Selected Presentation at the 2020 Agricultural & Applied Economics Association Annual Meeting, Kansas City, Missouri, July 26-28

Impact of Internet use on Health outcomes of Rural Residents: Evidence from China

Lili Li[1], Yiwu Zeng[2*], Zhonggen Zhang[1]

[1] China Academy of Rural Development, Zhejiang University, Hangzhou, Zhejiang, China

[2] School of Economics and Management, Hangzhou Normal University, Hangzhou, Zhejiang, China

[*] Corresponding author: 11420050@zju.edu.cn

Abstract

Health is an important component of human capital, and healthy human capital plays a huge role in promoting national economic growth. The internet has the potential to improve the health of rural residents by providing public access to large amount of health information, medical resources, and social support. However, most research has studied the impact of internet use for seeking health information on people with a specific illness or disease or explored factors that predict use of the internet for health purposes. Little is known about the impact of general internet use by general population on health outcomes. Moreover, little research has examined that issue in developing countries, especially in rural China, where medical resources are scarce and rural residents have low health literacy. To bridge these gaps in the literature, this study uses longitudinal data from a nationwide survey to investigate the causal relationship between internet use and health outcomes in rural China. The study finds that internet use has positive and significant impacts on health outcomes by increasing access to information, social interaction and physical abilities. We also find that internet use had heterogeneous effects on the health of rural residents of different gender, age groups and education levels.

Key words:
Internet use
Health outcomes
Information accessibility
Social interaction
Rural China

1. Introduction

Health is an important component of human capital, and healthy human capital plays a huge role in promoting national economic growth (Fogel, 1994). In particular, people with a higher level of healthy human capital can provide a stable, continuous, and high-quality labor, which is an important condition for high income and coordinated economic development. However, the

two-week prevalence rate of rural residents in China increased from 13.9% in 2003 to 20.2% in 2013. In addition, the two-week prevalence, chronic disease prevalence and other indicators indicated that the overall health status of Chinese rural residents was declining[1].

In developing countries, the inequality caused by the dual urban-rural structure in health care is widespread worldwide. In China, although medical and health expenditures continue to rise, there is a large gap between urban and rural health expenditures, both in terms of public expenditures and private expenditures. Moreover, some problems become prominent, such as the unreasonable regional allocation of medical resources, unfair medical services, and a large gap between urban and rural medical quality. Specifically, rural residents have poor public health facilities, poor access to health services, and the average level of education and health care in rural areas is far lower than that in cities. Results from the National Resident Health Literacy Survey show that in 2019, the health literacy level of urban residents was 24.81%, while that of rural residents was 15.67%. Although overall health literacy is on the rise, it still faces problems such as low overall level, uneven urban and rural development, and relatively low level of health literacy among rural residents[2]. In addition, due to the underdeveloped health care system in China, it is of high cost for rural residents to see a doctor in a regular hospital. In this context, how to improve health status of rural residents, reduce inequalities in health, and achieve the integration development of urban and rural areas has become one of the government's most concerned issues.

The internet has the potential to improve the health of rural residents by providing public access to large amount of health information, medical resources, and social support (Murero &

[1] Data source: National Bureau of Statistics, "China Health Statistics Yearbook 2019".
[2] Retrieved from National Health Commission of the People's Republic of China. http://www.nhc.gov.cn/xcs/s3582/202004/df8d7c746e664ad783d1c1cf5ce849d5.shtml

Rice, 2006; Jiang & Street, 2016). Specifically, internet use for health-related purposes could increase people's medical knowledge, improve their health literacy, and implement healthy behaviors, such as moderate exercise per week, healthy eating and avoiding unhealthy behaviors. Moreover, internet can provide opportunities for rural residents to interact with health care providers, especially for people in remote areas (Hao, 2015). For patients, the use of the internet can also enhance communication with relatives and friends in the distance, and provide them with social support from online support groups, which is conducive to their health-related knowledge and physical and mental health.

In recent years, China's internet has developed rapidly. As of June 2019, China had 854 million internet users, and the internet penetration rate was 61.2%. 33.2% of Chinese adults have sought health information from the internet (Wang et al., 2013). In addition, the Chinese government has successively proposed a series of informatization strategies such as "Internet +" and "digital village" to develop rural areas. Specifically, in 2018, the State Council issued the "Opinions on Promoting the Development of 'Internet + Medical Health' " in order to promote the deep integration of the internet and medical health, which helps to promote access to health information, adopt a healthy lifestyle, and improve the accessibility of medical and health services for rural residents by using the internet. Therefore, in this context, it is necessary and significant to explore whether the internet plays a role in improving the health of rural residents. Our study use longitudinal data from the China Family Panel Studies (CFPS) to investigate the relationship between internet use and health outcomes of rural residents. Moreover, we explore the pathways through which internet use could improve health outcomes to provide a theoretical basis for the popularity of the internet in rural areas, and the improvement of rural residents' health.

2. Literature review

In recent years, a number of research have empirically studied the impact of the internet on people's health, but results are mixed. On one hand, some studies found that internet use contributed to health outcomes of adults. Specifically, the internet has the potential to provide health information and behavioral advice to promote health status. For example, the internet is being used as a resource for health-related information in the general population (Trotter & Morgan, 2008; Kontos et al., 2012). Compared with other information resources, the internet has more advantages in obtaining health-related information because of its anonymity, convenience, and a wide variety of information (Cotton & Gupta, 2004). General internet users are more likely than non-internet users to have weekly moderate physical activity, healthy eating, and less likely to report smoking (Xavier et al., 2013). Moreover, the internet provides access to social interaction with others and social support resources, which have positive effects on health outcomes (Oh & Lee, 2012; Southwell, 2013; Jiang & Street, 2016). In particular, internet use for communicating with distant friends and family could lead to a larger social network (Wellman et al., 2001; Katz & Rice, 2002), which was associated with better psychological functioning and wellbeing benefits (Bessière et al., 2010). When people use the internet for health, they may find support, encouragement, and help from online communities and support groups sharing similar health-related experiences (Braithwaite et al., 1999; Kummervold et al., 2002).

On the other hand, some literature found that internet use had null or negative impacts on health outcomes. For example, internet use for health purposes was not related to breast cancer screening (Xavier et al., 2013). An internet-based intervention for diabetes patients had no impact on participants' knowledge, behavior, and health outcomes (Balsa & Gandelman, 2010).

Particularly, online health information could negatively affect health outcomes if people could not distinguish poor advice from correct health information and good advice and thus made hasty and ill-informed health decisions. In addition, health-related internet use was associated with small but reliable increases in depression (Bessière et al., 2010). Excessive use of the internet may also have negative impacts on people's health, since people sit too long and lack physical activity.

Although researchers have focused recently on the issues of the relationship between internet use and health outcomes, some gaps exist. First, most research has studied the impact of internet use for seeking health information on people with a specific illness or disease (Gustafson et al., 2008) or explored factors that predict use of the internet for health purposes (Jacobs et al., 2017). Little is known about the impact of general internet use by general population on health outcomes. Second, little research has examined the pathways through which internet use could affect health outcomes. Third, little research has examined that issue in developing countries, especially in rural China, where medical resources are scarce and rural residents have low health literacy. To bridge these gaps in the literature, this study uses longitudinal data from a nationwide survey to investigate the causal relationship between internet use and health outcomes in rural China. Specifically, we examine whether internet use have an impact on health outcomes of rural residents in China. How internet use can affect health outcomes of rural residents? We propose three pathways (information accessibility, social interaction, moderate exercise) through which internet use could improve health status. We also examine whether these effects differ for people with different characteristics (for example, gender, age and education). In addition, we employ instrumental variables and linear two-stage least squares method (2SLS) to address potential endogenous problems due to omitted variables or reverse causality.

3. Data and method

3.1. Sample Selection

The study uses data from the China Family Panel Studies (CFPS), which is a national and longitudinal survey conducted by the Institute of Social Science Survey (ISSS) at Peking University. The CFPS surveys economic and social development and changes in 25 provinces in China, excluding Tibet, Qinghai, Xinjiang, Ningxia, Inner Mongolia, Hainan, Hong Kong, Macau and Taiwan. It uses a stratified, multi-stage sampling strategy that ensures the sample represents 95 percent of the total population of China (Xie, 2013; Xie & Hu, 2014). Therefore, the CFPS sample can be regarded as a nationally representative sample.

The first round of data collection was carried out in 2010, followed by a further four rounds of data collection in 2012, 2014, 2016 and 2018. Our study primarily uses data from CFPS 2014, CFPS 2016 and CFPS 2018 since the CFPS contains an extensive set of measures of internet access and usage (Xie & Hu, 2014), and uses the same set of test questionnaires to measure health outcomes and sociodemographic characteristics needed in our study. As such, this study is based on rural residents who are 16 years old or older. The original sample of CFPS 2014 included 17883 villagers and 7042 rural households in 192 counties (districts) in 25 provinces. CFPS followed up 8091 villagers in 2014 and 2018. After eliminating the samples with missing values and inconsistent personal information of the respondents, the final sample included 7528 villagers.

3.2. Data Collection

Outcome measurement

The dependent variable in this study is the self-reported health outcomes of rural residents, as measured by a question about asking the health condition of respondents in the CFPS. The CFPS

measures it on a scale from 1 to 5, with 1 indicating very healthy, 2 indicating healthy, 3 indicating relatively healthy, 4 indicating general, and 5 indicating unhealthy. In the analysis of this paper, we transformed these answers into a dummy variable, defining self-reported health outcomes with values of 1 to 4 as “healthy" and assigning these to a new value of 1, and defining self-reported health outcomes with values of 5 as “unhealthy" and assigning it to a new value of 0.

Internet use measurement

The main set of independent variables that we use in the analysis is internet use, measured in two different ways: general internet usage, and weekly online time. General internet usage (1 = yes, 0 = no) is a binary variable that asks respondents whether they had access to the internet (in any capacity and via any means) in the CFPS survey. Weekly online time (hours) is a continuous variable measured by asking respondents how many hours they spend online each week in their spare time.

Control variables

The following covariates are used to control potential confounding in the relationship between the internet and health of rural residents. First, we control individual characteristics including gender, age, years of education, marriage, weekly exercise, smoke, drink, sleep quality (how many hours of sleep each night) and work (does the villager have a job). Second, several household characteristics are controlled for, including number of family members, household income per capita, number of houses owned, and family gift exchange (total gifts and cash for family social activities).

The baseline characteristics of rural residents in our sample are shown in Table 1. 48.2% of respondents were males and 51.8% of respondents were females. 95.7% of respondents were married, 80.0% had paid work. The average age of respondents was 49 years. With respect to education, the average years of education was 5.8, indicating that the education level of rural

residents was below primary school. We can see that 25.6% of respondents exercised weekly, 31.7% smoked and 16.4% drank more than three times a week in the past month. Moreover, the average sleep hours every night were 7.9 hours. The majority of the respondents (80.3%) were in good health. However, the percentage of rural residents using the internet was 13.7% in 2014, which was far lower than the national overall internet penetration rate (47.9%, CNNIC). Further information about the household characteristics of respondents are provided in Table 1.

Table 1. Baseline characteristics of rural residents.

	Definition	Mean	S.D.
Dependent variable			
Health	1 = healthy, 0 = unhealthy	0.803	0.398
Independent variables			
General internet usage	1 = yes, 0 = no	0.137	0.344
Weekly online time	Unit: hours	1.339	5.445
Individual characteristics			
Gender	1 = male, 0 = female	0.482	0.500
Age	Unit: years old	49.202	13.467
Years of education	Unit: years	5.799	4.249
Marriage	1 = yes, 0 = no	0.957	0.202
Exercise	1 = yes, 0 = no	0.256	0.437
Smoke	1 = yes, 0 = no	0.317	0.465
Drink	Yes=1, no=0	0.164	0.370
Sleep quality	Unit: hours per night	7.932	1.629
Work	1 = yes, 0 = no	0.800	0.400
Household characteristics			
Number of family members	Unit: persons	4.446	1.917
Household income per capita	Unit: yuan	10597.400	20725.550
Number of houses owned		1.160	0.445
Family gift exchange	The sum of gift income and gift expenditure, units: yuan	5282.456	10228.840

3.3. Empirical model

This subsection describes our analytical approach. Since health is a binary dummy variable, we construct a panel logit model with robust standard errors clustering at the village level to study the impact of internet use on health outcomes of villagers. The specific model is as follows:

$$P(y_{it}=1|x_{it},\beta,\mu_i)=\Phi(\mu_i+x_{it}\beta)=\frac{e^{\mu_i+x_{it}\beta}}{1+e^{\mu_i+x_{it}\beta}} \quad (1)$$

In this equation, Y_{it} is the outcome variable representing the health status of villager *i* at time *t*. X_{it} is a vector of variables that capture internet use measured by general internet usage and weekly online time, and other control variables including individual characteristics and household characteristics. μ_i is the individual effect, and $\Phi(\cdot)$ denotes cumulative distribution function subject to Logistic distribution.

4. Results

In this section, we seek to investigate whether internet use has an impact on health outcomes of rural residents. To do this, we first use multivariate analysis to examine the impacts of internet use (which are measured by general internet usage and weekly online time) while controlling for individual and household characteristics. Second, we conduct a potential path analysis for the impacts. Third, we examine whether internet use has different impacts on health outcomes of rural residents with different characteristics. Finally, we test the causal relationship between internet use and health outcomes after solving the endogenous problem.

4.1 The impacts of internet use

Results for the impact of internet use on health outcomes of rural residents are found in Table 2 and are reported in standard deviations. Both models 1 and 2 show the relationship between general internet usage and health outcomes after including control variables. In addition, Model 2 includes the year dummies. We find a positive and significant effect of general internet usage on health outcomes. Specifically, in Model 1, health status improves by 0.266 standard deviations at the 0.01 level of significance (row 1, column 1), if a rural resident uses the internet. In Model 2, health status improves by 0.351 standard deviations (row 1, column 2, $p < 0.01$), which is greater than the effect in Model 1. Moreover, we find similar results from another measure of internet use

(Weekly online time) in models 3 and 4. As in the case of models 1 and 2, Model 4 includes year dummies, compared with Model 3. We see that if rural residents increased their weekly online time by one hour, their health outcomes would improve by 0.010 standard deviations (Model 3, row 2, column 3) and 0.012 (Model 4, row 2, column 4), respectively. All four models show that internet use has a positive and significant impact on health outcomes of rural residents. After controlling the year dummies, internet use has a greater effect on health outcomes of rural residents. This is consistent with previous research that health outcomes of rural residents are improved by using the internet (Jiang & Street, 2017).

Table 2 also shows that being male, younger, more educated, and married were all more likely to report good self-rated health. Individuals, who exercised weekly, did not smoke and drink, were significantly more likely to report good health. Only after controlling for the year dummies, sleep quality was significantly correlated with health outcomes. The coefficients of models 2 (0.033) and 4 (0.032) were both positive and significant at the 0.1 level and small in size. In addition, people who had a job, more family members, more houses and higher household income reported better health conditions. All these effects are statistically different from zero.

Table 2. The impact of internet use on health outcomes of rural residents.

	Model 1	Model 2	Model 3	Model 4
(1) General internet usage	**0.266*****	**0.351*****		
	(0.088)	**(0.089)**		
(2) Weekly online time			**0.010***	**0.012****
			(0.006)	**(0.006)**
(3) Gender	0.624***	0.555***	0.630***	0.563***
	(0.104)	(0.106)	(0.105)	(0.106)
(4) Age	-0.070***	-0.065***	-0.072***	-0.067***
	(0.004)	(0.004)	(0.004)	(0.004)
(5) Years of education	0.068***	0.090***	0.070***	0.093***
	(0.009)	(0.010)	(0.009)	(0.010)
(6) Marriage	0.442*	0.419*	0.445*	0.424*
	(0.233)	(0.235)	(0.235)	(0.238)

(7) Exercise	0.145**	0.173***	0.156**	0.181***
	(0.061)	(0.062)	(0.061)	(0.062)
(8) Smoke	0.158*	0.174*	0.159*	0.177*
	(0.091)	(0.092)	(0.091)	(0.092)
(9) Drink	0.662***	0.660***	0.660***	0.657***
	(0.105)	(0.105)	(0.105)	(0.106)
(10) Sleep quality	0.031	0.033*	0.030	0.032*
	(0.019)	(0.019)	(0.019)	(0.019)
(11) Work	0.602***	0.610***	0.602***	0.610***
	(0.077)	(0.076)	(0.077)	(0.077)
(12) Number of family members	0.088***	0.084***	0.087***	0.083***
	(0.019)	(0.019)	(0.019)	(0.019)
(13) Household income per capita	0.037***	0.041***	0.038***	0.041***
	(0.013)	(0.013)	(0.013)	(0.013)
(14) Number of houses owned	0.218***	0.215***	0.219***	0.216***
	(0.072)	(0.072)	(0.072)	(0.072)
(15) Family gift exchange	-0.006	-0.008	-0.006	-0.008
	(0.005)	(0.005)	(0.005)	(0.005)
(16) Year dummies	NO	YES	NO	YES
(17) Constant	3.197***	2.618***	3.299***	2.799***
	(0.382)	(0.383)	(0.381)	(0.383)
(18) Number of observations	22584	22584	22584	22584

*** Statistically significant at 1%.
** Statistically significant at 5%.
* Statistically significant at 10%.

4.2. Pathway examination

We examine three potential pathways linking internet use to improved health outcomes. Specifically, one path is that internet use allows people to have access to a large amount of health information conveniently, which is positively associated with health status. Internet use for seeking health information helps to provide necessary informational support that improve people's ability to deal with health-related problems (Barak et al., 2008). Internet use also has direct positive paths to social interaction and social support, which, in turn, is positively associated with health outcomes (Beaudoin & Tao, 2008). Social interaction online can help rural residents get support and advice, improve health knowledge, and manage their health, leading to better health outcomes (Ahmadi, 2016). Moreover, internet use could expand people's social networks and increase their social interaction, which helps to obtain social support resources from distant friends, online support communities and healthcare experts. Another path is that internet use could

increase people's health knowledge, and promote health behavior change (Webb et al., 2010), such as weekly physical activities, not smoking. People who was a nonsmoker and who did exercise weekly were significantly more likely to report good self-rated health (Suka et al., 2015).

There are relevant questions to measure information accessibility, social interaction and exercise in the CFPS survey. Information accessibility was asked as a categorical variable captured by a question in the CFPS that asked how important people used the internet to obtain information. Respondents gave their answers on a 5-point Likert scale (1 = very unimportant to 5 = very important). Social interaction is a continuous variable (yuan) measured by family gift exchange, which is the sum of family gift income and family gift expenditure. Information about family gift income and expenditure is collected in the CFPS survey. Exercise is measured by asking respondents how many times a week they do exercise. In this paper, we transformed the answers into a dummy variable, defining weekly exercise with a value of 1, and defining no exercise a week with a value of 0. In the regression of information accessibility on internet use, this study uses an ordered logistic model, since information accessibility is an ordered discrete variable. In the regression of social support, random effects model is employed.

In Table 3, the analysis examines the impacts of internet use on three paths, information accessibility, social interaction and exercise. We see that general internet usage is positively associated with information accessibility (row 1, column 1, $p<0.01$), which means compared with people who do not use the internet, internet users claims that internet is an important channel for obtaining information. We find same results from the regression of the weekly online time (row 2, column 2). When rural residents are online for more time each week, they are more likely to obtain health information on the internet. All these coefficients are significantly different from

zero. Columns 3 and 4 show that internet use (general internet usage and weekly online time) also has a positive impact on social interaction. However, we see that the effect (row 2, column 4) of weekly online time is not statistically significant from zero. In other words, like a number of authors in the literature predict, the more time rural residents spend online to increase social networks, the less time they have to maintain and strengthen their social relationships offline, which may not necessarily improve overall social relationships. Finally, we examine the impact of internet use on rural residents' exercise. Similar to the regression of information accessibility, general internet usage (coefficient = 0.546; row 1, column 5) and weekly online time (coefficient = 0.017; row 2, column 6) both help to encourage rural residents to exercise moderately every week. Moreover, these effects are significant at the 0.01 level and moderate in size.

Table 3. Regression results of pathway examination.

	Information accessibility		Social interaction		Exercise	
	(1)	(2)	(3)	(4)	(5)	(6)
General internet usage	**2.359*****		**0.385*****		**0.546*****	
	(0.075)		**(0.132)**		**(0.068)**	
Weekly online time		**0.103*****		**0.000**		**0.017*****
		(0.005)		**(0.007)**		**(0.003)**
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	22584	22584	22584	22584	22584	22584

*** Statistically significant at 1%.
** Statistically significant at 5%.
* Statistically significant at 10%.

4.3. Heterogeneous effects

To further draw out important effects of internet use on health outcomes, we now examine important subgroups of rural residents, specifically gender, age and education. In the previous literature, female and older respondents were more likely to use the internet for seeking heath information (Ishikawa et al., 2012; Viswanath & Ackerson, 2011). Moreover, higher education

was positively associated with health-related internet use (Wang et al., 2013). More internet usage experiences could enhance users' attitudes toward the technology, and increase their capacity to take advantage of resources found online (Chang & Chen, 2008), which in turn positively affected health outcomes (Jiang & Street, 2016).

The effects of internet use on rural residents of different genders are presented in Table 4. We find that general internet usage (coefficient = 0.538; row 1, column 3) and weekly online time (coefficient = 0.027; row 2, column 4) both have a greater impact on health outcomes for female. These coefficients are positive and significant at the p = 0.01 level. For male, the coefficients of general internet usage and weekly online time are small and not statistically different form zero, although general internet usage has a positive impact on health outcomes (row 1, column 1). Moreover, weekly online time is negatively associated with men's health status (row 2, column 2).

Turning to the impact of internet use on the health outcomes of rural residents of different age groups, which are the 16-39 age group, the 40-59 age group and the age group over 60 years old, shown in Table 5. We find that general internet usage has a positive and significant effect on health outcomes for rural residents in the 16-39 age group (row 1, column 1) and 40- 59 age group (row 1, column 3). For people in the 16-39 age group, the magnitude of this effect (0.461) is larger than the effect of general internet usage on people in the 40-59 age group (0.212). However, the effect of general internet usage on people in the age group over 60 years old is not significant and smallest (coefficient = 0.081; row 1, column 5). In addition, the effects of weekly online time are not significant (row 2) for rural residents in all age groups.

In Table 6, which presents the impact of internet use on health outcomes of rural residents with different education levels, we find that the effects of internet use are significant only among

people with primary school education or below. Specifically, general internet usage is positively associated with health outcomes of rural residents with primary school education or below (row 1, column 1), although the coefficient (0.462) is smaller than the coefficient (0.521) for people with higher education than senior high school. We also find similar results of the effect of weekly online time (coefficient = 0.023; row 2, column 2), which is positive and significant. Again, the magnitude of this effect is smaller than the impact (coefficient = 0.028; row 2, column 6) of weekly online on the health of people with higher education than senior high school. For people with junior high school or senior high school, the effect of internet use is not significant, and the size is the smallest (columns 3 and 4).

Table 4. The impact of internet use on health outcomes of female and male.

Independent variable	**Dependent variable: health outcomes**			
	Male		**Female**	
	(1)	(2)	(3)	(4)
General internet usage	0.133		0.538***	
	(0.119)		(0.134)	
Weekly online time		-0.003		0.027***
		(0.008)		(0.009)
Year dummies	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
Number of observations	10896	10896	11688	11688

*** Statistically significant at 1%.
** Statistically significant at 5%.
* Statistically significant at 10%.

Table 5. The impact of internet use on health outcomes of rural residents in different age groups.

Independent variable	**Dependent variable: health outcomes**					
	16-39 age group		**40-59 age group**		**≧60 age group**	
	(1)	(2)	(3)	(4)	(5)	(6)
General internet usage	0.461*		0.212*		0.081	
	(0.242)		(0.110)		(0.287)	
Weekly online time		-0.005		0.010		-0.000
		(0.011)		(0.008)		(0.023)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4542	4542	11182	11182	6860	6860

*** Statistically significant at 1%.

** Statistically significant at 5%.
* Statistically significant at 10%.

Table 6. The impact of internet use on health outcomes of rural residents in with different educational backgrounds.

Independent variable	Dependent variable: health outcomes					
	Primary school and below		Junior high school and senior high school		Senior high school above	
	(1)	(2)	(3)	(4)	(5)	(6)
General internet usage	0.462***		0.273		0.521	
	(0.123)		(0.180)		(0.353)	
Weekly online time		0.023**		-0.004		0.028
		(0.009)		(0.010)		(0.021)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	15616	15616	5235	5235	1733	1733

*** Statistically significant at 1%.
** Statistically significant at 5%.
* Statistically significant at 10%.

4.4. Endogenous issues

Although this study uses longitudinal data and panel logit model, which allow stronger causal claims about the relationship between internet use and health outcomes than do cross-sectional data (Bessière et al., 2010), there may still be endogenous problems, thus we cannot draw a causal relationship between the two. The longitudinal design of the CFPS follows the same individuals through multiple years, helping to control stable characteristics such as demographic differences and personality, however, it cannot control for unmeasured variables that may change over time. There may also be reverse causality between internet use and health outcomes, since people with poor health are more likely to use the internet to seek health-related information or find support resources. Therefore, our study includes two instrumental variables (provincial internet penetration, whether there is a computer at home) to address endogenous problems. Internet penetration refers to the popularity and utilization of internet facilities in a region, which is not directly related to rural residents' health. However, internet penetration has an important impact on

individuals' online decision-making, which satisfies the exogenous requirements of instrumental variables. Whether people have computers in their homes is a prerequisite for people to use the internet, and has no direct relationship with their health. Having a computer at home (1 = yes, 0 = no) also meets the requirements of instrumental variables. Moreover, we use the two-stage least square (2SLS) method to estimate the effect of internet use on health outcomes.

Table 7 shows the results for the impact of general internet usage and weekly online time on health outcomes of rural residents. In the first-stage regression, both instrumental variables have a significant impact on the general internet usage (row 1, row 2, column 1), which is positively and significantly associated with health outcomes (row 3, column 2). The coefficient (0.099) decreased by 0.252 after addressing the endogenous problem. We also find similar results for the impact of weekly online time. Both instrumental variables are significantly associated with weekly online time (row 1, row 2, column 3) at a significance level of p = 0.01. Weekly online time has a significant effect on health outcomes (row 4, column 4). In addition, the effect (0.009) decreased by 0.003 after solving endogeneity.

Table 7. Results of instrumental variables regression.

	First-stage	IV regression	First-stage	IV regression
	(1)	(2)	(3)	(4)
Internet penetration	**0.109****		**2.549*****	
	(0.049)		**(0.760)**	
Having a computer at home (1 = yes, 0 = no)	**0.167*****		**2.015*****	
	(0.009)		**(0.193)**	
General internet usage		**0.099****		
		(0.039)		
Weekly online time				**0.009*****
				(0.003)
Year dummies	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
Number of observations	22584	22584	22584	22584

*** Statistically significant at 1%.
** Statistically significant at 5%.
* Statistically significant at 10%.

5. Conclusion

Although previous studies have investigated the impact of internet use on people's health outcomes, the causal relationship between the two is not yet fully understood. Most of the studies on this topic only use cross-sectional data to study this relationship. Moreover, little research has examined that issue in developing countries, especially in rural China, where rural residents are lack of high-quality medical resources. To bridge these gaps in the literature, we have examined the causal relationship between internet use and health outcomes of rural residents, using instrumental variables and longitudinal data from three waves of a nationally representative survey in rural China to identify the causal relationship. We also examined three pathways (information accessibility, social interaction, moderate exercise) through which internet use could improve health outcomes. In addition, we also investigated the heterogeneous effects of internet use on rural residents of different gender, ages and education levels.

The findings of this study are consistent with research, which shows that internet use has positive and significant impacts on health outcomes. After addressing endogenous problems, internet use still has a positive and significant impact on health outcomes. We find that internet use was positively associated with information accessibility, which means that the internet is an important source of information for rural residents to access to health information. We also find that internet use increases social interaction significantly where people can obtain social support and helpful advice from online doctors and peers sharing similar health experiences. In addition, internet use has a significant effect on people's health behaviors. In particular, rural residents would do exercise every week to improve their health status. The findings show that internet use provides more opportunities for rural residents to seek health information and draw more social

support resources, and have health behaviors, which help to improve their health literacy, and enhance their ability to manage health, leading to better health outcomes.

Another important finding is that internet use had heterogeneous effects on the health of rural residents of different gender, age groups and education levels. In the regression that shows gender heterogeneity, the results demonstrate that internet use has a positive and significant impact on women's health outcomes. The effects are greater than that of the internet on men' health outcomes. However, the effect of internet use on men's health is not significant. For rural residents of different age groups, we found that the health of rural residents in younger age groups would be improved if they had a general internet use. Specifically, the significant and positive effects were the effects of general internet usage on health outcomes of rural residents in the 16-39 age group and the 40-59 age group. Moreover, the effect size for the 16-39 age group is almost twice that of the 40-59 age group. However, we did not find a significant relationship between general internet use and health outcomes of people aged 60 or above. This may be because older people are less likely to use the internet, and they have low internet skills, leading to that they are unable to benefit from online health resources (Jacobs et al., 2017), especially for rural China. In addition, we found that internet use only had a positive and significant effect on the health outcomes of people with primary school or below. This may be because rural residents with more education generally have high socioeconomic status, and they have more opportunities to communicate directly with healthcare providers.

References

1. Shaohai Jiang & Richard L. Street. (2017). Pathway Linking Internet Health Information Seeking to Better Health: A Moderated Mediation Study, Health Communication, 32:8, 1024-1031, DOI: 10.1080/10410236.2016.1196514
2. Fogel, R. W. (1994). Economic growth, population theory and physiology: The bearing of long-term processes on the making of economic policy. American Economic Review, (3).
3. Wang, M. P., Viswanath, K., Lam, T. H., Wang, X., & Chan, S. S. (2013). Social determinants of health information seeking among Chinese adults in Hong Kong. Plos One, 8(8), e73049. doi:10.1371/journal.pone.0073049
4. Xavier AJ, d'Orsi E, Wardle J, Demakakos P, Smith SG, von Wagner C. (2013). Internet use and cancer-preventive behaviors in older adults: findings from a longitudinal cohort study. Cancer Epidemiol Biomarkers Prev, 22(11):2066-2074. doi:10.1158/1055-9965.EPI-13-0542
5. Kontos EZ, Emmons KM, Puleo E, Viswanath K. (2012). Contribution of communication inequalities to disparities in human papillomavirus vaccine awareness and knowledge. Am J Public Health, 102: 1911–1920.
6. Trotter M I，Morgan D W. (2008). Patients' use of the Internet for health related matters：a study of Internet usage in 2000 and 2006［J］.Health Informatics Journal, 14(3), 175-181.
7. Shelia R Cotton, Sipi S Gupta. Characteristics of online and offline health information seekers and factors that discriminate between them. Social Science & Medicine, Volume 59, Issue 9, November 2004, Pages 1795-1806.
8. Murero, M., & Rice, R. E. (2006). E-health research. In M. Murero & R. E. Rice (Eds.), T*he Internet and health care: Theory, research, and practice* (pp. 3–26). Mahwah, NJ: Lawrence Erlbaum Associates.
9. Hao, H. (2015). The development of online doctor reviews in China: An analysis of the largest online doctor review website in China. *Journal of Medical Internet Research*, 17(6), e134. doi:10.2196/jmir.4365
10. Oh, H. J., & Lee, B. (2012). The effect of computer-mediated social support in online communities on patient empowerment and doctor–patient communication. Health Communication, 27, 30–41. doi:10.1080/10410236.2011.567449
11. Southwell, B. G. (2013). Social networks and popular understanding of science and health: Sharing disparities. In Baltimore. MD and Research Triangle Park, NC: Johns Hopkins University Press and RTI Press.
12. Balsa, A. I., & Gandelman, N. (2010). The impact of ICT on health promotion: A randomized experiment with diabetic patients (IDB Working Paper No. 76). Washington, DC: Inter-American Development Bank.
13. Kummervold, P. E., Gammon, D., Bergvik, S., Johnsen, J-A. K., Hasvold, T., & Rosenvinge, J. H. (2002). Social support in a wired world: Use of online mental health forums in Norway. *Nordic Journal of Psychiatry*, 56(1), 59-65.
14. Katz J. E., Rice R. E. (2002). Social Consequences of Internet Use: Access, Involvement, and Interaction. Cambridge, MA: MIT Press.
15. Wellman B, Haase AQ, Witte J, Hampton K. (2001). Does the Internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment. Am Behav Sci, 45(3):436–455.
16. Katie Bessière, Sarah Pressman, Sara Kiesler, Robert Kraut. (2010). Effects of Internet Use on Health and Depression: A Longitudinal Study. J Med Internet Res, 12(1): e6.

17. Gustafson DH, Hawkins R, McTavish F, Pingree S, Chen WC, Volrathongchai K, Stengle W, Stewart JA, Serlin RC. (2008). Internet-Based Interactive Support for Cancer Patients: Are Integrated Systems Better? J Commun, 58, 238–257.
18. Wura Jacobs, Ann O. Amuta & Kwon Chan Jeon. (2017). Health information seeking in the digital age: An analysis of health information seeking behavior among US adults, Cogent Social Sciences, 3:1, 1302785
19. Beaudoin, C. E., & Tao, -C.-C. (2008). Modeling the impact of online cancer resources on supporters of cancer patients. New Media & Society, 10, 321–344. doi:10.1177/1461444807086477
20. Barak, A., Boniel-Nissim, M., & Suler, J. (2008). Fostering empowerment in online support groups. Computers in Human Behavior, 24, 1867-1883. doi:10.1016/j.chb.2008.02.004
21. Machi Suka, Takeshi Odajima, Masako Okamoto, Masahiko Sumitani, Ataru Igarashi, Hirono Ishikawa, Makiko Kusama, Michiko Yamamoto, Takeo Nakayama, & Hiroki Sugimori. (2015). Relationship between health literacy, health information access, health behavior, and health status in Japanese people. *Patient Education and Counseling*, 98, 660-668.
22. Webb, T., Joseph, J., Yardley, L., & Michie, S. (2010). Using the Internet to promote health behavior change: A systematic review and metaanalysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. Journal of Medical Internet Research, 12(1), e4. doi:10.2196/jmir.1376
23. Viswanath K, Ackerson L. K. (2011). Race, ethnicity, language, social class, and health communication inequalities: a nationally-representative cross-sectional study. PLoS ONE, 6, e14550.
24. Ishikawa Y, Nishiuchi H, Hayashi H, Viswanath K. (2012). Socioeconomic status and health communication inequality in Japan: a nationalwide cross-sectional survey. PLoS ONE, 7, e40664.
25. Wang MP, Viswanath K, Lam TH, Wang X, Chan SS. (2013). Social Determinants of Health Information Seeking among Chinese Adults in Hong Kong. PLoS ONE 8(8): e73049. doi:10.1371/journal.pone.0073049
26. Hernandez, L. M. (2013). Health literacy: Improving health, health systems and health policy around the world. Washington, DC: National Academies Press.
27. Rains, S. A. (2008). Health at high speed: Broadband internet access, health communication, and the digital divide. Communication Research, 35, 283–297. http://dx.doi.org/10.1177/0093650208315958