Check for
updates

Estimating the impact of social isolation on subjective health in Europe

Emiliano Sironi[1,2] ⬤ · **Amelie Nadine Wolff[1]**

Abstract
We investigate the relationship between social isolation and subjective health, considering that this relationship is potentially affected by endogeneity due to the presence of self-reported measures. Thus, if an increase in social isolation may impact the perception on health, alternative paths of causality may also be hypothesized. Using data from round 7 of the European Social Survey, we estimate an instrumental variable model in which isolation is explained as being a member of an ethnic minority and having experienced some serious family conflicts in the past. Our results confirm that changes in social isolation influence subjective general health. In particular, greater isolation produces a strong and significant deterioration of the perceived health status. With respect to the literature on social isolation and health, we try to advance it by supporting a path of causality running from social isolation to subjective health.

Keywords Social isolation · Subjective health · Instrumental variables

1 Introduction

Social isolation is defined in literature as a condition of deterioration of human relationships both from a quantitative and a qualitative point of view. The discussion concerning the role of social isolation is increasing in centrality nowadays, focusing on the context of industrialized societies, in which individuals experience new patterns in intergenerational ties, social mobility and living arrangements due also to dual career families, increased instability in unions and a decline in large families (Holt-Lunstad et al. 2010).

Recent literature has investigated deeply the consequences of social isolation, mainly documenting the relationship between a deprivation in social relationships and health (Holt-Lunstad et al. 2010; Steptoe et al. 2013; House et al. 1988). In more detail, the

✉ Emiliano Sironi
emiliano.sironi@unicatt.it

Amelie Nadine Wolff
wo.nadinewo@gmail.com

1   Center for Applied Statistics in Business and Economics, Università Cattolica del Sacro Cuore, Milan, Italy

2   Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

literature has underlined that low levels of social relationships have a negative impact on health conditions. The mechanism through which a deprivation in social contacts undermines health is well summarized in two theoretical streams: the "buffering" (Cohen et al. 2000) and the "direct effect" models. The "buffering" model states that social support is beneficial when individuals are exposed to the pathogenic effect of stressful events, by moderating the negative impact of those stressors.

On the other hand, the "direct effect" model suggests (Cohen and Wills 1985) that good social relationships have positive effects irrespective of whether individuals are under the effect of health stressors or not.

Empirical evidence (Cohen and Wills 1985) supports both the illustrated points of view, confirming in any case the positive impact of social networks on mental and physical health.

While this field of research has plenty of contributions which investigate and support (Ong et al. 2016; Cornwell and Waite 2009) the impact of social isolation (and/or loneliness) on health and, in particular, on the risk of death in the older cohorts, less attention has been devoted to the relationship between human relationships and health in large samples of adults of all ages.

In this paper, we consider a large representative dataset of more than 40,000 individuals aged sixteen and older and coming from 21 countries in Europe, with the aim of establishing a relationship between social isolation and health.

Data include self-reported responses as a suitable measure of health in a cross-sectional setting. Subjective health, being a self-reported measure of health, is commonly used in national representative surveys with the objective of measuring a health condition, but the literature shows that self-reported measures of health are in certain cases more subject to measurement errors if compared with more reliable objective measures (Butler et al. 1987; Johnston et al. 2009).

Nevertheless, subjective measures of health are widely used in the literature when objective measure of health are not available. In addition, self-reported scales are more reliable when individuals are asked to report on their current health and not over any long period of time (Bourne 2009), as in the case of our data.

Using data from the European Social Survey, we aim to assess whether social isolation increases or decreases subjective health. Due to the potential presence of measurement errors which may be present both in subjective measures of well-being and in the perceived component of the social isolation index adopted in the paper, the relationship between the two variables may be affected by endogeneity. In econometric practice, endogeneity may arise when we are in the presence of possible unobservable factors that may influence both the main predictor (social isolation) and the outcome (subjective health). In this framework, the presence of measurement errors in self-reported indices of social isolation and subjective well-being represent one of the most common sources of bias in estimates.

An instrumental variable approach, through the identification of exogenous variables that are strongly correlated with the main predictor (social isolation) but presumably not with the error term of the equation that explains subjective well-being as a function of social isolation, may protect estimates from the bias produced by endogeneity.

After this brief introduction, the structure of the paper is organized as follows: the section "Data and Methods" describes the peculiarity of the dataset and the statistical methodology implemented to evaluate the effect of social isolation on subjective well-being through instrumental variable estimation after having analytically defined the subjective well-being and provided full details about the construction of the social isolation index used in the analysis. The section "Results and Discussion" presents the outcome of the

regression analysis and discusses both the main and the corollary results, while the final section concludes.

2 Data and methods

2.1 Data

To analyze the relationship between social isolation and subjective general health, we employ data from round 7 of the ESS (2014). The ESS is an academically driven cross-national survey which issues a multidimensional questionnaire across several European countries every two years. In the seventh round, there were 21 countries surveyed: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Lithuania, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, Switzerland and the United Kingdom. We decided to choose this particular round not only because it is one of the most recent available, but also because it is the only one containing a specific module on social inequalities in health. Thus, it includes health measurements which are crucial for our analysis.

For this particular round the total sample consists of 40,185 individuals aged between 16 and older (Mean = 49.34; SE = 0.09). The analysis focuses on the individuals that answered all questions on subjective general health and social isolation, thus, 39,122 individuals.

2.2 Methods

The main model we use to estimate the impact of social isolation on the health status dummy variable is a probit regression with endogenous regressors specified as follows:

$$Y_i = \Pr\left(S_i = 1 | \begin{pmatrix} X_i \\ W_i \end{pmatrix}\right) = \phi\left(\beta_0 + \beta_1 X_i + \underline{\beta_2'} W_i + \varepsilon_i\right) \tag{1}$$

$$\phi^{-1}(Y_i) = \beta_0 + \beta_1 X_i + \underline{\beta_2'} W_i + \varepsilon_i \tag{2}$$

where S_i is a binary variable represents subjective general health for the individual i; Y_i, the dependent variable, represents the probability that the chosen health-level variable assumes value 1 for the individual i; X_i is the main explanatory variable, which is a measure of social isolation for the individual i; W_i is a vector consisting of all the other exogenous regressors, that is, all the other variables that have an influence on health, according to the literature. Finally, ε_i represents the error term and Φ is the cumulative distribution function (CDF) of the standard normal distribution. Since, as mentioned before, this relationship is likely to be affected by endogeneity problems, that is, social isolation is potentially correlated with the error term, we argue that an IV-probit regression model (Newey and West, 1987) would lead to more reliable estimates than a standard probit regression. Using an IV-probit estimation means implementing an auxiliary ordinary linear regression, where X_i is a function of the exogenous control variables W_i and of instruments, i.e. a set of predictors collected in the vector Z_i that have the requirement of not being correlated with the error component of the main equation. The auxiliary equation is specified in the following way:

$$X_i = \alpha_0 + \underline{\alpha_1'} Z_i + \underline{\alpha_2'} W_i + u_i \tag{3}$$

where $\mathbf{Z_i}$ is the vector of instrumental variables not correlated with the error term of Eq. (2). After estimating $[\alpha_0, \underline{\alpha_1}, \underline{\alpha_2}]$ using Eq. (3), the obtained prediction of X_i replaces X_i in Eqs. (1) and (2), giving robust estimates of the coefficient β_1.

2.2.1 Dependent variable

Subjective general health (SGH) is the dependent variable of the main equation. It has been obtained by generating a dummy from the following question:

> How is your health in general? Would you say it is… [Very good, good, fair, bad, or, very bad?].

The dummy SGH takes a value of 1 when respondents choose the answers "very good" or "good" (we denote this condition as characterizing individuals with high self-reported health) and 0 otherwise, denoting those respondents that we classify in a low self-reported health category. The choice to dichotomize the outcome variables is justified by the availability of econometric tests for supporting the choice of an instrumental variable procedure.

2.2.2 Endogenous variable and instruments

The main explanatory variable of our analysis is social isolation. The ESS has a specific section about social exclusion. In our analysis, we focus on the measures of social isolation. In order to construct a robust index of isolation, we follow the recent literature (Gustafsson et al. 2015) and build this measure by adding together the score assigned to each response category of three related ESS variables, namely, family status, social contacts and having close friends. These items have been used in many past studies (Gustafsson et al. 2013, 2014; Floderus 2009),

2.2.2.1 Family status This item, indicating whether the respondent cohabits with a partner and/or has children living at home, is obtained by the combination of the following two questions of the ESS survey:

> Respondent lives with husband/wife/partner? [Yes, No] and Respondent has children living at home? [Yes, No].

In particular, the variable family status has the following four response categories and to each possible answer a score between 1 to 4 points (p) was assigned:

- Lone without children (1p)
- Lone with children (2p)
- Cohabiting without children (3p)
- Cohabiting with children (4p)

2.2.2.2 Social contacts This variable indicates the frequency of the respondents' social meetings and is recorded after asking this question of the interviewee:

How often do you meet socially with friends, relatives or work colleagues?

- Never/less than once a month (1p)
- Once a month (2p)
- Several times a month/once a week (3p)
- Several times a week/every day (4p)

2.2.2.3 Having close friends This last item indicates whether the respondent has some close friends or not; it is thus coded as a binary variable.

How many people, if any, are there with whom you can discuss intimate and personal matters?

- None (1p)
- ≥ 1 (4p)

The index obtained following the past literature (Gustafsson et al. 2015) is reversely coded; that is, low values on the index mean that the individual experiences a high degree of social isolation, whereas a high score implies less social isolation-

Table 1 shows the social isolation items and the distribution of their response categories.

Hence, the social isolation index is obtained by adding together the score of the three individuals' answers. This index is decreasing with respect to the level of isolation and it varies between 3 and 12p. It means that individuals now scoring low on the index experience high levels of social isolation, while individuals scoring high on the index experience less social exclusion. Since this variable can assume 10 different values (10-point scale), it can be treated in practice as a continuous predictor (Rhemtulla et al. 2012) without a great loss of information.

Table 1 Descriptive statistics of social isolation items (percentages)

Social isolation (high = social integration)		Total	Total	Men	Women
Family status					
Lone without children	(1p)	30.00	31.65	28.47	
Lone with children	(2p)	6.44	2.75	9.88	
Cohabiting without children	(3p)	29.16	30.96	27.48	
Cohabiting with children	(4p)	34.40	34.65	34.18	
Social contacts					
Never/less than once a month	(1p)	8.44	7.73	9.11	
Once a month	(2p)	9.63	10.04	9.23	
Several times a month/once a week	(3p)	40.26	40.83	39.73	
Several times a week/every day	(4p)	41.67	41.40	41.93	
Having close friends					
No	(1p)	3.90	4.41	3.42	
Yes	(4p)	96.10	95.59	96.58	

The relationship between social isolation and subjective health is documented in Table 2. The first column shows percentages for the whole sample, while the remaining ones show them at different quartiles of social isolation.

Looking at the last four columns, we can immediately see that health increases for individuals who experience higher levels of social inclusion.

As outlined above, the main issue with the estimation strategy is endogeneity: social isolation is correlated with health through different mediating factors and the coefficient of the isolation factor would likely be a biased estimate of a causal effect. The reason may be: (i) omitted variable bias, if unobserved individual factors are affecting health; (ii) reverse causality, if bad health conditions lead people to isolate themselves; or (iii) persistent measurement error, if isolated individuals tend to underreport their health conditions. In order to overcome these issues, we specify an instrumental variable model, where the variance of the isolation factor exploited in the main regression comes from some exogenous variables. Our instruments are the membership in ethnic minorities and the experience of past conflicts within the family. The underlying assumption is that minority status and family conflicts determine the isolation of an individual, which in turn undermines her health. According to the literature (Smith and Calasanti 2005), individuals belonging to racial and ethnic minorities have less social support and feel that they do not fit properly into society. The link between these variables is further confirmed in the literature (Boden-Albala et al. 2005), since membership in an ethnic minority group is considered a proxy for differences in social resources, in particular, social isolation. On the other hand, ethnic minorities can have worse health not necessarily due to their stronger social isolation: minorities may have lower health care access and less information about services, have different cognitive and non-cognitive skills which may affect health. In order to prevent the robustness of the estimates we control in both the first stage and second stage regressions for the household income and for family and individual education, which are proxies of individual skills. In addition, in order to increase safety in estimates procedures econometric tests can help us to provide evidence of the validity of the instruments in an overidentified setting.

The exigence of having overidentification requires the introduction of a second instrumental variable, which is the presence of family conflicts. With respect to past family conflicts, literature shows how adolescents' unresolved feelings of loneliness linked with the tensions experienced between household members may be an obstacle to the formation of normative social relationships through the development of anxiety and social avoidance (Johnson et al. 2001; Lau and Kong 1999; Vernberg 1994).

This approach should be suspected of not being enough to fully overcome the problem of endogeneity. First, health may be related to ethnicities, if they have different health attitudes. In the ESS there is no information about ethnicities or area of origin: we do not know the immigration history of the respondent and her family; some may have moved to their current country to benefit from better services, including health care. The best we

Table 2 Descriptive statistics of social isolation and subjective general health (percentages)	By quartiles of social isolation index			
Base sample	(1)	(2)	(3)	(4)
Low health				
33.70	37.29	35.45	29.17	27.06
High health				
66.30	62.71	64.55	70.83	72.94

can do with these data is to control for area of residence (small or large town), occupation and country effects. Second, family conflict regarding the past is generally a problematic concept because it is a subjective variable like health. Thus, it is plausible that the bias in reporting the two variables is correlated, for instance, if perception about one's own health affects the level of trust in others (e.g. doctors) or vice versa. Third, social isolation is due to a set of factors, so it is reasonable to expect that family contrast, minority status and other observable covariates are able to explain only a limited part of its variability in the sample. Hence, we could run into weak instrument problems: instruments with small explanatory capacity would give artificially high IV coefficients and inflate inconsistency and finite sample bias (Bound et al. 1995). Although the structure of the data does not allow us to solve all these problems, the results of econometric tests displayed in the following section may help to support or reject the candidate instruments.

The questions of interest for building instruments are:

Do you belong to a minority ethnic group? [Yes, No];

Tell me how often there was serious conflict between the people living in your household when you were growing up? [Always, often, sometimes, hardly ever, never].

The first variable is a dummy indicating whether the respondent belongs to a minority ethnicity in his country or region. The second one is a categorical variable which we recode as a dummy, taking a value of 1 when the respondent answers "always" or "often" and 0 otherwise.

2.2.3 Exogenous explanatory variables

We use as controls some variables which affect health and social contacts. First, we include basic demographic variables: age classes and gender. Second, we include other socio-economic controls (Venetoklis 2019): paid work status, dummies for urban area of residence (big city, suburbs, village) and level of education attained by respondents and also by their parents. Finally, we use country dummies to rule out any country-level effect.

2.2.3.1 Age Older people are generally less healthy than younger people (Ecob 1996). We decided to use a set of dummies for age classes, rather than a continuous variable, because the relationship between age and isolation (or health) may be nonlinear.

2.2.3.2 Gender According to past studies (Wuest 2002), the inclusion of gender as a health determinant is crucial in order to control for sex-related social norms and structures that influence vulnerabilities to illness, health status, access to preventive and curative measures and quality of care.

2.2.3.3 Occupational status Work status may be associated with anxiety and other health-affecting symptoms (Benavides et al. 2000). In order to control for these effects, we included in our regression a set of dummies representing the following categories: employed, students, unemployed, retired and other. The latter category includes individuals such as houseworkers and those in community or military service.

2.2.3.4 Religiosity Scientific research neglected the connection between religiosity and health until recently; however, during the last few years the interest in the relationship

between these two variables has started to grow (Aukst-Margetic and Margetic 2005; See-man et al. 2003). Given the prevalence and importance of religiosity among the population, it is reasonable to consider the impact that religious beliefs, practices and traditions may have on physical and mental health outcomes. Although a large proportion of published work suggests a positive association between religion and health outcomes (Casar Harris 1995), some studies show a negative relationship (Trenholm et al. 1998). Considering the contradictory positions in the existing literature, further research is needed to investigate this relationship.

2.2.3.5 Domicile The area of residence may significantly impact health outcomes by influencing items such as healthcare access and behaviors (Humphreys and Carr-hill 1991; Macintyre et al. 1993; Mullen 1988). However, until now the impact of area characteristics on health has still not been completely clear because the existing studies show contradictory findings (Ecob 1996). By including a set of dummies representing residents of big cities, small cities, suburbs, country villages and the countryside, we try to capture these effects.

2.2.3.6 Education The subjects surveyed were asked: "What is the highest level of education you have successfully completed?" Answers were then recoded in the International Standard Classification of Education (ISCED) which was developed by UNESCO, and includes seven different levels. This coding frame facilitates comparisons of education statistics and indicators across countries. In our analysis, we reduced the possible answers to 3 different categories and generated four dummy variables accordingly:

- Primary education: assumes a value of 1 if the subject completed ISCED level 1 and 0 otherwise;
- Secondary education: takes a value of 1 if the subject completed level 2–5 of the ISCED and 0 otherwise;
- Tertiary education: a value of 1 is indicated if the individual successfully completed either level 6 or 7 of the ISCED and 0 otherwise.

Those who achieved a tertiary level of education represent the reference category.

In the literature, it is well documented that the well-educated have better health than the poorly educated, as indicated by high levels of self-reported health (Guralnik et al. 1993; Gutzwiller et al. 1989; Winkleby et al. 1992) with a directional path that goes from education to health and not vice versa (Doornbos et al. 1990; Fox et al. 1985; Wilkinson 1986).

2.2.3.7 Country dummies Finally, we included a dummy for each country in order to take into account the residual unobserved heterogeneity at country level, such as country-specific institutional effects.

3 Results and discussion

We now turn to the results of our analysis. Table 3 displays the results of the first-stage regressions using the social isolation index as the endogenous variable. Notice that due to missing values in the control variables, the total sample is reduced to 31,726 observations.

Table 3 Estimates for the determinants of subjective general health

Variables	Social isolation	Subjective general health	Subjective general health
	Regression OLS (1st stage)	Two-step probit (2nd stage)	ML probit
Social isolation (high = social integration)		1.291***	0.103***
		(0.005)	(0.005)
Ethnic minority	− 0.202***		
	(0.036)		
Family conflict	− 0.211***		
	(0.028)		
Age			
<20	− 0.040	1.001***	0.918***
	(0.064)	(0.101)	(0.065)
20–29	0.302***	0.579***	0.893***
	(0.048)	(0.085)	(0.045)
30–39	1.061***	− 0.569**	0.644***
	(0.045)	(0.179)	(0.041)
40–49	1.049***	− 0.871***	0.337***
	(0.044)	(0.178)	(0.039)
50–59	0.675***	− 0.673***	0.101**
	(0.042)	(0.123)	(0.037)
60–69	0.437***	− 0.347***	0.158***
	(0.032)	(0.084)	(0.028)
Over 70	–	–	–
	–	–	–
Gender			
Male	0.042*	0.023	0.081***
	(0.017)	(0.027)	(0.016)
Female	–		–
	–	–	–
Occupational status			
Employed	–	–	–
	–	–	–
Student	− 0.463***	0.604***	0.051
	(0.043)	(0.103)	(0.048)
Unemployed	−0.470***	0.277**	− 0.302***
	(0.039)	(0.099)	(0.036)
Retired	− 0.188***	− 0.208***	− 0.437***
	(0.035)	(0.061)	(0.031)
Other	0.100**	− 0.695***	− 0.592***
	(0.030)	(0.048)	(0.028)
Religiosity			
Religious	0.190***	- 0.213***	0.009
	(0.019)	(0.042)	(0.018)
Not religious	–	–	–

Table 3 (continued)

Variables	Social isolation	Subjective general health	Subjective general health
	Regression OLS (1st stage)	Two-step probit (2nd stage)	ML probit
	–	–	–
Domicile			
Big city	– 0.204***	0.206**	– 0.060
	(0.040)	(0.071)	(0.038)
Suburbs	– 0.158***	0.201**	– 0.004
	(0.041)	(0.069)	(0.039)
Small city	– 0.136***	0.150*	– 0.021
	(0.037)	(0.061)	(0.035)
Country village	– 0.051	0.041	– 0.025
	(0.037)	(0.058)	(0.036)
Countryside	–	–	
	–	–	–
Education			
Primary	– 0.471***	0.123	– 0.445***
	(0.037)	(0.096)	(0.035)
Secondary	– 0.213***	– 0.005	– 0.259***
	(0.022)	(0.048)	(0.021)
Tertiary	–	–	–
	–	–	–
Father's education			
Primary	– 0.014	– 0.089	0.107**
	(0.038)	(0.058)	(0.037)
Secondary	0.012	– 0.057	– 0.041
	(0.030)	(0.046)	(0.030)
Tertiary	–	–	–
	–	–	–
Mother's education			
Primary	– 0.195***	0.197**	– 0.053
	(0.041)	(0.072)	(0.041)
Secondary	– 0.086**	0.108*	– 0.001
	(0.033)	(0.055)	(0.034)
Tertiary	–	–	–
	–	–	–
Country dummies	Yes	Yes	Yes
Observations	31,726	31,726	31,726

SE in parentheses

*$p<0.050$; ** $p<0.010$; *** $p<0.001$

The results were in line with the expectations: individuals who belong to an ethnic minority and/or who grew up in a household with serious conflicts between family members are more likely to be socially isolated. We can claim that ethnic minority and family

conflict are strong determinants of social isolation, since the coefficients of the two instrumental variables are statistically significant at the 0.1%. Most of the socio-economic controls have significant coefficients—namely age class, gender, religion, area of residence and education. The only exception is provided by the education level of the father, even if the results are compensated with those regarding the education of the mother. All the variables used in the regression, i.e. the instruments, the demographic and socioeconomic controls and the country dummies, were altogether significant with a F-statistic greater than 10. Thus, they are highly correlated with our endogenous variables. This is a good argument in favor of the effectiveness of our IV-probit regression (Baum 2003), proving the relevance of the selected instruments.

The second column of Table 3 reports the results of the second-stage regression, where health status is the dependent variable and the index of social isolation is used as an approximation of a continuous endogenous variable. The coefficient of social isolation is positive and strongly significant, meaning that health benefits from a higher level of social integration. This effect is robust to the inclusion of all our controls and country dummies. This basic model shows evidence that social isolation has adverse effects on health as reported in the ESS.

The last column of Table 3 presents ML estimates (standard probit estimates), which show a similar sign for social isolation: in both the regressions, social isolation is positively linked with subjective health, even if in the IV model the coefficient is larger: this means that subjective health is better for individuals who do not belong to families with past conflicts and who are not part of ethnic minority.

With respect to the other explanatory variables, ML results are in line with expectations: self-assessed health is worse in older cohorts and in retired and unemployed individuals. Higher level of education are also connected to best performances in subjective health.

Table 4 displays country fixed effect deriving from IV-probit. Portugal, which is the country with the worst performance, has been chosen as the reference category.

Results are surprising and reserve to Ireland the best performance in terms of self-assessed health, being constant the effect of isolation and of the other covariate. This result partially confirms that obtained in literature (Winkleby et al. 1992), which confirms the high health standards of Ireland.

In order to further support the choice of the IV-probit method with respect to the standard probit regression and to show the validity of our instruments, the Amemiya-Lee-Newey test and the Wald test of exogeneity were run. The results, reported in Table 5, are in favor of the IV-probit method. The Amemiya-Lee-Newey statistic can be calculated for cases in which the model is "overidentified," that is, when the number of items exceeds the number of endogenous variables. This test is equivalent to the Sargan test for the standard IV regressions and is implemented in order to judge whether the items are valid (i.e. they are exogenous and they affect the dependent variable only indirectly) or not. Results supported accepting the null hypothesis that the variables we used to instrument the index of social isolation, i.e. ethnic minority and family conflict, are exogenous. The Wald test is instead an indirect test for the endogeneity of the instrumented variable. Results were against the null hypothesis of exogeneity; thus, the error terms of the first and second-stage regressions are correlated. This suggests that using the IV-probit estimator produces consistent and more efficient estimates for the impact of social isolation on subjective general health.

Our analysis has some objective shortcomings. As explained in the previous sections, there are endogeneity issues which still remain in our model, even after instrumenting.

First, the validity of our instruments must be tested since the exclusion restriction might not be satisfied. In fact, belonging to an ethnic minority might be correlated with health

Table 4 Country fixed effects in IV estimation

Countries	Codes	Coefficients of country fixed effects	S. E
Austria	AT	1.305***	(0.126)
Belgium	BE	0.767***	(0.094)
Switzerland	CH	1.069***	(0.105)
Czech Republic	CZ	1.351***	(0.159)
Germany	DE	0.494***	(0.101)
Denmark	DK	0.476***	(0.094)
Estonia	EE	0.874***	(0.171)
Spain	ES	0.253**	(0.088)
Finland	FI	0.592***	(0.092)
Frances	FR	0.589***	(0.099)
United Kingdom	UK	1.325***	(0.132)
Ireland	IE	2.021***	(0.158)
Israel	IL	1.211***	(0.128)
Lithuania	LT	1.633***	(0.229)
Netherlands	NL	0.480***	(0.092)
Norway	NO	0.698***	(0.101)
Poland	PL	1.289***	(0.174)
Portugal	PT	–	–
Sweden	SE	0.552***	(0.094)
Slovenia	SI	0.680***	(0.128)

*$p < 0.050$; ** $p < 0.010$; *** $p < 0.001$

Table 5 Instrumental variable estimation diagnostic

	Subjective general health
Validity of instruments	
Amemiya-Lee-Newey $\chi^2_{(1)}$	3.51
Relevance of instruments	
F-Statistic of first stage regression	146.85***
Test of endogeneity	
Wald test of exogeneity $\chi^2_{(1)}$	144.48***

*$p < 0.050$; **$p < 0.010$; ***$p < 0.001$

status, even if the direction of the correlation is ambiguous. A stream of literature in public economics focuses on overutilization and underutilization of welfare services, including healthcare, by immigrants and ethnic minorities. A possible explanation for underutilization is that ethnic minorities and immigrants face barriers to access healthcare services due to a lack of information. This could result in worse health conditions for those belonging to an ethnic minority. Nevertheless, immigrants tend to be younger and, on average, healthier; hence, the health effect of being part of an ethnic minority is ambiguous. However, by controlling for age classes, this problem should be overcome. Our second instrument, namely,

family conflict, is a lagged variable because it investigates possible tensions between family members in the past; thus, it is a good candidate for being an exogenous item. The goodness of both our instruments is confirmed by the result of the Amemiya-Lee-Newey overidentification test, which states that the instruments used in our model are not endogenous (thus, valid) and so the model seems to be specified correctly.

A second issue is that ethnic minority and family conflict are not the only determinants of social isolation. In fact, the coefficients from the first-stage regressions are quite small (Table 3), even if jointly strongly significant. The F-statistic on excluded instruments in the first stage is greater than 10, so following the rule of thumb by Staiger and Stock (1997), we do not worry about weak instrument problems, even if this does not guarantee that the coefficients are unbiased.

Nevertheless, our objective is to test a relationship which is outlined in the literature. In this respect, it is important to focus on the qualitative conclusion: there is strong evidence that isolation has a negative effect on health. Our findings are consistent with those reported in the medical literature.

Our analysis provides additional evidence to support the hypothesis that social isolation deteriorates people's health. Although our results are coherent overall with the existing literature, they are not fully comparable. The first difference is the measure for social isolation: most studies use the UCLA loneliness scale, while, as explained above, we used the social isolation index.

Secondly, while many papers focus on specific medical conditions, often related to the circulatory system, such as cardiovascular diseases, blood pressure, myocardial infarction etc. (Umberson 2010; Cacioppo et al. 2002; Hawkley et al. 2003), we resort to self-reported health status at a single moment in time. A similar measure was however employed in the literature (Luo et al. 2011), together with other measures: using a crossed-lagged model, studies found that loneliness has a modest negative impact on self-rated health over two years. As we have mentioned, researchers who focus on health behaviors obtain contrasting figures.

4 Conclusions

In this paper, we investigate the relationship between social isolation and health. In particular, we test the hypothesis that higher isolation worsens the health status of individuals. However, since this relationship is ambiguous and affected by endogeneity, a simple correlation among these variables does not prove our hypothesis. For this reason, in order to investigate the causal relationship, we implemented an instrumental variable approach, in which the instruments are given according to the variables "belonging to an ethnic minority" and "family conflict", both of which are highly correlated with social isolation.

Using different data from a cross-sectional survey and a different methodology from those employed before, we find results in line with past research in medicine and psychology but extended to self-reported health for a large, representative sample of European citizens (not necessarily belonging to the older cohorts). High levels of social isolation, defined as a lack of personal contacts with peers and an absence of profound relationships, are found to favor a decline in the perceived health of individuals. The validity of the model providing these results is supported by the implementation of statistical tests of validity and the relevance of the instruments.

All in all, our results show that a lack of social connections impacts the health status of individuals. This is an important result for policymakers, as exploring the channels through which social status affects people's health may help in designing preventative interventions. The relationship between isolation and health becomes even more important if one considers the peculiarities of modern European societies. On the one hand, an aging population is likely to increase the already high healthcare expenditures. On the other hand, family size is decreasing with plausible consequences on social ties and feelings of loneliness. Hence, addressing the issues of social isolation and loneliness might be an effective strategy to improve the population's health, which in turn will benefit governments. Finally, we cannot ignore the role of the internet and social media, the effects of which on social relationships and loneliness are still not completely clear.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest.

References

Aukst-Margetic, B., Margetic, B.: Religiosity and health outcomes: review of literature. Coll. Antropol. **29**, 365–371 (2005)

Baum, C.F., Schaffer, M.E., Stillman, S.: Instrumental variables and GMM: estimation and testing. Stata. J. **3**, 1–31 (2003)

Benavides, F.G., Benach, J., Diez-Roux, V., Roman, C.: How do types of employment relate to health indicators? Findings from the second European survey on working conditions. J. Epidemiol. Commun. Health **54**, 494–501 (2000). https://doi.org/10.1136/jech.54.7.494

Boden-Albala, B., Litwak, E., Elkind, M.S.V., Rundek, T., Sacco, R.L.: Social isolation and outcomes post stroke. Neurology **64**, 1888–1892 (2005). https://doi.org/10.1212/01.WNL.0000163510.79351.AF

Bound, J., Jaeger, D.A., Baker, R.M.: Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J. Am. Stat. Assoc. **90**, 443–450 (1995). https://doi.org/10.1080/01621459.1995.10476536

Bourne, P.A.: The validity of using self-reported illness to measure objective health. North Am. J. Med. Sci. **1**(5), 232–238 (2009). https://doi.org/10.4297/najms.2009.5232

Butler, J.S., Burkhauser, R.V., Mitchell, J.M., Pincus, T.P.: Measurement error in self-reported health variables. Rev. Econ. Stat. **69**(4), 644–650 (1987). https://doi.org/10.2307/1935959

Cacioppo, J.T., Hawkley, L.C., Crawford, L.E., Ernst, J.M., Burleson, M.H., Kowalewski, R.B., et al.: Loneliness and Health: Potential Mechanisms. Psychosom. Med. **64**, 407–417 (2002). https://doi.org/10.1097/00006842-200205000-00005

Casar Harris, R., Dew, M.A., Lee, A., Amaya, M., Buches, L., Reetz, D., et al.: The role of religion in heart-transplant recipients' long-term health and well-being. J. Relig. Health **34**, 17–32 (1995). https://doi.org/10.1007/BF02248635

Cohen, S., Wills, T.A.: Stress, social support, and the buffering hypothesis. P Psychol Bull **98**(2), 310–357 (1985). https://doi.org/10.1037/0033-2909.98.2.310

Cohen, S., Underwood, L., Gottlieb, B.: Social relationships and health. In: Cohen, S., Underwood, L., Gottlieb, B. (eds.) Social support measurement and intervention, pp. 3–25. Oxford University Press, Oxford (2000)

Cornwell, E.Y., Waite, L.J.: Social disconnectedness, perceived isolation, and health among older adults. J. Health Soc. Behav. **50**(1), 31–48 (2009). https://doi.org/10.1177/002214650905000103

Doornbos, G., Kromhout, D.: Educational level and mortality in a 32-year follow-up study of 18-year-old men in the Netherlands. Int. J. Epidemiol. **19**, 374–379 (1990). https://doi.org/10.1093/ije/19.2.374

Ecob, R.A.: Multilevel modelling approach to examining the effects of area of residence on health and functioning. J. R. Stat. Soc. Ser. A **159**, 61–75 (1996). https://doi.org/10.2307/2983469

Floderus, B., Hagman, M., Aronsson, G., Marklund, S., Wikman, A.: Work status, work hours and health in women with and without children. Occup. Environ. Med. **66**, 704–710 (2009). https://doi.org/10.1136/oem.2008.044883

Fox, A.J., Goldblatt, P.O., Jones, D.R.: Social class mortality differentials: artefact, selection or life circumstances? J. Epidemiol. Commun. Health. **39**, 1–8 (1985). https://doi.org/10.1136/JECH.39.1.1

Guralnik, J., Land, K., Blazer, D., Fillenbaum, C., Branch, L.: Educational status and active life expectancy among blacks and whites. N. Engl. J. Med. **329**, 110–116 (1993). https://doi.org/10.1056/NEJM199307083290208

Gustafsson, K., Aronsson, G., Marklund, S., Wikman, A., Floderus, B.: Does social isolation and low societal participation predict disability pension? A population based study. PLoS One **8**, e80655 (2013). https://doi.org/10.1371/journal.pone.0080655

Gustafsson, K., Aronsson, G., Marklund, S., Wikman, A., Hagman, M., Floderus, B.: Social integration, socioeconomic conditions and type of ill health preceding disability pension in young women: a Swedish population-based study. Int. J. Behav. Med. **21**, 77–87 (2014). https://doi.org/10.1007/s12529-012-9287-5

Gustafsson, K., Marklund, S., Aronsson, G., Wikman, A., Floderus, B.: Interaction effects of social isolation and peripheral work position on risk of disability pension: a prospective study of Swedish women and men. PLoS ONE **10**, e0130361 (2015). https://doi.org/10.1371/journal.pone.0130361

Gutzwiller, F., La Vecchia, C., Levi, F., Negri, E., Wietlisbach, V.: Education, disease prevalence and health service utilization in the Swiss National Health Survey "SOMIPOPS." Prev. Med. (Baltim). **18**, 452–459 (1989). https://doi.org/10.1016/0091-7435(89)90005-4

Hawkley, L.C., Burleson, M.H., Berntson, G.G., Cacioppo, J.T.: Loneliness in everyday life: cardiovascular activity, psychosocial context, and health behaviors. J. Pers. Soc. Psychol. **85**, 105–120 (2003). https://doi.org/10.1037/0022-3514.85.1.105

Holt-Lunstad, J., Smith, T.B., Layton, J.B.: Social relationships and mortality risk: a meta-analytic review. PLoS Med. **7**(7), e1000316 (2010)

House, J.S., Landis, K.R., Umberson, D.: Social relationships and health. Science **241**(4865), 540–545 (1988). https://doi.org/10.1126/science.3399889

Humphreys, K., Carr-hill, R.: Area variations in health outcomes: artefact or ecology. Int. J. Epidemiol. **20**, 251–258 (1991). https://doi.org/10.1093/ije/20.1.251

Johnson, H.D., Lavoie, J.C., Mahoney, M.: Interparental conflict and family cohesion: predictors of loneliness, social anxiety, and social avoidance in late adolescence. J. Adolesc. Res. **16**, 304–318 (2001). https://doi.org/10.1177/0743558401163004

Johnston, D.W., Propper, C., Shields, M.A.: Comparing subjective and objective measures of health: evidence from hypertension for the income/health gradient. J. Health Econ. **28**(3), 540–552 (2009). https://doi.org/10.1016/j.jhealeco.2009.02.010

Lau, S., Kong, C.K.: The acceptance of lonely others: effects of loneliness and gender of the target person and loneliness of the perceiver. J. Soc. Psychol. **139**, 229–241 (1999). https://doi.org/10.1080/00224549909598377

Luo, Y., Hawkley, L.C., Waite, L.J., Cacioppo, J.T.: Loneliness, health, and mortality in old age: a national longitudinal study. Soc. Sci. Med. **74**, 907–914 (2012). https://doi.org/10.1016/j.socscimed.2011.11.028

Macintyre, S., Maciver, S., Sooman, A.: Area, class and health: should we be focusing on places or people? J. Soc. Policy **22**, 213–234 (1993). https://doi.org/10.1017/S0047279400019310

Mullen, K.: Area and health in cities: a review of the literature. Int. J. Sociol. Soc. Policy **10**, 1–24 (1988). https://doi.org/10.1108/eb013119

Newey, W.K., West, K.D.: A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix. Econometrica **55**, 703–708 (1987). https://doi.org/10.3386/t0055

Ong, A.D., Uchino, B.N., Wethington, E.: Loneliness and health in older adults: a mini-review and synthesis. Gerontology **62**(4), 443–449 (2016). https://doi.org/10.1159/000441651

Rhemtulla, M., Brosseau-Liard, P.É., Savalei, V.: When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol. Methods **17**, 354–373 (2012). https://doi.org/10.1037/a0029315

Seeman, T.E., Dubin, L.F., Seeman, M.: Religiosity/spirituality and health: a critical review of the evidence for biological pathways. Am. Psychol. **58**, 53–63 (2003). https://doi.org/10.1037/0003-066X.58.1.53

Sironi E, Wolff AN (2018) The impact of social isolation on subjective health: an instrumental variable approach. Preprint at https://www.preprints.org/manuscript/201802.0164/v1, 2018020164. doi: https://doi.org/10.20944/preprints201802.0164.v1.

Smith, J.W., Calasanti, T.: The influences of gender, race and ethnicity on workplace experiences of institutional and social isolation: An exploratory study of university faculty. Sociol. Spectr. **25**, 307–334 (2005). https://doi.org/10.1080/027321790518735

Staiger, D., Stock, J.: Instrumental variables regression with weak instruments. Econometrica **65**, 557–586 (1997)

Steptoe, A., Shankar, A., Demakakos, P., Wardle, J.: Social isolation, loneliness, and all-cause mortality in older men and women. P. Natl. Acad. Sci. USA **110**(15), 5797–5801 (2013). https://doi.org/10.1073/pnas.1219686110

Trenholm, P., Trent, J., Compton, W.C.: Negative religious conflict as a predictor of panic disorder. J Clin Psychol. **54**, 59–65 (1998). https://doi.org/10.1002/(SICI)1097-4679(199801)54:1

Umberson, D., Montez, J.K.: Social relationships and health: a flashpoint for health policy. J. Health Soc. Behav. **51**, S54-66 (2010). https://doi.org/10.1177/0022146510383501

Venetoklis, T.: Do interactions cancel associations of subjective well-being with individual-level socioeconomic characteristics? An exploratory analysis using the European Social Survey. Qual Quan. **53**(6), 3033–3061 (2019). https://doi.org/10.1007/s11135-019-00919-0

Vernberg, E.M.E.: Sophistication of adolescents' interpersonal negotiation strategies and friendship formation after relocation: a naturally occurring experiment. J. Res. Adolesc. **4**(1), 5–19 (1994). https://doi.org/10.1111/1532-7795.ep11301464

Wilkinson, R.G.: Class and Health: Research and Longitudinal Data. Tavistock Publications, London (1986)

Winkleby, M.A., Jatulis, D.E., Frank, E., Fortmann, S.P.: Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. Am. J. Public Health **82**, 816–820 (1992). https://doi.org/10.2105/AJPH.82.6.816

Wuest, J., Merritt-Gray, M., Berman, H., Ford-Gilboe, M.: Illuminating social determinants of women's health using grounded theory. Health Care Women Int. **23**, 794–808 (2002). https://doi.org/10.1080/07399330290112326

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.