working paper
2003

Social isolation, health dynamics, and mortality: Evidence across 21 Europe countries

Yarine Fawaz
Pedro Mira

June 2020

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 2003
June 2020

Social isolation, health dynamics, and mortality: Evidence across 21 Europe countries

Abstract

We provide a comprehensive picture of the health effects of social isolation using longitudinal data over 21 European countries (SHARE) and seven waves (14 years) : first by looking at how social isolation at baseline impacts mortality at follow-up using Cox duration models, then looking at the dynamics of the health effects of social isolation, i.e. how social isolation at baseline affects functional, physical, mental and cognitive health at each future wave, when controlling for all these facets of health at baseline along with an extensive set of other covariates, in a standard linear regression framework. Our results suggest social isolation leads to worse health along all the dimensions we observe, and this effect is persistent. Being socially isolated at baseline is associated with a 20 to 30% increase in the mortality hazard, in line with other studies. Allowing for heterogeneity across countries, we find a remarkably strong association (up to a 45% increase) in Eastern countries. This association is not just picking up a correlation of social isolation with concurrent loneliness, health behaviors or health care utilization.

JEL Codes: I10, C41.

Keywords: Social isolation, loneliness, health, mortality, SHARE.

Yarine Fawaz
CEMFI
fawaz@cemfi.es

Pedro Mira
CEMFI
mira@cemfi.es

Acknowledgement

We are very grateful to participants at the 4th PHS Workshop at University of Pennsylvania, the 2020 Alp-Pop conference, the SHARE user conference, the BUDIE meeting at CEMFI, for the helpful feedback of many participants. We particularly thank Atheendar Venkataramani, Ryan Brown, Jerome Adda, Keith Head, for their constructive remarks and suggestions.

1 Introduction

As the Covid-19 outbreak has been shaking most of the world, public health efforts have globally focused on finding short-term strategies to limit casualties while working on longer-term cures. With 4.5 billion people confined in the world (as of April 17th 2020, according to an AFP database), i.e. almost 60% of the world population, social distancing has been praised as the only way to “flatten the curve”. One collateral damage of Covid-19 -or rather of social distancing- is social isolation, and the health hazards that have been growingly shown to be associated with it.

The current evidence about the interplay between social isolation and health points at social isolation -sometimes the subjective side of it, loneliness- being devastating to a person’s health. According to recent studies, being- i.e. feeling- lonely, and being socially isolated, i.e. lacking social connections, is at least as bad as being obese or a heavy smoker (Holt-Lunstad et al. (2015)). The impact of loneliness and social isolation on health and mortality has been increasingly investigated by physicians, epidemiologists, as well as social scientists over the last decade. The bulk of evidence points at social isolation and loneliness being linked to a variety of physical and mental conditions such as high blood pressure, cardiovascular diseases, obesity, a weakened immune system, anxiety, depression, cognitive decline, Alzheimer’s disease, and even death.

According to the latter meta-analysis of scientific literature on the subject (Holt-Lunstad et al. (2015)), social isolation, i.e. having few network ties, increases your risk of death over 7 years[1] by about 30%, while the effect of loneliness (i.e. feeling lonely) is estimated at around 26%, and living alone seems to be the utmost risk factor with a weighted average effect of 32%. This study also reviews a number of previous analyses that showed that individuals with less social connection have disrupted sleep patterns, altered immune systems, more inflammation and higher levels of stress hormones. Valtorta et al. (2016) -a more recent meta-analysis- found

[1]Seven years is the average of the follow-ups across the studies that were analyzed in the meta-analysis.

that isolation increases the risk of heart disease by 29 percent and stroke by 32 percent.

Those meta-analyses report results from a variety of articles that do not share a common level of rigor, e.g. out of the 70 studies analyzed in Holt-Lunstad et al. (2015), 31 are fully "unadjusted", meaning that they include no control of any sort, and 20% of the remaining studies do not control for baseline health, which according to the meta-analysis changes radically the findings. The remaining multivariate analyses that do control for baseline health and other factors rarely have background data on individuals, and are usually not based on random samples, as participants are often recruited from a medical setting. Even when studies recruit participants from the general community, they usually do not collect as much information as in multi-disciplinary surveys such as SHARE, and cannot claim to be fully representative. We rely on longitudinal data on a large representative population across 20 European countries. The SHARE (Survey of Health, Aging and Retirement in Europe) data allows us to follow individuals across time and mitigate part of the endogeneity concerns; it provides us with a comprehensive set of health indicators and social isolation and loneliness variables, which we observe every two years over 14 years. A few studies have exploited similar data -such as the American HRS or its UK equivalent ELSA- to look at correlations between social isolation, loneliness, and mortality or a specific health outcome. One particularly well carried-out study is Steptoe et al. (2013), which uses ELSA to investigate how social isolation and loneliness at baseline are associated with mortality over a seven year follow-up period.

We provide a more complete picture of the health effects of social isolation: we first follow the latter study in looking at how social isolation at baseline impacts mortality at follow-up (over a 14 year period), then dive into the dynamics of the health effects of social isolation, by studying how social isolation at baseline affects functional, physical, mental and cognitive health at each future wave, when controlling for all these facets of health at baseline. Going further, we investigate whether health behavior (such as smoking, drinking, or a sedentary lifestyle) and health care utilization are potential pathways between social isolation and adverse health outcomes. We find a significant and important association between our social isolation index and both mortality and future health outcomes, along almost all health dimensions. Remarkably, controlling for loneliness barely weakens this relationship- same as in Steptoe et al. (2013). This suggests that loneliness cannot be the only mechanism through which social isolation affects health. While we find that socially isolated individuals are more likely to adopt a worse lifestyle (particularly in terms of physical inactivity), including unhealthy behavior measures in our regressions does not affect the coefficient on social isolation. Health care utilization is another potential channel as we find that socially isolated individuals do not use more health care services (except for drugs consumption) than their non-socially-isolated counterparts, although their health keeps worsening. On top of the traditional robustness checks (using different specifications, over different samples, and showing the stability of our key coefficient), we provide evidence in favor of a causal interpretation of our estimates using Oster's test for selection on unobservables (Oster (2019)).

Last but not least, the impact of social isolation at older ages may have some cultural and/or institutional dimension, which should be examined in a cross-national framework. The prevalence of social isolation amongst the elderly is likely to differ a lot between, for instance, Mediterranean (relying heavily on a family network) and Northern (where giving autonomy to the individual was made a priority by many governments, sometimes at the cost of being isolated) countries. We find a much stronger association between social isolation and mortality in Eastern countries. While all of our pooled countries estimates ranged between a 20 to 30% increase in the mortality hazard for the socially isolated, that number jumps to 45% for the Eastern countries. That one same -objective- measure of social isolation does not lead to the same health consequences across countries, albeit using harmonized data, points at public health policies having a role to play in moderating the health risks posed by social isolation.

2 Social isolation and mortality

2.1 Data

We use longitudinal survey data from SHARE, over 7 waves from 2004 to 2017, and 20 European countries plus Israel[2]. SHARE is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of about 140,000 individuals aged 50 or older (around 380,000 interviews).

To carry out our empirical strategy, we create a set of health indicators that cover physical, functional, mental and cognitive health. Physical health is investigated along several dimensions: objective (i.e. number of diagnosed chronic diseases) and subjective (self-assessed health status); focusing on functional health (Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs)), and constructing an index of frailty (Fried et al. (2001)) which aggregates unintentional weight loss, self-reported exhaustion, weakness (grip strength), difficulties in walking, and low physical activity. Mental health is looked at using the EURO-D score, which is the sum of 12 items that can be relied on to diagnose depression in older adults. Cognitive health is an average of immediate and delayed word recall. We also make use of the so-called ”exit interviews”, which allow us to keep track of respondents' death.

Regarding the key regressor, following Steptoe et al. (2013), we create a social isolation index summing information on whether the individual lives alone, has infrequent social contact with his/her children (less than weekly contact, or does bit have children), and does not participate to any social activities (including political, sports, educational or voluntary work activities). We also use the short form of the R-UCLA loneliness scale, which was created by summing 3 items -how much of the

[2]Although SHARE now encompasses 28 countries and we do make use of all 7 waves, we can only exploit data on the 20 countries that entered SHARE before the last wave, and appeared at least twice across the 7 first waves, non including Wave 3, which was dedicated to constructing life histories of SHARE respondents.

time they felt a sense of being left out, a lack of companionship, and isolation- into one single measure of loneliness. We can therefore investigate the health effects both of objective social isolation and of the perception of social isolation.

Table 1 provides descriptive statistics at baseline (when individuals enter the data) on both the above-mentioned health outcomes, and the socio-demographic controls that will be used in our analyses. It also puts forward important differences between two populations, those that are not socially isolated at all according to our index, and those with at least 1 social isolation point. The more socially isolated exhibit worse health measures (along all dimensions), are more likely to be female, less educated, and poorer (in income and assets) than the non socially isolated. We also add an employment indicator as being working or retired may explain an important part of social isolation for the over-50 population.

Figures 1 and 2 show that there is a great deal of variation in social isolation across countries, with Eastern and Southern Europe countries having the highest average levels of social isolation, and Western and Northern Europe the lowest. The correlation between loneliness and social isolation at the individual level being of (only) 24%, a few countries have very high levels of loneliness compared to other countries with a similar level of social isolation, e.g. Italy, Greece and Israel (all part of Southern Europe), while others, such as Switzerland, Austria, or Denmark, have very low rates of loneliness in comparison with other countries with a similar level of social isolation.

2.2 Main results from Cox models

We first look at how social isolation at baseline is associated with future mortality, by estimating Cox proportional hazards regression models, from the date of an individual's entry to the data (February 2004 at the earliest), until that individual potentially dies or is followed-up in subsequent surveys up to October 2017. Out of the 94,029 individuals we follow, 10,739 die over the period.

$$h(i,t|SI_{i,t_0}, X_{i,t_0}) = h_0(t)exp(SI_{i,t_0}\beta + X_{i,t_0}\gamma) \quad (1)$$

The hazard of dying h at time t is a function of a fully flexible baseline hazard h_0 that is common to all individuals in our sample, which is shifted proportionally upwards or downwards by social isolation SI at baseline and the individual characteristics X_i introduced in the model at baseline t_0. We fit 6 models, each one adding more constraints to the relationship between social isolation and mortality.

Our results in Table 2 are very much in line with Steptoe et al. (2013), which finds a hazard ratio for a comparable discrete social isolation index that varies between 1.50 to 1.26. Our point estimates of the hazard ratio for the social isolation index - in its continuous version from 0 to 3- go down from 1.25 to 1.17 when controlling for all health indicators at baseline, as part of the association between social isolation and future health goes through initial health conditions, but the coefficient remains both strongly significant and big. Restricting follow-up to individuals who are still alive after the first 24 months, and even 48 months- in order to make sure they did not have a life-threatening condition that was not captured by our observable covariates- does not challenge our estimate either. Comparing individuals with at least 1 point of social isolation to those who do not (column (5)) leads to a higher hazard ratio than looking at the effect of one extra point of social isolation (column (4)), which suggests that going from 0 to 1 point matters more in terms of mortality risk than an extra point at higher levels of social isolation.

Sensitivity checks We perform a number of sensitivity checks to make sure our results do not depend heavily on the way our social isolation index is constructed , and are not driven by a very specific subsample (see Tables A1, A2 and A3 in the Appendix). Table A1 in the Appendix compares the results from different Cox models using different social isolation index: the first column reproduces the case of a discrete social isolation index where the cutoff is between 0 and values above 0 (HR of 1.26). Column (2) shows the results of a regression featuring three indicator variables for each of the 4 values that can take the social isolation index: the higher the social isolation index, the greater the corresponding hazard ratio, i.e. the more socially isolated individuals are, the more at risk of dying they are, although there is no statistical difference between having a social isolation index of 2 or 3. In the third column, we use a longer version of the social isolation index: instead of summarizing each of the three items used to a binary variable, we let them take on a wider range of values, so that the resulting index lies between 1 and 12 (the contact frequency with children between 1 and 7, where 7 corresponds to ”never”, and the participation to social activities between 0 and 4, where 4 is the case of an individual who does not take part in any political, sport, educational, or voluntary work association). Whether we impose a cutoff to this variable (column (4)) or we leave it as it is (column (3)), the results are robust. One extra point of social isolation is associated with a lower HR in the case of the “long” index, which makes sense as it does not imply the same increase in isolation as in the case of a 0-3 index.

All three items used in the index construction seem to matter for future mortality, as showed in Table A2 in the Appendix. Whether introduced separately or together, they all point at higher mortality. Having no social activities is associated with the greatest HR, and contact with children does not come up as significant, even though it is quite close to “living alone” in terms of both magnitude and precision.

Last, in order to make sure our results are not driven by a specific subsample, we estimate the most demanding specification (column (6) of Table 2) on several subgroups. The main concern is related to the construction of the social isolation index: if a particular population was more likely to have less contact with their children (say, males), to live alone (unmarried individuals), or to participate in social activities (working versus retired individuals), then the results we found on the whole representative sample could be misleading. As we already control for these characteristics in our regressions, this is less of a concern, but we still display the results of this sensitivity analysis in Table A3. Apart from a few exceptions, e.g. the employed have a higher mortality risk associated with social isolation than the unemployed (hazard ratios of 1.62 against 1.22), the hazard ratio remains remarkably stable around 1.25 almost across all subsamples.

2.3 Country heterogeneity

One of the most unique features of the SHARE datasets is that it is harmonized across all Europe, so that we can look at how social isolation affects health and mortality differentially across countries. Ideally we would interact our index of social isolation with each country in order to capture differences in the association between social isolation and mortality across all countries, but doing so we would not have enough power to find any significant association. We instead group countries into four culturally and geographically homogeneous subgroups: Western (Austria, Germany, the Netherlands, France, Switzerland, Belgium and Luxembourg), Northern (Sweden and Denmark), Southern (Spain, Italy, Greece, Portugal), and Eastern countries (Czech Republic, Poland, Hungary, Slovenia, Estonia, Croatia).

A first look at heterogeneity across these four groups of countries (see Table 3, columns (1) and (2)) suggests that the hazard ratio found for the socially isolated against the non-socially isolated (1.30) is hiding important differences across countries. While in Western, Southern, and Northern countries, social isolation (defined as at least one social isolation point) is associated the hazard ratio is around 1.2, it is more than twice this number in Eastern countries (1.45). Columns (3) to (5) introduce each country group against the three others in order to find out if an y other pattern would appear when pooling more countries together in the reference group. Again, only in Eastern countries are the socially isolated more at mortality risk compared with other countries.

This finding is important to the extent that if a same level of social isolation is associated with different mortality hazards across countries, there may be room for public policies to weaken that association. Our results do not say anything about what makes people in Eastern countries more vulnerable when they are socially isolated. Our models allow us to rule out several hypotheses: at the individual level, we are controlling for income and wealth quartiles within country, as well as educational attainment, we are therefore looking at the effect of social isolation for individuals with a similar socioeconomic status. On top of that, we are including indicator variables for country groups (or even country dummies for all countries when not looking specifically at the effect of countries), so whatever may make individuals more or less healthy in a country -i.e. economic conditions, generosity of the health care system, etc.- is already captured by these indicators. If these country-specific characteristics are still reflected in the interaction term, it would mean socially isolated individuals are more vulnerable to economic or health care conditions than non-socially isolated individuals, even controlling for their income and health.

One possibility we explore using ESS (European Social Survey) data is that social isolation may have a different impact on people's health depending on the level of social support in a country. We compute the share of people who never meet or meet less than once a month socially with friends, relatives, or work colleagues, in all SHARE countries in 2004. We create a binary variable for countries whose average is above the average of all individuals in the ESS (using country and individual weights). As shown in Table 4, social isolation is associated with more excess mortality in countries where people meet less socially than in other European countries. This is only suggestive as the "Low social support countries" dummy is equivalent to adding a dummy for a group of countries, namely Italy, Greece, Czech Republic, Poland, Hungary, Slovenia and Estonia. Nevertheless, there is no such pattern when looking at loneliness and how it affects differentially countries with higher or lower support. If anything, the adverse effects of loneliness on mortality could be higher in high social support countries, which is consistent with the "social comparison" hypothesis: the effect of loneliness depends on the behavior of societal members, or "relevant others". This has been found to be true with respect to income (the utility brought by income is relative to the income of a certain reference group), unemployment (the unemployed suffer less from unemployment when relevant others are also unemployed).[3]The fact that this does not seem to be the case for social isolation, i.e. social isolation is associated with more acute mortality where others tend to be socially isolated too, suggests that the main channel behind the relationship between

[3] Clark (2003)

social isolation and mortality is not mental health, but rather that whether “others” meet almost never with friends or colleagues is a proxy for alternative social support in the country. In other words, where people do not have possibility to meet socially easily with other people, being socially isolated is worse.

3 Pathways from social isolation to death

3.1 Loneliness, health care utilization, and health behavior

The association we uncovered between social isolation and mortality was found to be robust to several definitions of the social isolation index, to the inclusion of all the confounders we suspected could be correlated both with social isolation and mortality (health at baseline, education, income and wealth, country of residence, etc.), to the exclusion of the first 48 months after baseline, and to the restriction to several subgroups.

Once this relationship has been established, the main question is that of the underlying mechanisms. It is often argued that loneliness, as the subjective feeling corresponding to social isolation, could be the channel through which objective social isolation impacts mortality. We test regressions in which we add the RUCLA scale of loneliness as a control: whether we include it as a binary or continuous variable (columns (1) to (3) of Table 5) the coefficient of the social isolation index is untouched, and loneliness by itself is not associated significantly with mortality. If we allow the loneliness measure to vary over time (column (4)), instead of being fixed at baseline, then a one-point increase in the R-UCLA scale is associated with a 10% increase of mortality, but reverse causality is more a concern as there is less time between loneliness measured and death observed, and even like this, the coefficient of the social isolation index is unchanged.

It therefore seems loneliness does not take any of the explanatory power of social isolation. We explore two other hypotheses in this section: that socially isolated individuals may have worse lifestyle, e.g. smoke more, drink more, or have a poorer diet, which could lead to putting them at a higher risk of mortality; and/or that because these individuals tend to live alone, have less contact with their children or with friends within associations, etc., they may lack incentives to visit the doctor, go to the hospital, or have preventive screening (mammogram, bloodstool test, etc.) as no one is there to prompt them to see one if they think they should.

We run principal component analyses and extract one unique factor of health behavior (using smoking, drinking, and sedentarism) and one unique factor of health care utilization (using number of visits to the doctor in past 12 months, whether was hospitalized over night in past 12 months, and number of drugs currently taking). We also extract the main factor of a principal component analysis using all the health outcomes we consider as relevant to have a precise idea of an individual's status health, which includes mental, cognitive, and physical health (and will be more detailed in the following section).

When adding in turn each of these three factors, at baseline, in our main Cox specification, social isolation does not lose any explanatory power, neither in magnitude nor in significance (see Table 6). Health status is as expected very much associated with mortality, but including this factor into the regression doe snot add anything to the previous models since we were already controlling separately for all the corresponding health variables. The more they use health care, the higher their mortality hazard, even controlling for health (reverse causality is here extremely likely). Health behavior is negatively associated with mortality, suggesting that when comparing individuals with a similar health status, smoking and drinking are associated with good outcomes (e.g. possibly though "social" drinking or smoking).

These results make it highly unlikely for either loneliness, health care utilization, or health behavior, to be important channels of the association between social isolation and mortality.

3.2 The dynamic impact of social isolation on health outcomes

As a second step in digging into the potential pathways from social isolation to health, we explore the dynamics of the association between social isolation and all relevant dimensions of health, some of which should show a significant decline (since social isolation leads to heightened mortality).

We choose to focus on important health indicators belonging to three dimensions of health: physical (including functional), mental, and cognitive health. Physical health is measured through self-assessed health and the number of chronic diseases (we construct a binary indicator for whether the individual has 2 or more chronic diseases). How many limitations in ADLs and IADLs ((instrumental) activities of daily living such as walking 100 meters, climbing stairs, dressing, bathing, eating, walking across a room, making telephone calls, etc.) an individual has gives us a measure of his/her functional health. We also construct an index of frailty that aggregates unintentional weight loss, self-reported exhaustion, weakness (grip strength), difficulties in walking, and low physical activity. Depression is measured using the EURO-D score, which sums symptoms such as suicidal thoughts, sadness, no hopes for the future, excessive guilt, sleep issues, fatigue, irritability, loss of appetite, tearfulness, concentration issues, lack of enjoyment, and difficulties keeping up interest in things. Cognitive health is an average of immediate and delayed word recall (how well individuals can remember a list of words they have just heard, and then again a few minutes afterward).

We regress each future health outcome $Health$ for individual i at time $t+j$ on social isolation SI at baseline t (the binary indicator that is equal to 1 whenever the index is non null) and other baseline characteristics, including the complete vector of health characteristics (amongst which the outcome at baseline), social isolation and the same control variables as before (age, sex, education categories, income and wealth quartiles, an employment dummy, and country and wave indicators).

$$Health_{i,t+j} = \alpha_0 + \alpha_1 SI_{i,t} + \alpha_2 Health_{i,t} + \alpha_3 X_{i,t} + \epsilon_{i,t} \tag{2}$$

where $j = 1, 2, ...6$.

Our sample is made of all individuals who were observed at least in two consecutive waves between 2004 and 2017 (discarding Wave 3 which was dedicated to collecting respondents' life histories), regardless of when they entered the study.

Graphical illustrations of these regressions are to be found as Figure 3, which plot the coefficient of social isolation over time, where "Time" indexes future waves, 1 for wave $t+1$, and so on, up to 6, for wave $t+6$ (individuals are observed at most from wave 1 to wave 7). In all our regressions, the outcome is measured at one of these future waves, while the rest of the variables are fixed at baseline. All health outcomes are standardized so that their mean is 0 and standard deviation is 1, which makes the graphical representations of our regressions more comparable to each other. They are coded so that higher values mean worse health, and represented using the same scale on all graphs. The number of observations decreases over time as 90,000 individuals are observed through 2 consecutive waves, while only 8.000 are observed 6 waves after their entry (which does not imply participating to all waves in between). Confidence intervals are therefore larger over time, but we still get a clear picture of how social isolation correlates with health over time.

Social isolation is undeniably associated with worsening health. Their dynamics vary slightly across outcomes. Cognitive health starts worsening in association with social isolation after one wave, reaches a plateau after two waves, in line with Shankar et al. (2013), which finds poorer cognitive functioning amongst the socially isolated four years after baseline using the ELSA data. Other outcomes, such as frailty, or self-assessed health, follow a similar trend, while some go back to their initial level, e.g. depression. It therefore seems social isolation physical and cognitive health in the long run, but its association with mental health is only transitory. Some outcomes take longer to be affected by social isolation: chronic diseases (having 2 or more of those) are associated with social isolation only 4-5 waves after baseline, which makes sense as it usually takes some time for a cancer or other kinds of chronic diseases to develop. Functional health, measured as "suffering at least one limitation" becomes more and more correlated with social isolation as time goes by, before possibly going back to the baseline level (the precision of the estimates does not allow us to derive any conclusion after 6 waves of follow-up). It is noteworthy that the relationship between functional health and social isolation is quite sensitive to the definition of functional health: when it is defined as the sum of limitations with ADLs, our estimates are much closer to being non-significantly different from 0, in line with Shankar et al. (2017), which does not find a significant association between number of ADLs and social isolation using two waves of the ELSA data.

Digging into the association between social isolation and health over such a long horizon (up to 14 years) allows us to discard the "mental health" channel as main mechanism behind this association. Physical and cognitive health seem much more likely to be the pathways from social isolation to a higher mortality.

Apart from the "biological" channel- through mechanisms such as inflammation and hormone responses to social isolation- the literature puts forward health behavior and health care utilization as potential pathways from social isolation to worsened health and mortality. In the mortality section, we already showed neither health care utilization nor health behavior seemed to mediate the association between social isolation and mortality, but this does not mean there is no specific pattern of the socially isolated in terms of health behavior or health care. We apply the same dynamic analysis to the variables that were used to construct a health behavior and health care utilization factors. As shown in Figure 4, there is no significant relationship between social isolation and smoking, except after one and three waves, but the dynamic pattern is unclear. If anything, social isolation seems to be associated with less drinking. The one important behavior that is increasingly and importantly associated with social isolation over time is sedentarism, defined as engaging in vigorous (e.g. sports) or moderate (e.g. gardening, going on a walk) physical activity one to three times a month, or never. Sedentarism may then play a role in how socially isolated individuals become sicker, but it is also reasonable that as individuals get sicker they would engage less in physical activity. A study by Shankar et al. (2011) finds similar results on inactivity, and a more clear-cut association with smoking, concluding that loneliness and social isolation may affect health independently through their effects on health behaviors. One way to check whether sedentarism is a mechanism per se is to control for the health factor at future waves as well, on top of at baseline. When doing so, the trend looks the same, but the coefficient is no longer significant, so the association between social isolation and sedentarism could also be spurious due to their common correlation with a worsened health status.

Last, Figure 5 points at socially isolated people not using more or less health care than non socially isolated individuals. Whether we include future health as a control or not in the regressions, socially isolated individuals do not stay more or less at the hospital than the non-socially isolated, and neither do they visit more or less their physician. If anything, they use slightly more drugs, after a few waves, but the relationship is weak (both in terms of significance and magnitude). This finding echoes some of the literature that points at lonely or socially isolated individuals using more health care than individuals who do not suffer from loneliness or social isolation. One example is Gerst-Emerson and Jayawardhana (2015), which finds that the lonely are more likely to visit their physician (but not to be hospitalized), even controlling for their health, suggesting that individuals who suffer from chronic loneliness look for social support in their physician, but that lack of health care use or barriers to health care access do not seem to drive the social isolation-health relationship. Even though health care utilization does not seem to be the sole driver of that relationship, the fact that socially isolated individuals do not use more health care although their health worsens across all dimensions is quite striking, and points at health care utilization being a potential channel.

4 Discussion: the causality challenge and other potential caveats

We uncovered a strong association between social isolation and mortality (a 25% increase in the mortality hazard rate for individuals who are socially isolated at baseline). How causal can we prove this association to be? In order to make a stronger case for a causal association, we discard all people who die in the 48 months following baseline (when social isolation is observed). This way, we make sure our sample does not suffer any life-threatening health condition that would not be captured in our health controls (which would be hard given the extensive set of health information we use) and that would still be the reason why one is socially isolated. This restriction, coupled with a very long follow-up period (up to 13.5 years, with a median follow-up of 65 months), makes it hard to believe in reverse causality “causing” our estimate.

The main concern is the potential existence of omitted variables that would affect both social isolation and mortality (or health outcomes). It is actually hard to come up with potential confounders that are not controlled for in our regressions and would be correlated with both baseline social isolation and future health: we are already controlling extensively for health at baseline, but also for socio-economic status, through income, wealth, and education, and for other observable characteristics that could be related to both the main explanatory variable and the outcome, such as gender, whether married, whether working, whether has children. An example of unobservable that could determine social isolation is personality: introverted individuals are probably more at risk of social isolation, as they would be less prone to participate to social activities for instance. If introverted people are also less healthy, then we would be in a situation of a relevant omitted variable. Nevertheless, it is quite unlikely that introverted individuals are less healthy for reasons that would not be linked with depression, which we control for at baseline. And second, introversion is part of the “Big 5” personality traits that are measured in Wave 7. Supposing they are stable over time (Cobb-Clark and Schurer (2012) shows they are for at leats a 4-year period), we can including the Big 5 items as controls. This does not change our social isolation coefficient.

A second more convincing example of potential omitted confounder is genetics: what if the same genes that are overexpressed in socially isolated individuals are also responsible for activating the immune system and the inflammation mechanism in the body? This is what has been found in Cole et al. (2007) amongst a sample of 230 Americans aged 50-67 years, which explains why lonely people suffer from chronic inflammation in spite of their high levels of cortisol and are vulnerable to microbes, viruses, and other sources of tissue damage.[4] Then genes could be an important source of omitted variable bias in our study, if they determined both social isolation and adverse health outcomes.

Any non randomized experiment carries its lot of concerns about omitted variable bias. In practice, several approaches can help mitigate these concerns. The most straightforward way consists in including an appropriate set of observable controls (Angrist and Pischke (2010)), for instance when we include frailty and chronic diseases along with functional, self-assessed, mental, and cognitive health in order to capture the true health status of the individual. An additional approach that has been more than widely used in the empirical literature relies on demonstrating the stability of the key coefficient faced with the inclusion of additional controls. Table A4 show the coefficient of the social isolation index remains quite stable over different specifications when adding a different subset of controls at each specification. For this table we chose to regress self-assessed health five waves ahead on social isolation at baseline (at entry into the study) and other controls, so that the coefficient in column (4) corresponds to the point at $time = 5$ in the "self-assessed health" graph in Figure 3, but the same stability could be shown for the other outcomes at other times.

As put forward in Oster (2019), although very intuitive, this idea relies on the se-

[4]This study looks at chronically lonely individuals, according to the RUCLA scale of loneliness, rather than at socially isolated individuals.

lection on observables being informative about the selection on unobservables, which is not implied by the baseline assumptions of the linear model. We therefore appeal to Oster's use of coefficient stability as a test for selection on unobservables. The test assesses both the stability of the estimated social isolation treatment effect when adding key observables and the importance of these factors in explaining health outcomes. The estimate of the coefficient of proportionality proposed by Oster as a summary of the robustness of results is 1.11, i.e. higher that the proposed cutpoint of 1 Oster proposes as a standard for robustness). This value implies that the unobservables would need to explain 11% more than the observables in order for the treatment effect to be zero, which seems quite unlikely given the quality and richness of our data, and the goodness of fit of our models (R-squared over 30 in the health regressions corresponding to Figure 3). Furthermore, to obtain this value of 1.11, we chose a multiplicative factor for the R-squared of 1.25, which allows the maximum R-squared to be 25% higher than the actual R-squared, on the grounds that about "40% of results would not survive" that threshold in a sample of 76 results extracted from 26 highly-cited articles published in "Top-5" Economics journals. We are therefore confident our results are not biased by unobservable omitted variables that would determine selection in to the social isolation "treatment".

5 Conclusion

In this paper we find a strong association between social isolation and future mortality, which is not solely mediated by loneliness, health behavior, nor health care utilization. We also explore the dynamics of the health impact of social isolation, and find social isolation to lead to a worsening of all the facets of health we consider (self-assessed, frailty, cognitive, mental, functional). The finding that socially isolated individuals do not resort more to health care use although their health worsens across all dimensions compared to non socially isolated individuals suggests that health care utilization might be a channel underlying the relationship between social isolation and health. We use the richness of our data to investigate heterogeneity in the social isolation-health relationship across countries, and find a much stronger association between social isolation and mortality in Eastern countries. That one same -objective- measure of social isolation does not lead to the same health consequences across countries, albeit using harmonized data, points at public health policies having a role to play in moderating the health risks posed by social isolation.

References

Angrist, Joshua D and Jörn-Steffen Pischke, "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of economic perspectives*, 2010, *24* (2), 3–30.

Clark, Andrew, "Unemployment as a Social Norm: Psychological Evidence from Panel Data," *Journal of Labor Economics*, 02 2003, *21*, 289–322.

Cobb-Clark, Deborah A and Stefanie Schurer, "The stability of big-five personality traits," *Economics Letters*, 2012, *115* (1), 11–15.

Cole, Steve W, Louise C Hawkley, Jesusa M Arevalo, Caroline Y Sung, Robert M Rose, and John T Cacioppo, "Social regulation of gene expression in human leukocytes," *Genome biology*, 2007, *8* (9), R189.

Fried, Linda P., Catherine M. Tangen, Jeremy Walston, Anne B. Newman, Calvin Hirsch, John Gottdiener, Teresa Seeman, Russell Tracy, Willem J. Kop, Gregory Burke, and Mary Ann McBurnie, "Frailty in Older Adults: Evidence for a Phenotype," *The Journals of Gerontology: Series A*, 03 2001, *56* (3), M146–M157.

Gerst-Emerson, Kerstin and Jayani Jayawardhana, "Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults," *American journal of public health*, 2015, *105* (5), 1013–1019.

Holt-Lunstad, Julianne, Timothy B Smith, Mark Baker, Tyler Harris, and David Stephenson, "Loneliness and social isolation as risk factors for mortality a meta-analytic review," *Perspectives on Psychological Science*, 2015, *10* (2), 227–237.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Shankar, Aparna, Anne McMunn, James Banks, and Andrew Steptoe, "Loneliness, social isolation, and behavioral and biological health indicators in older adults.," *Health Psychology*, 2011, *30* (4), 377.

_ , _ , Panayotes Demakakos, Mark Hamer, and Andrew Steptoe, "Social isolation and loneliness: Prospective associations with functional status in older adults.," *Health psychology*, 2017, *36* (2), 179.

_ , Mark Hamer, Anne McMunn, and Andrew Steptoe, "Social isolation and loneliness: relationships with cognitive function during 4 years of follow-up in the English Longitudinal Study of Ageing," *Psychosomatic medicine*, 2013, *75* (2), 161–170.

Steptoe, Andrew, Aparna Shankar, Panayotes Demakakos, and Jane Wardle, "Social isolation, loneliness, and all-cause mortality in older men and women," *Proceedings of the National Academy of Sciences*, 2013, *110* (15), 5797–5801.

Valtorta, Nicole K, Mona Kanaan, Simon Gilbody, Sara Ronzi, and Barbara Hanratty, "Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies," *Heart*, 2016, pp. heartjnl–2015.

6 TABLES AND FIGURES

Figure 1: Social Isolation and Loneliness across Europe (18 countries)

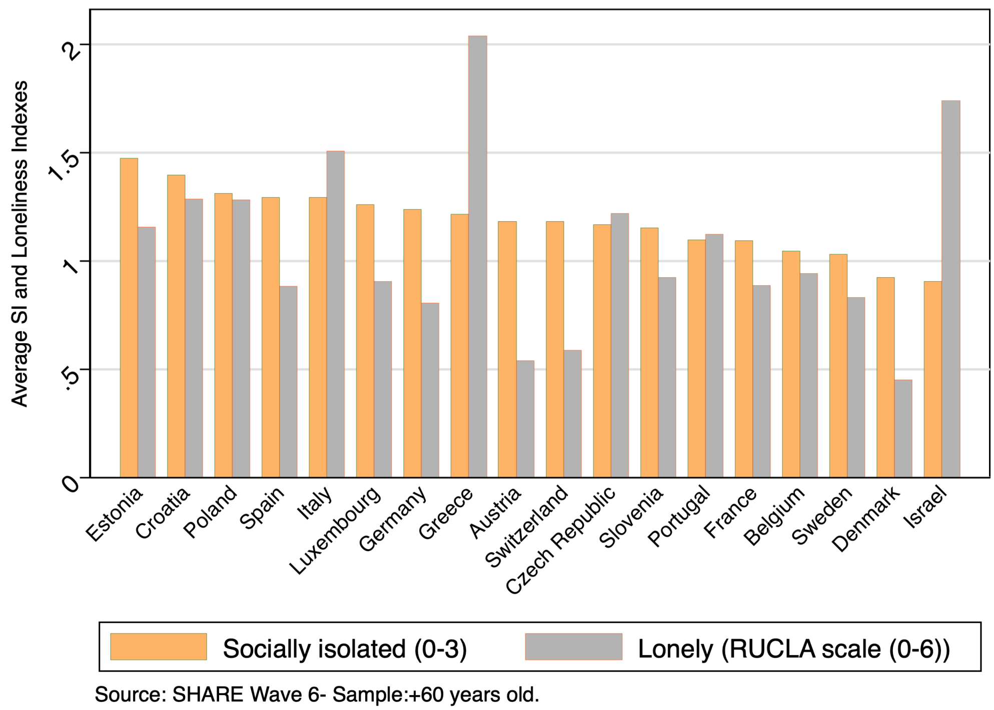
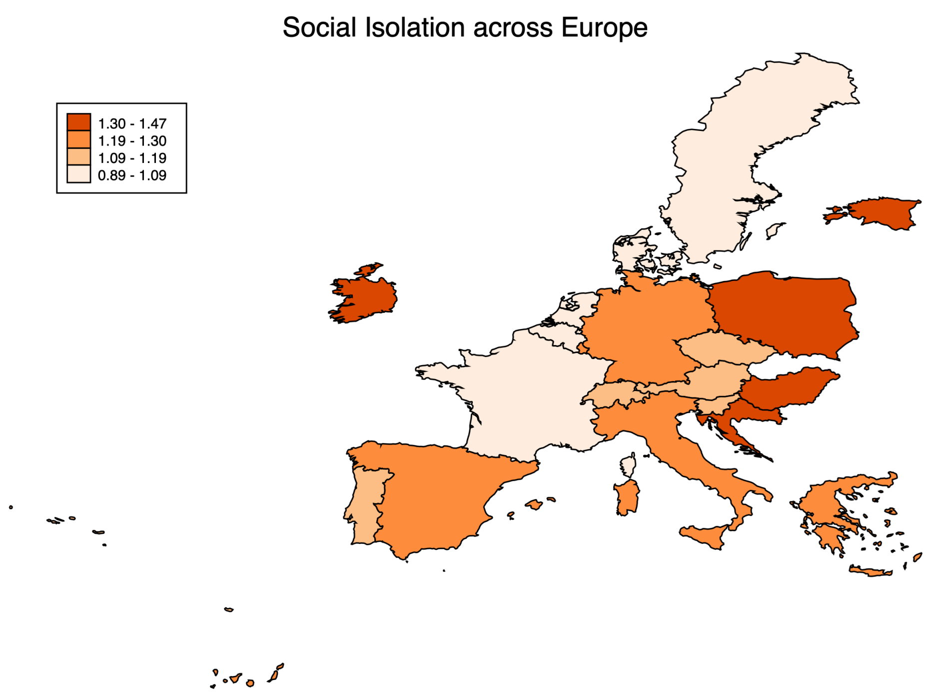


Figure 2: Social Isolation across Europe (20 countries)



Source: SHARE Wave 6 for 17 countries, Wave 2 for Ireland, Wave 4 for Hungary, Wave 5 for The Netherlands

Table 1: Descriptive Statistics: Health status, by Social Isolation level, at baseline

	SI Index= 0		SI Index> 0	
	mean	sd	mean	sd
Health				
Frailty index	0.56	0.81	1.07	1.20
Self-perceived health	2.80	1.01	3.29	1.07
Number of chronic diseases	1.41	1.36	1.76	1.60
Number of limitations	0.16	0.70	0.72	1.96
Depression score (EuroD)	1.98	1.87	2.77	2.47
Cognitive recall test	4.93	1.68	3.95	1.85
Socio-demographic controls				
Age in years	61.10	8.65	65.61	10.99
Female	0.45	0.50	0.56	0.50
No education	0.02	0.15	0.07	0.25
Education: Primary	0.10	0.30	0.26	0.44
Education: Secondary(i)	0.14	0.34	0.18	0.38
Education: Secondary(ii)	0.38	0.49	0.33	0.47
Education: Tertiary	0.35	0.48	0.16	0.37
Employed or Self-employed	0.46	0.50	0.27	0.44
Has at least one child	1	0	0.82	0.38
Income and Wealth				
Total household income	50787	82268	28193	63750
Household net worth	337531	500443	190419	427804
Observations	25540		73177	

Table 2: Cox models: Effect of Social Isolation at Baseline on Mortality up to Wave 7.

	(1)	(2)	(3)	(4)	(5)	(6)
Social Isolation index	1.257***	1.241***	1.153***	1.156***	1.308***	1.261***
	(0.017)	(0.020)	(0.021)	(0.024)	(0.051)	(0.058)
Basic Demo	yes	yes	yes	yes	yes	yes
Socio Demo	no	yes	yes	yes	yes	yes
Health	no	no	yes	yes	yes	yes
Follow-up	no	no	no	>24 mo	>24 mo	>48 mo
Discrete SI index	no	no	no	no	yes	yes
Observations	210205	205299	189936	153037	158796	103957
Individuals	83652	81695	74631	64728	66432	48721

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: "Basic Demo" controls: age, age squared, sex. "Socio Demo" controls: education categories, whether employed, whether has at least one child, income and wealth quartiles. "Health" controls: frailty, self-assessed health, number chronic diseases, number limitations, depression score, cognitive recall test. All regressions include wave and country-specific fixed effects.

Table 3: Cox models: Country Heterogeneity of the Impact of Social Isolation at Baseline on Mortality up to Wave 7.

	(1)	(2)	(3)	(4)	(5)
Socially Isolated: SI index>0	1.30***	1.18**	1.36***	1.30***	1.25***
	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)
Socially Isolated X Western Countries		ref.	ref.		
Socially Isolated X Southern countries		1.02	0.88		
		(0.12)	(0.09)		
Socially Isolated X Northern countries		1.20*		1.09	
		(0.12)		(0.10)	
Socially Isolated X Eastern countries		1.45***			1.36***
		(0.16)			(0.14)
Western Countries	ref.	ref.	ref.	ref.	ref.
Southern countries	1.14***	1.13	1.27**	1.13***	1.14***
	(0.04)	(0.12)	(0.13)	(0.04)	(0.04)
Northern countries	1.58***	1.43***	1.66***	1.55***	1.65***
	(0.07)	(0.13)	(0.07)	(0.13)	(0.07)
Eastern countries	1.77***	1.04	1.44***	1.44***	1.10
	(0.06)	(0.11)	(0.06)	(0.06)	(0.11)
Observations			152538		
Individuals			63788		

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and wave and country group fixed effects.

Follow-up restricted to being > 24months.

Table 4: Country Heterogeneity: Impact of Social Isolation on Future Mortality, in countries where social support is lower or higher than average.

	(1)	(2)	(3)	(4)
Index	1.31***	1.21***	1.21*	1.28*
	(0.05)	(0.05)	(0.12)	(0.18)
Index X Low Social Support Countries		1.33***		0.90
		(0.12)		(0.17)
Index used :	SI index>0	SI index>0	RUCLA index>6	RUCLA index>6
Observations	158796	158796	50527	50527
Individuals	66432	66432	32868	32868

Exponentiated coefficients; Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and wave and country fixed effects. .

Follow-up restricted to being > 24months.

Table 5: Cox models: Does Loneliness mediate the association between Social Isolation and Mortality

	(1)	(2)	(3)	(4)
Social Isolation index	1.27***	1.27***	1.27***	1.28***
	(0.10)	(0.10)	(0.10)	(0.06)
RUCLA Loneliness at Baseline		1.03		
		(0.02)		
RUCLA Loneliness at Baseline (d)			1.18	
			(0.12)	
RUCLA Loneliness- Time-Varying				1.09***
				(0.01)
Observations	48255	48255	48255	94662
Individuals	31272	31272	31272	54949

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and wave and country fixed effects.

Follow-up restricted to being > 24months.

Table 6: Cox models: Health, Health Care, and Health Behavior as Potential Mediators for the association between Social Isolation and Mortality

	(1)	(2)	(3)	(4)
Social Isolation index	1.29***	1.26***	1.26***	1.25***
	(0.07)	(0.07)	(0.07)	(0.07)
Health Factor		1.70***	1.61***	1.58***
		(0.04)	(0.04)	(0.04)
Health Care Factor			1.18***	1.18***
			(0.03)	(0.03)
Health Behavior Factor				0.92**
				(0.03)
Observations			121556	
Individuals			62029	

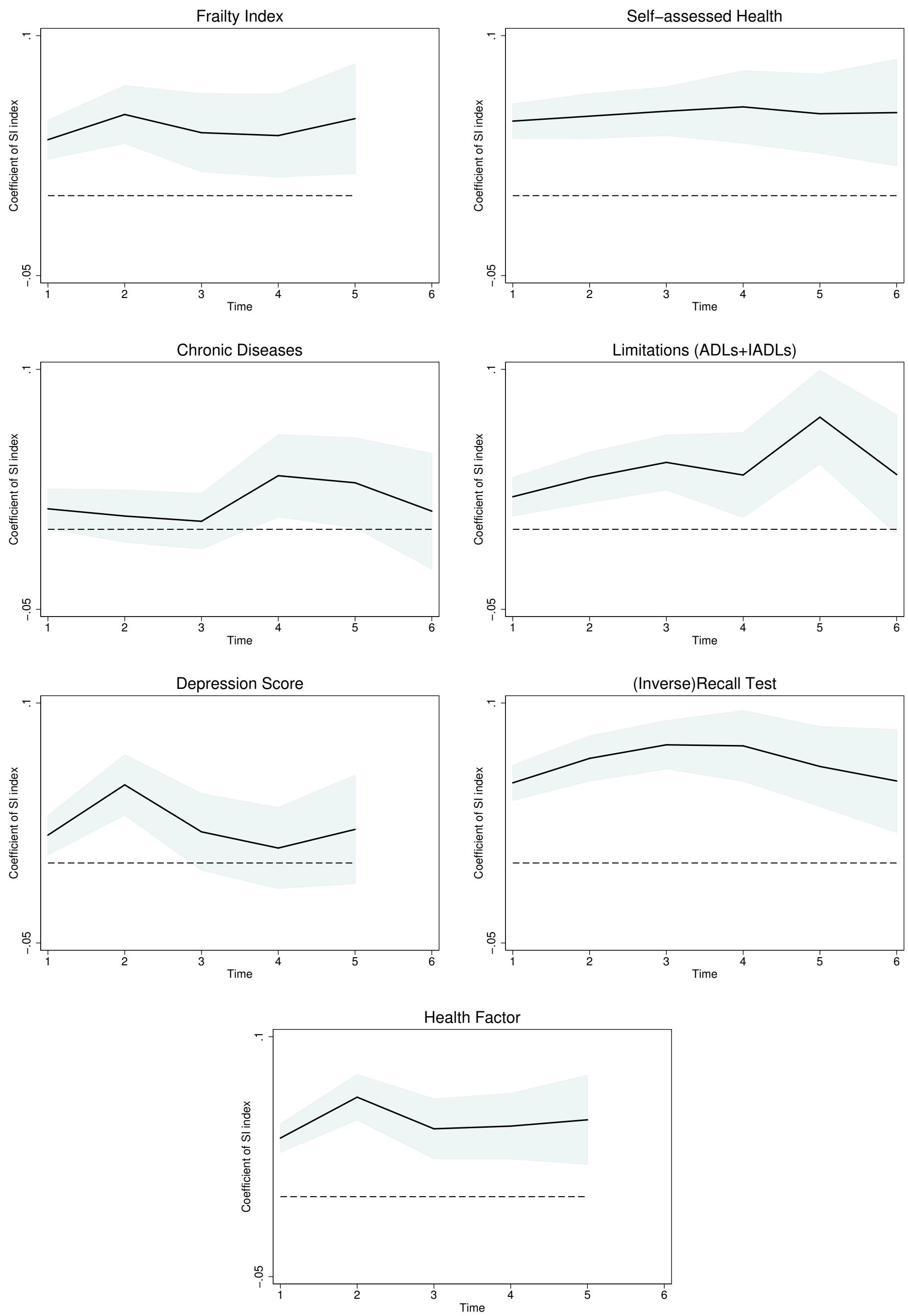
Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and wave and country dummies.

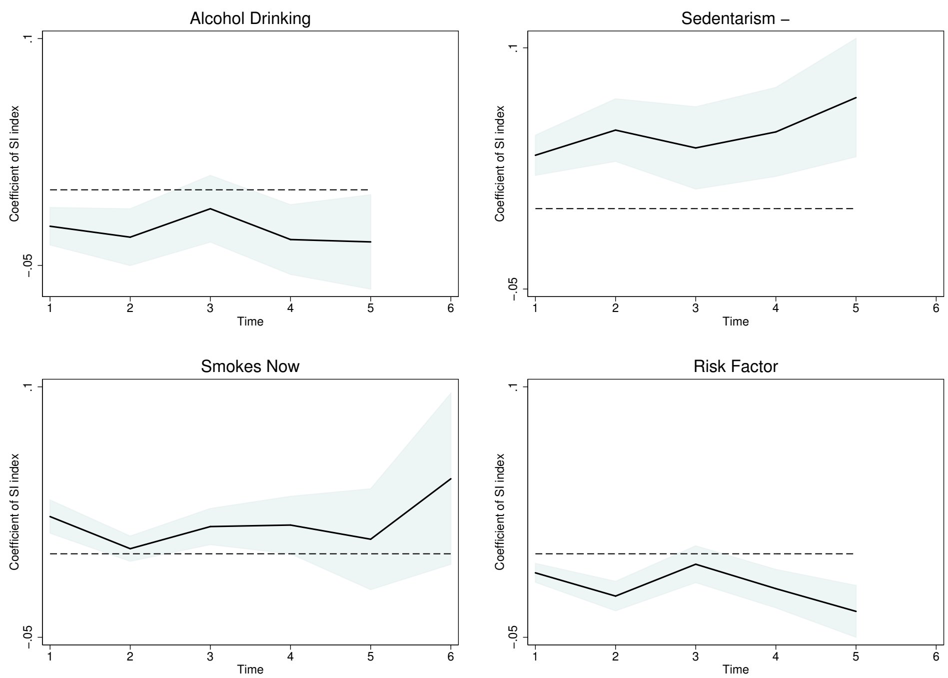
Follow-up restricted to being > 24months.

Figure 3: Social Isolation at Baseline and Health Dynamics



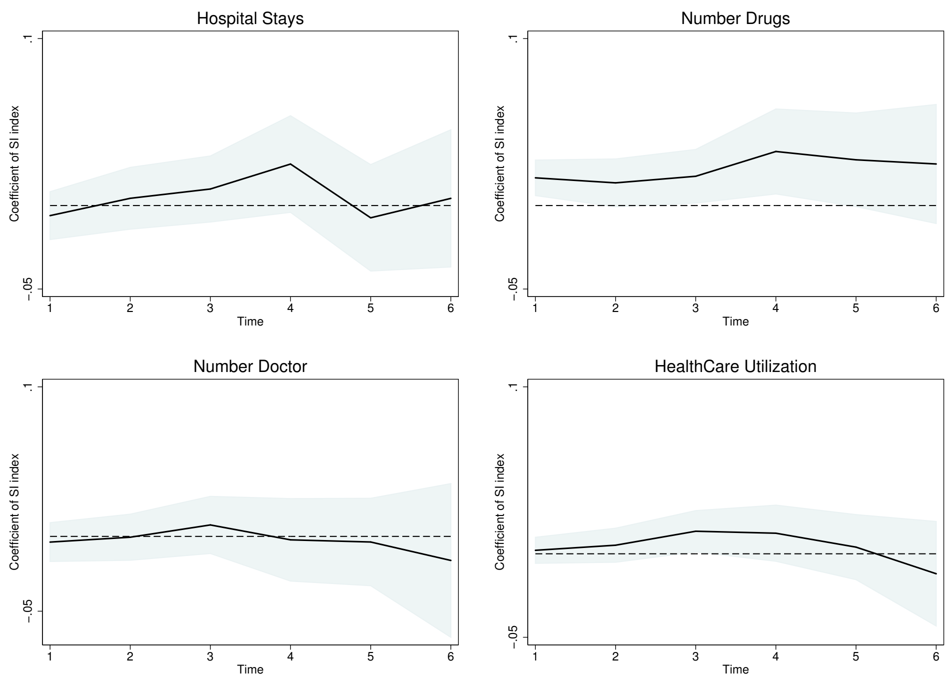
Note: The figures displays the coefficients and 95 percent confidence intervals for the effect of social isolation on different health outcomes. "Time" indexes future waves: 1 is wave $t+1$, 2 is wave $t+2$, etc.Regressions are done separately for each outcome and lag. All regressions include the variables displayed in Table 1 and wave (a dummy for which wave is baseline) and country dummies.

Figure 4: Social Isolation at Baseline and Health Behavior Dynamics



Note: The figures displays the coefficients and 95 percent confidence intervals for the effect of social isolation on different health behavior outcomes. “Time” indexes future waves: 1 is wave $t+1$, 2 is wave $t+2$, etc.Regressions are done separately for each outcome and lag. All regressions include the variables displayed in Table 1 and wave (a dummy for which wave is baseline) and country dummies.

Figure 5: Social Isolation at Baseline and Health Care Utilization Dynamics



Note: The figures displays the coefficients and 95 percent confidence intervals for the effect of social isolation on different health care utilization outcomes. "Time" indexes future waves: 1 is wave $t+1$, 2 is wave $t+2$, etc.Regressions are done separately for each outcome and lag. All regressions include the variables displayed in Table 1 and wave (a dummy for which wave is baseline) and country dummies.

APPENDIX

Table A1: Cox models: Sensitivity Analysis over the Social Isolation Index.

	(1)	(2)	(3)	(4)
SI index	1.261***	1.229***	1.057***	1.175***
	(0.058)	(0.057)	(0.012)	(0.039)
SI index 2		1.382***		
		(0.076)		
SI index 3		1.459***		
		(0.127)		
SI index	SI>0	SI=1	Long SI index	Long SI index>5
SI index 2		SI=2		
SI index 3		SI=3		
Observations	103957	103957	103957	103957
Individuals	48721	48721	48721	48721

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and country dummies.

Follow-up restricted to being > 48months.

Table A2: Cox models: Effects of every item of the Social Isolation Index at Baseline on Mortality up to Wave 7.

	(1)	(2)	(3)	(4)
Lives alone	1.08**			1.08**
	(0.04)			(0.04)
No social activities		1.29***		1.28***
		(0.05)		(0.05)
Infrequent contact with children			1.06	1.05
			(0.06)	(0.06)
Observations	114527	114474	104001	103957
Individuals	52918	52886	48748	48721

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and country dummies.

Follow-up restricted to being > 48months.

Table A3: Cox models: Sensitivity Analysis over specific subsamples.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SI index>0	1.35***	1.21***	1.26***	1.24	1.25***	1.25*	1.62***	1.22***
	(0.11)	(0.07)	(0.06)	(0.76)	(0.07)	(0.15)	(0.22)	(0.06)
Restricted to	Female	Male	Has children	No children	Married	Not married	Em-ployed	Not employed
Observations	58258	45699	95392	8565	59176	27755	31252	72705
Individuals	27093	21628	45737	4977	33254	15854	14585	34136

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls included: all variables displayed in Table 1, and country dummies.

Follow-up restricted to being > 48months.

Table A4: Stability of Key coefficient: Effect of Social Isolation at Baseline on Self-assessed Health Five waves Ahead.

	(1)	(2)	(3)	(4)
Social Isolation index	0.098***	0.086***	0.067***	0.076***
	(0.015)	(0.014)	(0.015)	(0.020)
Demo + baseline SA Health	yes	yes	yes	yes
All Health	no	yes	yes	yes
Socio Demo	no	no	yes	yes
Discrete SI index	no	no	no	yes
Observations	9750	9750	9750	9750

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$