I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15629

Social Isolation, Health Dynamics, and Mortality: Evidence across 21 European Countries

Yarine Fawaz
Pedro Mira

OCTOBER 2022

DISCUSSION PAPER SERIES

IZA DP No. 15629

Social Isolation, Health Dynamics, and Mortality: Evidence across 21 European Countries

Yarine Fawaz
CEMFI and IZA

Pedro Mira
CEMFI

OCTOBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Social Isolation, Health Dynamics, and Mortality: Evidence across 21 European Countries*

We provide a comprehensive picture of the health effects of social isolation using longitudinal data over 21 European countries (SHARE). First, using Cox regressions, we find a significant, strong and robust association between our social isolation index and mortality, which is much stronger 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. We then estimate linear regressions to study the dynamic "value added" effects of SI on health and other mediator outcomes, and find that social isolation at baseline leads to worsening health in the next waves along all the dimensions we observe. Up to 13 percent of the effect of baseline social isolation on mortality can be imputed to the combined one-wave-ahead impact of social isolation on increased frailty, reduced cognitive function and increased smoking.

JEL Classification: I10, C41
Keywords: social isolation, loneliness, health, mortality, SHARE

Corresponding author:
Yarine Fawaz
CEMFI
Casado del Alisal, 5
28014 Madrid
Spain
E-mail: fawaz@cemfi.es

* We are very grateful to participants at the 4th PHS Workshop at University of Pennsylvania, the 2020 Alp-Pop conference, the SHARE user conference, and seminars at CEMFI and CUNEF, for the helpful feedback of many participants. We particularly thank Atheendar Venkataramani, Ryan Brown, Jerome Adda, Keith Head, for their constructive remarks and suggestions. Yarine Fawaz and Pedro Mira acknowledge funding from the Spanish Ministry of Science and Innovation (grant PGC2018-097598-B-I00).

1 Introduction

One collateral damage of Covid-19 -or rather of social distancing- is social isolation, and the health hazards that have been shown to be associated with it. There should be much concern about the health toll of social isolation, especially in times of coronavirus pandemic, as those that are left isolated at home are likely to be those who were already quite isolated, i.e. those living alone, those without close family checking on them, those whose social ties were already few. Furthermore, as the virus hits more severely the elderly, the latter have undergone a stricter and longer period of social distancing than the younger, which puts them at risk of more acute social isolation, whose consequences in terms of mental and physical health should then be explored with the best existing data and scientific rigor.

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)). Ever since social sciences (Berkman and Syme (1979)) uncovered the importance of social ties as predictors of survival for persons aged less than 70 at baseline, the impact of loneliness and social isolation on health and mortality has been increasingly investigated in public health, medicine, epidemiology, gerontology and other health-related disciplines over the last decade. The bulk of evidence has pointed 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, fatigue, pain, cognitive decline, Alzheimer's disease, and even death (Steptoe et al. (2013); Cohen et al. (1997); Shankar et al. (2013); Heffner et al. (2011); Yu et al. (2020); Teguo et al. (2016); Powell et al. (2021)). In contrast, the economics literature on the topic has remained relatively scarce.

According to one 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) -another recent meta-analysis- found that social 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 21 European countries to investigate the association between social isolation and mortality and health.[2] The SHARE

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

[2] Loneliness will also be considered, but rather as a mediator in the association under study. Our focus is on objective social isolation, which we define according to objective criteria such as living alone, participating in social activities, and frequency of contact with family.

(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 18 years, from 2004 to 2021. 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 noteworthy 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.

Relative to those studies that use a representative longitudinal dataset, we make several contributions:

First, we present a rigorous analysis of the effects of social isolation at baseline on mortality over a long follow-up period, in a harmonized multi-country framework, including novel findings on the heterogeneous effects of SI across country groups.

Second, on top of the baseline health controls included in the main mortality analysis, we look into health behaviors, health care utilization, loneliness, and a poor social network as additional potential mediators of the relationship between SI and mortality.

Third, we run linear regressions to study the dynamic "value added" effects of SI on health and other mediator outcomes, i.e. the effect of baseline SI on all observable dimensions of health, health behaviors and health care utilization at each future wave when controlling for their baseline levels. We then combine the Cox analysis with the dynamic regressions to compute a simple metric of how much of the SI effects on mortality can be imputed to the one-wave ahead effect of baseline SI on each dimension of health, health behaviors and health care utilization. This metric can serve as a guide to where it is more necessary to intervene in order to curb the detrimental effects of social isolation.

We find a significant, strong and robust association between our social isolation index and mortality. A striking finding is in uncovering heterogeneous effects of social isolation on mortality across countries. 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. 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.

Remarkably, controlling for loneliness barely weakens the relationship between our social isolation index and mortality - 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 at baseline in our regressions does not affect the coefficient on social isolation. Likewise, health care utilization does not seem a major channel for the effect of social isolation on future health. On the one hand we find that the socially isolated individuals do not use more health care services than their non-socially-isolated counterparts, with the exception of prescription drug consumption. This is so in spite of the fact that their health keeps worsening, which suggests that social isolation might inhibit the use of some health care services. But on the other hand including measures of current health care use in our regressions does not affect the coefficient on social isolation.

We also find that social isolation at baseline leads to worsening health in the next waves along all the dimensions we observe, and these effects are quite persistent. Up to 13 percent of the effect of baseline social isolation on mortality can be imputed to the combined one-wave-ahead impact of social isolation on increased frailty, reduced cognitive function and increased smoking. On top of the traditional robustness checks (using different specifications, over different samples, and showing the stability of our key coefficient), we provide more evidence in support of a causal interpretation of our estimates using Oster's test for selection on unobservables (Oster (2019)). We also use education as a benchmark for the health effects of social isolation. We find the education gradient in mortality is smaller than the social isolation gradient, but the association of education with future health is stronger than the one we find for social isolation in dynamic value added regressions.

The rest of the paper is structured as follows: section 2 establishes a link between social isolation at baseline and future mortality, controlling for a rich set of socioeconomic and health indicators that cover physical, functional, mental and cognitive health at baseline. In section 3 we check for more potential mediators of the social isolation-mortality association, by adding them as baseline controls in the Cox regressions. Section 4 presents the dynamic regressions of health and other mediating variables on the SI index, with the same baseline controls as in the Cox regression and the baseline value of the mediating variable, and we present a metric of how much of the SI effects on mortality can be imputed to the effect of SI at baseline on each dimension of health, health behaviors and health care utilization, in the next wave. Last, section 5 discusses the causality challenge, Oster's test of selection on unobservables, and education as a benchmark of the SI effects on mortality and health outcomes. Section 6 concludes.

2 Social isolation and mortality

2.1 Data

We use longitudinal survey data from SHARE, over 8 waves from 2004 to 2019, plus the two "Corona" waves of Spring 2020 and Summer 2021, across 20 European countries plus Israel. 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 530,000 interviews). Our sample is made of 67,676 non-institutionalized (i.e. not living in nursing home, at least at baseline) individuals, corresponding to 243,515 observations, whom we observe at least twice over the 10 waves (the second time might be an "exit" interview, i.e. a post-mortem interview) and whose information on the set of variables we use is non missing (see Table 2 for more details). [3] Individuals enter the study at any wave between wave 1 (in 2004) and wave 6 (in 2015), and are followed a minimum period of 24 months, due to the minimum follow-up restriction we impose.[4] The maximum follow-up time is 207 months, i.e 17 years and 3 months. Median follow-up time is 79 months, i.e. about 6 years and a half. Over the course of our study, we observe 9,802 deaths, which corresponds to 14 per cent of our sample.

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

[3] Although SHARE now encompasses 29 countries and we do make use of all waves, we can only exploit data on the 21 countries that entered SHARE before the last wave (we need more than one observation per individual), and appeared at least twice across the 8 first waves, non including Wave 3, which was dedicated to constructing life histories of SHARE respondents.

[4] Otherwise, minimum follow-up time until death would be 1 month, which seems too little for obvious reverse causality concerns. We argue in favor of a 24 months period when presenting the Cox model in Section 2.2.

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 summarized by the EURO-D score, which is the sum of 12 items that can be relied on to diagnose depression in older adults, 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 functioning is an average of immediate and delayed word recall (i.e. the number of words an individual is able to remember out of a list of ten words). 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 not have children), and does not participate in any social activities (including political, sports, educational or voluntary work activities).[5] The resulting index lies between 0 and 3, with a 1.03 mean and 0.79 standard deviation for our sample at baseline. Higher values indicate increasing social isolation. Table 1 displays the distribution of our SI index. In our sample, 20 per cent of individuals live alone, 61 per cent participate in no social activities, and 20 per cent have infrequent contact with their children (13 per cent due to not having children). More than half our sample is massed at $SI = 1$, and very few individuals have the maximum value of 3 (4 per cent). More interesting is that the jump between a 0 and a 1 value of the social isolation index is driven mostly by the non-participation to social activities, while the jump from 1 to 2 is due to a shift in both the other two components, i.e. living alone and having low contact with children (or not having any children).

We also use the short form of the R-UCLA loneliness scale, which was created by summing 3 items -how much of the time they felt a sense of being left out, a lack of companionship, and isolation- into one single measure of loneliness.[6] We can therefore investigate the effects both of objective social isolation and of the perception of social isolation, on mortality.

Table 2 provides descriptive statistics at baseline (when individuals enter the data) on the SI-related variables, 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 who 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 (note that cognitive functioning is the only health measure where a higher value means better health), are more likely to be female, less educated, and childless than the non socially isolated. Regarding gender and social isolation, the reality is more complex, as women become socially isolated at a higher rate than men as they grow older (past 60), which is likely to reflect gender imbalances in the probability of widowhood after age 60. We also add an employment indicator as being working or retired may explain an important part of social isolation for the over-50 population.

Figure 1 and A2 (the latter in Appendix) 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

[5] Note that this index is quite close to the original Berkman-Syme Social Network Index developed in Berkman and Syme (1979) for a population aged less than 70, which included 1) marital status, 2) contacts with close friends and relatives 3) membership in a church group and 4) memberships in other types of groups). We do not include contact with other family or friends because these items were absent from SHARE until Wave 4, when a Social Networks module was introduced for the first time (it appears again in waves 6 and 8). We will use that module when creating an index of connectedness, but we do not include any item from that module in our main SI index so that we can follow respondents for a much longer time span.

[6] See Hughes et al. (2004) for a validation of the short version of the RUCLA scale of loneliness.

loneliness and social isolation at the individual level being of (only) 24%, it seems that objective and subjective measures of social isolation capture different aspects of social experience, as suggested in Hughes et al. (2004). As a consequence, 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 July 2021. Out of the 67,676 individuals we follow, 9,802 die over the period. Although our longitudinal data would allow us to let our explanatory variables vary across time, we keep them fixed at baseline, which is important in order to introduce some distance between the covariates, more particularly social isolation, and the outcome, i.e. mortality. In our prefereed specification, we impose a minimum of 24 months of follow-up between the moment social isolation is measured, and mortality, following the robustness checks performed in Steptoe et al. (2013): "we repeated the analysis excluding deaths within 24 months of baseline, and the results were very similar results to those for the full cohort, suggesting that existing terminal illness is not the primary explanation.". This helps to alleviate reverse causality concerns.

A potential concern may arise as to how stable our measure of social isolation is across time: if SI was to vary a lot from one wave to another, picking *ad hoc* its first observation might lead us to overestimate (or underestimate) the health effects of SI, if that observation was particularly low (or high) that precise year.

We present evidence in favor of the stability over time of the SI index in Figure A1, where each line represents the average SI over time for individuals who were followed for two waves, three waves, and so on, up to seven waves. Out of the 67,676 individuals who enter our survival analysis, 60,454 are represented on this graph, those who are not have a SI value at baseline and other value(s) at some future waves. Table A1 gives more detail about how many individuals have only a baseline SI index, how many have it at t_0 and $t_0 + 1$, how many at t_0, $t_0 + 1$, and $t_0 + 2$, and so on until t_0 to $t_0 + 6$. We can observe at most 7 values of the SI index, i.e. waves 1 to 8 except for wave 3. The items that are part of the SI index were not present in the questionnaires of the two "Corona" waves.

This graph informs us about two things: (i) those who "disappear" earlier have a (slightly) higher social isolation index than those who are followed over 4, 5, 6 consecutive waves. Since higher social isolation leads to higher mortality, it is reasonable that the "survivors" exhibit lower social isolation ; (ii) nevertheless, for each of these categories, the SI index seems quite stable over time.

Another way to have an idea of the stability of the (binary) SI index over time is by looking at transitions in and out of being socially isolated between t and $t + 1$, which is what we do in Table A2. Those who are socially isolated at t remain socially isolated at $t + 1$ with an almost 90% probability, while those who were not remain in that state with an almost 70% probability. The transition from not being socially isolated to being socially isolated is around 30%, and that from being socially isolated to not being socially isolated is lower, at around 11%.

We estimate the following Cox model:

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

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 5 models, each one adding more constraints to the relationship between social isolation and mortality.

Our results in Table 3 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 first two models suffer from obvious omitted variable bias as we do not include any background information on the individual in model (1), and no health information at baseline in model (2), while sex, age, or health at baseline are potentially both correlated with social isolation at baseline (it is hard to be socially connected to people when in bad health, for instance) and future mortality. We still display the results in columns (1) and (2). The point estimates of the hazard ratio for the social isolation index - in its continuous version from 0 to 3- go down from 1.21 to 1.13 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. Models (4) and (5) restrict follow-up to individuals who are still alive after the first 24 months, as a way to mitigate reverse causality: this way, we make sure our respondents did not have a life-threatening condition that was not captured by our observable covariates at baseline. Doing so does not challenge our estimate either. Comparing individuals with at least 1 point of social isolation to those who do not (column (5): the social isolation index is a binary variable) leads to a higher hazard ratio than looking at the effect of one extra point of social isolation (column (4)), meaning that a change from 0 to 1, 0 to 2 or 0 to 3, has more impact than the average increase of one point of SI over all the possibilities of SI increasing by 1. In all the following analyses, we will stick to the latter specification, with a binary SI index and excluding follow-up times inferior to 24 months.

Sensitivity checks We perform a number of sensitivity checks to 1) open the black box of the SI index and check whether a specific component is driving most of the SI effects on mortality 2) make sure that our results are not driven by a specific subsample 3) verify if marital status is a confounder in our analysis (we have not included it yet because of its high correlation with living alone).

As the three components that make the SI index may reflect different dimensions of social interaction and support, we provide some evidence regarding the relative importance of each component. In Table A3 (in Appendix) we display the results of a Cox regression including each of the eight possible combination of the three items. The only item that does not impact significantly mortality by itself is “living alone”. Even considered jointly with one or two other SI items, living alone does not seem to add (at least significantly) to the effect of the other items. For instance, living alone and not participating in any associations leads to a 24% higher mortality hazard than not being socially isolated at all (SI=0), but not participating in any associations, alone, leads to a 28% higher hazard. When associated with having few contacts with children, although living alone seems to have a higher impact than when considering “few contacts” alone, there is no statistical difference between the two estimated hazard ratios. Both non-participating in associational activities and having few contacts with children (or no children) -considered separately- have a great impact on mortality, with associational activities possibly even more important (and the most widespread across the sample, with 41 per cent of our sample having only that item). When considered jointly, whether this means a value of 2 or of 3 for the SI, their impact on mortality is even stronger (with the hazard being shifted upwards by 42 per cent).

This exercise is informative as to the contribution of each component to the index, and hence useful in terms of policy implications, since it allows us to understand better where to intervene: if policymakers were to incentivize older individuals to participate into associational activities (on which local authorities for instance have agency, contrary to the frequency of contact with individuals' children), they might be able to curb at least one of the two most harmful dimensions of SI for individuals' health.

Second, in order to make sure our results are not driven by a specific subsample, we estimate our specification (column (5) of Table 3) 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 A4 Panel A (in Appendix). Apart from a few exceptions, e.g. the employed have a higher mortality risk associated with social isolation than the non-employed (hazard ratios of 1.49 against 1.22), the hazard ratio remains remarkably stable around 1.22-1.29 almost across all subsamples. Individuals with no children are mechanically assigned to the "socially isolated" group when the SI index is binary and defined as "SI>0", since the item "infrequent contact with children" is set to 1 for individuals with no children. Hence, we cannot identify the coefficient of the (binary) SI index for the childless.

One solution to get an idea of whether SI affects mortality differentially for the childless and those with children is to re-run these regressions using the continuous index of social isolation: we will learn how a one-unit increase of the SI index affects the two groups, albeit the estimate of the impact of the SI index will not incorporate the effect of going from 0 to 1 for the childless group. Table A4 Panel B shows the impact of the (continuous) SI index is remarkably stable across the childless and "with children" subsamples. One difference with the binary case is the difference between the married and non-married individuals: when we use the continuous measure, the married seem to be more at mortality risk when more socially isolated than those who are not married. Again, this could be due to the married being much less socially isolated than the unmarried (a SI average of 0.79 versus 1.71), partly due to the very high correlation between one of the SI components -living alone- and the SI index.

Last, so far we have not controlled by marital status, because of its very high correlation with "live alone" (between 75 and 83 per cent depending on the definition of marital status). It might still be an important confounder, which we check by re-running the main Cox estimation 1) using marital status instead of "'live alone" in the definition of the SI index (columns (2) and (3) of Table A5) 2) controlling for marital status in the main regression (columns (4) and (5)) 3) adding marital status together with the other two items in a regression (columns (6) and (7)). We define being married as: "married and living with spouse" or "in a registered partnership", versus "married, not living with spouse", "never married", "divorced", and "widow"; we define "in couple" as being in couple and living with partner regardless of the official marital status.

As shown in Table A5 in the Appendix, our SI index (column (1)) seems to have a greater impact on mortality than when replacing "living alone" with marital status (columns (2) and (3)), which was not straightforward as it seemed the "living alone" item had no particular importance in the SI index. Second, controlling for either version of marital status does not change anything to the estimated impact of our SI index. Third, we find (almost) no evidence of the well-known protective effect of being married on individuals' health, which we interpret as evidence that our baseline controls do a good job at capturing health at baseline. When dropping these (columns (8) and (9)), married individuals face a 7 per cent lower hazard of death over the follow-up period, which is close to the lower bound of the effects of marital status found in the literature (12 per cent according to a meta-analysis consisting of more than 250,000 elderly subjects(Manzoli et al. (2007))), as one could expect given the richness of the set of baseline controls we use.

All in all, marital status does not seem to drive the effect of social isolation on mortality.

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.

We group countries into four culturally and geographically consistent 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), excvluding Ireland and Israel of this part of the analysis.

A first look at heterogeneity across these four groups of countries (see Table 4, columns (1) and (2)) suggests that the hazard ratio found for the socially isolated against the non-socially isolated is hiding important differences across countries. While in Western, Northern and Southern countries, social isolation (defined as at least one social isolation point) is associated to a 1.19 hazard ratio (Northern countries exhibit a higher HR but not statistically different at traditional thresholds), the social isolation HR is much higher in Eastern countries (1.45=1.19x1.22). Put in another way, social isolation has a similar impact in Western, Southern, and to some extent in Northern countries, but there is a very strong and significant difference between these countries and Eastern countries. Columns (3) to (5) introduce each country group against the other three in order to find out if any 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, and strikingly so: in all non Eastern countries, being socially isolated is associated with a 23% extra likelihood of dying over the follow-up period; in Eastern countries that likelihood increases by 18% more.

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, but our model allows 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 for each country when not looking specifically at the effect of a certain group of countries), so whatever may make individuals more or less healthy in a country -i.e. aggregate 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 aggregate economic or health care conditions than non-socially isolated individuals, even controlling for their income and health.

We attempt to shed light on the Eastern countries specificities that could explain this result by first re-estimating the same model separately for each country: even though we may lack power to find significant results, this will allow us to see if a particular country might be driving the results of the group it belongs too. As shown in Table A6 (in Appendix), Poland for instance, where the socially isolated are 83 per cent more likely to die over the period than the non socially isolated, could be an outlier. With a significant hazard ratio of .59, Portugal could also be the reason why the socially isolated die more over the period in Eastern than in Southern countries. We then reestimate the Cox regressions with interactions presented in Table 4 (more specifically Column (2)) by excluding one country at a time, in order to see whether the hazard ratio for the interaction is stable or whether an outlier country might be driving our result for Eastern countries. Table A7 (in Appendix) confirms that regardless of which country is excluded from the analysis, Western, Southern, and Northern countries do not differ in the mortality pattern of their socially isolated, while Eastern countries face an additional mortality hazard of between 17 and 27 per cent for the socially isolated.

On top of looking at differences in the observables between the four group of countries in Table A8, we also look into potential cultural and policy factors, using data from the European Social Survey (ESS). Regarding social isolation, Eastern and Southern countries are much more socially isolated than Northern and Western countries, mostly due to their high rate of non participation to associational activities; they are also more lonely and less connected in terms of social networks (using the scale defined in Section 3).

Eastern and Southern countries also have in common that they perform worse than Western and Northern countries across all health dimensions, with Eastern countries doing worse specially in terms of self-assessed health, number of chronic diseases, and number of limitations, and much better than Southern countries in terms of cognitive functioning. Notably, our Eastern and Southern countries samples differ greatly along the education dimension, with older Eastern Europeans being much more educated.

Eastern countries seem to differ from the rest of countries by a combination of high social isolation and bad health, although they are similar to Southern countries in many features. One potential explanation to the heterogeneous mortality effects of SI we find is that conditional on both Southern and Eastern countries' older individuals suffering from poorer health, Eastern countries' health care system might be worse than Southern countries'.

Table A9 shows suggestive evidence that this could be the case: while the proportion of people who declare they ever suffered of symptoms of depression which lasted at least two weeks, is remarkably stable across the four groups of countries (around 26 per cent), the share of those who were ever treated for depression by a doctor or psychiatrist (amongst those who were ever depressed) is much lower in Eastern countries (40 per cent against 54 per cent in Southern countries). The ESS also points toward the same direction, with individuals from Eastern countries rating the "state of health services in [their] countries nowadays" as worse than in the rest of the countries (on a 0-10 scale, where 0 is extremely bad and 10 is extremely good).

Another possibility is that our social isolation index lacks the friendship dimension that is present in Steptoe et al. (2013), and that the frequency of meeting friends be positively correlated with other items of our index. The proportion of individuals who never meet with their friends, relatives or colleagues, is much higher in Eastern Europe (24 per cent) than in Southern Europe (19 per cent) and Western (10 per cent) or Northern countries (5 per cent). If this is an important dimension of social isolation, and if it correlates with for instance participation to associational activities or frequency of contact with children, then what we capture in the interaction with Eastern Europe could actually be due to that precise dimension.

3 Pathways from social isolation to death: Loneliness, social connectedness, 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 24 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. We will look in turn at loneliness, social connectedness, health behavior, and health care utilization, as potential mediators in the relationship between social isolation and health.

How does social isolation lead to adverse health outcomes? Apart from the biological pathways -the inflammatory and antiviral processes are suspected to be at the very core of this association (see Leschak and Eisenberger (2019))-, health behavior, such as smoking, drinking, or a sedentary lifestyle, and health care utilization (doctor visits, preventive screenings, etc.) might explain why social isolation is so monotonically associated with worse health. These two mechanisms are in turn very much linked with one's social network, as "individuals who are socially engaged and connected are exposed to stronger normative pressures from and control by friends and loved ones to perform healthy behaviors and to access health care when needed" (Cacioppo and Hawkley (2003)). We therefore create an index of "connectedness", taking into account both the quantity and quality of one's social network, and check whether either of these three pathways mediate the association between social isolation and mortality.[7] It is also often argued that perceived social isolation, which is also referred to as the feeling of loneliness, could be the channel through which objective social isolation impacts mortality. Another question is to understand what contributes more to an individual's (bad) health: the objective or the subjective dimension of social isolation. We hence test whether loneliness mediated or mitigates this relationship.

We first estimate our preferred specification 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 hazard ratio of the social isolation index is unchanged, even though loneliness by itself is positively and significantly associated with mortality. If we allow the loneliness measure to vary over time (column (4)), instead of being fixed at baseline, then the impact of the SI index drops from 27 to 20 per cent, but reverse causality is also more of a concern as there is less time between loneliness measured and death observed. It therefore seems loneliness does not take much of the explanatory power of social isolation.

We then construct a social connectedness scale to test whether a poor social network could mediate the relationship between social isolation and mortality. We use rich information from the social networks modules introduced in Waves 4 and 6, in which respondents are asked to name up to 7 confidants, or people with whom [they] most often discussed important things, and to provide information about their relations to each of them. Following Malter and Börsch-Supan (2017), our measure of social connectedness uses (1) the number of persons cited (network size), (2) the number of cited social network members living within 25 km (proximity), (3) the number of cited persons with weekly or more contact (contact frequency), (4) the number of cited persons with very or extremely close emotional ties (support), and (5) the number of different types of relationships present within the network (diversity).[8] We then reverse the connectedness scale into a "disconnectedness" scale. The resulting scale lies between 0 and 4 (with a mean of 2), with higher values indicating a poorer social network. As shown in Table 6, similarly to the loneliness scale, social network disconnectedness is associated with higher mortality, but does not seem to mediate the relationship between social isolation and mortality, as it barely takes away anything from the impact of SI on mortality.

We then explore health behavior as a potential pathway between social isolation and higher mortality: socially isolated individuals may have worse lifestyle, e.g. smoke more, drink more, or

[7] This index of connectedness intends to summarize the richness of the social networks modules of SHARE waves 4, 6 and 8- which use name generators to construct respondents' networks of confidants- into one measure.

[8] See Malter and Börsch-Supan (2017) for details about the construction of the connectedness scale.

move less, which could lead to putting them at a higher risk of mortality. When adding health behavior information into the model (column (2) of Table 7), the hazard ratio of the social isolation index slightly drops from 1.23 to 1.19. All three variables come up as very much significant: smoking at baseline increases by 83 per cent the mortality hazard, being sedentary at baseline (i.e. not engaging in neither vigorous nor moderate activity, ever) does so by 14 per cent. Our measure of alcohol consumption, on the other hand, does not capture the harmful effects of alcohol, since it is associated with lower mortality. This has to be due either to the definition we had to use, i.e. having drunk any alcohol over the last three months or the last seven days, depending on the wave, or to the fact that controlling for health and socio-economic status, alcohol can be associated with positive outcomes. Had we been able to observe more extreme forms of alcohol consumption, such as binge drinking or pure alcoholism (which we only observe from wave 4), the result might have been different.

Regarding health care utilization as another potential pathway, adding health care utilization information (column (3)) does not seem to alter the SI coefficient at all. Neither the number of drugs an individual takes at baseline, nor the number of doctor visits one has had over the last twelve months, seem to matter a lot once health is taken into account (when health controls are not included in the equation, the number of drugs does). On the contrary, having stayed overnight at a hospital over the last twelve months is associated with a 13 per cent higher hazard, even though we are controlling for all observable dimensions of health at baseline. This latter finding highlights one dimension that is not being well captured by all our health controls, that is, the severity of one's condition: although we control for self-assessed health, and the number of chronic diseases for instance, how severely ill an individual is might be better proxied by adding the number of nights stayed at the hospital in the past 12 months.

Last, when we allow these potential "mediators" to vary wave by wave, they tend to take away more of the SI impact on mortality, as sedentarism and hospital stays become more important, which is consistent with them capturing some unobserved part of the health deterioration process.

Although loneliness, social disconnectedness and health behavior show some correlation with both social isolation and mortality, none of these appear as important channels of the association between social isolation and mortality. In the next section, we look into how social isolation affects the dynamics of health (in its observable dimensions), health behavior and health care utilization, and attempt to assess how much of its impact on mortality goes through each of these.

4 The dynamic impact of social isolation on health, health behavior and health care utilization outcomes

4.1 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).

As before, we focus on important health indicators belonging to both the objective and subjective health spectrum, and summarizing all relevant dimensions of health: physical (including frailty and functional health/limitations), mental, and cognitive health.

Our sample is exactly the same as before, but note that the number of observations decreases over time as 41,821 individuals are part of our sample and observed through 2 consecutive waves, while only 7,506 are observed 6 waves after their entry (which does not imply participating in all waves in between).

We estimate the following equations, and plot the relevant coefficient, $alpha_1$, in the dynamic graphs displayed in Figure 2:

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

where $j = 1, 2, ...6$. At $j = 0$ the coefficient of the SI index is mechanically 0, which is why we do not show it and start plotting coefficients at $j = 1$. 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 the same exact set of baseline characteristics we used in the Cox model, and which we summarize in Table 2, including the complete vector of health characteristics (amongst which the outcome at baseline). Hence, the α_1 coefficient measures the correlation between SI at baseline and the deterioration (or the change), rather than the level, of health. Again, we choose to fix covariates at baseline, in order to introduce some distance between the covariates and the outcomes.

For each health outcome, we run six regressions. We do have more than 6 waves, but contrary to mortality, which is known at each wave, and for which the date of death is known even if it happens between two waves, here, some outcomes are not informed at all waves, e.g. at wave 3 for some such as depression or frailty (SHARELIFE), or during the two "Corona" waves. "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. As mentioned before, sample size shrinks over time , so that confidence intervals are larger over time. We still get a clear picture of how social isolation correlates with health over time.

Social isolation is undeniably associated with worsening health. Nevertheless, there is heterogeneity across outcomes: cognitive health (recall test) starts worsening in association with social isolation after one wave, and the effect of social isolation remains at that same level 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. Since we are controlling for baseline cognitive functioning, our results point at a higher rate of decline for the socially isolated, in accordance with Ertel et al. (2008), which finds a higher rate of memory loss using word recall for individuals with lower social integration (which is very close to our measure of social isolation) using HRS 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 after four waves. It therefore seems social isolation worsens both physical and cognitive health in the short and long run, but its association with mental health is only transitory. Functional health, when measured as "suffering at least one limitation", instead of the number of limitations as we had done so far, 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). Nonetheless, 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.

This empirical exercise allows us to uncover a relationship between social isolation and the deterioration of health in almost all its facets. Even when considering mental health, for which the association does not persist in the long run, there is still a deterioration occurring two waves after baseline. Besides, a high correlation at baseline between social isolation and a poor mental health could lead to higher mortality, but would not show in our estimates. In other words, if the socially isolated at baseline suffer from more depression symptoms, even in the absence of further deterioration, a poor mental health that remains poor over the follow-up period could also be a channel leading to higher mortality, which is important to bear in mind when interpreting our results.

4.2 Health behavior and health care utilization outcomes

Apart from the "biological" channel, 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 health behaviors seemed to play some role in the SI-mortality relationship, while health care utilization did not. In the present section we investigate whether there is a specific pattern of the socially isolated in terms of health behavior or health care. We apply the same dynamic analysis to the set of health behavior and health care utilization variables used in section 3. Again, in each regression we control for the outcome at baseline, so that looking at how social isolation associates with smoking at later waves is equivalent to looking at how it correlates with changes in smoking.

As shown in Figure 3, there is no significant relationship between social isolation and smoking (when controlling for smoking at baseline), 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 "hardly ever, 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. Shankar et al. (2011) and Kobayashi and Steptoe (2018) find similar results on inactivity, and a more clear-cut association with smoking and drinking, without controlling for these variables at baseline, 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.

Regarding the relationship between social isolation and health care utilization, there are two (or more) possible directions in which social isolation might affect health care utilization: (1) because social isolation is associated to worse health (see previous subsection) and higher mortality (see section 2.2), the socially isolated might use more health care; (2) social isolation might make individuals less inclined to use health care, as they are less "pressured" by loved ones to realize medical checkups, prompt them to seek medical help when needed, or even to accompany them to a doctor visit.[9] Our results help to shed light on this discussion: Figure 4 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 use more medicines than the non-socially isolated, and neither do they visit more or less their physician. If anything, they spend more nights at the hospital, after a few waves, but the relationship is weak

[9] Socially isolated individuals might also resort less to health care due to lack of information, as put forward in Devillanova (2008), which documents a lower time to visit for immigrants with a strong social tie knowing about health care opportunities.

(both in terms of significance and magnitude). [10]

How is this "null" result compatible with the two directions above mentioned? First, as mentioned for the case of health behaviors, any baseline correlation between social isolation and health care utilization levels is already factored in the baseline controls. Second, mechanisms (1) and (2) could cancel each other out. The fact that we do not find a positive effect of SI on health care utilization due to the health deterioration (i.e. direction (1)) is striking and consistent with mechanism (2) being a potential channel.

4.3 Quantifying the contribution of each health dimension to the social isolation-mortality relationship

After estimating the mortality effects of SI, we dive into the dynamics of the social isolation effects in order to answer how social isolation might lead to higher mortality. In this section we ask two questions: (1) which mediator has the biggest impact on mortality hazards? (2) Which one is the most affected by social isolation? And then combine the answers to the two questions to compute how much of the SI effects on mortality can be imputed to the effect of SI at baseline on each dimension of health, at next period. In Table 8 we present the results of the same Cox regression as in the main table (Table 3 column(5)) but displaying the hazard ratios that correspond to the health controls. In column (1) the variables are introduced as in the descriptive statistics shown in Table 2, while in column (2) we standardize all the health variables so that their effects are more comparable to each other. As an answer to the first question, we find that self-assessed health, frailty, and cognitive functioning have a great impact -and of similar magnitude as that of social isolation: 28, 22 and 16 per cent respectively- on mortality. When controlling for all other health dimensions, depression is not associated with higher mortality. This was to be expected as it shares 57 per cent correlation with the frailty dimension, and 39 per cent with self-assessed health. When taking away those two dimensions from the equation, depression is significantly associated with higher mortality. Neither the number of chronic diseases, nor the number of limitations in ADLs and IADLs have a sizeable effect. We report these coefficients (as coefficients instead of hazard ratios this time), as $\beta(H)$ in Table 9 (first row), along with $\alpha(S)$, on the second row: the coefficient of social isolation in the health outcome regression (at $t+1$, i.e. the first point plotted on each graph of Figure2). The second row provides an answer to the second question: self-assessed health, frailty, and cognitive health -same as in question 1)- are the health dimensions more affected by social isolation, at least at wave $t+1$.

We then multiply one by the other, in order to obtain a coefficient that is generated by SI at $t+1$, through the health channels between t and $t+1$ (see third row). Last, we compare this coefficient with $\beta(S)$, the coefficient of SI at baseline in the Cox model, which is equal to 0.227. On the last row, we show how much of the SI effect on mortality can be imputed to the dynamic effect of social isolation on health one wave ahead: while chronic diseases, functional health (limitations), and depression, account for less than one per cent of the SI effect on mortality, the impact of SI on self-assessed and cognitive health at next wave, reported on the effect of these health variables on mortality, is around 6-7 per cent (4 per cent in the case of frailty). Although this does not seem like a very high figure, it is informative with respect to policy, i.e. where and how to intervene in

[10] This finding goes against 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.

order to curb the SI effects on mortality. A policy designed to target social isolation directly, for instance promoting associational activities for older people, could therefore be thought of as a way to allow those who else would have been socially isolated to live longer, but also as a way to slowing their cognitive decline. Reducing social isolation could hence be part of the recommendations such as those emitted by the WHO in order to reduce the risk of dementia and cognitive decline (Chowdhary et al. (2021)), and help " Understand the influence and interactions of non-modifiable (e.g., gender, genetics, age) and modifiable (e.g., physical activity, diet, and cognitive stimulation) risk and protective factors for dementia in population-based samples".

When reproducing the same exercise for health behaviors and health care utlization (see Table 10), we find that a small part of the impact of SI on mortality goes through the impact of SI on increased smoking (3 per cent). On the other hand, the coefficient of smoking in the Cox regression is as big as that of self-assessed health for instance, so that our "chain rule" exercise still yields a non-null percentage for smoking. By comparison, none of the impact of SI on mortality goes through its impact on sedentarism at $t+1$.

5 Discussion: the causality challenge and the education benchmark

5.1 Causality and other caveats

We uncovered a strong association between social isolation and mortality, i.e. a 25% increase in the mortality hazard rate for individuals who are socially isolated at baseline. The magnitude of this association is close to the estimates found in the literature for the causal impact of education on mortality. 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 24 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 and that would still be the reason why one is socially isolated. This restriction, coupled with a very long follow-up period (up to 17.25 years, with a median follow-up of 79 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 not straightforward 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 individuals are working, or whether they have children. We also control for housing variables, as there could be a link between living in a rural versus urban area, or in a house versus in a building, and future health, if for instance it is harder for older individuals living in a rural area to seek medical attention when needed. At the same time, everything else held equal, living in a house with no neighbors, or in an isolated area, could also be correlated with social isolation. An example of unobservable that could determine social isolation is personality: Cacioppo et al. (2000) shows that individuals from the lowest quintile on the UCLA loneliness scale were found to score lower on neuroticism and higher on surgency (extraversion), conscientiousness, and social agreeableness than individuals from the highest quintile, who in turn did not differ on any of these dimensions from individuals from the middle quintile. Introverted or neurotic individuals are probably more at risk of social isolation, as they would be less prone to participate in social activities for instance. Regarding personality and health, the latest evidence using SHARE's "Big 5" data that was collected for the first time at Wave 7 points at personality being associated with an array of health indicators in older European adults, more particularly the strongest and most consistent personality-level correlates of good health are high conscientiousness and low neuroticism (Shemesh et al. (2019)).[11] In any case, these would be factored in our health indicators at baseline in our main Cox model. Nevertheless, as neuroticism for instance could also be correlated with a more acute deterioration of health, personality traits should be part of our controls. This is not without problems since they are measured at wave 7, while our baseline measures of health and social isolation might come from previous waves. Arguing for the stability of (at least some of) the Big 5 over time in later life, we include them in our Cox regressions. [12] Table A10 shows that the same traits that correlated highly with several health indicators, i.e. high conscientiousness (both as a continuous and binary variable) and low neuroticism (only as a binary variable) are associated with lower mortality, even when controlling for health at baseline. Individuals who have low conscientiousness (i.e. in the first quintile of that measure), which supposedly captures having a high propension to be self-controlled and to delay gratification, to be task and goal directed, organized, efficient, precise and deliberate (John et al. (1999)), are exposed to a 25 per cent higher mortality. [13]Even such a high correlation does not take away any effect from our social isolation index.

One could think of one potential omitted confounder we can unfortunately not include in our analysis: 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.[14] Then genes could be an important source of omitted variable bias in our study, if they determined both social isolation and mortality.

Once established a list of potential confounders one can possibly think about, what else could be done to assess how causal a relationship can be? In practice, several approaches can help mitigate omitted variable 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 A11 shows 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 six 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 $time = 6$ in the "self-assessed health" graph in Figure 2, but the same stability could be shown for the other outcomes at other times.

[11]Often referred to by the acronym OCEAN, these are: openness to experience (vs. closedness), conscientiousness (vs. lack of direction), extraversion (vs. introversion), agreeableness (vs. antagonism) and neuroticism (vs. emotional).

[12]Cobb-Clark and Schurer (2012) shows they are stable for at least a 4-year period, and Erlich and Litwin (2019) establishes using the SHARE Big 5 data that two personality attributes - conscientiousness and neuroticism - hardly vary across age.

[13]Conscientiousness is derived in SHARE positively from answers to the statement "I see myself as someone who does a thorough job" and negatively from answers to "I see myself as someone who tends to be lazy".

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

As put forward in Oster (2019), although very intuitive, this idea relies on the selection 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 1.37, 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 37% 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 2). Furthermore, to obtain this value of 1.37, 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".

5.2 Education as a benchmark of social isolation effects on mortality

Education and mortality The association we find between social isolation and mortality in Table 3 seems both significant and robust to many checks, but how "big" exactly is it? We compare the effects of social isolation on mortality to those of education, which seems an ideal benchmark candidate as there exists a compelling literature establishing significant positive associations between education and several dimensions of adult health, and negative associations between education and mortality. Whether these associations can be qualified as "causal" has been subject to debate, and several studies have reached diverging conclusions. Using changes in education legislation as a source of exogenous variation in educational attainment, Lleras-Muney (2005) in the US, Crespo et al. (2014) in Europe, found support for a causal link. No such causal effect is found in Behrman et al. (2011), which uses comparisons of twins in Denmark. On the contrary, Halpern-Manners et al. (2020) - a more recent assessment of this causal relationship based on representative US "twins data", supports a causal interpretation of the education-mortality gradient. Hayward et al. (2015) reminds us of the importance of assessing the magnitude and shape of this association, rather than focusing on the causal nature of that association, which is more likely to depend on historical and social contexts.

We reproduce the same specification as before (see equation (1)), and display both the hazard ratio associated with our social isolation index, and those associated with our education controls (see Panel A of Table A12 in Appendix). Education is introduced as 4 categories, the higher educational attainment group being omitted as the reference group (the "other" category was dropped here, to make the interpretation more straightforward, hence the slight discrepancies in sample sizes with respect to Table 3). Controlling for socio-demographic information at baseline, lower education levels are associated with higher hazard ratios (ranging from 1.35 for the lowest education level to 1.22 for the upper secondary group), with greater magnitude than social isolation (1.21).

Note that when adding baseline health as a control most of the effect of lower education goes away, and the magnitude of the effect is now much more similar to that of social isolation. This is less the case for the upper secondary education group. This suggests that most of the effect of education on mortality has already been channeled into individuals' health status when they enter the study, except for those who are the closest to the higher education group. Thus the more controls and horizon constraints we add to the model, the less significant the association between education and mortality is. The only remaining significant difference is between the upper secondary and the tertiary education group. In short, the education-mortality gradient has a similar magnitude as the one between social isolation and mortality, but is less robust to the inclusion of the same additional controls and constraints. This result does not question the causal relationship between education and health obtained from the quasi-experimental studies cited above, which do not control for health when looking at education and mortality. When taking out social isolation from the equation, and looking at years of education instead of education categories (see Panel B of Table A5), to be closer to Lleras-Muney (2005) for instance, we find that one year of education is associated with a 2.4% drop in the hazard of dying (over the period), which lies in the 1.3-3.6 interval found in that study, over a 10-year period. When adding baseline health into the equation, this result does not hold anymore.

Dynamic effects of education on health outcomes Again, in order to gauge the magnitude of the effects of social isolation on future health, we construct the same graphs as before, but displaying the coefficient of the lowest education category (with respect to the highest category) instead of that of the social isolation index (see Figure A3). As before, controlling for baseline health implies we are looking at the effect of education on changes in health (or on future health given health at entry), so one possibility would be that all of the impact of education is already factored in health at entry, and is not reflected in deviations of health from that point on. We find a very similar pattern as in Figure 2, except that the magnitude of the coefficients is much bigger, e.g. coefficients are twice those of social isolation for self-assessed health, and 6 times those of social isolation for cognitive health (word recall). Looking at the health factor that summarizes all the observed dimensions of health, being socially isolated takes a toll on individuals' health, approximately half the toll of being in the lowest education group.

6 Conclusion

In this paper we find a strong association between social isolation and future mortality, which is not solely mediated by concurrent loneliness, health behavior, health care utilization, nor social connectedness. We also explore the dynamics of the health impact of social isolation, and find social isolation to lead to a persistent worsening of all the facets of health we consider (self-assessed, frailty, cognitive, mental, functional).

Although previous studies have identified similar associations, we add to the existing literature by looking at health across many dimensions instead of focusing on a single health outcome, and by doing so in a harmonized multi-country longitudinal framework that allows us for a long follow-up period and alleviates endogeneity concerns. We 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.

We test several potential underlying mechanisms, and find that socially isolated individuals do not resort to more health care use in subsequent waves although their health worsens across all dimensions compared to non socially isolated individuals, which suggests that health care utilization might be a channel underlying the relationship between social isolation and health.

When we combine our mortality, health and health behavior models in an accounting exercise we obtain that up to 13 percent of the effect of baseline social isolation on mortality can be imputed to the combined one-wave-ahead impact of social isolation on increased frailty, reduced cognitive function and increased smoking.

Last but not least, we provide evidence in favor of a causal interpretation of our estimates using Oster's test for selection on unobservables. We also compare the social isolation-health gradient to the much studied education-health gradient. Using the same models, the education gradient in mortality is smaller than the social isolation gradient, but the association of education with future health is stronger than the one we find for social isolation in dynamic value added regressions.

We believe this study has important implications for public policy in a pandemic context, as confinement measures, by affecting each of the three items of our social isolation index, might have pushed those who were already socially isolated into deeper isolation, and those who were still connected into some degree of social isolation. It is therefore urgent, in a post-confinement world, to help reconstructing social connections, with family and peers, and through clubs and associations, in order to curb down what could be indirect health effects of the pandemic.

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.

Behrman, Jere R, Hans-Peter Kohler, Vibeke Myrup Jensen, Dorthe Pedersen, Inge Petersen, Paul Bingley, and Kaare Christensen, "Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins," *Demography*, 2011, *48* (4), 1347–1375.

Berkman, Lisa F and S Leonard Syme, "Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents," *American journal of Epidemiology*, 1979, *109* (2), 186–204.

Cacioppo, John T and Louise C Hawkley, "Social isolation and health, with an emphasis on underlying mechanisms," *Perspectives in biology and medicine*, 2003, *46* (3), S39–S52.

_, John M Ernst, Mary H Burleson, Martha K McClintock, William B Malarkey, Louise C Hawkley, Ray B Kowalewski, Alisa Paulsen, J Allan Hobson, Kenneth Hugdahl et al., "Lonely traits and concomitant physiological processes: The MacArthur social neuroscience studies," *International Journal of Psychophysiology*, 2000, *35* (2-3), 143–154.

Chowdhary, Neerja, Corrado Barbui, Kaarin J Anstey, Miia Kivipelto, Mariagnese Barbera, Ruth Peters, Lidan Zheng, Jenni Kulmala, Ruth Stephen, Cleusa P Ferri et al., "Reducing the risk of cognitive decline and dementia: WHO recommendations," *Frontiers in Neurology*, 2021, *12.*

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

Cohen, Sheldon, William J Doyle, David P Skoner, Bruce S Rabin, and Jack M Gwaltney, "Social ties and susceptibility to the common cold," *Jama*, 1997, *277* (24), 1940–1944.

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.

Crespo, Laura, Borja López-Noval, and Pedro Mira, "Compulsory schooling, education, depression and memory: New evidence from SHARELIFE," *Economics of Education Review*, 2014, *43*, 36–46.

Devillanova, Carlo, "Social networks, information and health care utilization: evidence from undocumented immigrants in Milan," *Journal of Health Economics*, 2008, *27* (2), 265–286.

Erlich, Bracha and Howard Litwin, "Personality, age and the well-being of older Europeans," *Health and socioeconomic status over the life course: first results from SHARE waves*, 2019, *6*, 35–41.

Ertel, Karen A, M Maria Glymour, and Lisa F Berkman, "Effects of social integration on preserving memory function in a nationally representative US elderly population," *American journal of public health*, 2008, *98* (7), 1215–1220.

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.

Halpern-Manners, Andrew, Jonas Helgertz, John Robert Warren, and Evan Roberts, "The effects of education on mortality: Evidence from linked US Census and administrative mortality data," *Demography*, 2020, *57* (4), 1513–1541.

Hayward, Mark D, Robert A Hummer, and Isaac Sasson, "Trends and group differences in the association between educational attainment and US adult mortality: Implications for understanding education's causal influence," *Social Science & Medicine*, 2015, *127*, 8–18.

Heffner, Kathi L, Molly E Waring, Mary B Roberts, Charles B Eaton, and Robert Gramling, "Social isolation, C-reactive protein, and coronary heart disease mortality among community-dwelling adults," *Social science & medicine*, 2011, *72* (9), 1482–1488.

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.

Hughes, Mary Elizabeth, Linda J Waite, Louise C Hawkley, and John T Cacioppo, "A short scale for measuring loneliness in large surveys: Results from two population-based studies," *Research on aging*, 2004, *26* (6), 655–672.

John, Oliver P, Sanjay Srivastava et al., "The Big-Five trait taxonomy: History, measurement, and theoretical perspectives," 1999.

Kobayashi, Lindsay C and Andrew Steptoe, "Social isolation, loneliness, and health behaviors at older ages: longitudinal cohort study," *Annals of Behavioral Medicine*, 2018, *52* (7), 582–593.

Leschak, Carrianne J and Naomi I Eisenberger, "Two distinct immune pathways linking social relationships with health: inflammatory and antiviral processes," *Psychosomatic medicine*, 2019, *81* (8), 711.

Lleras-Muney, Adriana, "The relationship between education and adult mortality in the United States," *The Review of Economic Studies*, 2005, *72* (1), 189–221.

Malter, Frederic and Axel Börsch-Supan, "SHARE wave 6: panel innovations and collecting dried blood spots," *Munich: Munich Center for the Economics of Aging (MEA)*, 2017.

Manzoli, Lamberto, Paolo Villari, Giovanni M Pirone, and Antonio Boccia, "Marital status and mortality in the elderly: a systematic review and meta-analysis," *Social science & medicine*, 2007, *64* (1), 77–94.

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

Powell, Victoria D, Nauzley C Abedini, Andrzej T Galecki, Mohammed Kabeto, Navasuja Kumar, and Maria J Silveira, "Unwelcome Companions: Loneliness Associates with the Cluster of Pain, Fatigue, and Depression in Older Adults," *Gerontology and Geriatric Medicine*, 2021, *7*, 2333721421997620.

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.

Shemesh, Jonathan, Ella Schwartz, and Howard Litwin, "Personality and physical health among older Europeans," *Health and socioeconomic status over the life course: first results from SHARE waves*, 2019, p. 43.

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.

Teguo, Maturin Tabue, Nadine Simo-Tabue, Ralitsa Stoykova, Céline Meillon, Mélanie Cogne, Hélène Amiéva, and Jean-François Dartigues, "Feelings of loneliness and living alone as predictors of mortality in the elderly: the PAQUID study," *Psychosomatic medicine*, 2016, *78* (8), 904–909.

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.

Yu, Bin, Andrew Steptoe, Li-Jung Chen, Yi-Huei Chen, Ching-Heng Lin, and Po-Wen Ku, "Social isolation, loneliness, and all-cause mortality in patients with cardiovascular disease: a 10-year follow-up study," *Psychosomatic medicine*, 2020, *82* (2), 208–214.

TABLES

Table 1: Distribution of SI index and its components

	SI=0	SI=1	SI=2	SI=3	Total
Lives Alone	0.00	0.09	0.58	1.00	0.20
No participation to any social activities	0.00	0.82	0.87	1.00	0.61
Less than weekly contact with children	0.00	0.09	0.56	1.00	0.19
Observations	18256	33992	12581	2847	67676
%	26.98	50.23	18.59	4.21	100

Table 2: Descriptive Statistics on Observables, by social isolation status

	SI index=0		SI index>0		All	
	mean	sd	mean	sd	mean	sd
Social Isolation						
SI> 0	0.00	(0.00)	1.00	(0.00)	0.73	(0.44)
Lives Alone	0.00	(0.00)	0.27	(0.44)	0.20	(0.40)
No participation to any social activities	0.00	(0.00)	0.84	(0.37)	0.61	(0.49)
Less than weekly contact with children	0.00	(0.00)	0.26	(0.44)	0.19	(0.39)
Loneliness (RUCLA scale)	3.49	(0.96)	3.89	(1.36)	3.76	(1.26)
SN disconnectedness	1.88	(0.91)	2.17	(0.86)	2.09	(0.88)
Health behavior						
Currently Smokes	0.15	(0.36)	0.21	(0.40)	0.19	(0.39)
Sedentary	0.03	(0.16)	0.10	(0.30)	0.08	(0.28)
Drank alcohol past 3 months/7 days	0.82	(0.38)	0.64	(0.48)	0.69	(0.46)
Health care utilization						
Number of drugs/medicine	1.17	(1.28)	1.58	(1.54)	1.47	(1.49)
Number of doctor visits	5.41	(7.91)	6.65	(9.68)	6.32	(9.25)
Nb of overnight stays past 12 mo	0.11	(0.32)	0.14	(0.35)	0.13	(0.34)
Health						
Self-perceived health	2.71	(1.03)	3.20	(1.07)	3.06	(1.08)
Frailty index	0.56	(0.81)	1.02	(1.16)	0.90	(1.09)
Number of chronic diseases	1.47	(1.38)	1.80	(1.58)	1.71	(1.54)
Number of limitations	0.17	(0.71)	0.47	(1.35)	0.39	(1.22)
Depression score (EuroD)	1.91	(1.86)	2.56	(2.30)	2.39	(2.21)
Cognitive recall test (higher is good)	5.09	(1.62)	4.24	(1.76)	4.47	(1.76)
Socio-demographic controls						
Age	62.14	(8.33)	65.45	(9.73)	64.56	(9.49)
Female	0.49	(0.50)	0.57	(0.50)	0.55	(0.50)
Education: None	0.01	(0.12)	0.05	(0.22)	0.04	(0.20)
Education: Primary	0.09	(0.29)	0.23	(0.42)	0.20	(0.40)
Education: Secondary(i)	0.14	(0.35)	0.19	(0.39)	0.18	(0.38)
Education: Secondary(ii)	0.35	(0.48)	0.31	(0.46)	0.32	(0.47)
Education: Tertiary	0.39	(0.49)	0.20	(0.40)	0.25	(0.44)
Education: Other	0.01	(0.07)	0.01	(0.08)	0.01	(0.08)
Employed or Self-employed	0.42	(0.49)	0.26	(0.44)	0.30	(0.46)
Has at least one child	1.00	(0.00)	0.82	(0.39)	0.87	(0.34)
Lives in rural area	0.34	(0.47)	0.30	(0.46)	0.31	(0.46)
Lives in a house (not a building)	0.77	(0.42)	0.63	(0.48)	0.67	(0.47)
Total household income	65688	(2144930)	28498	(41142)	38530	(1114689)
Household net worth	376684	(664093)	618897	(64990243)	553558	(55537994)
N	18,256		49,420		67,676	

Note: Sample with a follow-up of 24 months minimum.

Table 3: Cox models: Effect of Social Isolation at Baseline on Mortality up to Second wave Covid-19.

	(1)	(2)	(3)	(4)	(5)
Social Isolation index	1.188***	1.208***	1.127***	1.118***	1.254***
	(0.013)	(0.016)	(0.017)	(0.018)	(0.038)
Basic	yes	yes	yes	yes	yes
Socio-Demo	no	yes	yes	yes	yes
Health	no	no	yes	yes	yes
Follow-up	no	no	no	>24 mo	>24 mo
Binary SI index	no	no	no	no	yes
Observations	341806	325406	273201	243515	243515
Individuals	97751	92558	72659	67676	67676

Exponentiated coefficients; Standard errors in parentheses

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

Note: Basic controls: age, age squared, sex. Socio-demo controls: education categories, whether employed, whether has at least one child, income and wealth quartiles, living in a house (vs building), rural (versus urban). 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 4: Cox models: Country Heterogeneity of the Impact of Social Isolation at Baseline on Mortality up to Wave 2 Covid-19.

	(1)	(2)	(3)	(4)	(5)
SI: SI index>0	1.27***	1.19***	1.29***	1.26***	1.23***
	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Western Countries	ref.	ref.	ref.	ref.	ref.
Southern countries	1.07**	1.06	1.13	1.07**	1.08**
	(0.03)	(0.09)	(0.09)	(0.03)	(0.03)
Northern countries	1.58***	1.46***	1.58***	1.54***	1.58***
	(0.06)	(0.10)	(0.06)	(0.10)	(0.06)
Eastern countries	1.49***	1.25***	1.48***	1.49***	1.29***
	(0.05)	(0.09)	(0.05)	(0.05)	(0.09)
SI X Western		ref.	ref.	ref.	ref.
SI X Southern		1.02	0.94		
		(0.09)	(0.08)		
SI X Northern		1.10		1.04	
		(0.09)		(0.08)	
SI X Eastern		1.22**			1.18**
		(0.10)			(0.09)
Observations	235154				
Individuals	65210				

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave FE.

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***	1.20***
	(0.07)	(0.07)	(0.07)	(0.08)
RUCLA Loneliness at Baseline		1.05***		
		(0.02)		
RUCLA Loneliness at Baseline (d)			1.26***	
			(0.09)	
RUCLA Loneliness- Time-Varying				1.07***
				(0.02)
Observations	101894	101894	101894	61677
Individuals	34544	34544	34544	33347

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave and country FE.

Waves 1 to 3 are excluded from the analysis (no RUCLA information).

Follow-up restricted to being > 24months.

Table 6: Cox models: Does social disconnectedness mediate the association between Social Isolation and Mortality

	(1)	(2)	(3)	(4)
Social Isolation index	1.30***	1.28***	1.29***	1.26**
	(0.07)	(0.07)	(0.07)	(0.11)
SN disconnectedness at Baseline		1.09***		
		(0.02)		
SN disconnectedness at Baseline (d)			1.16***	
			(0.04)	
SN disconnectedness- Time-Varying				1.16***
				(0.04)
Observations	84818	84818	84818	38707
Individuals	25234	25234	25234	22949

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave and country FE.

Follow-up restricted to being > 24months.

Table 7: Cox models: Do Health Care and Health Behavior mediate the association between Social Isolation and Mortality

	(1)	(2)	(3)	(4)	(5)
Social Isolation	1.23***	1.19***	1.23***	1.19***	1.17***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Currently smokes		1.83***		1.85***	1.71***
		(0.05)		(0.05)	(0.07)
Sedentary		1.14***		1.14***	1.92***
		(0.04)		(0.04)	(0.06)
Had alcohol last 3mo/last week		0.92***		0.93***	0.79***
		(0.02)		(0.02)	(0.02)
Number of medicaments			1.00	1.00	1.00
			(0.01)	(0.01)	(0.01)
Number of doctor visits			1.00***	1.00***	1.01***
			(0.00)	(0.00)	(0.00)
Stay overnight hospital last 12 mo			1.13***	1.13***	1.62***
			(0.03)	(0.03)	(0.05)
Observations	224691	224691	224691	224691	148512
Individuals	62553	62553	62553	62553	60458
Controls		at baseline	at baseline	at baseline	time-varying

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave and country FE.

Follow-up restricted to being > 24months.

Table 8: Cox models: how does health impact mortality?

	(1)	(2)
Social Isolation index	1.254***	1.246***
	(0.038)	(0.040)
Self-perceived health	1.239***	1.278***
	(0.016)	(0.020)
Frailty index	1.203***	1.216***
	(0.013)	(0.015)
Number of chronic diseases	1.027***	1.041***
	(0.007)	(0.012)
Number of limitations	1.035***	1.051***
	(0.006)	(0.012)
Depression score (EuroD)	0.965***	0.923***
	(0.005)	(0.012)
Cognitive recall test (higher is good)	0.906***	0.839***
	(0.007)	(0.012)
Health variables	non-standardized	standardized
Observations	243515	225779
Individuals	67676	67127

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave and country FE.

Follow-up restricted to being > 24months.

Table 9: Chain rule effect: How much of the SI impact on mortality hazards can be explained through dynamic health effects?

	SA Health	Frailty	Chronic diseases	Limitations	Depression	Cognitive (bad) health
Coeff of Health variable in Cox model $\beta(H)$	0.245	0.195	0.040	0.050	-0.080	0.176
Coeff of SI in health outcome (at $t+1$) regression $\alpha(S)$	0.063	0.043	0.019	0.020	0.017	0.078
$(\beta(H) * \alpha(S))$	0.016	0.008	0.001	0.001	-0.001	0.014
Coeff of SI in Cox model $\beta(S)$			0.227			
Chain rule effect $(\beta(H) * \alpha(S))/\beta(S) * 100$	6.866	3.710	0.331	0.446	-0.596	6.039

Note: 0.227 is the coefficient of SI in a Cox model with all the health and socio-demo variables displayed in Table 2, and wave and country FE.

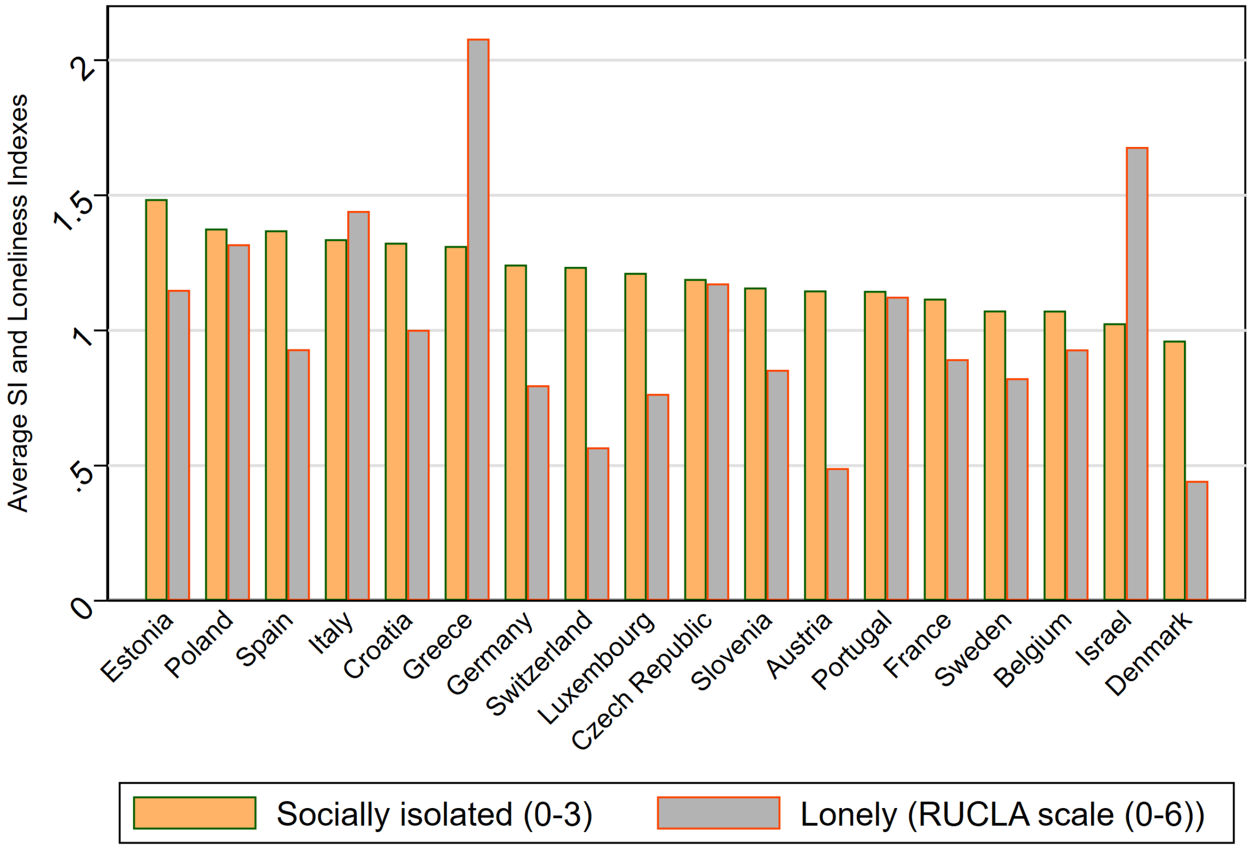
Table 10: Chain rule effect: How much of the SI impact on mortality hazards can be explained through health behavior and health care utilization?

	Smokes	Sedentarism	Drinking	Number of drugs	Doctor visits	Hospital stays
Coeff of behavior/hcu variable in Cox model $\beta(H)$	0.208	0.005	-0.067	-0.087	0.069	0.111
Coeff of SI in behavior/hcu outcome (at $t+1$) regression $\alpha(S)$	0.040	0.035	-0.017	0.023	0.003	-0.006
$(\beta(H) * \alpha(S))$	0.008	0.000	0.001	-0.002	0.000	-0.001
Coeff of SI in Cox model $\beta(S)$	0.257	0.289	0.285	0.257	0.287	0.286
Chain rule effect $(\beta(H) * \alpha(S))/\beta(S) * 100$	3.195	0.065	0.406	-0.778	0.081	-0.252

Note: $\beta(S)$ is the coefficient of SI (binary) in a Cox model with all the health and socio-demo variables, wave and country FE, plus the health behavior or health care utilization variable that corresponds to each column.

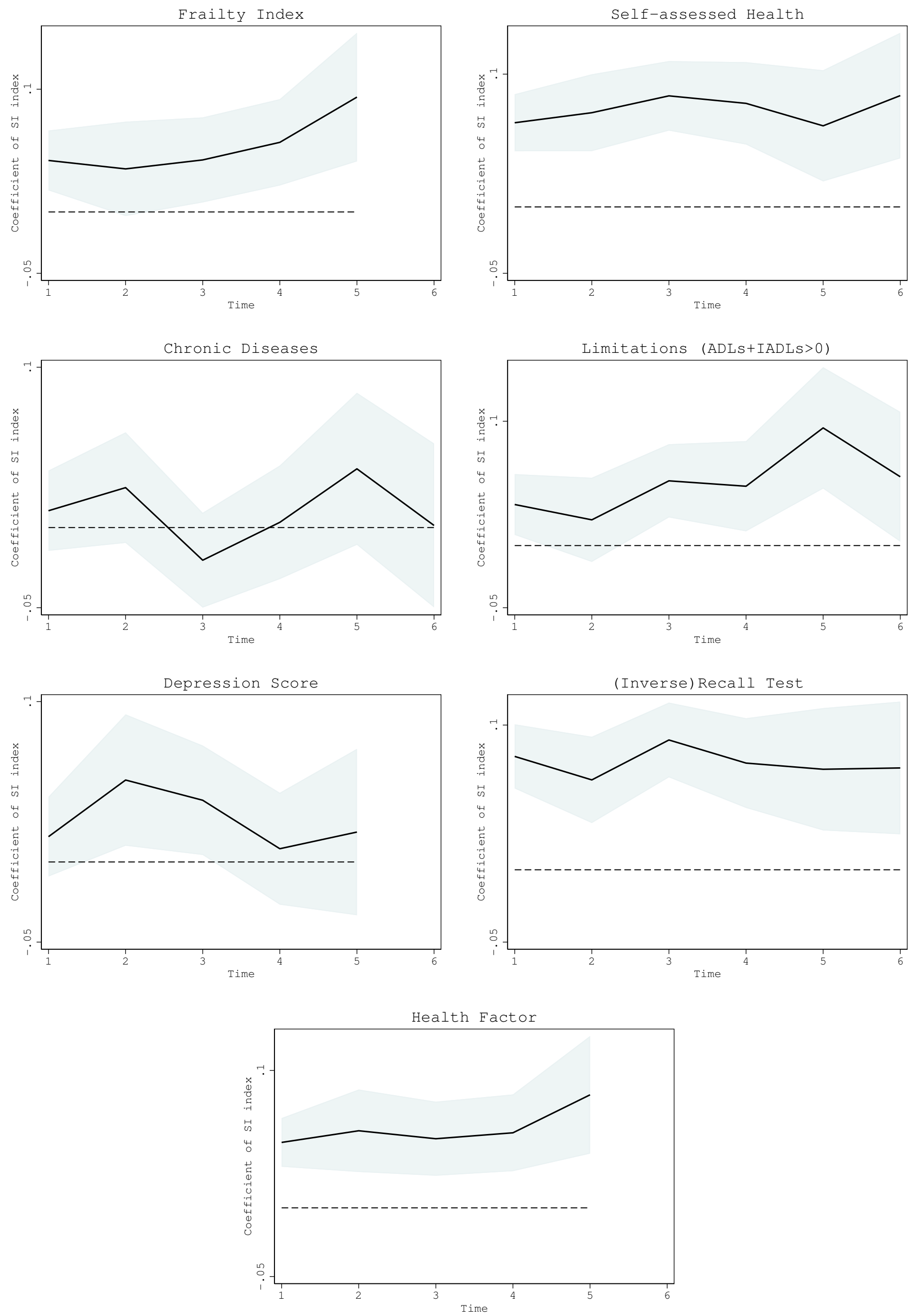
FIGURES

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



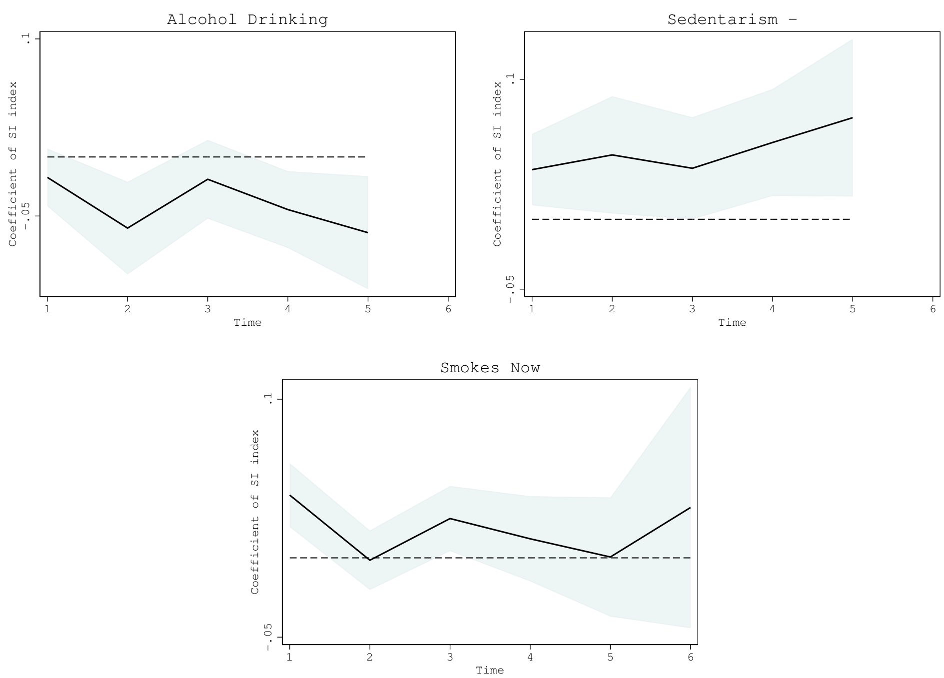
Source: SHARE Wave 6- Sample:+60 years old.

Figure 2: Social Isolation at Baseline and Health Dynamics



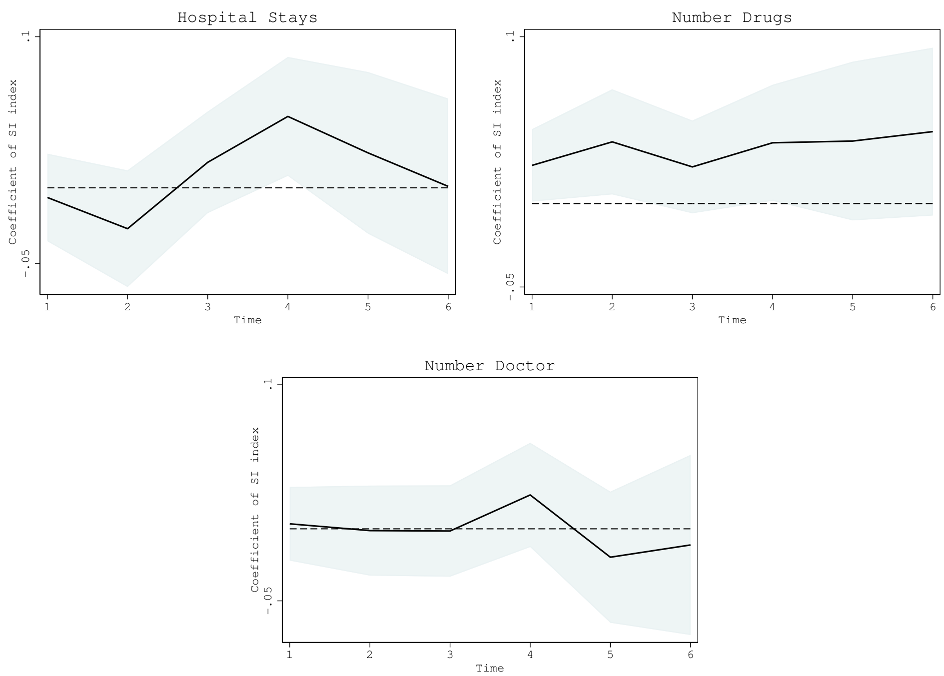
Note: The figures display the coefficients and 95 percent confidence intervals for the effect of social isolation on all (standardized) health outcomes. "Time" indexes future waves: 1 is wave $t+1$, 2 is wave $t+2$, etc. Regressions are done separately for each outcome and lag. All regressions include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

Figure 3: Social Isolation at Baseline and Health Behavior Dynamics



Note: The figures display 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 2, the outcome at baseline, and wave (a dummy for which wave is baseline) and country FE, and control for the outcome at baseline.

Figure 4: Social Isolation at Baseline and Health Care Utilization Dynamics



Note: The figures display 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 2, the outcome at baseline, and wave (a dummy for which wave is baseline) and country FE.

APPENDIX

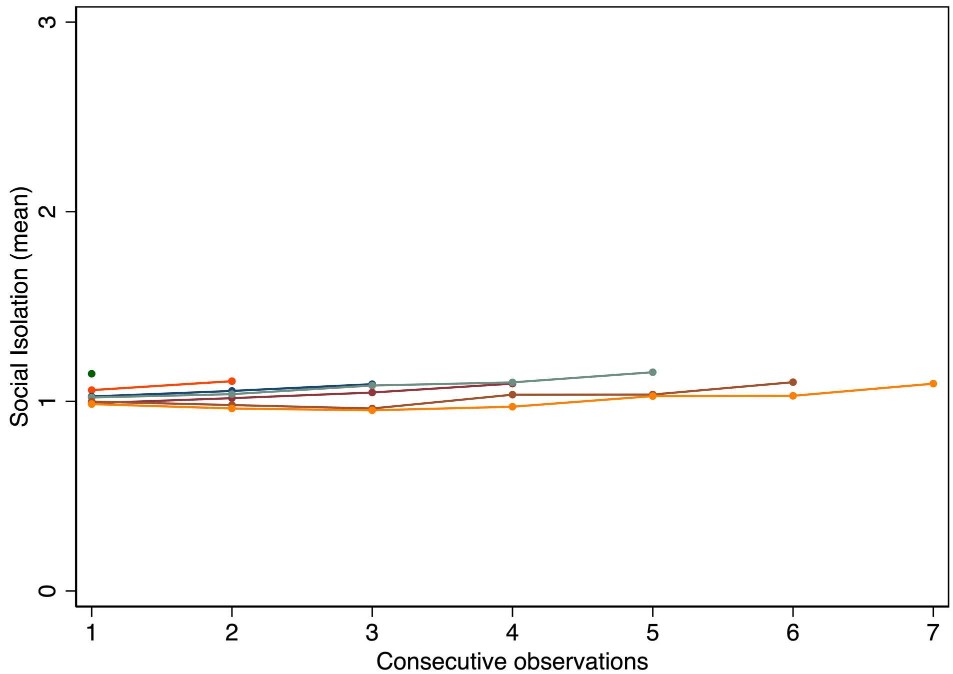


Figure A1: How stable is social isolation over time and across groups with different follow-up periods?

Table A1: Number of observations summarized on each line of Figure

time	No.	%
1	6705	3.46
2	24347	12.56
3	35372	18.25
4	46640	24.07
5	45463	23.46
6	19407	10.02
7	15843	8.18
Total	193777	100.00

Table A2: Transition matrix of SI (0-1) between t and $t+1$

	SI (binary) at $t+1$					
SI (binary) at t	0		1		Total	
	No.	%	No.	%	No.	%
0	20710	69.16	9235	30.84	29945	100.00
1	7761	9.93	70417	90.07	78178	100.00
Total	28471	26.33	79652	73.67	108123	100.00

Note: the number of observations is lower than in our sample because the transition matrix requires for SI to be observed at two consecutive waves, and also therefore not to be dead by $t+1$.

Table A3: Cox models: Effects of all combinations of the Social Isolation Index at Baseline on Mortality.

	Cox hazard ratios	Mean of Xs
No SI items	ref.	0.28
SI=1 (Lives alone)	1.04	0.05
	(0.06)	
SI=1 (No associations)	1.28***	0.41
	(0.04)	
SI=1 (Few contacts)	1.19**	0.05
	(0.09)	
SI=2 (Lives alone+No associations)	1.24***	0.08
	(0.05)	
SI=2 (No associations+Few contacts)	1.42***	0.08
	(0.07)	
SI=2 (Lives alone+Few contacts)	1.28***	0.03
	(0.11)	
SI=3 (Lives alone+No associations+Few contacts)	1.45***	0.04
	(0.09)	
Observations	243515	
Individuals	67676	

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave and country FE.

Follow-up restricted to being > 24months.

Table A4: Cox models: Sensitivity Analysis over specific subsamples.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Binary SI								
SI index>0	1.29***	1.24***	1.27***	.	1.27***	1.33**	1.49***	1.22***
	(0.07)	(0.05)	(0.04)	(.)	(0.04)	(0.15)	(0.13)	(0.04)
Panel B: Continuous SI								
SI (cont.)	1.10***	1.12***	1.12***	1.12**	1.21***	1.08***	1.27***	1.10***
	(0.03)	(0.02)	(0.02)	(0.06)	(0.03)	(0.03)	(0.07)	(0.02)
Restricted to	Female	Male	Has child	No children	Married	Not married	Employed	Not employed
Observations	136186	107329	211638	31877	176405	66294	77762	165753
Individuals	36934	30742	58716	8960	48655	18806	20557	47119

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, wave and country FE.

Follow-up restricted to being > 24 months.

Table A5: Cox models: Does marital status drive the effect of social isolation?.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SI index (binary)	1.25***			1.25***	1.26***				
	(0.04)			(0.04)	(0.04)				
SI including married		1.13***							
		(0.02)							
SI including in couple			1.13***						
			(0.02)						
Married				0.99		0.96*		0.93***	
				(0.02)		(0.02)		(0.02)	
Couple					1.02		0.98		0.95**
					(0.03)		(0.02)		(0.02)
No social activities						1.24***	1.24***	1.47***	1.46***
						(0.03)	(0.03)	(0.04)	(0.04)
Few contacts with children						1.13***	1.14***	1.16***	1.17***
						(0.04)	(0.04)	(0.04)	(0.04)
controls	All	All	All	All	All	All	All	No health	No health
Observations	243515	242699	243515	242699	243515	242699	243515	290999	292044
Individuals	67676	67461	67676	67461	67676	67461	67676	85805	86118

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables, and country and wave FE, except in columns (8) and (9).

Follow-up restricted to being > 24months.

Table A6: Cox models: Effects of SI index at Baseline on Mortality, by Country.

	Southern				Western							Northern		Eastern					
	Spain	Italy	Greece	Portugal	Austria	Germany	Netherl.	France	Switz.	Belgium	Lux.	Sweden	Denmark	Czech Rep.	Poland	Hungary	Slovenia	Estonia	Croatia
SI	1.09	1.59***	1.48**	0.59**	1.54***	1.35**	1.29*	1.00	1.05	1.09	0.93	1.30***	1.25**	1.57***	1.83**	1.33	1.39**	1.39***	2.21
	(0.14)	(0.24)	(0.24)	(0.16)	(0.23)	(0.18)	(0.20)	(0.12)	(0.16)	(0.10)	(0.44)	(0.12)	(0.13)	(0.19)	(0.55)	(0.29)	(0.20)	(0.18)	(1.17)
Obs.	16453	17564	12069	3861	13767	17656	9295	17705	11716	21920	2726	15210	14890	16366	6130	3635	10934	20672	2585
Ind.	4671	4354	3366	1315	3717	4803	3463	4705	2753	5707	987	4186	3820	4584	1674	1644	3299	5078	1084

Exponentiated coefficients; Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included: all health and socio-demo variables displayed in Table 2.
Follow-up restricted to being > 24months.

Table A7: Cox models: Eastern countries Vs Other countries, taking out each country at a time.

	Southern				Western							Northern		Eastern					
EXCLUDING:	Spain	Italy	Greece	Portugal	Austria	Germany	Netherl.	France	Switz.	Belgium	Lux.	Sweden	Denmark	Czech Rep.	Poland	Hungary	Slovenia	Estonia	Croatia
SI	1.18***	1.19***	1.19***	1.19***	1.14**	1.17***	1.17***	1.24***	1.19***	1.22***	1.19***	1.19***	1.19***	1.19***	1.18***	1.19***	1.19***	1.17***	1.19***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
SI X Southern	1.13	0.92	0.96	1.08	1.07	1.03	1.04	0.98	1.02	1.00	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.02
	(0.12)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
SI X Northern	1.10	1.10	1.10	1.10	1.15*	1.11	1.12	1.06	1.10	1.07	1.10	1.08	1.13	1.11	1.10	1.11	1.10	1.11	1.10
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.11)	(0.11)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
SI X Eastern	1.22**	1.22**	1.21**	1.22**	1.27***	1.23**	1.24***	1.17*	1.22**	1.20**	1.22**	1.22**	1.22**	1.17*	1.22**	1.23**	1.22**	1.29***	1.21**
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.10)
Obs.	218701	217590	223085	231293	221387	217498	225859	217449	223438	213234	232428	219944	220264	218788	229024	231519	224220	214482	232569
Ind.	60539	60856	61844	63895	61493	60407	61747	60505	62457	59503	64223	61024	61390	60626	63536	63566	61911	60132	64126

Exponentiated coefficients; Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included: all health and socio-demo variables displayed in Table 2.
Same specification as in column (4) of Table 4.

Table A8: Descriptive Statistics on Observables, by group of countries

	Southern		Western		Northern		Eastern	
	mean	sd	mean	sd	mean	sd	mean	sd
Social Isolation								
SI> 0	0.86	(0.34)	0.66	(0.47)	0.58	(0.49)	0.79	(0.41)
Lives Alone	0.15	(0.36)	0.22	(0.41)	0.21	(0.41)	0.20	(0.40)
No participation to any social activities	0.81	(0.39)	0.50	(0.50)	0.42	(0.49)	0.71	(0.45)
Less than weekly contact with children	0.19	(0.39)	0.21	(0.41)	0.17	(0.37)	0.17	(0.38)
Loneliness (RUCLA scale)	3.88	(1.42)	3.67	(1.19)	3.44	(0.95)	3.93	(1.30)
SN disconnectedness	2.22	(0.86)	1.92	(0.91)	1.72	(0.86)	2.19	(0.85)
Health behavior								
Currently Smokes	0.19	(0.39)	0.18	(0.39)	0.20	(0.40)	0.21	(0.40)
Sedentary	0.12	(0.32)	0.06	(0.25)	0.03	(0.18)	0.10	(0.30)
Drank alcohol past 3 months/7 days	0.54	(0.50)	0.81	(0.40)	0.88	(0.32)	0.58	(0.49)
Health care utilization								
Number of drugs/medicine	1.52	(1.49)	1.38	(1.42)	1.16	(1.31)	1.64	(1.57)
Number of doctor visits	6.59	(10.06)	6.56	(9.18)	3.68	(6.03)	6.46	(8.80)
Nb of overnight stays past 12 mo	0.10	(0.29)	0.15	(0.36)	0.11	(0.31)	0.14	(0.35)
Health								
Self-perceived health	3.15	(1.02)	2.91	(1.02)	2.47	(1.09)	3.52	(1.00)
Frailty index	1.14	(1.18)	0.76	(1.01)	0.66	(0.91)	1.00	(1.14)
Number of chronic diseases	1.67	(1.51)	1.61	(1.49)	1.60	(1.46)	1.94	(1.61)
Number of limitations	0.39	(1.30)	0.33	(1.07)	0.26	(0.92)	0.51	(1.39)
Depression score (EuroD)	2.64	(2.46)	2.22	(2.07)	1.81	(1.84)	2.70	(2.27)
Cognitive recall test (higher is good)	3.82	(1.71)	4.76	(1.75)	4.92	(1.63)	4.38	(1.73)
Socio-demographic controls								
Age	65.04	(9.54)	64.00	(9.57)	64.37	(9.75)	65.03	(9.20)
Female	0.53	(0.50)	0.54	(0.50)	0.52	(0.50)	0.58	(0.49)
Education: None	0.12	(0.33)	0.03	(0.16)	0.00	(0.05)	0.01	(0.11)
Education: Primary	0.42	(0.49)	0.14	(0.35)	0.20	(0.40)	0.11	(0.31)
Education: Secondary(i)	0.18	(0.39)	0.17	(0.37)	0.11	(0.32)	0.23	(0.42)
Education: Secondary(ii)	0.14	(0.35)	0.37	(0.48)	0.30	(0.46)	0.42	(0.49)
Education: Tertiary	0.13	(0.33)	0.29	(0.45)	0.38	(0.48)	0.23	(0.42)
Education: Other	0.01	(0.08)	0.01	(0.08)	0.01	(0.08)	0.00	(0.05)
Employed or Self-employed	0.24	(0.43)	0.32	(0.47)	0.43	(0.50)	0.25	(0.44)
Has at least one child	0.84	(0.37)	0.86	(0.35)	0.89	(0.32)	0.90	(0.30)
Lives in rural area	0.21	(0.41)	0.36	(0.48)	0.20	(0.40)	0.39	(0.49)
Lives in a house (not a building)	0.57	(0.49)	0.75	(0.43)	0.76	(0.43)	0.60	(0.49)
Total household income	19824	(22791)	63466	(1793152)	50255	(31250)	11220	(15238)
Household net worth	957123	(85420437)	741221	(64501992)	318756	(449797)	92082	(146661)

Note: Sample with a follow-up of 24 months minimum.

Table A9: Descriptive Statistics on variables excluded from the analysis, by group of countries

	Southern		Western		Northern		Eastern	
	mean	sd	mean	sd	mean	sd	mean	sd
Depression (individual level)								
Depression ever	0.26	(0.44)	0.27	(0.44)	0.26	(0.44)	0.28	(0.45)
Ever treated for depression by doctor or psychiatrist	0.54	(0.50)	0.59	(0.49)	0.54	(0.50)	0.40	(0.49)
Other dimensions taken from ESS (country level)								
State of health services in country (high is better)	5.06	(0.69)	6.05	(0.92)	5.94	(0.63)	4.46	(0.73)
Meets never or so with friends etc.	0.19	(0.12)	0.10	(0.04)	0.05	(0.01)	0.24	(0.08)

Note: Sample with a follow-up of 24 months minimum.

Table A10: Cox model: Effects of SI controlling for the Big 5.

	(1)	(2)	(3)
SI index (binary)	1.20***	1.19***	1.19***
	(0.08)	(0.08)	(0.08)
Extraversion		1.02	0.97
		(0.03)	(0.05)
Agreeableness		0.98	0.99
		(0.03)	(0.05)
Conscientiousness		0.88***	1.25***
		(0.02)	(0.06)
Neuroticism		1.03	0.90**
		(0.03)	(0.04)
Openness		0.95*	1.07
		(0.02)	(0.05)
Big 5	None	Continuous	Binary: LOW (vs high)
Observations	180939	180939	180937
Individuals	41340	41340	41338

Exponentiated coefficients; Standard errors in parentheses

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

Controls included: all health and socio-demo variables displayed in Table 2, and wave and country FE.

Follow-up restricted to being $>$ 24months.

Table A11: Stability of Key coefficient: Effect of Social Isolation at Baseline on Self-assessed Health Six waves Ahead.

	(1)	(2)	(3)	(4)
Social Isolation Index	0.079***	0.070***	0.057***	0.084***
	(0.016)	(0.016)	(0.016)	(0.024)
Demo + baseline SA Health	yes	yes	yes	yes
All Health	no	yes	yes	yes
Socio Demo	no	no	yes	yes
Binary SI index	no	no	no	yes
Observations	7800	7800	7800	7800

Standard errors in parentheses

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

Table A12: Cox models: Education as a benchmark for the effect of social isolation on mortality.

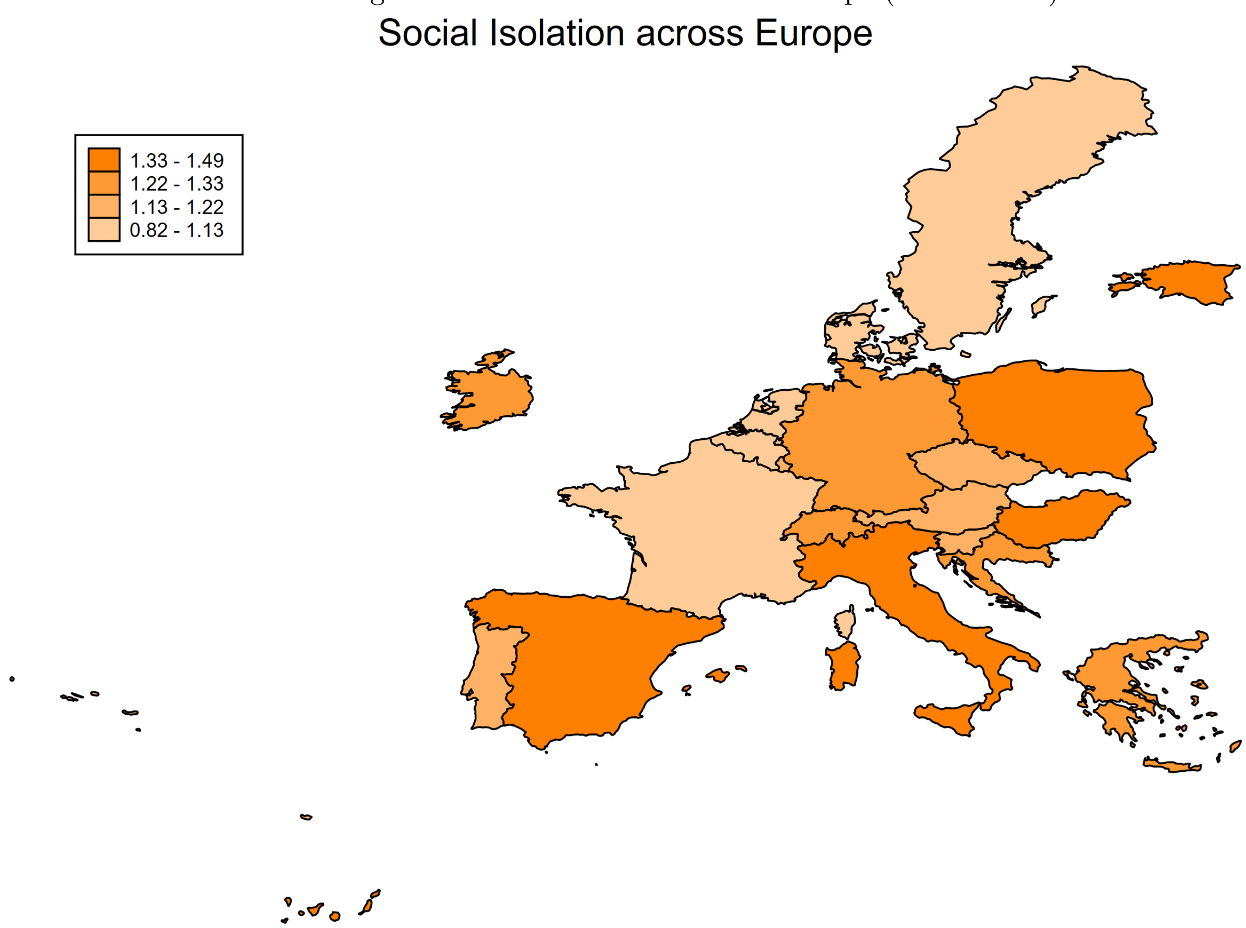
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A (with SI)						
Social Isolation index	1.189***	1.210***	1.128***	1.118***	1.260***	
	(0.013)	(0.016)	(0.017)	(0.018)	(0.039)	
Education: Tertiary		ref.	ref.	ref.	ref.	
Education: None or Primary		1.353***	1.068*	1.077**	1.068*	
		(0.039)	(0.036)	(0.039)	(0.039)	
Education: Secondary		1.226***	1.060*	1.068*	1.063*	
		(0.037)	(0.036)	(0.040)	(0.039)	
Education: Upper Secondary		1.221***	1.140***	1.129***	1.124***	
		(0.033)	(0.035)	(0.037)	(0.037)	
Panel B (without SI)						
Years of education		0.973***	0.996	0.997	0.997	
		(0.002)	(0.003)	(0.003)	(0.003)	
Basic	yes	yes	yes	yes	yes	
Socio-Demo	no	yes	yes	yes	yes	
Health	no	no	yes	yes	yes	
Follow-up	no	no	no	>24 mo	>24 mo	
Binary SI index	no	no	no	no	yes	
Observations	339738	323449	271600	242051	242051	
Individuals	97104	91950	72207	67247	67247	

Exponentiated coefficients; Standard errors in parentheses

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

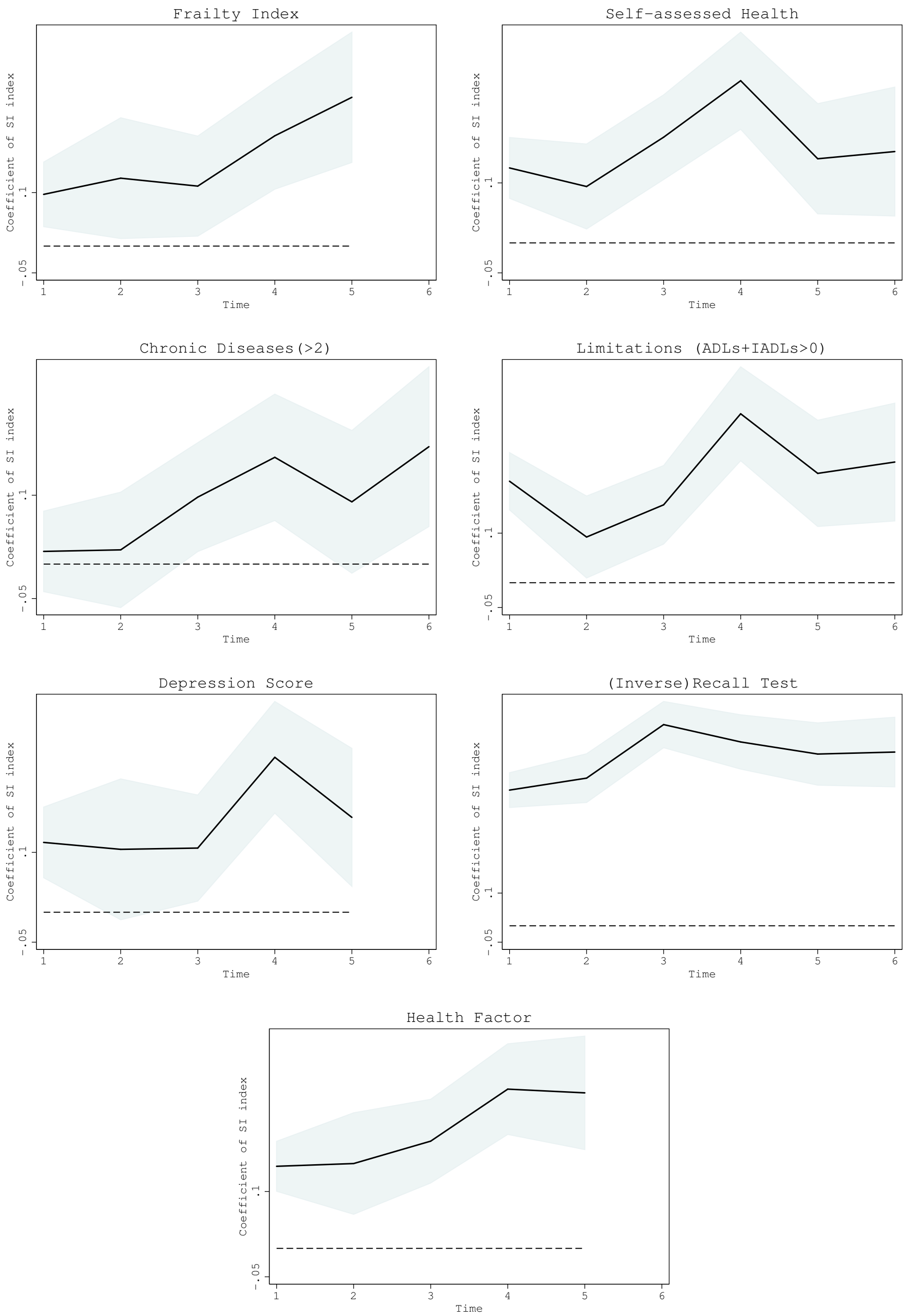
Note: "Basic Demo" controls: age, age squared, sex. "Socio Demo" controls: education categories, whether employed, whether has at least one child, income and wealth quartiles, living in a house (vs in a building), rural (vs urban). "Health" controls: frailty, self-assessed health, number chronic diseases, number limitations, depression score, cognitive recall test. All regressions include wave and country-specific fixed effects. Panel A includes social isolation, Panel B does not.

Figure A2: 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

Figure A3: Benchmarking Social Isolation Effects: Low Education at Baseline and Health Dynamics



Note: The figures display the coefficients and 95 percent confidence intervals for the effect of none or primary education (compared with the highest category of education) 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 2, the outcome at baseline, and wave (a dummy for which wave is baseline) and country FE.