Y.S., Gualtieri, L., Gallagher, S.S.: The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J. Health Commun. **15**(sup3), 279–293 (2010)
36. Pauly, M.V., Satterthwaite, M.A.: The pricing of primary care physicians services: a test of the role of consumer information. Bell J. Econ. **12**(2), 488–506 (1981)
37. Eurostat. Health Care Resources Statistics. Eurostat, Luxembourg (2017)
38. OECD: Health at a Glance: Europe 2016. OECD, Paris (2016)
39. Kenkel, D.: Health behaviour, health knowledge, and schooling. J. Polit. Econ. **99**(2), 287–305 (1991)
40. Schultz, T.W.: The value of the ability to deal with disequilibria. J. Econ. Lit. **13**(3), 827–846 (1975)
41. Eurostat. Regional ICT Statistics. Eurostat, Luxembourg (2017)
42. Heckman, J.J.: Sample selection bias as a specification error. Econometrica **47**(1), 153–161 (1979). https://doi.org/10.2307/1912352

43. Greene, W.H., Hensher, D.A.: Modeling Ordered Choices: A Primer. Cambridge University Press, New York (2010)
44. Terza, J.V., Basu, A., Rathouz, P.J.: Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. J. Health Econ. **27**(3), 531–543 (2008). https://doi.org/10.1016/j.jhealeco.2007.09.009
45 Roodman, D.: Fitting fully observed recursive mixed-process models with cmp. Stata J. **11**(2), 159–206 (2011)
46. Sindelar, J.L.: Differential use of medical care by sex. J. Polit. Econ. **90**(5), 1003–1019 (1982)
47. Hay, J., Leahy, M.J.: Physician-induced demand: an empirical analysis of the consumer information gap. J. Health Econ. **1**(3), 231–244 (1982)
48. Fang, H., Miller, N.H., Rizzo, J., Zeckhauser, R.: Demanding customers: Consumerist patients and quality of care. The B.E. J. Econ. Anal. Policy **11**(1) (2011). https://doi.org/10.2202/1935-1682.2966
49. Murray, E., Lo, B., Pollack, L., Donelan, K., Catania, J., Lee, K., Zapert, K., Turner, R.: The impact of health information on the Internet on health care and the physician-patient relationship: national US Survey among 1050 US physicians. J. Med. Internet Res. **5**(3), e17 (2003). https://doi.org/10.2196/jmir.5.3.e17
50. Johnson, E.M., Rehavi, M.M.: Physicians treating physicians: Information and incentives in childbirth. Am. Econ. J. Econ. Pol. **8**(1), 115–141 (2016). https://doi.org/10.1257/pol.20140160
51. Frakes, M.D., Gruber, J., Jena, A.: Is great information good enough? Evidence from physicians as patients. J. Health Econ. **75**, 102406 (2021). https://doi.org/10.1016/j.jhealeco.2020.102406