Online health information seeking behaviour, healthcare access and health status during exceptional times

Cinzia Di Novi, Matija Kovacic, Cristina Elisa Orso*

**European Commission, Joint Research Centre (JRC), Via Enrico Fermi 2749, TP 267, 21027 Ispra (VA), Italy.*

Email: Cinzia.DI-NOVI@ec.europa.eu

Introduction

In the last decades, with the propagation of smartphones, tablets and laptops, the share of individuals who are turning to the Internet to obtain information about health has raised in the vast majority of the European countries.

According to the Eurostat, in 2022, about 52 % of individuals have searched for health-related information and symptoms online and this proportion is still growing (Eurostat, 2022).

Despite the Internet might be an important source of health-related information, with negligible monetary and opportunity (time) costs concerns about the quality of information available on the Internet and the ability of individuals to understand its content are raising questions about the implications of expansions in its use.

Aims

Despite e-HISB has become a global trend, only a few empirical investigations on how it may affect healthcare access and individuals' health exist.

In this paper, we tried to fill this gap and investigated whether access to health information on the Internet is likely to affect individuals' health and healthcare related decisions in exceptional time such as the COVID-19 pandemic which has had an enormous impact on people worldwide, subjecting the global population to health risks, fear, anxiety and to an incredible amount of health information.

To assess both the potential merits and shortcomings of seeking health information online and how doing so may affect individuals' healthcare access and health status, we used data collected in the second SHARE Corona Survey supplemented with data from the previous 8th wave of SHARE.

Data

The first SHARE Corona Survey was implemented as a quick response within the SHARE study to understand the effects of the COVID-19 pandemic. The interviews took place between June and September 2020 partly to collect a set of basic information as in the regular SHARE questionnaire, and partly to elicit information on life circumstances in the presence of COVID-19.

Respondents who participated in the first SHARE Corona Survey were interviewed again and participated in the second SHARE Corona Survey from June to August 2021 which contains questions on their use of the Internet, including its use for information about matters pertaining to health.

Besides the second SHARE Corona Survey dataset, we use data from the regular 8th wave of SHARE which collects information on the health, demographic and socio-economic status of respondents aged 50 years and over. The interviews took place between October 2019 and March 2020.

Sample

Once conditioning on having no missing values on any dependent variable and/or covariate, the final sample consisted of 13.829 observations across 18 European countries, namely:

- ✓ Austria,
- ✓ Belgium,
- ✓ Bulgaria,
- ✓ Croatia,
- ✓ Cyprus,
- ✓ Czech Republic,
- ✓ Denmark,
- ✓ Estonia,
- ✓ France,
- ✓ Italy,
- ✓ Latvia,
- ✓ Lithuania,
- ✓ Netherlands,
- ✓ Romania,
- ✓ Slovakia,
- ✓ Slovenia,
- ✓ Spain, and
- ✓ Sweden.

Sample

Since our **identification strategy** is based on information collected from the

- **Eurostat survey on the use of Information and Communication Technologies (ICT) in households and by individuals** and
- **the Eurostat data on the density of physicians at the Nomenclature of territorial units for statistics (NUTS) 2 level**

the sample was restricted excluding respondents living in European countries from for which sub-national geographies were not available

Older Adults

In this study, we specifically focused on adults aged 50 and over.

- Although older adults show lower rates of Internet adoption as compared to younger adults, online health information seeking is becoming increasingly common among them and this trend accelerated as a result of the COVID-19 pandemic (Lee & Jang, 2022; Symeonaki et al., 2022).

- Health deteriorates with age, so older adults may be more motivated in seeking health related information to cope with their uncertainty.

Hypothesis and Previous Literature

According to Lee (2008) there are two the potential and contrasting hypothesis regarding the effects of e-HISB on physician visits:

(i) online information seeking, by responding to patients' needs of health information, may negatively affect the likelihood of visiting health professionals and the frequency of visits;

(ii) on the other hand, e-HISB, by making individuals more aware about their health conditions, might increase their health concerns and consequently their demand for physician visits and other medical services.

Suziedelyte (2012) and Suenaga and Vicente (2021) examined the relationship between e-HISB and the demand for physician services going in these directions.

Hypothesis and Previous Literature

However, they failed to take into account that e-HISB, individuals' health and healthcare access might be simultaneously determined. For instance:

i. individuals who need to deal with health conditions, such as a chronic disease, tend to be much more motivated in engaging in e-HISBs(see for instance Ayers et al, 2007; Weaver III et al, 2010);

ii. e-HISB may affect (negatively or positively) the demand for healthcare services that in turn may influence individuals' health status.

iii. Moreover, e-HISB is likely to be correlated to other variables that can also affect individuals' demand for health and healthcare. Individuals who are more efficient producers of health, such as those who are highly educated and with a higher level of health literacy, for instance, have also greater ability to find health information but at the same time are also more likely to have greater demand for health and healthcare services (Bundorf et al., 2006; Costa-Font, et al., 2009).

Empirical Strategy

This discussion suggests that a step toward a complete understanding of the effects described requires a complex model that considers the simultaneous relationships between e-HISB, healthcare access, and individuals health status.

As such, we used a simultaneous equation model for binary variables. Specifically:

- we constructed a joint model of e-HISB, healthcare access, and individuals' health status that considers an individual’s unobserved characteristics that are likely to be correlated with health information seeking, individuals’ health status and healthcare utilization.

Empirical Strategy

Using a recursive multivariate probit design, we constructed and estimated a system of three equations with one reduced-form equation and two

$$\textit{Health Status}_i = \delta_1 y \textit{Healthcare Access}_i + \delta_2 \mathrm{e} - \mathrm{HISB}_\mathrm{i} + \alpha_1' Z_{1i} + \varepsilon_{1i}$$

$$\textit{Healthcare Access}_i = \gamma_2 e{-}\textit{HISB}_I + \alpha_2' Z_{2i} + \varepsilon_{2i} \qquad (1)$$

$$e{-}\textit{HISB}_j = \alpha_3' Z_{3i} + \varepsilon_{3i}$$

where $\boldsymbol{Z}_{hi}$ (with h = 1, 2, 3) are vectors of exogenous variables, α_h are parameter vectors, and δ_o (with o = 1, 2) and γ_2 are scalar parameters. The error terms distributed as multivariate normal are ε_{hi}, each with a mean zero and variance covariance matrix Σ. Σ has values of 1 on the leading diagonal and correlations $\rho_{jk} = \rho_{kji}$ on the off-diagonal elements (where ρ_{jk} is the covariance between the error terms of equation *j* and *k*).

Outcome Variables

For the empirical model, we identified three classes of dependent variables:

- **e-HISB** was defined as a binary indicator of whether respondents have looked for information on health-related issues on the Internet since the COVID-19 outbreak. According to their answer, they were classified as e-HI seekers and non-e-HI seekers.

- As a measure of **health professional visits access**, we created a binary variable indicating whether respondents went to a doctor’s office or a medical facility in the last twelve months before the interview.

- As a measure of individuals' health status, we used the **self-assessed health (SAH).** We dichotomized the multiple-category responses and constructed a binary indicator with a value of 1 if individuals reported that their health was fair or poor, and 0 otherwise (i.e., excellent, very good, or good).

Explanatory Variables

In our model, we controlled for a rich set of individuals' demographic and socio-economic characteristics, general health literacy, computer skills and health conditions collected from the 8th wave of SHARE.

For demographics, we included the respondent's sex, age, family size, geographic location (rural vs urban area) and a dummy variable for the region of residence.

For socioeconomic characteristics we included individuals' level of education, marital status, occupation and income.

All these variables were constructed according to the information included in the 8th wave of SHARE.

Exclusion Restrictions

e-HISB equation

To deal with the potential endogeneity of e-HISB, we exploited the heterogeneity in regional NUTS-2 on **broadband coverage**. Specifically, we used data from the **Eurostat database on Information and Communication Technology (ICT) usage in households and by individuals** and measure broadband internet diffusion with the variable that refers to the percentage of households with broadband internet access (isoc_r_broad_h).

We assumed that the high-speed connection increases the frequency of Internet use and the engagement with Internet activities facilitating information search, moreover the high-speed connection might also mean that individuals can access more content in a given amount of time.

Exclusion Restrictions

Health professional visits equation

In order to address the potential endogeneity of health professional visits binary indicator in the health equation we included, in the vector z_{2i}, an indicator of healthcare supply at the regional level (NUTS-2), namely **the number of medical doctors**, including generalists and specialist medical practitioners per 1.000.000 inhabitants provided by **the Eurostat**.

We expected that the number of doctors and their geographic distribution might influence the likelihood of accessing a health professional in normal circumstances and even more so during exceptional times such as the COVID-19 pandemic.

Results - Multivariate Probit Model – Estimated Marginal Effects

	e-HISB	Health Professional Visits	SAH
SAH_{t-1}	-0.026	0.041***	0.350***
	(0.037)	(0.031)	(0.040)
chronic_conditions	0.034***	0.132***	0.110***
	(0.034)	(0.029)	(0.037)
rural	-0.055***	-0.001	0.018*
	(0.049)	(0.041)	(0.028)
health_literacy	0.058***	0.021	-0.080***
	(0.069)	(0.040)	(0.038)
d_high_pcskill	0.192***	-0.001	-0.036***
	(0.019)	(0.016)	(0.015)
female	0.023***	0.016**	-0.003
	(0.031)	(0.023)	(0.028)
hhsize	-0.016*	-0.009*	0.002
	(0.022)	(0.015)	(0.018)
age2021	-0.013***	-0.001	0.003***
	(0.002)	(0.002)	(0.002)
local_spread_covid	0.052***	0.039***	0.020***
	(0.031)	(0.030)	(0.028)
decline_stringency_index	0.012*	0.033***	-0.009
	(0.033)	(0.034)	(0.031)
marital_status	0.038***	0.008	-0.003
	(0.038)	(0.035)	(0.046)
retired	0.047***	0.040***	-0.013
	(0.049)	(0.042)	(0.043)
employed	0.038***	0.008	-0.043**
	(0.046)	(0.046)	(0.058)
med_ed	0.097***	0.012	-0.025**
	(0.036)	(0.032)	(0.033)
high_ed	0.205***	0.045***	-0.049***
	(0.041)	(0.041)	(0.039)
2°quartile	0.030**	0.024**	-0.021
	(0.054)	(0.034)	(0.049)
3°quartile	0.072***	0.027*	-0.030**
	(0.055)	(0.042)	(0.048)
4°quartile	0.077***	0.012	-0.033
	(0.055)	(0.057)	(0.064)
tech01	0.017***		
	(0.008)		
n. of physicians		0.004***	
		(0.001)	
e-HISB		0.061***	-0.017***
		(0.052)	(0.062)
health professional visits			0.057***
			(0.072)
N	13829	13829	13829

Indirect Effect of eHISB on Health through Healthcare Access

	Marginal Effects	*SE*
Health Care Utilization (total effect)	0.104***	0.013
Direct effect	-0.008*	0.006
Indirect Effect	0.112***	0.013

Results: correlation coefficients between the error terms

	e-HISB	Health professional visits	SAH
e-HISB	1	0.008(0.028)	0.070*(0.029)
Health professional visits		1	-0.026(0.036)
SAH			1

The null hypothesis of exogeneity is rejected in only one case. According to our results, there exists a positive and statistically significant correlation between the disturbance of the e-HISB equation and the structural equation for individuals' health status — i.e., unobservable variables that increase the likelihood of bad health also increase the probability of searching health information online.

Conclusions

- Consistently with the previous literature, the results of this analysis show that the effect of Internet health information seeking on health care utilization is positive, and statistically significant. So, patients do not see the Internet as a replacement for the health professional, but as **complement.**

- The effect of e-HISB on health appeared much more complex:

  - the **estimated direct effect** of e-HISB on individuals poor self-perceived health, is negative, the **indirect effect** of e-HISB on individuals' poor health status through healthcare access is positive.

  - These results are consistent with the interpretation that, in principle, **patients lack the ability to understand medical information they find on the Internet**. Accordingly, **they may decide to make health-related decisions independently of a health professional**. However, **incorrect online health information and a wrong self-diagnosis increases for them the likelihood to doctor visits, that in turn increases also the likelihood of perceiving poor health status.**

Thank you

European Commission