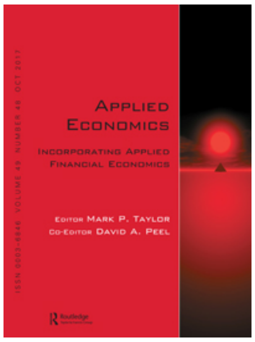

Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: http://www.tandfonline.com/loi/raec20

Does sports activity improve health? Representative evidence using local density of sports facilities as an instrument

Marc Brechot, Stephan Nüesch & Egon Franck

To cite this article: Marc Brechot, Stephan Nüesch & Egon Franck (2017) Does sports activity improve health? Representative evidence using local density of sports facilities as an instrument, Applied Economics, 49:48, 4871-4884, DOI: 10.1080/00036846.2017.1296548

To link to this article: http://dx.doi.org/10.1080/00036846.2017.1296548

Published online: 24 Feb 2017.

APPLIED ECONOMICS, 2017
VOL. 49, NO. 48, 4871–4884
https://doi.org/10.1080/00036846.2017.1296548

Does sports activity improve health? Representative evidence using local density of sports facilities as an instrument

Marc Brechot[a], Stephan Nüesch[b] and Egon Franck[a]

[a]University of Zurich, Zürich, Switzerland; [b]University of Münster, Münster, Germany

ABSTRACT

Using representative and geocoded data from the Swiss Household Panel and the *Swiss Business Census*, we estimate the effect of sports activity on health. OLS models show that sports activity significantly decreases overweight, sleeping problems, headaches, back problems, and perceived health impediments in everyday activities and significantly increases health satisfaction. Because sports activity is likely to be misreported and correlated with unobserved determinants of health, we use the number of sports facilities within 6 miles of the individual's residence as an instrument for sports activity. Although the instrument is powerful in explaining sports activity, the second-stage effects on health are mostly statistically insignificant due to the high SEs of the IV estimates.

KEYWORDS
Sports activity; health; instrumental variable; sports facilities

JEL CLASSIFICATION
I10; I12; H51; C26

I. Introduction

Physical inactivity is widely acknowledged as a global health problem in the twenty-first century. The proportion of inactive people is rising in many countries, creating risks for individual health, health care utilization and ultimately public health care costs (World Health Organization 2010). Therefore, exercise and intervention programmes that target an increase of individual physical activity are a recurring theme on the agenda of policy makers around the world (Heath et al. 2012). Such programmes are supported by a rich body of cross-sectional epidemiological research that shows a positive correlation between physical inactivity and a wide variety of detrimental health outcomes such as obesity, hypertension, osteoporosis, osteoarthritis, diabetes mellitus, colon and breast cancer, depression (Warburton, Nicol, and Bredin 2006; Katzmarzyk and Janssen 2004) and the use of health care services from physicians and hospitals (Sari 2009; Manning et al. 1991; Katzmarzyk, Gledhill, and Shephard 2000).

However, because sports activity is an endogenous, and in most data sets also a self-reported, choice variable, evidence from such correlational studies cannot be given a causal interpretation for three reasons: First, an individual's probability of doing sports likely correlates with unobserved health determinants such as a person's healthy or unhealthy lifestyle, for example, their nutrition, sleeping behaviour or personal hygiene. A healthy lifestyle tends to be positively correlated with sports activity and negatively correlated with health issues. Second, correlational studies do not identify the direction of causality. There may be negative correlation between headaches and sports activity because people with severe headaches do not participate in sports rather than vice versa. Third, self-reported sports activity information likely suffers from misreporting (Ferrari, Friedenreich, and Matthews 2007), which biases the correlational estimates towards zero (Wooldridge 2002).

Randomized control trials can solve these issues by assigning individuals to either a treatment group with an intervention programme or a control group. Field experiments on physical activity and health-related outcomes have been conducted with Texaco employees (Baun, Bemacki, and Tsai 1986), employees of insurance companies (Shephard 1992), Bank of America retirees (Leigh et al. 1992) and Johnson and Johnson employees (Ozminkowski et al. 2002). But because samples in these studies are small and derived from very specific settings, results from these

CONTACT Stephan Nüesch nueesch@wwu.de University of Münster, Georgskommende 26, D-48143 Münster, Germany

experimental studies cannot be generalized to the rest of the population (Sari 2009).

In this study, we use nationally representative survey data to provide evidence with high external validity. The common issues of correlational studies are addressed by using an instrumental variable approach. More specifically, we employ geographic coordinates of individual home addresses and units of sports facilities and use the local density of sports facilities to predict individual sports activity. The reason behind this strategy is that living close to sports facilities implies easier access to sports infrastructure (Huang and Humphreys 2012) and reduces the 'costs' of doing sports. Both monetary costs (in terms of transportation costs) and time costs (for travelling) indicate a positive relation between short distances to sports facilities and sports activity (Felfe, Lechner, and Steinmayr 2011).

Indeed, we find that our instrument has high power in predicting individual sports activity. After controlling for neighbourhood effects (such as average community-level income and age) that may influence residential sorting and health, the local density of sports facilities is likely to be uncorrelated to unobservable factors affecting individual health (exclusion restriction). Our identification strategy is related to the work of Huang and Humphreys (2012) and Ruseski et al. (2014), who use proximity to sports facilities to identify the effect of sports activity on happiness, and Bowblis and McHone (2013) and Grabowski et al. (2013), who use proximity to nursing homes with different ownership to test the influence of nursing ownership on care quality. We are the first to use the density of sports facilities as an instrument in the context of the influence of sports activity on health.

Few studies (Humphreys, McLeod, and Ruseski 2014; Sarma et al. 2014, 2015) address the endogeneity of sports activity when testing the influence of sports activity on health. Sarma et al. (2014) use longitudinal data and fixed-effect models to control for unobserved heterogeneity in the relationship between sports activity and health such as genetic predispositions. However, fixed-effect modelling only accounts for time-constant heterogeneity and does not control for unobserved time-varying confounders such as opportunity cost of time. Sarma et al. (2015) use average local temperatures surrounding the respondent´s interview month as an instrument for sports activity to identify the causal effect of sports activity on health in Canada. While the first-stage effects of temperature on sports activity are plausible and strong, the instrument's validity is questionable due to potential direct effect of temperature on health (Greenough et al. 2001).

Humphreys, McLeod, and Ruseski (2014) use an individual's 'sense of belonging' to the local community as an instrument to examine the link between sports activity and health. While their instrument is a self-reported variable about an individual's perception, our instrument is based on geocoded data from a mandatory and written business census. This also provides a solution to the errors-in-variables problem and the resulting attenuation bias. Moreover, because the sports facilities in our data set are mostly publicly funded by the community, we control for socio-economic variables at the community level. Once controlling for factors such as community-level average income, the influence of individual sports activity on the supply of sports facilities is small. Thus, our instrumental variable (IV) approach also addresses potential issues of reverse causation.

The remainder of this article is structured as follows: In Section II, we outline our data and the empirical strategy; in Section III, we present the results, and Section IV concludes.

II. Data and empirical strategy

In this section we first describe our data sources, then discuss the dependent and independent variables that we investigate in our analysis and finally present and discuss our instrumental variable.

Description of data sources

The data on sports activity, health and socio-economic characteristics were taken from the Swiss Household Panel (SHP). A key advantage of SHP is that the sample includes a stratified random sample of households representing the resident population of Switzerland. Originally, the randomization of the sample was constructed under the guidance of the *Swiss Federal Statistical Office* based on the major statistical regions in Switzerland (for detailed information about the sample design, see Voorpostel et al. (2012)).

While the SHP is available as an annual panel, we only use the tenth wave, collected between September 2008 and February 2009, because for this wave, we could perfectly match information on the local density of sports facilities.[1] After dropping a small number of individuals that did not respond correctly to all of the items of our analysis, the final sample consists of 6,645 out of the original 6,872 individuals included in the SHP survey.

To construct the local density of sports facilities as our instrument, we obtained additional data from the *Swiss Business Census*. The *Business Census* is a mandatory survey of workplaces and businesses in Switzerland and collects data on businesses' economic activity, the number of persons employed and their exact geographic location via 'Swiss grid coordinates'.[2] These coordinates pinpoint the location of a sports facility within a few meters of the building's midpoint and allow us to draw a very precise map of the geographic distribution of sports facilities in Switzerland. Sports facilities include indoor and outdoor facilities, for example, football grounds, athletics grounds, swimming pools, golf courses and so on. For our analysis, we use the 2008 data from the *Swiss Business Census*.

In the standard version of SHP, the most accurate geographic information on an individual's home location is the canton of residence. However, to obtain an accurate link between SHP individuals and sports facilities, we needed more detailed geographic information. We gratefully acknowledge SHP's provision of exact home addresses for each individual in the data set, after we signed a special confidentiality agreement. The provided home addresses included information on the community, zip code, street name and street number.[3] We used the public webpage http://tools.retorte.ch/map/to transform this address data into 'Swiss grid coordinates'. Using these home address coordinates, we are able to pinpoint linear distances between the residence of an individual and all sports facilities obtained from the *Swiss Business Census* with a precision of a few meters.

Health measures

Regarding the health information, we are restricted to the information the SHP survey offers. Following the questions included in the SHP survey, we consider five specific indicators for health problems. Most notably, we include a discrete indicator for overweight, which has been argued to be both a consequence of physical inactivity due to a disrupted energy balance (Katzmarzyk and Janssen 2004) *and* a risk factor for chronic health problems (Dixon 2010) and health care utilization (Cawley and Meyerhoefer 2012). To identify overweight individuals, we converted height and weight data into a discrete measure of overweight via WHO Body Mass Index guidelines (World Health Organization 2000). Other specific indicators available from the SHP survey include regular suffering from sleeping problems, headaches, back problems and weariness. They were obtained from questions of the type: 'During the last 4 weeks, have you suffered from one of the following disorders or health problems?' While respondents were allowed to choose between three categories (not at all, somewhat, very much), we used a binary yes/no coding that only treats serious incidences (i.e. very much) as a specific health problem.

The SHP survey also includes two measures concerning the perceived overall health status of a person: health satisfaction and health impediments in everyday activities. Both variables are measured on an 11-point Likert scale.

Sports activity measure

To identify individual sports activity, we draw on an SHP question from the leisure time section. Respondents were asked: 'How frequently do you

[1]While data on individual sports activity are available for the years 1999, 2001, 2005, 2008, 2010 and 2013, data on sports facilities are available for the years 1998, 2001, 2005, 2008 and 2011. Except in the year 2008, information of the two data sources do not cover the same year. Thus, the only way to use the panel structure of the data is to interpolate the values of sports activity and sports facilities. When doing so, the variation of the number of sports facilities over time is too low to have any statistical power in the first-stage regression (see Table A1 in the Appendix). Thus, an IV-panel model with individual fixed effects is not feasible in our context.

[2]Participation in the survey is compulsory for all targeted workplaces and businesses. However, there is a minimum of 20 h of weekly work for a business unit to be targeted by the survey. Therefore, the data do not include very small sports facilities that do not employ at least one person with an engagement of 50% or more.

[3]SHP was not able to provide the complete address for 40 individuals (either no street name was provided or the provided street name was not identifiable). In these cases, we were not able to obtain exact 'Swiss grid coordinates'. Hence, we were not able to match these individuals with the sports infrastructure data and were forced to exclude them from our sample.

practice an individual or team sport (e.g. fitness, jogging, football, volley ball, tennis)?' Respondents were free to provide any description of their sports activity level but interviewers were directed to help respondents provide a reasonable answer if necessary. Afterwards, interviewers had to assign the responses to five different levels of sports activity: every day, at least once a week, at least once a month, less than once a month, never. Large proportions of the respondents reported doing sports activities at least once a week (58.2%) or not at all (25.3%). Each of the other three categories contained only a small proportion of the respondents: 7.0% reported daily sports activity, and 9.6% reported some occasional sports activity but not every week (at least once a month: 7.2%; less than once a month: 2.4%).

To allow for a straightforward interpretation of the results, we aggregate the different categories into a discrete measure of sports activity 'at least once a week', equalling 1 for individuals who reported to do sports 'every day' or 'at least once a week' and 0 for individuals who reported to do sports 'at least once a month', 'less than once a month' or 'never'.[4]

Instrumental variable: local density of sports facilities

We use the local density of sports facilities as an instrument for sports activity. From the sports management literature, we know that not only individual variables such as age, gender or income influence sports activity but also infrastructure variables such as the availability of sports facilities and/or park areas (Wicker, Breuer, and Pawlowski 2009; Hallmann et al. 2012; Wicker, Hallmann, and Breuer 2013).

A key issue in the construction of the instrument is to identify an appropriate radius up to which sports facilities potentially affect a person's sport activity. We use the number of sports facilities within a radius of 6 miles as an instrument in the main specification because the use of 6 miles as a distance boundary had one of the highest explanatory powers in the first-stage regression (see Table A3 in the Appendix) and is also consistent with an empirical finding by Pawlowski et al. (2009) that people are on average willing to spend a maximum of 28 min to travel to sport facilities. However, the *F*-test of instrument exclusion is above the threshold level of 10 (Staiger and Stock 1997) for all radii between 5 and 10 miles. Moreover, our results are widely robust to the use of alternative distance boundaries to construct the instrument (see Table A4 in the Appendix).[5]

To be valid, IV regressions require that an instrument has to be both powerful in predicting the endogenous variable *and* uncorrelated with the error term in the second-stage regression (Stock and Watson 2003; Murray 2006). If the local density of sports facilities was randomly assigned, we could just predict sports activity using the density of sports facilities and then relate the predicted values to health. But it is very reasonable to believe that individuals self-select into locations based on individual preferences and neighbourhood characteristics that might correlate with the local density of sports facilities and individual health. For example, affluent people are likely to self-select into affluent neighbourhoods whose communities have sufficient finances to build sports facilities. If individual wealth is also correlated with health, this will bias our results.

Therefore, we control for both neighbourhood and individual characteristics in our estimations. Neighbourhood characteristics include the average income, squared average income, the Gini coefficient of income in the community, average age, squared average age and the SD of age in the community. Average community-level income is a good proxy of the financial power of a community, which is positively related to the number of public sports facilities. Average age and SD of age characterize the population structure in a community. The squared terms of average income and average age are included to test for potential non-linearity of the effects. The community-level variables are obtained from the *Swiss Federal Statistical Office*.

Because doctor density is often positively correlated with health care utilization (McGuire 2000) and may also influence perceived health, we control for the doctor density within 6 miles of residence.

[4]The dichotomization avoids any functional form assumptions for different subgroup effects (Lechner 2009). Table A2 in the Appendix of this article shows that our results are unaffected by alternative definitions of 'active' versus 'inactive' individuals.

[5]The only findings that are not consistent for all boundaries between 4 and 8 miles are the effects of sports activity on overweight and sleeping problems. While the coefficients remain negative throughout all specifications, the effects become more significant for the 4- and 5-mile boundaries and insignificant for the 7- and 8-mile boundaries.

Because urban areas with a high population density naturally provide a higher number of sports facilities, but may be less attractive for other reasons (e.g. noise pollution), we also control for the population density within 6 miles of residence.[6] Here again we also include the squared term to control for potential non-linearity.

To account for time-constant regional heterogeneity, we include two groups of regional fixed effects. First, we include dummies for each community typology following the official SHP categorization (e.g. centres, suburban, preurban, touristic, industrial, rural, agricultural). Second, we include dummies for the language regions. Switzerland has three main languages: German (spoken in the north and the centre of Switzerland), French (spoken in the west of Switzerland) and Italian (spoken in the south of Switzerland). The different language areas in Switzerland come along with substantial differences in political and social attitudes (Eugster et al. 2011) and labour market outcomes (Eugster et al. forthcoming) and sports activity (Stamm and Lamprecht 2011).[7]

In addition to the neighbourhood controls, we also include demographic and socio-economic control variables that may reflect individual characteristics that have often been used in studies on residential choice and in studies on health-related outcomes (Winkelmann 2004; Sari 2009; Lee and Waddell 2010; Kim, Pagliara, and Preston 2005). Individual characteristics include age, sex, marital status, education, work status, household earnings[8] and household with children. Because Huang and Humphreys (2012) and Farrell and Shields (2002) show that the effects of age on sports activity are non-linear, we also include the squared terms of age.

III. Results

Summary statistics

Descriptive statistics for all our variables are shown in Table 1. While 35.7% of the individuals are overweight, each of the other four health issues (sleeping problems, headaches, back problems and weariness) is reported to afflict between 8% and 10% of individuals. The average satisfaction with the health status is 7.9 and the average perceived health impediment in everyday activities is 1.9 on an 11-point Likert scale. Thus, the Swiss population is generally satisfied with its health status. The proportion of individuals in our sample who do sports at least once a week is 65.2%, lower than the 76.5% of people found to be 'physically active' in a US sample (Huang and Humphreys 2012), but higher than the around 50% of people found to be 'physically active' in Canadian studies (Sari 2009; Humphreys, McLeod, and Ruseski 2014).

Our instrument (number of sports facilities within 6 miles) has a mean value of 23.4. This indicates that individuals in our sample have on average around 23 sports facilities within 6 miles of their place of residence. The measure has substantial variation as the number of sports facilities ranges from 0 to 106 with a SD of 23.4.

The average community-level household income is 79,086 Swiss Francs (approximately US$76,000 at that time). The community-level Gini coefficient of income is 35.1%. The average community-level age is 40.8 and the community-level SD of age is 22.5. There are 222 doctors within a radius of 6 miles on average. The population density within a radius of 6 miles is 178,135 on average. Most study subjects live in centres and suburban communities. 71.7% live in the German-speaking area, 25.6% in the French-speaking area and 2.7% in the Italian-speaking area.

The average age in the sample is 46.2 years.[9] A little under half the sample is male (44.3%) and a little over half the sample is married (53.8%). Individual education splits into five categories with shares between 10% and 36%. The high share of apprenticeships (36.5%) reflects the importance of occupational training in the Swiss education system. Most of the individuals are employed, namely 69.2%. 25.6% are not in the labour force and only 2.7% are unemployed. In terms of household income, individuals are divided into four income levels

[6]The data on home-address-specific population density is obtained from the 1990, 2000 and 2010 *Swiss Population Census* through extrapolation for the year 2008. The *Population Census* counts all individuals in each and every hectare in Switzerland. We linked this hectare-based population data to the SHP households via 'Swiss grid coordinates'. Population density is measured based on the same distance boundary that is used for our density of sports facilities measure (i.e. 6 miles in the main specification).

[7]We are thankful to an anonymous referee for suggesting to control for regional fixed effects.

[8]Unfortunately, 7.9% of the individuals did not provide valid data for household earnings. In order to keep these observations in the sample, we classify respondents into five different income groups, one of which is 'unknown'.

[9]The average age is higher than the community-level average age because the SHP only considers individuals that are at least 14 years old.

Table 1. Summary statistics.

Variables	Mean	SD	Min.	Max.
Health				
Overweight	0.357	0.479	0	1
Sleeping problems	0.087	0.281	0	1
Headaches	0.078	0.268	0	1
Back problems	0.103	0.304	0	1
Weariness	0.091	0.288	0	1
Health satisfaction (0–10)	7.884	1.797	0	10
Perceived health impediments in everyday activities	1.876	2.500	0	10
Individual sports activity				
Doing sports at least once a week	0.652	0.476	0	1
Instrumental variable				
Local density of sports facilities (within 6 miles)	23.42	23.392	0	106
Regional controls				
Community-level average income (in thousands)	79.09	22.39	40.05	318.05
Community-level income Gini	0.351	0.061	0.243	0.732
Community-level average age	40.78	1.930	30.69	50.99
Community-level SD of age	22.531	0.693	19.25	25.42
Local doctor density (within 6 miles)	222.17	248.128	0	926
Local population density (within 6 miles)	178,135	154.725	1,046	648.816
Community typology: Centres	0.268	0.443	0	1
Community typology: Suburban	0.305	0.461	0	1
Community typology: Wealthy	0.038	0.192	0	1
Community typology: Periurban	0.114	0.318	0	1
Community typology: Touristic	0.023	0.149	0	1
Community typology: Industrial	0.088	0.284	0	1
Community typology: Rural	0.079	0.270	0	1
Community typology: Agricultural	0.084	0.278	0	1
Country region: German	0.717	0.451	0	1
Country region: French	0.256	0.436	0	1
Country region: Italian	0.027	0.163	0	1
Demographics and socio-economic controls				
Age	46.19	18.373	14	96
Male	0.443	0.497	0	1
Married	0.538	0.499	0	1
Education: Compulsory	0.229	0.420	0	1
Education: Apprenticeship	0.365	0.481	0	1
Education: University entrance diploma	0.099	0.299	0	1
Education: Post-apprenticeship diploma	0.161	0.367	0	1
Education: University degree	0.147	0.354	0	1
Work status: Employed	0.692	0.462	0	1
Work status: Unemployed	0.015	0.121	0	1
Work status: Not in labour force	0.294	0.455	0	1
Household income: < 50,001 Swiss Francs	0.110	0.314	0	1
Household income: 50,001–100,000 Swiss Francs	0.326	0.469	0	1
Household income: 100,001–150,000 Swiss Francs	0.279	0.448	0	1
Household income: > 150,000 Swiss Francs	0.206	0.404	0	1
Household income: Unknown	0.079	0.269	0	1
Household with children	0.372	0.483	0	1
Number of households	4,049			
Number of individuals	6,645			

Data on sports facilities and doctors are drawn from the 2008 *Swiss Business Census*. Data on population density are interpolated from the 1990, 2000 and 2010 *Swiss Population Census*. Community-level average income, income Gini age and SD of age are from the *Swiss Federal Statistical Office*. All other variables are directly drawn from the 2008 SHP survey.

and a non-response group, with most individuals (32.6%) living in households with an income between 50,000 and 100,000 Swiss Francs (reflecting roughly US $48,000–US$96,000 based on the currency rate of 2008). 37.2% live in a household with children.

First-stage results

Table 2 shows three specifications of the first-stage results: the first model predicts individual sports activity using the number of sports facilities within 6 miles of residence only. The second model additionally includes neighbourhood controls such as the financial power of the community, the age structure of the community and the population density. The third model also includes individual demographics and socio-economic controls.

The number of sports facilities within 6 miles of residence significantly increases sports activity in

Table 2. First-stage results: probability of an individual to do sports at least once a week.

Variables	LPM (M1)	LPM (M2)	LPM (M3)
Instrumental variable			
Local density of sports facilities (within 6 miles)	0.0006**	0.0054***	0.0053***
–	(0.0002)	(0.0012)	(0.0011)
Neighbourhood controls			
Community-level average income (in thousands)	–	0.002*	0.002*
		(0.001)	(0.001)
Squared community-level average income (in thousands)		−0.000	−0.000
		(0.000)	(0.000)
Community-level income Gini		−0.236	−0.283
		(0.223)	(0.245)
Community-level average age	–	−0.001	−0.026
		(0.076)	(0.076)
Squared community-level average age		0.000	0.000
		(0.001)	(0.001)
Community-level SD of age	–	−0.029**	−0.025**
		(0.013)	(0.012)
Local doctor density (within 6 miles)	–	0.0002	0.0002*
		(0.0001)	(0.0001)
Local population density/1000 (within 6 miles)	–	−0.001***	−0.001***
		(0.0002)	(0.0003)
Squared local population density/1000 (within 6 miles)		−0.000	−0.000
		(0.000)	(0.000)
Community typology dummies	–	Yes	Yes
Country region dummies	–	Yes	Yes
Demographics and socio-economic controls			
Age	–	–	−0.002
			(0.002)
Squared age	–	–	−0.000
			(0.000)
Male	–	–	−0.025**
			(0.012)
Married	–	–	−0.012
			(0.015)
Household with children	–	–	−0.005
			(0.014)
Education dummies	–	–	Yes
Work status dummies	–	–	Yes
Household income dummies	–	–	Yes
Observations	6,645	6,645	6,645
F-test of excluded instrument	5.15	21.32	21.52

Models M1, M2 and M3 display OLS estimates with robust SEs in parentheses. The dependent variable is a dummy variable that takes a value of 1 for individuals that do sports at least once a week and a value of 0 otherwise. All estimations also include a constant (not reported). Robust SEs are given in parentheses.
*, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

all three models. However, the estimates and the power of the instrument substantially change across the three models. While the number of sports facilities increase the likelihood of doing sports at least once a week by 0.06% in Model 1, the estimated coefficient increases to 0.54% in Model 2 (only neighbourhood controls) and to 0.53% in Model 3 (neighbourhood and individual controls). The *F*-statistic for excluding the number of sports facilities in the regression is 5.15 in Model 1, 21.32 in Model 2 and 21.52 in Model 3, indicating that the *F*-statistic exceeds the critical value of 10 suggested by Staiger and Stock (1997) only when controlling for neighbourhood characteristics. If we do not control for neighbourhood characteristics, the statistical power of our instrument is weak and its effect is negatively biased due to omitted variables. The population density, for example, is such a confounder because it correlates positively with the number of sports facilities and significantly decreases individual sports activity (see Model 3 in Table 2).

While some of the individual demographic and socio-economic characteristics significantly affect sports activity (e.g. gender, educational background (not shown)), they hardly change the effect of the instrument when including them in Model 3. Interestingly, we observe males to be less active than women in the Swiss context, whereas an earlier study from England showed the opposite relationship between gender and sports activity (Farrell and Shields 2002).

Table 3. Regression results: Effects of sports activity on health.

	Non-IV (1)	IV (2)
Panel A: measures of specific health issues (yes/no)		
Overweight	−0.103***	−0.316†
	(0.012)	(0.222)
Sleeping problems	−0.030***	−0.187†
	(0.008)	(0.137)
Headaches	−0.027***	−0.073
	(0.007)	(0.130)
Back problems	−0.047***	0.021
	(0.008)	(0.152)
Weariness	−0.061***	0.076
	(0.008)	(0.141)
Panel B: measures of perceived overall health (0–10)		
Health satisfaction	0.428***	0.002
	(0.048)	(0.857)
Perceived health impediments in everyday activities	−0.630***	0.784
	(0.067)	(1.241)
Observations	6,645	6,645
F-test of excluded instrument	–	21.52

Non-IV estimates for sports activity with robust SEs in parentheses are displayed in column (1). 2SLS estimates for sports activity with robust SEs in parentheses are displayed in column (2). All models control for average local income, squared average local income, average local income Gini, average local age, squared average local age, local SD of age, local density of doctors, local population density, squared local population density, local community typology, country region, age, squared age, gender, marital status, education, work status, household income and household with children. All estimations also include a constant (not reported).
*, ** and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.
†denotes *p*-values between 0.10 and 0.20.

Regression results

We estimate the effect of sports activity on health with two-stage least squares (2SLS) regressions that use the number of sports facilities within 6 miles as an instrument and both neighbourhood and individual characteristics as control variables.[10]

Table 3 presents the estimates of the effect of sports activity on specific health issues such as overweight, sleeping problems, headaches, back problems and weariness in Panel A and on measures of perceived overall health in Panel B. For the purpose of comparison, column (1) shows the estimated coefficients of sports activity from non-IV models and column (2) shows the estimated coefficients of sports activity using the IV approach.

The results from the non-IV models show that sports activity significantly reduces overweight, sleeping problems, headaches, back problems, weariness and perceived health impediments in everyday activities and significantly increases health satisfaction (see Table 3, column 1). Comparing the non-IV results with the IV results in column (2), we find that the effects of sports activity on headaches, back problems, weariness, health satisfaction and perceived health impediments in everyday activities are no longer statistically significant due to the high SEs. The effects of sports activity on overweight and sleeping problems are higher in magnitude but also lower in statistical significance in the IV models than in the non-IV models. The *p*-values in the IV models are between 0.1 and 0.2, which can be considered as marginally significant given that inference is based on two-sided tests.

Reverse causation provides a plausible explanation for the significantly negative effects in the non-IV models: people with pre-determined headaches, back problems and weariness do not participate in sports, causing a negative correlation between sports activity and these specific health issues, with the direction of causality running from health issues to sports activity rather than vice versa.

IV. Discussion and conclusion

This article uses a representative sample of the Swiss population and geocoded data on sports facilities, sports activity and health to estimate the causal effect

[10]As we have binary outcome variables and a binary endogenous variable, the question arises whether discrete choice models or linear probability models (LPM) should be applied. Angrist (2001) suggests to use LPM whenever the underlying error distribution is unknown as is the case in our models. If the first and second stages are estimated by non-linear Probit models, the effect of sport activity on overweight becomes statistically significant at the 5% level. The other results do not change in any significant way (see Table A5 in the Appendix).

of sports activity on health. Unlike previous correlational studies, this article uses the number of sports facilities within 6 miles of the individual's place of residence as an instrument for sports activity. We show that the local density of sports facilities is powerful in explaining individual sports activity even after controlling for numerous community-level determinants and individual controls. Whereas the non-IV models show that sports activity significantly improves health, the IV results on health are mostly statistically insignificant due to high SEs. We only find marginally significant effects on overweight and sleeping problems. Thus, our IV results question the idea of providing sports facilities to encourage sports activity as a means of reducing health problems. The OLS results are likely to be negatively biased due to reverse causation. Individuals with health problems such as back problems, headaches and weariness tend to reduce sports activity (rather than vice versa).

Of course, one must always be cautious with regard to the exogeneity condition of IV models because it is impossible to prove the null hypothesis of no correlation between instruments and the (unobserved) error term in the second stage. We argue that the exogeneity condition of our instrument is plausible for three reasons: first, because sports facilities are provided by communities and not by individuals, reverse causality can be excluded. Second, community-level variables (e.g. average local income, local age distribution, population density) control for potential confounders that are likely to be correlated with both the number of sports facilities and health. Third, individuals typically self-select into neighbourhoods based on housing prices, housing quality, commuter distance, school quality and/or environmental factors such as noise. Non-work-related travel distances such as geographic proximity to sports facilities are found to play only a negligible role in selecting a neighbourhood to live in (Lee et al. 2010; Chatman 2009).

Unfortunately, we only have self-reported data on sports activity and health. Although the IV method helps to correct for reporting errors in the sports activity measure, it does not eliminate reporting errors in the outcome variables. Another limitation of our article is that we have data on sports activity only at the consolidated level for all different types of sports and at an ordinal level for five different frequencies of doing sports. Ideally, we would have data on subgroups of sports (e.g. football, tennis, jogging) and hours of weekly sports activity, which would allow us to estimate the marginal effects of an additional hour of doing different types of sports.

Despite these limitations, this article makes an important contribution by providing IV estimates on the effects of sports activity on health based on a representative sample. Although the instrument is powerful in explaining sports activity, the second-stage effects on health are mostly statistically insignificant due to the high SEs of the IV estimates.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Angrist, J. D. 2001. "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors." *Journal of Business Economic Statistics* 19 (1): 2–16. doi:10.1198/07350010152472571.

Baun, W. B., E. J. Bemacki, and S. P. Tsai. 1986. "A Preliminary Investigation: Effect of a Corporate Fitness Program on Absenteeism and Health Care Cost." *Journal of Occupational and Environmental Medicine* 28 (1): 18–22. doi:10.1097/00043764-198601000-00007.

Bowblis, J. R., and H. S. McHone. 2013. "An Instrumental Variables Approach to Post-Acute Care Nursing Home Quality: Is There a Dime's Worth of Evidence that Continuing Care Retirement Communities Provide Higher Quality?" *Journal of Health Economics* 32 (5): 980–996. doi:10.1016/j.jhealeco.2013.06.007.

Cawley, J., and C. Meyerhoefer. 2012. "The Medical Care Costs of Obesity: An Instrumental Variables Approach." *Journal of Health Economics* 31 (1): 219–230. doi:10.1016/j.jhealeco.2011.10.003.

Chatman, D. G. 2009. "Residential Choice, the Built Environment, and Nonwork Travel: Evidence Using New Data and Methods." *Environment and Planning. A* 41 (5): 1072–1089. doi:10.1068/a4114.

Dixon, J. B. 2010. "The Effect of Obesity on Health Outcomes." *Molecular and Cellular Endocrinology* 316 (2): 104–108. doi:10.1016/j.mce.2009.07.008.

Eugster, B., R. Lalive, A. Steinhauer, and J. Zweimüller. 2011. "The Demand for Social Insurance. Does Culture Matter?" *Economic Journal* 121: F413–F448. doi:10.1111/j.1468-0297.2011.02479.x.

Eugster, B., R. Lalive, A. Steinhauer, and J. Zweimüller. forthcoming. "Culture, Work Attitudes and Job Search: Evidence from the Swiss Language Border." *Journal of the European Economic Association*.

Farrell, L., and M. A. Shields. 2002. "Investigating the Economic and Demographic Determinants of Sporting Participation in England." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165 (2): 335–348. doi:10.1111/rssa.2002.165.issue-2.

Felfe, C., M. Lechner, and A. Steinmayr. 2011. "Sports and Child Development." IZA Discussion Paper No. 6105, Bonn: Institute for the Study of Labor.

Ferrari, P., C. Friedenreich, and C. E. Matthews. 2007. "The Role of Measurement Error in Estimating Levels of Physical Activity." *American Journal of Epidemiology* 166 (7): 832–840. doi:10.1093/aje/kwm148.

Grabowski, D. C., Z. Feng, R. Hirth, M. Rahman, and V. Mor. 2013. "Effect of Nursing Home Ownership on the Quality of Post-Acute Care: An Instrumental Variables Approach." *Journal of Health Economics* 32 (1): 12–21. doi:10.1016/j.jhealeco.2012.08.007.

Greenough, G., M. McGeehin, S. M. Bernard, J. Trtanj, J. Riad, and D. Engelberg. 2001. "The Potential Impacts of Climate Variability and Change on Health Impacts of Extreme Weather Events in the United States." *Environmental Health Perspectives* 109: 191–198. doi:10.2307/3435009.

Hallmann, K., P. Wicker, C. Breuer, and L. Schönherr. 2012. "Understanding the Importance of Sport Infrastructure for Participation in Different Sports–Findings from Multi-Level Modeling." *European Sport Management Quarterly* 12 (5): 525–544. doi:10.1080/16184742.2012.687756.

Heath, G. W., D. C. Parra, O. L. Sarmiento, L. B. Andersen, N. Owen, S. Goenka, F. Montes, and R. C. Brownson. 2012. "Evidence-Based Intervention in Physical Activity: Lessons from around the World." *The Lancet* 380 (9838): 272–281. doi:10.1016/S0140-6736(12)60816-2.

Huang, H., and B. R. Humphreys. 2012. "Sports Participation and Happiness: Evidence from US Microdata." *Journal of Economic Psychology* 33 (4): 776–793. doi:10.1016/j.joep.2012.02.007.

Humphreys, B. R., L. McLeod, and J. E. Ruseski. 2014. "Physical Activity and Health Outcomes: Evidence from Canada." *Health Economics* 23 (1): 33–54. doi:10.1002/hec.v23.1.

Katzmarzyk, P. T., N. Gledhill, and R. J. Shephard. 2000. "The Economic Burden of Physical Inactivity in Canada." *Canadian Medical Association Journal* 163 (11): 1435–1440.

Katzmarzyk, P. T., and I. Janssen. 2004. "The Economic Costs Associated with Physical Inactivity and Obesity in Canada: An Update." *Canadian Journal of Applied Physiology* 29 (1): 90–115. doi:10.1139/h04-008.

Kim, J. H., F. Pagliara, and J. Preston. 2005. "The Intention to Move and Residential Location Choice Behaviour." *Urban Studies* 42 (9): 1621–1636. doi:10.1080/00420980500185611.

Lechner, M. 2009. "Long-Run Labour Market and Health Effects of Individual Sports Activities." *Journal of Health Economics* 28 (4): 839–854. doi:10.1016/j.jhealeco.2009.05.003.

Lee, B. H., and P. Waddell. 2010. "Residential Mobility and Location Choice: A Nested Logit Model with Sampling of Alternatives." *Transportation* 37 (4): 587–601. doi:10.1007/s11116-010-9270-4.

Lee, B. H., P. Waddell, L. Wang, and R. M. Pendyala. 2010. "Reexamining the Influence of Work and Nonwork Accessibility on Residential Location Choices with a Microanalytic Framework." *Environment and Planning. A* 42 (4): 913–930. doi:10.1068/a4291.

Leigh, J. P., N. Richardson, R. Beck, C. Kerr, H. Harrington, C. L. Parcell, and J. F. Fries. 1992. "Randomized Controlled Study of a Retiree Health Promotion Program: The Bank of America Study." *Archives of Internal Medicine* 152 (6): 1201–1206. doi:10.1001/archinte.1992.00400180067010.

Manning, W. G., E. B. Keeler, J. P. Newhouse, E. M. Sloss, and J. Wasserman. 1991. *The Costs of Poor Health Habits*. Cambridge, MA: Harvard University Press.

McGuire, T. G. 2000. "Physician Agency." In *Handbook of Health Economics*. Vol. 1, edited by A. Culyer and J. Newhouse, 461–536. Amsterdam: Elsevier.

Murray, M. P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives* 20 (4): 111–132. doi:10.1257/jep.20.4.111.

Ozminkowski, R. J., D. Ling, R. Z. Goetzel, J. A. Bruno, K. R. Rutter, F. Isaac, and S. Wang. 2002. "Long-Term Impact of Johnson & Johnson's Health & Wellness Program on Health Care Utilization and Expenditures." *Journal of Occupational and Environmental Medicine* 44 (1): 21–29. doi:10.1097/00043764-200201000-00005.

Pawlowski, T., C. Breuer, P. Wicker, and S. Poupaux. 2009. "Travel Time Spending Behaviour in Recreational Sports: An Econometric Approach with Management Implications." *European Sport Management Quarterly* 9 (3): 215–242. doi:10.1080/16184740903023971.

Ruseski, J. E., B. R. Humphreys, K. Hallmann, P. Wicker, and C. Breuer. 2014. "Sport Participation and Subjective Well-Being: Instrumental Variable Results from German Survey Data." *Journal of Physical Activity and Health* 11 (2): 396–403. doi:10.1123/jpah.2012-0001.

Sari, N. 2009. "Physical Inactivity and Its Impact on Healthcare Utilization." *Health Economics* 18 (8): 885–901. doi:10.1002/hec.v18:8.

Sarma, S., R. A. Devlin, J. Gilliland, M. K. Campbell, and G. S. Zaric. 2015. "The Effect of Leisure-Time Physical Activity on Obesity, Diabetes, High BP and Heart Disease among Canadians: Evidence from 2000/2001 to 2005/2006." *Health Economics* 24 (12): 1531–1547. doi:10.1002/hec.v24.12.

Sarma, S., G. S. Zaric, M. K. Campbell, and J. Gilliland. 2014. "The Effect of Physical Activity on Adult Obesity: Evidence from the Canadian NPHS Panel." *Economics & Human Biology* 14: 1–21. doi:10.1016/j.ehb.2014.03.002.

Shephard, R. J. 1992. "Twelve Years Experience of a Fitness Program for the Salaried Employees of a Toronto Life

Assurance Company." *American Journal of Health Promotion* 6 (4): 292–301. doi:10.4278/0890-1171-6.4.292.

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586. doi:10.2307/2171753.

Stamm, H., and M. Lamprecht. 2011. "Swiss Sports Participation in an International Perspective." *European Journal of Sport and Society* 8: 15–29. doi:10.1080/16138171.2011.11687867.

Stock, J. H., and M. W. Watson. 2003. *Introduction to Econometrics*. Vol. 104. Boston: Addison Wesley.

Voorpostel, M., R. Tillmann, F. Lebert, U. Kuhn, O. Lipps, V.-A. Ryser, and B. Wernli. 2012. "Swiss Household Panel Userguide (1999–2011), Wave 13, October 2012." Lausanne: FORS.

Warburton, D. E., C. W. Nicol, and S. S. Bredin. 2006. "Health Benefits of Physical Activity: The Evidence." *Canadian Medical Association Journal* 174 (6): 801–809. doi:10.1503/cmaj.051351.

Wicker, P., C. Breuer, and T. Pawlowski. 2009. "Promoting Sport for All to Age-Specific Target Groups: The Impact of Sport Infrastructure." *European Sport Management Quarterly* 9 (2): 103–118. doi:10.1080/16184740802571377.

Wicker, P., K. Hallmann, and C. Breuer. 2013. "Analyzing the Impact of Sport Infrastructure on Sport Participation Using Geo-Coded Data: Evidence from Multi-Level Models." *Sport Management Review* 16 (1): 54–67. doi:10.1016/j.smr.2012.05.001.

Winkelmann, R. 2004. "Health Care Reform and the Number of Doctor Visits: An Econometric Analysis." *Journal of Applied Econometrics* 19 (4): 455–472. doi:10.1002/jae.764.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

World Health Organization. 2000. *Obesity: Preventing and Managing the Global Epidemic (No. 894)*. Geneva: World Health Organization.

World Health Organization. 2010. *Global Recommendations on Physical Activity for Health*. Geneva: World Health Organization.

Appendix

Table A1. First-stage results: individual fixed-effect model with panel data (1999–2008).

	Linear probability model (LPM)
Variables	FE1
Number of sports facilities within 6 miles	0.0001
	(0.0007)
Demographic and socio-economic controls	Yes
Regional controls	Yes
Individual fixed effects	Yes
F-test of excluded instrument in the first stage	0.04
Number of observations	65,909
Number of individuals	14,574

The estimation included 10 years of individual panel data from 1999 to 2008. The data on sports facilities are drawn and interpolated from the 1998, 2001, 2005 and 2008 *Swiss Business Census*. Data on population density are drawn and interpolated from the 1990, 2000 and 2010 *Swiss Population Census*. All other variables are directly drawn from the SHP surveys 1999–2008. In Model FE1, OLS estimates for M3 with individual fixed effects are displayed (not included are the regional controls for local average income, squared local average income, local average income Gini, local average age, squared local average age and SD of average age, because these measures are not available to us for other years than 2008). The dependent variable is a dummy variable that takes a value of 1 for individuals that do sports at least once a week and a value of 0 otherwise. The estimation also included a constant (not reported).
*, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table A2. Effects of sports activity on health: Alternative cut-offs for being 'active'.

	IV (1)	IV (2)	IV (3)
	(at least weekly (orig.))	(at least monthly)	(not never)
Panel A: measures of specific health issues (yes/no)			
Overweight	−0.316†	−0.386†	−0.413†
	(0.222)	(0.274)	(0.296)
Sleeping problems	−0.187†	−0.229†	−0.245†
	(0.137)	(0.169)	(0.181)
Headaches	−0.073	−0.090	−0.096
	(0.130)	(0.160)	(0.171)
Back problems	0.021	0.027	0.028
	(0.152)	(0.186)	(0.199)
Weariness	0.076	0.093	0.100
	(0.141)	(0.174)	(0.186)
Panel B: measures of perceived overall health (0–10)			
Health satisfaction	0.002	0.003	0.003
	(0.857)	(1.048)	(1.121)
Perceived health impediments in everyday activities	0.784	0.960	1.025
	(1.241)	(1.526)	(1.635)
Observations	6,645	6,645	6,645
F-test of excl. instrument in the first stage	21.52	16.74	15.83

Non-IV estimates for sports activity with robust SEs in parentheses are displayed in column (1). 2SLS estimates for sports activity with robust SEs in parentheses are displayed in column (2). All models control for average local income, squared average local income, average local income Gini, average local age, squared average local age, local SD of age, local density of doctors, local population density, squared local population density, local community typology, country region, age, squared age, gender, marital status, education, work status, household income and household with children. All estimations also include a constant (not reported).
*, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.
†denotes *p*-values between 0.10 and 0.20.

Table A3. Comparison of different measures of local density of sports facilities.

Distance boundaries	Mean	SD	First-stage *F*-test of excluded instrument
Sports facilities within 1 mile	1.67	2.29	F = 0.78
Sports facilities within 2 miles	5.00	6.40	F = 2.02
Sports facilities within 3 miles	9.11	10.85	F = 8.08
Sports facilities within 4 miles	13.69	15.40	F = 8.79
Sports facilities within 5 miles	18.44	19.52	F = 12.29
Sports facilities within 6 miles	23.42	23.39	F = 21.52
Sports facilities within 7 miles	28.67	27.60	F = 17.03
Sports facilities within 8 miles	34.39	31.91	F = 22.23
Sports facilities within 9 miles	40.55	36.34	F = 23.21
Sports facilities within 10 miles	47.19	40.74	F = 15.97
Number of households	4,049		
Number of individuals	6,645		

Data on sports facilities are drawn from the 2008 *Swiss Business Census* and are linked to the home addresses of SHP individuals. The *F*-test of excluded instrument reflects the power of the number of sports facilities. All models control for average local income, squared average local income, average local income Gini, average local age, squared average local age, local SD of age, local density of doctors, local population density, squared local population density, local community typology, country region, age, squared age, gender, marital status, education, work status, household income and household with children.

Table A4. Effects of sports activity on health: Alternative estimation approach.

	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)
	(4 miles)	(5 miles)	(6 miles)	(7 miles)	(8 miles)
Panel A: measures of specific health issues (yes/no)					
Overweight	−0.576†	−0.530*	−0.316†	−0.314	−0.245
	(0.233)	(0.307)	(0.222)	(0.250)	(0.211)
Sleeping problems	−0.445*	−0.379*	−0.187†	−0.065	−0.023
	(0.256)	(0.202)	(0.137)	(0.148)	(0.125)
Headaches	−0.223	−0.065	−0.073	−0.114	−0.031
	(0.206)	(0.170)	(0.130)	(0.148)	(0.131)
Back problems	−0.076	−0.146	0.021	0.153	0.107
	(0.227)	(0.196)	(0.152)	(0.177)	(0.149)
Weariness	−0.059	−0.034	0.076	0.139	0.098
	(0.227)	(0.187)	(0.141)	(0.159)	(0.138)
Panel B: measures of perceived overall health (0–10)					
Health satisfaction	−0.894	0.282	0.002	−0.826	−0.541
	(1.390)	(1.128)	(0.857)	(0.985)	(0.834)
Perceived health impediments in everyday activities	0.456	−0.127	0.784	1.526	1.149
	(1.846)	(1.571)	(1.241)	(1.437)	(1.206)
Observations	6,645	6,645	6,645	6,645	6,645
Number of sport facilities within distance boundary	13.69	18.44	23.42	28.67	34.39
Instrument coefficient in the first stage	0.0040***	0.0043***	0.0053***	0.0045***	0.0049***
F-test of excl. instrument in the first stage	8.79	12.29	21.52	17.03	22.23

2SLS estimates for sports activity are displayed with robust SEs in parentheses. Column (1)–(5) refer to different distance boundaries for density of sports facilities. All models control for average local income, squared average local income, average local income Gini, average local age, squared average local age, local SD of age, local density of doctors, local population density, squared local population density, local community typology, country region, age, squared age, gender, marital status, education, work status, household income and household with children. All estimations also include a constant (not reported).

*, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

†denotes *p*-values between 0.10 and 0.20.

Table A5. Effects of sports activity on health: Alternative estimation approach.

	IV (First-S: LPM)	IV (First-S: Probit)	
	(1) (LPM/OLS (orig.))	(2) (LPM/OLS)	(3) (Probit)
Panel A: measures of specific health issues (yes/no)			
Overweight	−0.316†	−0.495**	−1.252** [−0.457]
	(0.222)	(0.211)	(0.623)
Sleeping problems	−0.187†	−0.206†	−1.192† [−0.177]
	(0.137)	(0.127)	(0.816)
Headaches	−0.073	−0.074	−0.686 [−0.089]
	(0.130)	(0.119)	(0.904)
Back problems	0.021	−0.019	−0.107 [−0.018]
	(0.152)	(0.147)	(0.807)
Weariness	0.076	0.072	0.376 [0.814]
	(0.141)	(0.133)	(0.058)
Panel B: measures of perceived overall health (0–10)			
Health satisfaction	0.002	0.650	-
	(0.857)	(0.836)	
Perceived health impediments in everyday activities	0.784	0.227	-
	(1.241)	(0.846)	
Observations	6,645	6,645	6,645

In column (1), the predicted values of sports activity are obtained from a linear probability first-stage regression. In columns (2) and (3), the predicted values of sports activity are obtained from a Probit first-stage regression. In columns (1) and (2), LPM estimation results are displayed in Panel A and OLS estimation results are displayed in Panel B. In column (3), Probit estimates are displayed (with marginal effects at the mean in brackets) for Panel A. All models control for average local income, squared average local income, average local income Gini, average local age, squared average local age, local SD of age, local density of doctors, local population density, squared local population density, local community typology, country region, age, squared age, gender, marital status, education, work status, household income and household with children.

*, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

†denotes p-values between 0.10 and 0.20.