The Causal Effect of Social Activities on Cognition: Evidence from 20 European Countries[#]

Dimitris Christelis
University of Glasgow, CSEF, CFS and Netspar

Loretti I. Dobrescu
University of New South Wales and CEPAR

This version: December 27, 2019

Abstract

Using harmonized data from 20 European countries, we examine the causal effect of being socially active on old age cognition. To address the endogeneity of social participation, we employ nonparametric partial identification methods that bound the average treatment effect for the population under fairly plausible, and thus credible, assumptions. We find strong evidence that social activities have a positive impact on all cognitive dimensions we analyze. At their upper bound these effects are quite large, while at their lower bound they are more modest but still salient. Additionally, we show that ignoring the endogeneity of social activities severely underestimates the uncertainty about their causal effect on cognition.

Keywords: social activities, cognition, aging, partial identification, SHARE
JEL Codes: I10, J14, C14

[#] We would like to thank the Editors, Lisa Brüggen, Denise Doiron, Randall Ellis, Denzil Fiebig, Charles Manski, Anna Sanz-de-Galdeano as well as seminar participants at the University of New South Wales, the 12th CRETE conference, and the Netspar International Pension Workshop for helpful comments. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life). Further support by the European Commission through the 6th framework programme (projects SHARE-I3, RII-CT-2006-062193, as an Integrated Infrastructure Initiative, COMPARE, CIT5-CT-2005-028857, as a project in Priority 7, Citizens and Governance in a Knowledge Based Society, and SHARE-LIFE (CIT4-CT-2006-028812)) and through the 7th framework programme (SHARE-PREP (No 211909) and SHARE-LEAP (No 227822)) is gratefully acknowledged. Substantial co-funding for add-ons such as the intensive training programme for SHARE interviewers came from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064). This research was supported by the Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR), under project CE110001029. Christelis also acknowledges financial support from the European Union and the Greek Ministry of Education under program Thales (Grant MIC 380266), as well as from CEPAR and the School of Economics at the University of New South Wales. All errors remain our own.

1. Introduction

The strong positive association between cognitive abilities and social engagement has been widely documented in recent decades (see Bassuk et al. 1999, Fratiglioni et al. 2004, Ertel et al. 2008 and references within). What remains unclear is the causal interpretation of this association (i.e., whether an active social life does in fact preserve cognitive skills). This is due to the likely endogeneity of social activity, specifically to its correlation with unobservables that also affect cognition and that may be both time-invariant (e.g., personality traits like intellectual curiosity or zest for life) and time-varying (e.g., health, financial, and family problems).

We address this problem by using for the first time in the related literature nonparametric partial identification (PI henceforth) methods that bound the estimate of the causal effect of interest. These methods allow for arbitrary correlations of social activities with unobservables that can also impact cognition, while making much more plausible assumptions than those used in ordinary least square (OLS), panel data and instrumental variables (IV) estimation. As a result, our estimates are considerably more likely to pin down the causal effect of interest than those found in related literature up to now.

Assessing the cognitive effect of social activities is especially important in older age, as the extent to which people maintain their cognitive ability has a major impact on how well they age. For instance, better cognition is associated with better physical health (Der et al., 2009) and can influence one's behaviour in directions beneficial to health (Cutler and Lleras-Muney, 2010). In addition, it is a good predictor of mental health (especially dementia – Fratiglioni et al. 2004, Verghese et al. 2003) and lifetime length (Deary et al. 2009). On the economic side, higher cognition leads to higher productivity at work, thus helping older workers remain active for longer and build up their retirement savings (Engelhardt et al. 2010). There is also a strong positive association of cognitive abilities with financial literacy (Delavande et al. 2008), wealth

and risky portfolio holdings (Smith et al. 2010), occupational rank, job performance and income (Murnane et al. 1995; Judge et al. 2010), and consumption smoothing and life-satisfaction in retirement (Banks et al., 2010).

Previous studies have proposed five main channels through which social involvement may help preserve cognitive functions. First, a 'weak' social network can cause loneliness (Andersson, 1992), a predictor of mental problems (Prince et al. 1997). Second, by providing meaningful social roles and a sense of purpose in old age (Berkman, 2000), social activities might have direct neurohormonal effects, including stress reduction (Fratiglioni et al. 2004). Third, social involvement is likely to inhibit cognitive decline by challenging ones' communication skills and encourages complex interpersonal exchanges (Berkman, 2000). Fourth, staying socially engaged might also require a certain degree of physical activity (beyond regular exercise and walking) that could enhance physical health (Colcombe and Kramer, 2003; Fratiglioni et al. 2004). Finally, an active social life may induce greater self-esteem and better self-care practices, e.g., regular exercise, smoking abstention (Hurst, 1997).

Despite the potential gains to be had from a socially active life, so far only a few studies have attempted to investigate the nature of the relation between social involvement and old age cognition. Hu et al. (2012) use cross-sectional data from the China Health and Retirement Longitudinal Study (CHARLS) to show a positive association between social activities and cognition (especially for short-term memory). The authors try to address the endogeneity of social activities via IV methods, but their instruments are not related strongly enough to their social involvement measures. Furthermore, Engelhardt et al. (2010) use wave 1 of the Survey of Health, Ageing and Retirement in Europe (SHARE) and a stochastic frontier approach (implemented via OLS) to document the same positive relation across 11 European countries. Both these observational studies, however, are likely affected by the existence of unobservables that can impact both social activities and cognition.

Dodge et al. (2015) conduct a randomized control trial on 83 residents of retirement centers in Portland, Oregon (selected after eliminating from 383 candidates those with various physical and mental ailments) in which the intervention consisted of having conversations with trained interviewers 5 days a week for 6 weeks. The authors find a larger improvement compared to the control group in fluency test scores for those with no dementia at baseline, but no difference in the scores in the computer assessment of mild cognitive impairment, memory or word tests. Similarly, Hikichi et al. (2015) assessed the effect of government intervention program in Taketoyo, Japan that created 'community salons' to foster social interactions among local older residents. With a sample of 2,421 individuals, they use propensity score methods (that rely on the assumption of balanced unobservables among control and treated groups) and exogenous IV methods to find that salon participants had a lower chance of functional disability.

On the other hand, our study does not rely on an assumption of balanced unobservables (which is also needed for randomized control trials, see Deaton and Cartwright, 2018), or on the existence of exogenous IVs. Moreover, we use large samples, constructed to be representative of the 50+ population in 20 countries, hence raising fewer external validity concerns.

We thus believe we can provide the first robust causal estimates of social activities for 20 European countries across several cognitive dimensions (numeracy, fluency, immediate and delayed recall capacity). Specifically, we find that being socially active in older age has an important positive impact on cognition. As expected, the lower bound estimates are smaller than those derived by assuming that social participation is exogenous to cognition, while the upper bounds are quite sizeable. For instance, increasing social participation from zero to two or more activities leads on average to an increase in cognitive scores from about 0.03 to about 0.67 standard deviations (SDs henceforth).

2. Data

We use data from SHARE waves 1-2 and 4-6 that were conducted in a total of 20 European countries in 2004-5, 2006-7, 2010-11, 2013 and 2015, respectively (SHARE wave 3 – called SHARELIFE - is a retrospective lifetime survey with a completely different questionnaire compared to standard waves). Sweden, Denmark, Germany, the Netherlands, Belgium, France, Switzerland, Austria, Italy, and Spain participated in all five waves, while Greece was present in the first two and in the sixth one. Ireland was present only in the second wave. Czech Republic and Poland joined from wave 2 onwards, although Poland skipped wave 5. Hungary, Portugal, Slovenia, and Estonia joined SHARE in wave 4, with Estonia and Slovenia continuing into wave 5, when Luxembourg also entered. Finally, Croatia entered the survey in wave 6, in which Portugal also participated.

SHARE surveys those aged 50 and above and their partners and collects data on demographics, physical and mental health (including biomarkers like grip strength and lung capacity), cognition, social activities, housing, employment, income, housing, assets, and expectations (Börsch-Supan et al. 2005). Its questionnaire is modeled after the U.S. Health and Retirement Study (HRS) and importantly, it is harmonized, and thus comparable across all participant countries, as well as over time. SHARE is a longitudinal data set, but there is also a refresher sample in most countries in most waves in order to make up the loss of observations via respondents' death or attrition. Some countries appear only once in SHARE, and obviously for these countries we have a single cross-section.

The SHARE questionnaire provides detailed information on the social activities in which respondents engage. Specifically, to construct our measure of social participation we use questions asked in all five waves on whether respondents i) engage in voluntary or charity work; ii) follow an educational or training course; iii) attend a sport, social or other kind of club; and

iv) are involved in a political or community-related organization. We then define a variable that takes three possible values denoting engagement in zero, one, and two or more activities from the aforementioned four.

It is important to note a change in the SHARE questionnaire after wave 2 that regards the reference period in which social activities are performed. While in waves 1-2 respondents are asked about activities performed in the month prior to the interview, in waves 4-6 the relevant time frame is the previous year. It is not clear, however, whether respondents in waves 4-6, who are asked about activities performed as far back as one year, manage (or even try) to remember what happened during the whole year (as opposed to thinking about activities performed closer to the interview date). To check whether our results are sensitive to the change in the social activities timeline put in place in waves 4-6, we perform a robustness check in which only observations from waves 1-2 are used. Our results remain essentially unchanged.

To obtain information on respondents' cognitive abilities we use questions that are meant to test their recall capacities, as well as their fluency and numeracy skills. *Recall capacity* was captured via a modified Rey's Auditory Verbal Learning Test (Schmidt, 1996), one of the oldest (and extensively validated) mental tests of memory in continuous use albeit in modified form (see Strauss et al., 2006 and references within). Specifically, respondents were read a 10-word list. During the immediate memory test, they were asked to recall aloud as many of these words as possible, just after the interviewer finished reading them. To assess delayed memory, they were asked to recall the same words after five other questions. The two recall scores equal to the number of words correctly recalled. For *fluency*, we use a Retrieval Fluency variable generated via the Woodcock Johnson III (WJ III) Test of Cognitive Abilities (Woodcock et al., 2001) that shows the number of animals respondents named in one minute, excluding repetitions and proper nouns. Finally, *numeracy* is based on the WJ III Test of Achievement (Woodcock et al., 2001) and captured via four maths questions: i) how many people out of 1,000 would be

expected to get sick if the chance of getting a disease is 10 percent; ii) what is the sale cost of a sofa, given an initial price of €300 and a 50 percent discount; iii) what is the initial price of a car if two-thirds of what it costs new is €6,000; iv) what is the final balance of a savings account that initially holds €2,000, at 10 percent interest after 2 years. If respondents answer i) correctly they are then asked iii) and if they are correct again, they are asked iv). If, however, answer i) is incorrect they are directed to ii). Given this sequence, the numeracy score – considered a standard indicator of crystallized intelligence (Salthouse, 1985) - is the number of correctly answered questions plus one. Note that WJ III is the 3rd revision of the first conormed IQ/achievement test ever published (i.e., the 1977 Woodcock-Johnson Psycho-Educational Battery; Mather et al., 2001). This is a multifaceted and technically sophisticated test (McGrew & Woodcock, 2001) with high validity and generalizability of scores (Schrank et al., 2002).

SHARE performs multiple imputation of the missing values of several variables, but not of those that make up our measure of social activities. We thus need to discard 6,773 observations (about 2.8 percent of the sample) with such missing information. Moreover, the rates of missing values for the four measures of cognition range from about 3.5 to 4 percent of the sample. These missing values are, however, multiply imputed in SHARE. As discussed in Section 5, our results are not affected by whether we include in our estimation sample the imputed values of the cognitive test variables.

After combining waves 1-2 and 4-6 of the SHARE data we ended up with a sample of 237,644 observations with non-missing information for social activities across the five waves. Additionally, the numeracy questions were not asked to respondents in the panel subsamples in wave 4-6; consequently, there are 107,026 fewer observations for this cognitive score. Table 1 provides information by country on both social activities and our measures of cognition. We note that Switzerland exhibits the highest average numeracy score (2.8 questions answered correctly out of 4), and Spain the lowest (1.6 correct answers). Spain represents the lowest

extreme also for immediate and delayed recall, with respondents managing to remember on average only 4.20 and 2.82 words out of 10, respectively. The opposite is true of Switzerland, where respondents remembered 5.67 words immediately and 4.48 words after some time. As for fluency, Greece had the lowest average score (13.74 words), while Sweden the highest (23.09 words).

With respect to social activities, we note that only 34 percent of respondents are socially active. The countries with the highest prevalence of social participation are Denmark and the Netherlands (63 percent and 58 percent respectively), while the lowest prevalence is found in Poland (11 percent). Appendix Fig. A.1 plots the mean of each test score by the level of social participation. In line with previous empirical findings, we note a strong positive relation between social activities and all four cognitive test scores. Interestingly, this association appears to be nonlinear, with the rise in mean cognitive scores when going from 0 to 1 activity being larger than when going from 1 to 2+ activities. In the remainder of the paper, we investigate the extent to which this strong association can be interpreted as a causal effect.

3. Empirical Methodology

When estimating the causal effect of social activities on cognition one must account for the likely endogeneity of social activities. Such endogeneity may be due to several factors. First, time-invariant unobservable factors likely affect both the propensity to be socially active and cognitive abilities. For instance, a privileged family background, specific character traits (e.g. inquisitiveness, zest for life) or intellectual abilities can make it easier to be socially active and are positively associated with cognition. Second, time-varying unobservable factors can also act as confounders, with a family crisis, an adverse professional development or financial constraints potentially reducing social activities and causing psychological distress, which in turn can negatively impact cognition.

These unobservable factors imply that socially active people are likely systematically different from those less so, even after controlling for observable characteristics. Hence, traditional OLS or panel data methods that disregard the effect of such unobservables might produce inconsistent estimates of the causal impact of social activity on cognition. A possible solution would be to use IV estimation but finding variables that are exogenous to the outcome (cognition) while being associated with the treatment (social activity) is not easy. Moreover, as the cognitive effect of social activities is very likely heterogeneous across the population, IV estimation is problematic because it identifies only the local average treatment effect (LATE henceforth), that is, the effect of social activity on cognition for those who become socially active due to a change in the instrument value (Imbens and Angrist, 1994). These respondents cannot be identified in the data, and in any case, we would prefer to estimate, if possible, the causal impact of social activity on cognition across the whole population.

To address these issues, we use PI methods that can accommodate all possible sources of endogeneity and partially identify the causal effect of interest for the whole population. PI methods, introduced by Manski (1989, 1990), are nonparametric and produce bounds on the average treatment effect (ATE henceforth), meaning they locate it in an identification region instead of producing a point estimate. Importantly, they use much more plausible assumptions than OLS, panel or IV estimation methods. Below we give a brief overview of the use of PI methods in our case, with more details in the Appendix.

As in Manski (1997), let every individual i have a response function $y_i(\bullet): D \to Y$ that maps mutually exclusive and exhaustive treatments $d \in D$ into outcomes $y_i(d) \in Y$. Importantly, the response functions $y_i(\bullet)$ can differ across individuals in arbitrary ways, thus allowing for unlimited response heterogeneity. Let w_i denote the realized treatment received by i, and $y_i \equiv y_i(w_i)$ the associated observed outcome. In our case, the outcomes are the cognitive

test scores, while the treatment variable (social participation) takes three possible values denoting engaging in zero, one, and two or more activities (as discussed in Section 2).

Let $y_i(d_1)$ and $y_i(d_2)$ be two possible values of the outcome for individual i as a function of two different levels of activities d_1 and d_2, with $d_2 > d_1$. We would like to estimate the ATE of an increased social engagement on the cognitive test score, i.e.,

$$ATE = E[y(d_2)] - E[y(d_1)] \quad (1)$$

Note that the ATE in (1) represents the difference in the two mean outcomes, which are both evaluated using all population units while taking the distribution of all other observable and unobservable variables as given (Manski 1997, p. 1322). This situation can be thought of as a non-randomized experiment in which all participants take two different treatment values. Clearly, this is a counterfactual setup, as respondents are actually observed taking only one treatment value at any given point in time. But it does have some desirable features. First, as the whole population is assumed to take two different treatment values, the control and treated groups coincide with the population. Thus, the problem of systematic differences between the control and treated groups not due to different treatment values is ruled out by construction, as the distribution of all variables other than the outcome and the treatment is taken as given, and thus all relevant factors other than the treatment are controlled for. Second, since this counterfactual non-randomized experiment takes place in a large sample that is representative of the population, the external validity of the results is less of a concern.

A counterfactual experiment, however, necessarily generates counterfactual outcomes. Hence, by the law of iterated expectations, and denoting $E[y(d)|w = d]$ by $E(y|w = d)$, the expected potential outcome $E[y(d)]$ is given by

$$E[y(d)] = E(y|w = d)P(w = d) + E[y(d)|w \neq d]P(w \neq d) \quad (2)$$

where $P(w = d)$ denotes the probability that $w = d$. Note that the term $E[y(d)|w \neq d]$ on the right-hand side of (2) is an unobserved counterfactual one. The remaining three terms on the

right-hand side of (2), however, have sample analogues that are observed in the data. Given that $E[y(d)|w \neq d]$ is unobserved, the unconditional expectation $E[y(d)]$ is also unobserved. Hence the ATE in (1) is equal to the difference between two unobserved average outcomes, and thus cannot be calculated without further assumptions.

If one assumes that the counterfactual conditional expectation $E[y(d)|w \neq d]$ is equal to the observed one when the treatment received is equal to d, i.e., if

$$E[y(d)|w \neq d] = E(y|w = d) \tag{3}$$

then from (2) it follows that

$$E[y(d)] = E(y|w = d) \tag{4}$$

Equation (4) states that the unobserved potential outcome under d is equal to the mean outcome when the treatment in fact received is d. As the sample analogue of the latter is observed in the data, one can estimate the unobserved potential outcome $E[y(d)]$, and then the ATE in (1) as

$$ATE = E(y|w = d_2) - E(y|w = d_1) \tag{5}$$

We refer to the ATE estimate in (5) as the one under exogenous treatment selection (ETS henceforth) because it is derived under the assumption that (3) holds, which in turn implies that respondents receiving different treatments are not systematically different from one another. In other words, (3) implies that selection into treatment is exogenous.

Equation (3) is likely to be true in the case of a randomized control trial, in which treatment assignment is indeed exogenous. In observational data, however, (3) is unlikely to hold because treatment assignment is not random, especially when the treatment variable reflects a respondent's decision. In our context, respondents decide whether to be socially active or not, and, as already mentioned, those socially active are likely to be systematically different from those that are not. Hence, the expected value of the outcome is likely to differ among population groups defined by different social activity levels, and this holds for any value d of the treatment. In other words, the fact that equation (3) is unlikely to hold in our data is due to

the endogeneity of the decision to be socially active. Such endogeneity can have any source (e.g. time-invariant and time-varying unobservables, or selectivity), and all such sources eventually lead to non-random treatment assignment, that is, to the violation of (3).

The starting point of PI is precisely this violation of (3), and thus estimation proceeds under the assumption that the treatment is endogenous. The aim of PI is to manage the endogeneity problem by making credible assumptions that can mitigate its consequences and lead to informative estimates of the effect of interest.

Once one rules out the application of (3), the problem of estimating the unobservable potential outcome $E[y(d)]$ arises. As a solution, Manski (1989) suggested bounding this outcome from above and below by bounding the counterfactual potential outcome $E[y(d)|w \neq d]$ in (2). Let $LB^M(d)$ and $UB^M(d)$ denote the lower and upper bounds on $E[y(d)]$, computed using a set of assumptions M. Given $LB^M(d) \leq E[y(d)] \leq UB^M(d)$, Manski (1990) points out that (1) implies that one can bound the ATE using M as follows:

$$LB^M(d_2) - UB^M(d_1) \leq ATE \leq UB^M(d_2) - LB^M(d_1) \quad (6)$$

The interval between the lower and the upper bound on the ATE is its identification region. Since it is an interval, the ATE is partially identified.

When calculating the upper and lower bounds on $E[y(d)]$, a natural starting point is to assume that, for any value d of the treatment, the outcome space Y is bounded below and above by finite values Y_{min} and Y_{max}, respectively. These values can be used to bound $E[y(d)|w \neq d]$. The assumption of the existence of these finite values is an appropriate one in our case, as all our cognition measures have a minimum of zero, while i) the numeracy score has a maximum equal to four; ii) the two memory scores have maxima equal to 10; and iii) the fluency score is derived based on a one-minute time limit for respondents to provide as many words as possible, which implies a finite maximum score that we set to 100 words - the maximum score observed across all countries in our data.

Given that Y_{min} and Y_{max} are obvious bounds on $E[y(d)|w \neq d]$, we consider the resulting identification regions of $E[y(d)]$ and the ATE as ones derived under no assumptions (NA henceforth). The NA identification regions are typically very wide, thus uninformative, as one would expect when using uninformative assumptions. Hence, we make some additional assumptions to narrow the ATE identification range (see below) and provide more details in the Appendix.

The first assumption we make is that of monotone treatment response (MTR henceforth; see Manski, 1997), which states that a higher level of social participation weakly increases cognition on average (i.e., not necessarily increases it for every individual in the sample; see the Appendix for more details). This is a reasonable assumption, as engaging in more social activities is highly unlikely to have a widespread negative effect on cognition. There are, however, as already discussed, quite a few reasons why one would expect being socially active to be beneficial to cognition. In any case, social activities should at worst have no effect on cognition on average, which is fully allowed by the MTR assumption. This implies that the MTR identification regions always include zero (as shown by Manski, 1997), and thus by using MTR on its own one cannot reject of the null of no effect of the treatment on the outcome.

The use of the MTR assumption can also be justified by the results of several meta-analyses (see, e.g. Kuiper et al., 2016; Yates et al., 2016; Kelly et al., 2017; Evans et al., 2019) that uniformly show that the overwhelming majority of related studies find either a positive or a non-existent association between social activities and cognitive and other health outcomes. Both results are allowed by the MTR assumption.

Our second assumption is that of monotone treatment selection (MTS henceforth; see Manski and Pepper, 2000), which states in our context that those who are more socially active have weakly higher cognitive capacities than those who are less socially active, and this will be true for any common level (possibly counterfactual) of social participation. This assumption

could be justified, for example, if being socially active is due to personality traits such as higher intelligence, intellectual curiosity or zest for life. These characteristics could be associated with (weakly) higher cognition, which would manifest itself under any circumstances, as defined by the level of social participation.

Importantly, Manski and Pepper (2000) show that the combination of MTR and MTS assumptions can be tested, as it implies that the average level of cognition in the sample is weakly positively correlated with the level of social activity (see the Appendix for more details). This clearly holds in our sample, as is apparent from Appendix Fig. A.1. Hence, we cannot reject the combined MTR and MTS assumptions.

The third assumption we use is the monotone instrumental variable (MIV henceforth; see Manski and Pepper, 2000) one, which states that, for any given number of social activities, respondents with higher MIV values have weakly higher cognitive scores on average. Thus, a MIV can be associated with the outcome, but in a weakly monotonic manner. This assumption can be used to narrow the ATE identification region (see Appendix for details). We choose as MIV a variable denoting respondents' body mass index (BMI), with values divided in 15-tiles (as MIVs need to be discretized) and excluding about 2.6 percent of our sample due to BMI<=18.5 as such low values are considered indicative of health issues. In our case, the MIV assumption implies that BMI is negatively associated with the cognitive score, or not associated at all with it. As is the case with exogenous instruments (XIVs), this weak monotonicity assumption is imposed on the unobserved mean potential outcome $E[y(d)]$; it cannot be thus tested using the observed data.

In particular, the MIV assumption for BMI implies that if one examines any two subsamples whose members counterfactually have the same level of social engagement but two different BMI levels (constant in each subsample), the mean cognitive score in the lower BMI subsample will be weakly higher than that in the higher BMI one. First, note that this assumption

refers to the mean outcome, thus it does not require that the weak inequality hold for every individual in the two subsamples. Second, this inequality is assumed to hold for all treatment values (i.e., all levels of social engagement). Finally, as it holds weakly, it also allows for no cognitive effect of social activities.

Clearly this assumption is unverifiable, as in practice respondents with different BMIs do not all engage in the same number of social activities. However, given the strong evidence of a negative association of BMI with cognition (see below), this assumption seems reasonable. Crucially, it is much milder than the (also unverifiable) one of random treatment assignment that is necessary to make social activities exogenous.

Our choice of MIV is based on an extensive literature documenting their negative association with old age cognition. For instance, numerous studies have shown poor cognitive outcomes to occur more likely in overweight and obese individuals (BMI $\geq$ 25 and BMI $\geq$ 30, respectively) who systematically fall in the lowest quartile of global cognition, immediate and delayed recall, fluency and intelligence (Benito-León et al., 2013). Moreover, BMI has been related to a range of chronic problems (Bray, 2004) that have been associated with worsened cognition, such as: i) cardiovascular diseases (e.g., hypertension, myocardial infarction - Waldstein and Elias, 2001), which also predispose to stroke (Kannel, 1992); ii) high or low levels of cholesterol (Muldoon et al., 1997); iii) pancreatic diseases (e.g., diabetes - Ryan, 2001); and iv) various cancers (Berg, 1988). Most of the mechanisms through which these conditions affect cognition are physiological, ranging from reduced cerebral blood flow (due to cardiovascular diseases) or reductions in hippocampal volume (due to obesity).

Importantly, there is independent evidence on the negative association between BMI and cognition that comes directly from the SHARE data used in this study (Ziegler and Doblhammer, 2010; Memel et al., 2016).

These results are confirmed in our sample as well, as BMI is negatively correlated with all our four measures of cognition. As discussed, this is not proving the validity of our MIV because these correlations refer to the observed outcomes y not the potential ones $y(d)$. However, they do suggest that our MIV assumptions are not unreasonable.

The SHARE data are multiply imputed using five different imputed datasets, and thus we use the results in Rubin (1987) to combine results from the five datasets and compute the point estimates of the bounds and standard errors. To conduct inferences on $E[y(d)]$ and the ATE we compute (using 1,000 bootstrap replications) two different kinds of confidence intervals (CIs): i) for results not using MIV, we compute Imbens-Manski (2004) CIs; while ii) for results using MIV, we compute bootstrap percentile CIs as in de Haan (2011) because Imbens-Manski CIs, while very similar to the bootstrap ones, are invalid when using MIVs (Manski and Pepper, 2009). Both kinds of CIs cover $E[y(d)]$ and the ATE with 95 percent and 90 percent probability. To account for multiple imputation, we pool the five bootstrapped datasets (Schomaker and Heumann, 2018). Finally, since MTR+MTS+MIV bounds involve optimizations, the bootstrapped bounds estimates can be biased. We thus apply the bias correction procedure suggested by Kreider and Pepper (2007) (see also Manski and Pepper, 2009). As de Haan (2011) notes, however, the estimate of the bias can be volatile, and thus we show results both with and without the bias correction.

There are many reasons that make PI methods preferable to OLS, IV or paned data estimation (see Appendix A.5). PI methods are nonparametric, they estimate the ATE allowing for its unlimited heterogeneity in the population, and they rely on plausible assumptions while making the identifying power of each one completely transparent. On the other hand, PI can sometimes lead to wide identification regions and so, lack of strong conclusions. However, the point estimates obtained via standard estimation methods may give one a false confidence about results, as the reduction in uncertainty is obtained through non-testable assumptions that might

not hold in the data. This trade-off between point identification obtained via strong assumptions and PI obtained via mild assumptions shows up very clearly in our results below.

Moreover, while it is easier to find MIVs compared to XIVs, there is no guarantee that identification regions that use MIVs will be narrower than those who don't. In any case, as discussed in Appendix A.4 and similarly to XIVs, any valid MIV will generate a valid identification region for $E[y(d)]$, although not necessarily the narrowest possible one.

4. Empirical Results

We begin by presenting our results for the ATEs of being socially active on respondents' cognitive scores, as obtained from the whole sample. Table 2 shows the related estimates that capture a change in social participation from no activities to two or more activities. For every estimation method, we show the estimated lower and upper bounds on the ATE (or, for ETS, the point estimate), as well as the 95 percent and 90 percent CIs of the ATE.

The first method assumes ETS, that is, social participation is completely exogenous to cognition. In practice, these estimates are equal to those obtained by running a weighted OLS regression on a constant and two dummy variables denoting one and two or more social activities. The ETS results show that being socially active has a strong positive effect on cognition, with point estimates amounting to 0.67-0.84 SDs for all four cognitive scores when engaging in two or more activities compared to none. In addition, the 95 percent CIs around the ETS estimates are very narrow, implying that little uncertainty affects these estimates.

As already mentioned, however, the ETS estimates are likely to be upwardly biased due to various unobservables that are positively correlated with both social participation and cognition. We thus address the endogeneity of social participation with the PI methods discussed in Section 3, starting with the NA method that produces, as expected, the widest identification regions by making no assumptions other than the existence of finite extrema. We

obtain lower NA bounds of the ATE that are well below zero, and upper NA bounds are well above; thus, the identification ranges are uninformative about the ATEs.

Adding the MTS assumption leads to ATE lower bounds that are somewhat larger than under NA, but still well below zero. On the other hand, upper bounds are considerably smaller compared to the NA ones. Using the MTR assumption on its own implies that the lower bounds of the ATE cannot be smaller than zero, while the upper bounds when compared to NA do not change at all. Finally, using the MIV assumption on its own results in lower (upper) bounds that are somewhat larger (smaller) than the NA ones, but identification regions remain uninformative.

Combining the MTS and MTR assumptions leaves the lower bounds equal to zero and makes the upper bounds considerably smaller than the NA ones. The ATE upper bound derived assuming a change from zero to two or more activities equals the corresponding ATE under ETS (i.e., the difference in observed mean outcomes – see González, 2005). This value is thus likely overestimating the true ATE as it is derived under the very strong assumption of social participation exogeneity. Moreover, the 95 percent CIs of the ATEs under MTR+MTS lie between zero and values slightly above the upper bounds; hence, the ATEs' uncertainty under MTR+MTS is considerably higher than that under ETS.

When combining the MTS and MIV assumptions (using BMI as a MIV), the identification regions become somewhat narrower compared to those obtained under only MTS, but the ATE lower bounds remain well below zero. On the other hand, combining the MTR with the MIV assumptions yields ATE lower bounds larger than zero that are statistically significant (as showed by the lower bounds of the 95 percent or the 90 percent CIs) when one goes from zero to two or more activities. For this particular treatment change, the ATE lower bounds without a bias correction are equal to 0.031 words (numeracy), 0.189 words (fluency), 0.051 words (immediate recall), and 0.062 words (delayed recall). Hence, the MIV assumption

clearly leads to the rejection of the null hypothesis of no effect only when combined with the MTR assumption. Correspondingly, the MTR assumption is necessary but not sufficient for the rejection of the null, as the ATE lower bound that it produces when used on its own is zero.

To get a better idea of the magnitude of the MTR+MIV results for lower bounds we can express them in terms of the SDs of the associated test scores. We find that, with bias correction, when one goes from zero to two or more activities the ATE lower bounds equal 0.028 SDs for numeracy, 0.025 SDs for fluency, 0.028 SDs for immediate recall, and 0.029 SDs for delayed recall. While modest, these lower bounds still indicate that social activity has a meaningful positive impact on cognition. Upper bounds remain uninformative.

Finally, when using MIV together with the MTR and MTS assumptions lower bounds do not change compared to MTR+MIV, while upper bounds become considerably smaller than the MTR+MIV ones. Specifically, these bounds are equal to 0.727 words (0.65 SDs), 6.013 words (0.80 SDs), 1.265 words (0.69 SDs), and 1.44 words (0.68 SDs) respectively, and are all quite lower than both the ETS point estimates and their 95 percent CI lower bounds. Hence, ETS leads to inflated estimates of the cognitive effect of social engagement, while PI estimates, though smaller, still provide strong evidence of the positive causal impact of social activities on cognition.

The overall result patterns are unchanged when not applying the bias correction. Lower bounds are somewhat larger and equal to about 0.035 SDs, while upper bounds are slightly smaller and about 0.61 to 0.78 SDs.

Additionally, we note that the ATEs corresponding to going from zero to one activity are also bounded away from zero when using MTR+MTS+MIV, save for fluency (results are shown in Panel A of Appendix Table A.1). These ATE lower bounds are, however, not statistically significant. As expected, both the lower and the upper bounds of the ATEs are smaller than those derived when the number of social activities increases from zero to two or

more. On the other hand, when activities change from one to two or more, the ATE identification regions (shown in Panel B of Appendix Table A.1) always contain zero. This result could be due to the nonlinear relation between social activities and cognition (Appendix Fig. A.1), with the marginal cognitive impact of an extra activity falling with the number of activities. However, the increased uncertainty associated with PI estimation may also cause this pattern of results.

We also illustrate how much weaker the MIV assumption is compared to the XIV one by computing results using the latter (see Appendix Table A.2; we provide in Appendix A.4 details on how the XIV assumption operates in PI). We note that in almost all cases ATE lower bounds become much larger while upper bounds much smaller. Hence, the ATE becomes considerably stronger, and the uncertainty about it is significantly reduced. Given that the exogeneity of BMI is highly unlikely (as discussed in Section 3), this is an example of implausible assumptions producing misleading results.

5. Robustness Checks

To check the robustness of our results we perform various tests. Due to space constraints, we show our results in the Appendix.

We first examine whether there are differences in PI results in regions across Europe. To do so, we group the 20 countries in our sample in two groups, namely i) Northern and Middle Europe, consisting of Sweden, Denmark, Germany, the Netherlands, Belgium, France, Switzerland, Austria, Ireland and Luxembourg; and ii) Southern and Eastern Europe, consisting of Italy, Spain, Portugal, Greece, the Czech Republic, Poland, Hungary, Slovenia, Estonia, and Croatia. We show results based on this sample split in Appendix Tables A.3 and A.4 and note that they are very similar to the whole sample ones. Specifically, lower bounds of the ATEs are strictly positive under MTR+MTS+MIV for both country groups and for all four cognitive

measures when the number of social activities changes from zero to two or more, with magnitudes similar to those from the whole sample. We also find a consistently statistically significant lower bound for numeracy in Southern and Eastern Europe when the number of activities changes from zero to one.

We then derive PI regions separately for SHARE waves 1-2 and waves 4-6 to check whether our results are affected by the change in the definition of the reference period in which social activities are undertaken, as discussed in Section 2. Our results are in Appendix Tables A.5 - A.6, with the only notable difference between the two groups of results coming from the small and generally statistically insignificant numeracy lower bounds in waves 1-2.

Finally, we compute PI regions using only observations without imputed values, which leads to samples that are lower by about 3,500 to 4,500 observations, depending on the score. Results are shown in Appendix Table A.7 and they are clearly very similar to those from our mainline sample that includes imputed values. This is to be expected, given the small, relatively to our sample size, number of missing values.

6. Summary

This study provides, to the best of our knowledge, the first empirically robust causal estimates of the positive effects of social activities on cognition later in life. Using survey data on individuals aged 50+ from 20 European countries, we address the issue of social participation endogeneity by means of nonparametric partial identification methods that use plausible assumptions. The assumption we use that has the highest identifying power is MTR, which states that social activities do not harm cognition on average and can also leave it unchanged. The use of MTR can be justified both through formal statistical testing and by the preponderance of evidence found in the related literature.

We first find that social engagement has a positive impact on all cognitive dimensions we analyze (numeracy, fluency, immediate and delayed recall). At their upper bound these effects are quite large, but smaller than those obtained by assuming away the endogeneity of social activities. Naturally, the lower bounds of the ATE identification regions are also smaller but still indicate that social engagement has a meaningful positive impact on cognition. Second, we show that ignoring the endogeneity of social activities severely underestimates the uncertainty about their causal effect on cognition. Finally, our results hold true regardless of respondents' geographic region.

From a policy perspective, our findings suggest that promoting active ageing should be a priority. For instance, access to transport and to technologies supporting self-sufficiency and independent life, opportunities for social, cultural and leisure activities, as well as educational courses targeted to older individuals, could be subsidized. Similarly, policies targeting the reduction of the employment gap between older women and men could be implemented to provide equal chances to longer and better working life (e.g., increasing delayed retirement provisions, flexible hours, part-time work). Finally, insurance companies could implement schemes that reward those who engage in activities that enhance cognitive fitness, and public campaigns could raise awareness about the benefits of active ageing. Such interventions may provide older individuals with higher quality of life and be cost-effective by keeping them in better health, mental and physical, deeper into their lives.

One interesting future work direction relates to the cognitive impact of the timing and duration of social participation along one's life course. For instance, it could be worthwhile to examine whether the positive impact of social activities on cognition extends to ages younger than 50. Or whether being socially active in older age reflects long-life behavioural patterns or lifestyle changes that occurred rather late in life. Such findings would provide policymakers

with valuable information on the mechanisms that could potentially change the dynamics of cognitive decline and bring considerably higher quality of life and financial gains in old age.

References

Andersson, L. (1992) 'Loneliness and perceived responsibility and control in elderly community residents.' *Journal of Social Behavior and Personality* 3, 431- 43.

Banks, J., O'Dea, C., Oldfield, Z. (2010) 'Cognitive function, numeracy and retirement saving trajectories.' *Economic Journal* 120, 381–410.

Bassuk, S.S., Glass, T.A., Berkman, L.F. (1999) 'Social disengagement and incident cognitive decline in community-dwelling elderly persons.' *Annals of Internal Medicine* 131, 165–73.

Benito-León, J., Mitchell, A. J., Hernández-Gallego, J., et al. (2013) 'Obesity and impaired cognitive functioning in the elderly: A population-based cross-sectional study (NEDICES).' *European Journal of Neurology* 20(6), 899–e77.

Berg, R.A. (1988) 'Cancer. In:' *Medical Neuropsychology*, Tarter, RE, Van Thiel, DH, Edwards, KL, eds. New York: Plenum Press.

Berkman, L.F. (2000) 'Which influences cognitive function: Living alone or being alone?' *Lancet* 355, 1291-92.

Börsch-Supan A., Brugiavini A., Jürges H., Mackenbach J., Siegriest J., Weber G., eds. (2005) *Health, Ageing and Retirement in Europe: first results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim Research Institute for the Economics of Aging: Mannheim.

Bray, G.A. (2004) 'Medical consequences of obesity.' *Journal of Clinical Endocrinology and Metabolism* 89(6), 2583–89.

Colcombe, S., Kramer, A. F. (2003) 'Fitness effects on the cognitive function of older adults: A meta-analytic study.' *Psychological Science* 14, 125–130.

Cutler, D., Lleras-Muney, A. (2010) 'Understanding differences in health behaviors by education.' *Journal of Health Economics* 29, 1-28.

Deary, I.J., Corley, J., Gow, A.J., et al. (2009) 'Age-associated cognitive decline.' *British Medical Bulletin* 92, 135–52.

de Haan, M. (2011) 'The effect of parents' schooling on child's schooling: A nonparametric bounds analysis.' *Journal of Labor Economics* 29(4), 859-92.

Deaton, A., Cartwright, N. (2018), 'Understanding and misunderstanding randomized controlled trials,' *Social Science and Medicine*, 210, 2-21.

Delavande, A., Rohwedder, S., Willis, R.J. (2008) 'Preparation for retirement, financial literacy and cognitive resources.' Michigan Retirement Research Center No. 2008-190.

Der, G., Batty, G.D, Deary, I.J. (2009) 'The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth.' *Intelligence* 37, 573–80.

Dodge, H.H, Zhu, J., Mattek, N.C., Bowman, M., Ybarra, Wild, K.V., Loewenstein, D.A., Kayea, J.A (2015), ' Web-enabled conversational interactions as a method to improve cognitive functions: Results of a 6-week randomized controlled trial,' *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 1, 1-12.

Engelhardt, H, Buber, I., Skirbekk, V., et al. (2010) 'Social involvement, behavioural risks and cognitive functioning among the aged.' *Ageing and Society* 30(5), 779-809.

Evans. I.E.M., Martyr, A., Collins, R., Brayne, C., Clare, L. (2019), 'Social Isolation and Cognitive Function in Later Life: A Systematic Review and Meta-Analysis.' *Journal of Alzheimer's Disease* 70, 119–144.

Ertel, K.A., Glymour, M.M., Berkman, L.F. (2008) 'Effects of social integration on preserving memory function in a nationally representative US elderly population.' *American Journal of Public Health* 98(7), 1215–20.

Fratiglioni, L., Paillard-Borg, S., Winblad, B. (2004) 'An active and socially integrated lifestyle in late life might protect against dementia.' *Lancet Neurology* 3(6), 343–53.

González, L. (2005). 'Nonparametric bounds on the returns to language skills.' *Journal of Applied Econometrics* 20, 771–95.

Hikichi, H., Kondo, N., Kondo, K., Aida, J., Takeda, T., Kawachi, I. (2015), 'Effect of a community intervention programme promoting social interactions on functional disability prevention for older adults: propensity score matching and instrumental variable analyses, JAGES Taketoyo study,' *Journal of Epidemiology and Community Health*, 69, 905-10.

Hu, Y., Lei, X., Smith, J.P., et al. (2012) 'Effects of social activities on cognitive functions: Evidence from CHARLS.' In: National Research Council (US) Panel on Policy Research and Data Needs to Meet the Challenge of Aging in Asia; Smith JP, Majmundar M, eds. *Aging in Asia: Findings From New and Emerging Data Initiatives.* Washington (DC): National Academies Press (US).

Hurst, D.F., Boswell, D.L., Boogaard, S.E., et al. (1997) 'The relationship of self-esteem to the health-related behaviors of the patients of a primary care clinic.' *Archives of Family Medicine* 6, 67-70.

Imbens, G.W., Angrist, J. (1994), 'Identification and estimation of local average treatment effects.' *Econometrica* 62, 467–76

Imbens, G.W., Manski, C. (2004) 'Confidence intervals for partially identified parameters.' *Econometrica* 72, 1845–57.

Judge, T.A, Klinger, R.L, Simon, L.S. (2010) 'Time is on my side: Time, general mental ability, human capital, and extrinsic career success.' *Journal of Applied Psychology* 95, 92–107.

Kannel, W.B. (1992) 'The Framingham experience.' In: *Coronary Heart Disease Epidemiology*; Marmot, M, Elliott, P, eds. New York: Oxford.

Kelly, M.E., Duff, H., Kelly, S., McHugh Power, J.E., Brennan, S., Lawlor, B.A., Loughrey, D.G. (2017) 'The impact of social activities, social networks, social support and social relationships on the cognitive functioning.' Systematic Reviews 6:259.

Kreider, B., Pepper, J.V. (2007) 'Disability and employment: Re-evaluating the evidence in light of reporting errors.' *Journal of the American Statistical Association* 102, 432–41.

Kuiper, J.S., Zuidersma, M., Zuidema, S.U. (2016) 'Social relationships and cognitive decline: a systematic review and meta-analysis of longitudinal cohort studies.' *International Journal of Epidemiology* 45, 1169–1206.

Manski, C.F. (1989) 'Anatomy of the selection problem.' *Journal of Human Resources* 24, 343–60.

Manski, C. (1990) 'Nonparametric bounds on treatment effects.' *American Economic Review: Papers and Proceedings* 80, 319-23.

Manski, C. (1997) 'Monotone treatment response.' *Econometrica* 65, 1311–34.

Manski, C., Pepper, J. (2000) 'Monotone instrumental variables: With an application to the returns to schooling.' *Econometrica* 68, 997–1010.

Manski, C., Pepper, J. (2009) 'More on monotone instrumental variables.' *Econometrics Journal* 12, S200 – 16.

Mather, N., Woodcock, R.W. (2001) *Woodcock-Johnson III Tests of Cognitive Abilities examiner's manual.* Itasca, IL: Riverside Publishing.

McGrew, K.S., Woodcock, R.W. (2001) *Woodcock-Johnson III technical manual.* Itasca, IL: Riverside Publishing.

Memel, M., Bourassa, K., Woolverton, C., et al. (2016) 'Body mass and physical activity uniquely predict change in cognition for aging adults.' *Annals of Behavioral Medicine* 50(3), 397-408.

Muldoon, M.F., Ryan, C.M., Matthews, K.A., et al. (1997) 'Serum cholesterol and intellectual performance.' *Psychosomatic Medicine* 59, 382-7.

Murnane, R.J., Willett, J.B., Levy, F. (1995) 'The growing Importance of cognitive skills in wage determination.' *Review of Economics and Statistics* 77(2), 251-66.

Prince, M.J., Harwood, R.H., Blizard, R.A., et al. (1997) 'Social support deficits, loneliness and life events as risk factors for depression in old age. The Gospel Oak Project VI.' *Psychological Medicine* 27, 323-32.

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*, New York, NY: John Wiley & Sons.

Ryan, C.M. (2001) 'Diabetes-associated cognitive dysfunction.' In: *Neuropsychology of Cardiovascular Disease*; Waldstein, SR, Elias, MF, eds. Mahwah, NJ: L. Erlbaum Assoc.

Salthouse T.A. (1985) *A Theory of Cognitive Aging*. Amsterdam, North-Holland.

Schmidt, M. (1996) *Rey Auditory and Verbal Learning Test. A handbook*. Los Angeles: Western Psychological Assoc.

Schomaker, M. and C. Heumann (2018), "Bootstrap inference when using multiple imputation," *Statistics in Medicine*, 37, 2252–66.

Schrank, F.A., Flanagan, D.P., Woodcock, R.W., Mascolo, J.T. (2002). *Essentials of WJ III Cognitive Abilities assessment.* New York: John Wiley & Sons, Inc.

Strauss, E., Sherman, E.M.S., Spreen, S. (2006) A compendium of neuropsychological tests administration, norms, and commentary (3rd edition). New York: Oxford University Press.

Smith, J.P., McArdle, J.J., Willis, R.J. (2010) 'Financial decision making and cognition in a family context.' *Economic Journal* 120, 363–80.

Verghese, J., Lipton, R.B., Katz, M.J. *et al*. (2003) 'Leisure activities and the risk of dementia in the elderly.' *North England Journal of Medicine* 348, 2508–16.

Waldstein S.R., Elias, M.F. (2001) *Neuropsychology of Cardiovascular Disease*. Mahwah, NJ: L. Erlbaum Assoc.

Woodcock, R.W., McGrew, K.S., Mather, N. (2001) *Woodcock- Johnson III Tests of Cognitive Abilities / Achievement.* Itasca, IL: Riverside Publishing.

Yates, L.A., Ziser, S., Spector, A. (2016) ‘Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis.’ *International Psychogeriatrics* 28, 1791–1806

Ziegler, U., Doblhammer-Reiter, G. (2010) ‘Trends and determinants of severe cognitive impairment – longitudinal results from the Survey of Health, Ageing and Retirement in Europe (SHARE),’ presented at 2010 European Population Conference.

Table 1. Descriptive Statistics

Variable	Austria	Germany	Sweden	Netherlands	Spain	Italy	France
Number of correctly answered numeracy questions (max=4)	2.72	2.63	2.67	2.71	1.62	2.00	2.24
Number of words in fluency test	22.43	21.31	23.09	20.44	15.30	14.95	18.97
Number of words recalled immediately (max=10)	5.59	5.49	5.38	5.38	4.19	4.56	5.04
Number of words recalled with a delay (max=10)	4.28	4.07	4.26	4.11	2.81	3.05	3.75
Number of social activities engaged in	0.53	0.59	0.77	0.82	0.22	0.25	0.58
No participation in any social activity	0.61	0.57	0.44	0.42	0.82	0.79	0.57
Participation in one social activity	0.26	0.28	0.34	0.34	0.14	0.17	0.27
Participation in two or more social activities	0.14	0.16	0.22	0.24	0.04	0.04	0.16
Number of observations	14,872	16,701	15,760	12,096	19,489	18,324	19,076

	Denmark	Greece	Switzerland	Belgium	Czech Rep.	Poland	Ireland
Number of correctly answered numeracy questions (max=4)	2.61	2.39	2.80	2.40	2.58	2.04	2.40
Number of words in fluency test	23.06	13.73	21.06	20.52	22.51	16.55	15.66
Number of words recalled immediately (max=10)	5.64	4.86	5.66	5.27	5.48	4.52	5.21
Number of words recalled with a delay (max=10)	4.45	3.34	4.47	3.88	3.95	2.91	4.06
Number of social activities engaged in	0.89	0.19	0.78	0.67	0.39	0.13	0.53
No participation in any social activity	0.37	0.84	0.45	0.52	0.69	0.89	0.60
Participation in one social activity	0.37	0.13	0.32	0.29	0.24	0.08	0.27
Participation in two or more social activities	0.26	0.03	0.23	0.19	0.08	0.03	0.13
Number of observations	13,746	10,385	11,650	22,451	17,785	5,787	975

	Luxembourg	Hungary	Portugal	Slovenia	Estonia	Croatia	Total
Number of correctly answered numeracy questions (max=4)	2.45	2.33	1.83	2.23	2.37	2.27	2.29
Number of words in fluency test	18.37	17.31	14.37	21.78	21.62	18.61	18.59
Number of words recalled immediately (max=10)	5.42	5.08	4.30	5.02	5.34	5.17	5.00
Number of words recalled with a delay (max=10)	4.40	3.59	3.07	3.36	3.92	3.49	3.60
Number of social activities engaged in	0.70	0.24	0.26	0.50	0.39	0.20	0.46
No participation in any social activity	0.52	0.81	0.80	0.61	0.71	0.84	0.66
Participation in one social activity	0.27	0.13	0.14	0.28	0.19	0.12	0.23
Participation in two or more social activities	0.21	0.06	0.06	0.11	0.10	0.04	0.12
Number of observations	3,019	2,924	3,369	9,537	17,323	2,375	237,644

Notes: All figures denote weighted averages, except for the number of observations.

Table 2. Average treatment effect of social activities on cognitive test scores, due to a change from no social activities to two or more activities, whole sample

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Exogenous treatment selection	0.747		0.711	0.778	0.718	0.773	6.366		6.192	6.542	6.218	6.509	1.309		1.268	1.350	1.275	1.343	1.531		1.480	1.581	1.489	1.576
No assumptions	-2.423	2.442	-2.438	2.456	-2.435	2.454	-21.813	26.110	-21.907	26.196	-21.893	26.181	-5.805	6.483	-5.828	6.506	-5.825	6.503	-4.996	7.292	-5.028	7.314	-5.024	7.310
MTS	-2.423	0.747	-2.438	0.778	-2.435	0.773	-21.813	6.366	-21.907	6.542	-21.893	6.509	-5.805	1.309	-5.828	1.350	-5.825	1.343	-4.996	1.531	-5.028	1.581	-5.024	1.576
MTR	0.000	2.442	0.000	2.456	0.000	2.454	0.000	26.110	0.000	26.196	0.000	26.181	0.000	6.483	0.000	6.506	0.000	6.503	0.000	7.292	0.000	7.314	0.000	7.310
MIV (bias corr.)	-2.397	2.380	-2.424	2.413	-2.418	2.406	-21.633	24.523	-21.822	24.826	-21.781	24.759	-5.745	6.095	-5.795	6.174	-5.784	6.157	-4.929	7.017	-4.988	7.091	-4.975	7.074
MTR + MTS	0.000	0.747	0.000	0.778	0.000	0.773	0.000	6.366	0.000	6.542	0.000	6.509	0.000	1.309	0.000	1.350	0.000	1.343	0.000	1.531	0.000	1.581	0.000	1.576
MTS + MIV (bias corr.)	-2.397	0.727	-2.424	0.795	-2.418	0.780	-21.633	6.013	-21.822	6.272	-21.781	6.215	-5.745	1.265	-5.795	1.317	-5.784	1.305	-4.929	1.441	-4.988	1.515	-4.975	1.498
MTR + MIV (bias corr.)	0.031	2.380	0.001	2.413	0.007	2.406	0.189	24.523	0.047	24.826	0.078	24.759	0.051	6.095	0.014	6.174	0.023	6.157	0.062	7.017	0.022	7.091	0.031	7.074
MTR + MTS + MIV (bias corr.)	0.031	0.727	0.001	0.795	0.007	0.780	0.189	6.013	0.047	6.272	0.078	6.215	0.051	1.265	0.014	1.317	0.023	1.305	0.062	1.441	0.022	1.515	0.031	1.498
MTR + MTS + MIV (no bias corr.)	0.042	0.679	0.012	0.746	0.019	0.731	0.249	5.876	0.109	6.133	0.140	6.077	0.066	1.227	0.030	1.279	0.038	1.267	0.077	1.400	0.039	1.474	0.047	1.458
Number of observations	124,381						231,407						231,407						231,407					

Note: MTR: monotone treatment response, MTS: monotone treatment selection, MIV: monontone instrumental valuabe (reversed 15-tiles of body mass index).

Appendix (to be posted on the journal web site only)

This Appendix presents the results related to the bounds calculations originally discussed in Manski (1990, 1997) and Manski and Pepper (2000).

A.1 No Assumptions (NA) Bounds

As in Manski (1990), one can replace the counterfactual term $E[y(d)|w \neq d]$ in equation (2) by Y_{min} and Y_{max}, and thus bound $E[y(d)]$ from below and above as follows:

$$\begin{matrix} E[y|w=d]P(w=d) + Y_{min}\,P(w \neq d) \\ \leq E[y(d)] \leq \\ E[y|w=d]P(w=d) + Y_{max}P(w \neq d) \end{matrix} \quad \text{(A.1)}$$

The bounds in (A.1) are obtained without imposing any assumptions on the data, other than the existence of finite Y_{min} and Y_{max} The NA bounds can be readily calculated using their sample analogues, as these are observed in the data. As Manski (1989) points out, taking sample averages leads to consistent estimates of $E[y|w=d]$, $P(w=d)$ and $P(w \neq d)$.

A.2 Monotone Treatment Response (MTR)

The NA identification region for the ATE is typically very wide, and always includes zero. Hence, one must make additional assumptions to narrow it. The first such assumption is that of monotone treatment response (see Manski, 1997). The MTR assumption states that for all sample units i, and for any two treatment values $d_1 \in D$ and $d_2 \in D$ such that $d_2 > d_1$,

$$y_i(d_2) \geq y_i(d_1) \quad \text{(A.2)}$$

In our context, the MTR assumption implies that being socially active has a weakly increasing effect on cognition for all individuals in the sample. Importantly, (A.2)

holds irrespective of the treatment in fact received. Given that at each point in time one observes only one outcome for every individual in the sample, one cannot test for the validity of (A.2) in isolation using the data at hand. As already discussed in the Introduction, however, there are various reasons - also supported by considerable evidence - why one would expect social activity to have a positive effect on cognition.

In practice, we use a weaker, and thus more conservative, version of the MTR assumption than the one in (A.2), which states that for any treatment value $d \in D$, and any two values $d_1 \in D$ and $d_2 \in D$ such that $d_2 > d_1$,

$$E[y(d_2)|w = d] \geq E[y(d_1)|w = d] \tag{A.3}$$

Equation (A.3) implies that social activities have a weakly positive effect on cognition on average, but not necessarily for every individual in the sample. Furthermore, this average weak monotonicity holds for all subsamples that are defined by the treatment in fact received. (Given that (A.3) holds for all values d of the observed treatment w, it is clearly the case that the weak monotonicity in (A.3) applies also to the unconditional expectation, hence (A.3) implies $E[y(d_2)] \geq E[y(d_1)]$. However, the converse need not be true.) Clearly, (A.2) implies (A.3), but the converse is not necessarily true. As discussed below, we test the joint validity of (A.3) and of another hypothesis, and we are unable to refute this joint hypothesis.

As Manski (1997) shows, the MTR assumption (A.2) implies that the bounds on $E[y(d)]$ can be expressed as follows:

$$\begin{gathered} E[y|w < d]P(w < d) + E[y|w = d]P(w = d) + Y_{min}\,P(w > d) \\ \leq E[y(d)] \leq \\ Y_{max}P(w < d) + E[y|w = d]P(w = d) + E[y|w > d]P(w > d) \end{gathered} \tag{A.4}$$

However, it is easy to see that one can obtain (A.4) also under the weaker assumption (A.3) instead of (A.2), using the same reasoning as in Manski (1997).

Specifically, note that (A.3) implies that $E[y|w < d] \leq E[y(d)|w < d]$. Hence, $E[y|w < d]$ can be used in (A.1) as a lower bound for the counterfactual term $E[y(d)|w < d]$ instead of Y_{min}, and this leads to the lower bound of (A.4). In an analogous fashion, (A.3) implies that $E[y|w > d] \geq E[y(d)|w > d]$. Hence, $E[y|w > d]$ can be used in (A.1) as an upper bound for the counterfactual term $E[y(d)|w > d]$ instead of Y_{max}, which leads to the upper bound of (A.4).

Importantly, Manski (1997) shows that the identification region of the ATE under MTR has a lower bound equal to zero. This is to be expected, as the MTR assumption in (A.2) or (A.3) rule out the possibility that a higher value of the treatment induces a lower mean outcome, while allowing for the possibility of a zero effect.

A.3 Monotone Treatment Selection (MTS)

One can further narrow down the identification regions of $E[y(d)]$ by adding another assumption to the MTR one, namely that of MTS as introduced by Manski and Pepper (2000, MP henceforth). The MTS assumption states that for any treatment value $d \in D$, and any two values $d_1 \in D$ and $d_2 \in D$ such that $d_2 > d_1$,

$$E[y(d)|w = d_2] \geq E[y(d)|w = d_1] \qquad \text{(A.5)}$$

One can think about the MTS assumption as a particular form of non-random selection into treatment, i.e., a particular form of violation of equation (3). If (3) does not hold, then those who choose different levels of the treatment are also systematically different with respect to the outcome in general. The MTS assumption pins down the direction of this difference, as it states that higher observed treatment levels lead to weakly higher potential outcomes on average, for every value of the treatment.

One can test the joint validity of the MTR and MTS hypotheses using a result from MP (p.1004, footnote 9), namely that these two hypotheses jointly imply that for any two treatment values $d_1 \in D$ and $d_2 \in D$ such that $d_2 > d_1$,

$$E[y|w = d_2] \geq E[y|w = d_1] \tag{A.6}$$

Equation (A.6) states that the MTR and MTS assumptions jointly imply that the observed mean outcomes (i.e. the levels of $E[y(d)]$ under ETS) are weakly increasing in the value of the treatment. This is clearly the case in our context, in which observed mean cognition scores increase with the number of activities (see Fig. 1).

As shown by MP, the MTR+MTS assumption implies that $E[y(d)]$ can now be bounded as follows:

$$\begin{array}{c} E[y|w < d]P(w < d) + E[y|w = d]P(w \geq d) \\ \leq E[y(d)] \leq \\ E[y|w = d]P(w \leq d) + E[y|w > d]P(w > d) \end{array} \tag{A.7}$$

The identification region of the ATE under MTR+MTS, while narrower than the MTR one, has still a lower bound equal to zero. On the other hand, González (2005) shows that the ATE upper bound of a change in the treatment between its minimum to its maximum value (i.e., from no activities to two or more activities) is equal to the ATE under ETS (i.e., to the difference in mean outcomes between the two treatment levels).

A.4 Monotone Instrumental Variables (MIVs)

One can further narrow the identification region of the ATE by using a considerably weaker kind of IV than the usual exogenous one, namely the MIV. MIVs were introduced by MP, and they satisfy the following requirement for any pair of values z_1, z_2 of Z such that $z_2 > z_1$,

$$E[y(d)|Z = z_2] \geq E[y(d)|Z = z_1] \tag{A.8}$$

Equation (A.8) states that the MIV can influence the outcome in a particular direction, but also allows for the possibility of no influence whatsoever. Hence, this requirement is much weaker than that of an exogenous instrument which requires no direct relationship between the instrument and the outcome. It is important to note that (A.8) captures only a positive association of Z with Y; a causal relationship is neither implied nor required.

To better understand how MIVs operate, we first note that we can always express the lower bound on $E[y(d)]$ under a set of assumptions M as

$$LB^M(d) = \sum_z LB^M(d|Z=z)\, P(Z=z) \tag{A.9}$$

Clearly, $P(Z=z)$ is given by the data and thus cannot be changed. Hence, to increase the overall lower bound $LB^M(d)$ one needs to increase the conditional lower bounds $LB^M(d|Z=z)$. Similar arguments hold for the upper bound $UB^M(d)$.

Let us first examine how an exogenous IV (XIV) – the IV type typically used in treatment effect estimation - can help narrow the identification range. Following Manski (1990), a variable Z is a XIV if $\forall\ d \in D, \forall\ z \in Z$,

$$E[y(d)|\ Z=z] = E[y(d)] \tag{A.10}$$

Equation (A.10) implies that conditioning on any value of the XIV does not change the mean potential outcome. Hence, all identification regions conditional on values of Z should provide identical lower and upper bounds on $E[y(d)]$. Therefore, the identification region of $E[y(d)]$ is the intersection of all identification regions conditional on Z. This intersection is contained between the maximum of all conditional lower bounds and the minimum of all conditional upper bounds. Hence, we have

$$\max_z LB^M(d|Z=z) \le E[y(d)] \le \min_z UB^M(d|Z=z) \tag{A.11}$$

Using XIVs implies that one searches for the maximum lower bound and the minimum upper bound on $E[y(d)]$ by partitioning the sample in cells defined by the XIV values and then comparing the extrema calculated in each cell. This search for the extrema is analogous to the search for extrema of objective functions in a dynamic program, or of likelihood functions in econometric estimation, which occurs in subsets of the parameter space defined by the chosen grid and/or the optimization method. Clearly, different XIVs will define different partitions of the sample space, and thus likely yield different extrema.

There are, however, a couple of key difference between searching for extrema in sample space versus parameter space partitions: i) the size of the sample partitions is in practice constrained by the number of observations in each cell, whereas there is no such constraint when partitioning the parameter space; and ii) local extrema of the bounds on $E[y(d)]$ define perfectly valid identification regions, similarly to less informative bounds computed using weaker assumptions. In other words, using different valid XIVs and various possible combinations of their values will always produce valid identification regions, albeit not necessarily the most informative ones. On the contrary, local extrema of objective functions in a dynamic program or of likelihood functions will typically yield estimates that are inconsistent and thus potentially misleading. Hence, PI optimization delivers considerably more robust results than dynamic programming or likelihood optimization.

When using an MIV, equation (A.11) does not hold because (A.8) implies that the MIV is weakly monotonically correlated with the outcome. As a result, one cannot calculate the overall identification region as the intersection of all conditional identification regions, as was the case with XIVs. On the other hand, it is possible to exploit the fact that, by (A.8), a lower bound on $E[y(d)|\, Z = z_1]$ is also a lower bound

on $E[y(d)|\ Z = z]$ for $z \geq z_1$, and, correspondingly, an upper bound on $E[y(d)|\ Z = z_2]$ is also a upper bound on $E[y(d)|\ Z = z]$ for $z \leq z_2$. Hence, one can potentially increase the lower bound $LB^M(d|Z = z)$ in (A.9) by taking the maximum lower bound $LB^M(d|Z = z_1)$ over all $z_1 \leq z$. Correspondingly, one can potentially decrease the upper bound $UB^M(d|Z = z)$ by taking the minimum upper bound $UB^M(d|Z = z_2)$ over all $z_2 \geq z$. Hence, we obtain

$$\max_{z_1 \leq z} LB^M[d|Z = z_1] \leq E[y(d)|Z = z] \leq \min_{z \leq z_2} UB^M[d|Z = z_2] \qquad \text{(A.12)}$$

Once the bounds in (A.12) have been computed for all z, one can take their weighted average over all z and bound the potential outcome $E\big(Y(d)\big)$ as follows:

$$\begin{aligned} &\sum_z P(Z = z) \max_{z_1 \leq z} LB^M[d|Z = z_1] \\ \leq &\sum_z P(Z = z) E[y(d)|Z = z] = E[y(d)] \leq \\ &\sum_z P(Z = z) \min_{z \leq z_2} UB^M[d|Z = z_2] \end{aligned} \qquad \text{(A.13)}$$

Hence, by integrating Z out of the conditional expectation $E[y(d)|Z = z]$, one can obtain bounds on $E[y(d)]$. Clearly, the optimization operations in (A.12) take place over a restricted range of values of Z compared to (A.11), and thus the identifying power of the MIV assumption is smaller than that of the XIV one. This is to be expected, as the weak monotonicity of a MIV in (A.8) is a weaker assumption than the exogeneity of an XIV in (A.10). As is the case with XIVs, this weak monotonicity assumption is imposed on the unobserved potential outcome $E[y(d)]$; hence, it cannot be tested using the observed data without imposing further assumptions.

We note that when using MIVs one partitions the sample space to look for extrema of the bounds conditional on each MIV value, that is, of $LB^M(d|Z = z)$ and

$UB^M(d|Z = z)$. Once these conditional extrema have been computed, they can be used to calculate the unconditional bounds on $E[y(d)]$ by integrating out the conditioning MIV as in (A.13). As is the case with XIVs, all valid MIVs will generate valid identification regions, although not necessarily the most informative ones.

In our context, the MIV used (i.e., BMI), is assumed to have a weakly negative effect on cognition. To adhere to the positive weak positive monotonicity in (A.8) we redefine it by using 15-tiles of its values in reverse order.

A.5 Advantages of PI

All in all, there are many reasons why one would prefer PI methods to other more commonly used ones (e.g. OLS-, IV- or panel data-based) when trying to estimate the causal effect of interest. First, PI methods are completely nonparametric, as they require only the calculation of sample averages of the outcome and the prevalence of the treatment.

Second, PI methods produce estimates of the ATE across all sample units, and not of the LATE as is the case with IV estimation when the treatment is heterogeneous. Thus, they allow for arbitrary forms of heterogeneity of the treatment effect because the ATE is just an average magnitude across sample units. Such unlimited heterogeneity of the treatment effect is not typically allowed for, as in most estimation methods one makes specific assumptions about how the treatment variable enters the specification. Moreover, if one is interested in the heterogeneity of the treatment effect in specific dimensions, then one can simply apply PI methods to subsamples defined by particular combinations of values of control variables.

Third, in PI one bounds the unconditional expectation $E[y(d)]$, taking as given the distribution of all observables and unobservables (other than the treatment) that might

affect the outcome. Hence, one does not need to worry about i) which variables to add in the empirical specification; ii) the manner in which they appear; and iii) whether they are endogenous.

Fourth, PI methods accommodate any form of endogeneity (e.g., due to both time-varying and time-invariant unobservables or selectivity), as they allow for any form of non-random selection into treatment. This also implies that one does not need to assume specific properties of the error term, as is the case with regression methods.

Fifth, PI uses very few and quite mild assumptions to narrow the identification region of the estimates, and some of them can be tested (e.g. MTR+MTS). Importantly, it is completely transparent about how adding each assumption affects the identification region. In contrast, most commonly used estimation methods impose simultaneously many assumptions on the empirical model, and thus it is typically unclear how each of them affects estimates.

Sixth, PI methods allow the use MIVs that can tighten the identification regions. As is the case with standard IV estimation, the assumptions behind those variables cannot be tested without making further assumptions. However, MIVs - unusable in standard IV estimation - are required to be weakly monotonically related to the outcome, which is a much weaker assumption than the exogeneity required of standard IVs.

Seventh, PI uses the data as a cross-section, and so panel data are not required. One can accommodate dependencies among sample units (e.g. due to repeated observation or features of the sampling process) through the appropriate clustering and stratification when bootstrapping standard errors.

Table A.1. Average treatment effect of social activities on cognitive test scores, whole sample,

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
Exogenous treatment selection	0.493		0.465	0.519	0.469	0.515	3.884		3.729	4.018	3.763	3.999	0.846		0.814	0.881	0.819	0.875	0.965		0.926	1.007	0.932	1.000
No assumptions	-2.134	2.249	-2.147	2.261	-2.145	2.259	-19.778	23.780	-19.860	23.857	-19.847	23.845	-5.243	5.925	-5.265	5.947	-5.262	5.943	-4.596	6.572	-4.625	6.595	-4.621	6.592
MTS	-1.885	0.627	-1.898	0.652	-1.896	0.648	-17.351	6.014	-17.418	6.156	-17.406	6.133	-4.601	1.372	-4.616	1.407	-4.614	1.401	-4.110	1.647	-4.131	1.688	-4.129	1.682
MTR	0.000	2.139	0.000	2.150	0.000	2.149	0.000	21.939	0.000	21.997	0.000	21.987	0.000	5.454	0.000	5.470	0.000	5.466	0.000	5.956	0.000	5.977	0.000	5.974
MIV (bias corr.)	-2.102	2.174	-2.130	2.208	-2.124	2.201	-19.557	22.025	-19.743	22.358	-19.702	22.285	-5.176	5.512	-5.227	5.595	-5.216	5.576	-4.521	6.261	-4.579	6.339	-4.566	6.322
MTR + MTS	0.000	0.517	0.000	0.542	0.000	0.538	0.000	4.174	0.000	4.306	0.000	4.286	0.000	0.900	0.000	0.932	0.000	0.929	0.000	1.031	0.000	1.071	0.000	1.065
MTS + MIV (bias corr.)	-1.860	0.604	-1.888	0.651	-1.881	0.641	-17.135	5.691	-17.324	5.920	-17.283	5.870	-4.535	1.305	-4.586	1.356	-4.575	1.345	-4.035	1.556	-4.094	1.618	-4.081	1.605
MTR + MIV (bias corr.)	0.000	2.057	0.000	2.092	0.000	2.084	0.000	19.963	0.000	20.318	0.000	20.240	0.000	4.984	0.000	5.072	0.000	5.053	0.000	5.557	0.000	5.646	0.000	5.626
MTR + MTS + MIV (bias corr.)	0.006	0.488	0.000	0.538	0.000	0.527	0.000	3.864	0.000	4.092	0.000	4.042	0.004	0.829	0.000	0.874	0.000	0.864	0.002	0.940	0.000	0.995	0.000	0.983
MTR + MTS + MIV (no bias corr.)	0.017	0.461	0.000	0.510	0.000	0.499	0.000	3.728	0.000	3.954	0.000	3.905	0.016	0.806	0.000	0.851	0.000	0.841	0.013	0.915	0.000	0.969	0.000	0.957
Panel B. ATE of a change from one activitiy to two or more activities																								
Exogenous treatment selection	0.254		0.212	0.291	0.219	0.284	2.482		2.303	2.678	2.329	2.652	0.462		0.415	0.506	0.423	0.498	0.566		0.508	0.617	0.518	0.610
No assumptions	-3.424	3.329	-3.435	3.341	-3.433	3.339	-32.112	32.408	-32.216	32.501	-32.202	32.486	-8.274	8.270	-8.301	8.295	-8.296	8.291	-8.112	8.432	-8.139	8.456	-8.136	8.452
MTS	-2.459	2.041	-2.480	2.074	-2.476	2.067	-20.183	16.074	-20.280	16.266	-20.268	16.237	-5.329	4.061	-5.355	4.106	-5.350	4.099	-4.292	3.290	-4.324	3.339	-4.318	3.332
MTR	0.000	1.881	0.000	1.894	0.000	1.892	0.000	21.358	0.000	21.425	0.000	21.415	0.000	5.225	0.000	5.242	0.000	5.239	0.000	6.339	0.000	6.360	0.000	6.357
MIV (bias corr.)	-3.392	3.302	-3.418	3.325	-3.412	3.320	-31.689	32.111	-31.893	32.253	-31.848	32.222	-8.183	8.198	-8.229	8.236	-8.219	8.228	-7.999	8.346	-8.053	8.390	-8.041	8.380
MTR + MTS	0.000	0.593	0.000	0.624	0.000	0.619	0.000	5.024	0.000	5.199	0.000	5.167	0.000	1.016	0.000	1.055	0.000	1.050	0.000	1.197	0.000	1.244	0.000	1.238
MTS + MIV (bias corr.)	-2.439	2.024	-2.483	2.092	-2.474	2.077	-19.997	15.821	-20.193	16.064	-20.150	16.011	-5.300	4.050	-5.345	4.102	-5.335	4.090	-4.242	3.234	-4.295	3.304	-4.283	3.289
MTR + MIV (bias corr.)	0.000	1.855	0.000	1.881	0.000	1.875	0.000	20.811	0.000	21.042	0.000	20.991	0.000	5.129	0.000	5.174	0.000	5.164	0.000	6.240	0.000	6.290	0.000	6.279
MTR + MTS + MIV (bias corr.)	0.000	0.573	0.000	0.642	0.000	0.627	0.000	4.737	0.000	4.973	0.000	4.921	0.000	0.996	0.000	1.044	0.000	1.034	0.000	1.132	0.000	1.202	0.000	1.187
MTR + MTS + MIV (no bias corr.)	0.000	0.530	0.000	0.597	0.000	0.582	0.000	4.609	0.000	4.844	0.000	4.792	0.000	0.959	0.000	1.007	0.000	0.996	0.000	1.091	0.000	1.161	0.000	1.146
Number of observations	124,381						231,407						231,407						231,407					

Note: MTR: monotone treatment response, MTS: monotone treatment selection, MIV: monontone instrumental valuabe (reversed 15-tiles of body mass index).

Table A.2. Average treatment effect of social activities on cognitive test scores, whole sample, using BMI as an exogenous instrument

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
XIV (bias corr.)	-1.800	2.170	-1.859	2.208	-1.846	2.200	-17.215	22.016	-17.630	22.351	-17.539	22.277	-4.549	5.511	-4.658	5.594	-4.634	5.576	-3.805	6.261	-3.926	6.340	-3.899	6.322
MTS + XIV (bias corr.)	-1.504	0.482	-1.568	0.542	-1.554	0.529	-14.277	5.037	-14.737	5.386	-14.636	5.309	-3.767	1.052	-3.888	1.134	-3.861	1.116	-3.195	1.217	-3.325	1.318	-3.297	1.295
MTR + XIV (bias corr.)	0.000	2.053	0.000	2.092	0.000	2.083	0.000	19.945	0.000	20.305	0.000	20.226	0.000	4.981	0.000	5.070	0.000	5.050	0.000	5.552	0.000	5.642	0.000	5.622
MTR + MTS + XIV (bias corr.)	0.208	0.405	0.163	0.465	0.173	0.451	0.854	3.415	0.610	3.728	0.664	3.659	0.275	0.601	0.207	0.683	0.222	0.665	0.316	0.671	0.242	0.763	0.258	0.743
MTR + MTS + XIV (no bias corr.)	0.210	0.364	0.166	0.421	0.176	0.408	0.955	3.175	0.719	3.473	0.771	3.408	0.304	0.571	0.241	0.651	0.255	0.633	0.353	0.641	0.282	0.731	0.297	0.711
Panel B. ATE of a change from no activities to two or more activities																								
XIV (bias corr.)	-2.181	2.373	-2.228	2.411	-2.218	2.403	-19.781	24.498	-20.152	24.807	-20.070	24.739	-5.263	6.089	-5.356	6.170	-5.336	6.152	-4.352	7.014	-4.462	7.092	-4.438	7.075
MTS + XIV (bias corr.)	-2.181	0.572	-2.228	0.661	-2.218	0.641	-19.781	5.305	-20.152	5.753	-20.070	5.654	-5.263	1.045	-5.356	1.148	-5.336	1.125	-4.352	1.212	-4.462	1.329	-4.438	1.303
MTR + XIV (bias corr.)	0.246	2.373	0.199	2.411	0.210	2.403	1.174	24.498	0.937	24.807	0.989	24.739	0.326	6.089	0.261	6.170	0.276	6.152	0.379	7.014	0.304	7.092	0.321	7.075
MTR + MTS + XIV (bias corr.)	0.234	0.586	0.189	0.676	0.199	0.656	1.174	5.305	0.937	5.753	0.989	5.654	0.326	1.045	0.261	1.148	0.276	1.125	0.379	1.212	0.304	1.329	0.321	1.303
MTR + MTS + XIV (no bias corr.)	0.236	0.551	0.193	0.634	0.202	0.616	1.271	5.116	1.041	5.554	1.092	5.458	0.355	0.986	0.295	1.087	0.308	1.064	0.413	1.148	0.343	1.262	0.358	1.237
Panel B. ATE of a change from one activity to two or more activities																								
XIV (bias corr.)	-3.351	3.173	-3.383	3.220	-3.376	3.210	-31.248	31.165	-31.511	31.454	-31.453	31.390	-8.071	7.935	-8.136	8.015	-8.122	7.998	-7.909	8.116	-7.974	8.192	-7.959	8.175
MTS + XIV (bias corr.)	-2.340	1.752	-2.393	1.848	-2.381	1.827	-19.124	13.888	-19.455	14.410	-19.382	14.295	-5.031	3.527	-5.108	3.651	-5.091	3.624	-3.922	2.760	-4.012	2.885	-3.992	2.858
MTR + XIV (bias corr.)	0.000	1.808	0.000	1.844	0.000	1.836	0.000	20.823	0.000	21.038	0.000	20.991	0.000	5.091	0.000	5.149	0.000	5.136	0.000	6.133	0.000	6.202	0.000	6.187
MTR + MTS + XIV (bias corr.)	0.000	0.378	0.000	0.465	0.000	0.445	0.000	3.672	0.000	4.117	0.000	4.019	0.000	0.720	0.000	0.819	0.000	0.797	0.000	0.809	0.000	0.925	0.000	0.899
MTR + MTS + XIV (no bias corr.)	0.000	0.341	0.000	0.423	0.000	0.405	0.000	3.527	0.000	3.961	0.000	3.866	0.000	0.658	0.000	0.755	0.000	0.734	0.000	0.737	0.000	0.850	0.000	0.825
Number of observations	124,381						231,407						231,407						231,407					

Note: MTR: monotone treatment response, MTS: monotone treatment selection, XIV: exogenous instrumental valuabe (reversed 15-tiles of body mass index).

Table A.3. Average treatment effect of social activities on cognitive test scores, Northern and Central Europe

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
Exogenous treatment selection	0.350		0.318	0.382	0.323	0.377	2.693		2.513	2.870	2.548	2.842	0.591		0.548	0.630	0.556	0.625	0.703		0.648	0.754	0.660	0.745
No assumptions	-2.299	2.269	-2.316	2.287	-2.314	2.284	-21.988	23.768	-22.106	23.888	-22.089	23.873	-5.699	6.033	-5.729	6.063	-5.725	6.059	-5.308	6.424	-5.343	6.453	-5.338	6.450
MTR	0.000	2.113	0.000	2.130	0.000	2.127	0.000	21.108	0.000	21.202	0.000	21.184	0.000	5.341	0.000	5.361	0.000	5.358	0.000	5.523	0.000	5.549	0.000	5.546
MTR + MTS	0.000	0.378	0.000	0.408	0.000	0.404	0.000	3.026	0.000	3.194	0.000	3.163	0.000	0.661	0.000	0.699	0.000	0.692	0.000	0.790	0.000	0.837	0.000	0.829
MTR + MTS + MIV (bias corr.)	0.016	0.345	0.000	0.391	0.000	0.381	0.007	2.778	0.000	3.023	0.000	2.969	0.034	0.582	0.000	0.646	0.000	0.632	0.009	0.708	0.000	0.782	0.000	0.765
MTR + MTS + MIV (no bias corr.)	0.033	0.310	0.000	0.355	0.003	0.345	0.087	2.596	0.000	2.840	0.000	2.786	0.054	0.541	0.000	0.604	0.010	0.590	0.037	0.656	0.000	0.729	0.000	0.713
Panel B. ATE of a change from no activities to two or more activities																								
Exogenous treatment selection	0.552		0.516	0.590	0.522	0.585	4.614		4.419	4.857	4.450	4.801	0.995		0.950	1.043	0.957	1.034	1.202		1.142	1.261	1.151	1.250
No assumptions	-2.619	2.466	-2.637	2.485	-2.634	2.482	-24.111	26.053	-24.234	26.176	-24.213	26.159	-6.262	6.601	-6.292	6.633	-6.287	6.627	-5.708	7.155	-5.744	7.184	-5.738	7.181
MTR	0.000	2.466	0.000	2.485	0.000	2.482	0.000	26.053	0.000	26.176	0.000	26.159	0.000	6.601	0.000	6.633	0.000	6.627	0.000	7.155	0.000	7.184	0.000	7.181
MTR + MTS	0.000	0.552	0.000	0.590	0.000	0.585	0.000	4.614	0.000	4.857	0.000	4.801	0.000	0.995	0.000	1.043	0.000	1.034	0.000	1.202	0.000	1.261	0.000	1.250
MTR + MTS + MIV (bias corr.)	0.048	0.508	0.009	0.564	0.018	0.552	0.307	4.429	0.065	4.711	0.118	4.649	0.102	0.947	0.045	1.014	0.058	0.999	0.099	1.127	0.039	1.219	0.053	1.198
MTR + MTS + MIV (no bias corr.)	0.062	0.462	0.025	0.518	0.033	0.505	0.398	4.200	0.159	4.481	0.211	4.419	0.120	0.892	0.063	0.959	0.076	0.944	0.121	1.060	0.061	1.151	0.074	1.131
Panel C. ATE of a change from one activity to two or more activities																								
Exogenous treatment selection	0.202		0.164	0.242	0.171	0.234	1.921		1.721	2.127	1.761	2.103	0.404		0.356	0.449	0.363	0.442	0.499		0.439	0.561	0.448	0.552
No assumptions	-3.234	3.111	-3.250	3.128	-3.248	3.125	-29.959	30.122	-30.103	30.268	-30.080	30.246	-7.700	7.705	-7.737	7.743	-7.731	7.737	-7.537	7.868	-7.576	7.904	-7.570	7.897
MTR	0.000	1.734	0.000	1.750	0.000	1.748	0.000	19.834	0.000	19.923	0.000	19.910	0.000	4.998	0.000	5.020	0.000	5.016	0.000	5.924	0.000	5.950	0.000	5.945
MTR + MTS	0.000	0.407	0.000	0.444	0.000	0.437	0.000	3.376	0.000	3.582	0.000	3.550	0.000	0.723	0.000	0.764	0.000	0.758	0.000	0.879	0.000	0.934	0.000	0.924
MTR + MTS + MIV (bias corr.)	0.000	0.375	0.000	0.429	0.000	0.418	0.000	3.215	0.000	3.476	0.000	3.419	0.000	0.707	0.000	0.765	0.000	0.752	0.000	0.813	0.000	0.898	0.000	0.879
MTR + MTS + MIV (no bias corr.)	0.000	0.332	0.000	0.386	0.000	0.374	0.000	3.026	0.000	3.284	0.000	3.227	0.000	0.657	0.000	0.715	0.000	0.702	0.000	0.760	0.000	0.845	0.000	0.826
Number of observations	66,981						126,289						126,289						126,289					

Note: See note to Table A.1. Northern and Middle Europe includes Sweden, Denmark, Germany, the Netherlands, Belgium, France, Switzerland, Austria, Ireland and Luxembourg

Table A.4. Average treatment effect of social activities on cognitive test scores, Southern and Eastern Europe

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
Exogenous treatment selection	0.472		0.423	0.523	0.433	0.514	3.384		3.160	3.610	3.204	3.580	0.914		0.848	0.978	0.857	0.966	0.974		0.903	1.048	0.914	1.039
No assumptions	-1.918	2.222	-1.938	2.242	-1.935	2.238	-16.798	23.795	-16.920	23.891	-16.896	23.874	-4.629	5.780	-4.659	5.807	-4.654	5.802	-3.637	6.772	-3.676	6.804	-3.669	6.798
MTR	0.000	2.173	0.000	2.191	0.000	2.188	0.000	23.060	0.000	23.155	0.000	23.143	0.000	5.605	0.000	5.631	0.000	5.627	0.000	6.540	0.000	6.575	0.000	6.569
MTR + MTS	0.000	0.481	0.000	0.530	0.000	0.521	0.000	3.493	0.000	3.722	0.000	3.690	0.000	0.933	0.000	0.994	0.000	0.985	0.000	0.995	0.000	1.067	0.000	1.058
MTR + MTS + MIV (bias corr.)	0.061	0.362	0.000	0.524	0.009	0.489	0.128	3.085	0.000	3.543	0.000	3.442	0.039	0.845	0.000	0.924	0.000	0.906	0.040	0.911	0.000	1.001	0.000	0.981
MTR + MTS + MIV (no bias corr.)	0.078	0.314	0.014	0.473	0.028	0.437	0.266	2.806	0.041	3.257	0.091	3.157	0.067	0.795	0.014	0.874	0.025	0.857	0.068	0.855	0.007	0.945	0.021	0.925
Panel B. ATE of a change from no activities to two or more activities																								
Exogenous treatment selection	0.713		0.628	0.795	0.640	0.780	6.056		5.652	6.487	5.734	6.428	1.399		1.297	1.504	1.314	1.489	1.488		1.368	1.616	1.388	1.603
No assumptions	-2.166	2.411	-2.188	2.431	-2.184	2.427	-18.714	26.187	-18.862	26.293	-18.836	26.278	-5.188	6.325	-5.225	6.355	-5.220	6.351	-4.037	7.476	-4.082	7.507	-4.075	7.501
MTR	0.000	2.411	0.000	2.431	0.000	2.427	0.000	26.187	0.000	26.293	0.000	26.278	0.000	6.325	0.000	6.355	0.000	6.351	0.000	7.476	0.000	7.507	0.000	7.501
MTR + MTS	0.000	0.713	0.000	0.795	0.000	0.780	0.000	6.056	0.000	6.487	0.000	6.428	0.000	1.399	0.000	1.504	0.000	1.489	0.000	1.488	0.000	1.616	0.000	1.603
MTR + MTS + MIV (bias corr.)	0.069	0.720	0.002	0.979	0.017	0.922	0.218	5.654	0.000	6.282	0.044	6.144	0.057	1.259	0.004	1.424	0.016	1.388	0.055	1.276	0.000	1.503	0.009	1.453
MTR + MTS + MIV (no bias corr.)	0.086	0.552	0.022	0.793	0.036	0.740	0.355	5.193	0.135	5.816	0.183	5.679	0.084	1.172	0.032	1.336	0.043	1.300	0.084	1.191	0.025	1.411	0.038	1.362
Panel C. ATE of a change from one activity to two or more activities																								
Exogenous treatment selection	0.241		0.153	0.338	0.166	0.322	2.672		2.250	3.108	2.312	3.045	0.485		0.379	0.603	0.399	0.582	0.513		0.390	0.656	0.412	0.634
No assumptions	-3.671	3.612	-3.686	3.627	-3.683	3.625	-35.015	35.491	-35.140	35.600	-35.122	35.583	-9.047	9.032	-9.075	9.063	-9.071	9.058	-8.887	9.191	-8.920	9.220	-8.916	9.215
MTR	0.000	2.072	0.000	2.092	0.000	2.089	0.000	23.413	0.000	23.509	0.000	23.493	0.000	5.531	0.000	5.559	0.000	5.554	0.000	6.900	0.000	6.932	0.000	6.929
MTR + MTS	0.000	0.628	0.000	0.712	0.000	0.697	0.000	5.406	0.000	5.823	0.000	5.769	0.000	1.224	0.000	1.330	0.000	1.312	0.000	1.300	0.000	1.426	0.000	1.408
MTR + MTS + MIV (bias corr.)	0.000	0.629	0.000	0.882	0.000	0.827	0.000	4.962	0.000	5.594	0.000	5.455	0.000	1.095	0.000	1.258	0.000	1.222	0.000	1.099	0.000	1.325	0.000	1.275
MTR + MTS + MIV (no bias corr.)	0.000	0.466	0.000	0.692	0.000	0.642	0.000	4.524	0.000	5.149	0.000	5.011	0.000	1.010	0.000	1.172	0.000	1.136	0.000	1.017	0.000	1.235	0.000	1.187
Number of observations	57,400						105,118						105,118						105,118					

Note: See note to Table A.1. Southern and Eastern Europe includes Italy, Spain, Portugal, Greece, the Czech Republic, Poland, Hungary, Slovenia, Estonia, and Croatia.

Table A.5. Average treatment effect of social activities on cognitive test scores, waves 1 and 2

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
Exogenous treatment selection	0.516		0.481	0.550	0.486	0.544	4.045		3.844	4.266	3.880	4.233	0.797		0.744	0.853	0.753	0.845	0.756		0.695	0.820	0.705	0.809
No assumptions	-2.101	2.197	-2.117	2.212	-2.115	2.210	-18.663	23.237	-18.772	23.344	-18.754	23.328	-4.938	5.806	-4.967	5.833	-4.962	5.829	-4.172	6.572	-4.207	6.605	-4.202	6.598
MTR	0.000	2.113	0.000	2.128	0.000	2.125	0.000	22.066	0.000	22.158	0.000	22.145	0.000	5.491	0.000	5.514	0.000	5.510	0.000	6.146	0.000	6.179	0.000	6.175
MTR + MTS	0.000	0.535	0.000	0.566	0.000	0.561	0.000	4.232	0.000	4.448	0.000	4.419	0.000	0.833	0.000	0.888	0.000	0.879	0.000	0.794	0.000	0.854	0.000	0.845
MTR + MTS + MIV (bias corr.)	0.000	0.521	0.000	0.561	0.000	0.552	0.055	3.976	0.000	4.281	0.000	4.214	0.071	0.698	0.005	0.804	0.019	0.781	0.028	0.729	0.000	0.815	0.000	0.796
MTR + MTS + MIV (no bias corr.)	0.005	0.484	0.000	0.524	0.000	0.515	0.160	3.741	0.000	4.046	0.001	3.979	0.096	0.656	0.030	0.761	0.044	0.738	0.053	0.668	0.000	0.754	0.011	0.735
Panel B. ATE of a change from no activities to two or more activities																								
Exogenous treatment selection	0.768		0.721	0.812	0.728	0.806	6.570		6.244	6.859	6.304	6.803	1.282		1.204	1.362	1.217	1.348	1.272		1.179	1.366	1.191	1.349
No assumptions	-2.411	2.388	-2.427	2.405	-2.425	2.402	-21.074	25.714	-21.206	25.829	-21.190	25.809	-5.563	6.434	-5.597	6.466	-5.592	6.461	-4.605	7.392	-4.647	7.420	-4.641	7.416
MTR	0.000	2.388	0.000	2.405	0.000	2.402	0.000	25.714	0.000	25.829	0.000	25.809	0.000	6.434	0.000	6.466	0.000	6.461	0.000	7.392	0.000	7.420	0.000	7.416
MTR + MTS	0.000	0.768	0.000	0.812	0.000	0.806	0.000	6.570	0.000	6.859	0.000	6.803	0.000	1.282	0.000	1.362	0.000	1.348	0.000	1.272	0.000	1.366	0.000	1.349
MTR + MTS + MIV (bias corr.)	0.007	0.744	0.000	0.813	0.000	0.797	0.223	6.054	0.027	6.576	0.070	6.461	0.107	1.233	0.040	1.432	0.055	1.388	0.065	1.207	0.011	1.346	0.023	1.316
MTR + MTS + MIV (no bias corr.)	0.024	0.689	0.000	0.756	0.003	0.742	0.325	5.798	0.129	6.316	0.172	6.202	0.130	1.108	0.064	1.305	0.079	1.261	0.090	1.104	0.036	1.242	0.048	1.211
Panel C. ATE of a change from one activity to two or more activities																								
Exogenous treatment selection	0.252		0.202	0.299	0.208	0.292	2.526		2.188	2.846	2.243	2.805	0.486		0.396	0.562	0.408	0.552	0.516		0.411	0.619	0.430	0.600
No assumptions	-3.511	3.393	-3.524	3.407	-3.522	3.405	-33.622	33.689	-33.748	33.814	-33.728	33.795	-8.628	8.631	-8.659	8.663	-8.654	8.658	-8.436	8.823	-8.472	8.850	-8.466	8.846
MTR	0.000	1.865	0.000	1.879	0.000	1.877	0.000	21.574	0.000	21.666	0.000	21.650	0.000	5.381	0.000	5.404	0.000	5.401	0.000	6.641	0.000	6.667	0.000	6.663
MTR + MTS	0.000	0.626	0.000	0.670	0.000	0.662	0.000	5.462	0.000	5.752	0.000	5.696	0.000	1.064	0.000	1.140	0.000	1.126	0.000	1.064	0.000	1.152	0.000	1.137
MTR + MTS + MIV (bias corr.)	0.000	0.605	0.000	0.673	0.000	0.658	0.000	4.950	0.000	5.467	0.000	5.353	0.000	1.033	0.000	1.231	0.000	1.187	0.000	1.011	0.000	1.149	0.000	1.118
MTR + MTS + MIV (no bias corr.)	0.000	0.554	0.000	0.620	0.000	0.606	0.000	4.711	0.000	5.225	0.000	5.112	0.000	0.910	0.000	1.105	0.000	1.062	0.000	0.913	0.000	1.050	0.000	1.020
Number of observations	57,509						57,509						57,509						57,509					

Note: See note to Table A.1.

Table A.6. Average treatment effect of social activities on cognitive test scores, waves 4, 5 and 6

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
Exogenous treatment selection	0.459		0.415	0.500	0.424	0.494	3.765		3.603	3.954	3.636	3.923	0.832		0.790	0.873	0.799	0.867	1.023		0.975	1.077	0.982	1.067
No assumptions	-2.196	2.346	-2.218	2.367	-2.215	2.364	-20.427	24.095	-20.531	24.193	-20.515	24.175	-5.421	5.995	-5.447	6.022	-5.443	6.017	-4.844	6.572	-4.876	6.602	-4.871	6.596
MTR	0.000	2.188	0.000	2.206	0.000	2.204	0.000	21.865	0.000	21.944	0.000	21.931	0.000	5.432	0.000	5.451	0.000	5.448	0.000	5.845	0.000	5.871	0.000	5.866
MTR + MTS	0.000	0.495	0.000	0.533	0.000	0.528	0.000	4.114	0.000	4.294	0.000	4.261	0.000	0.891	0.000	0.928	0.000	0.924	0.000	1.095	0.000	1.146	0.000	1.137
MTR + MTS + MIV (bias corr.)	0.038	0.383	0.000	0.502	0.000	0.476	0.000	3.714	0.000	4.017	0.000	3.951	0.000	0.812	0.000	0.867	0.000	0.855	0.008	0.985	0.000	1.058	0.000	1.042
MTR + MTS + MIV (no bias corr.)	0.067	0.340	0.000	0.458	0.009	0.432	0.000	3.549	0.000	3.846	0.000	3.781	0.007	0.780	0.000	0.835	0.000	0.823	0.029	0.945	0.000	1.018	0.000	1.002
Panel B. ATE of a change from no activities to two or more activities																								
Exogenous treatment selection	0.727		0.681	0.775	0.689	0.767	6.228		6.018	6.424	6.049	6.398	1.249		1.204	1.297	1.210	1.289	1.537		1.474	1.594	1.482	1.585
No assumptions	-2.445	2.543	-2.468	2.566	-2.464	2.562	-22.242	26.340	-22.363	26.433	-22.344	26.418	-5.946	6.511	-5.975	6.539	-5.971	6.534	-5.224	7.233	-5.262	7.260	-5.255	7.256
MTR	0.000	2.543	0.000	2.566	0.000	2.562	0.000	26.340	0.000	26.433	0.000	26.418	0.000	6.511	0.000	6.539	0.000	6.534	0.000	7.233	0.000	7.260	0.000	7.256
MTR + MTS	0.000	0.727	0.000	0.775	0.000	0.767	0.000	6.228	0.000	6.424	0.000	6.398	0.000	1.249	0.000	1.297	0.000	1.289	0.000	1.537	0.000	1.594	0.000	1.585
MTR + MTS + MIV (bias corr.)	0.080	0.663	0.004	0.799	0.021	0.769	0.144	5.837	0.000	6.172	0.016	6.098	0.037	1.200	0.000	1.263	0.006	1.249	0.075	1.428	0.026	1.521	0.037	1.501
MTR + MTS + MIV (no bias corr.)	0.106	0.587	0.032	0.720	0.048	0.691	0.246	5.621	0.081	5.951	0.118	5.878	0.058	1.151	0.019	1.213	0.028	1.199	0.095	1.373	0.047	1.465	0.057	1.445
Panel C. ATE of a change from one activity to two or more activities																								
Exogenous treatment selection	0.268		0.215	0.321	0.225	0.313	2.463		2.247	2.678	2.284	2.645	0.418		0.369	0.469	0.376	0.458	0.514		0.452	0.577	0.463	0.567
No assumptions	-3.262	3.209	-3.280	3.228	-3.277	3.226	-31.233	31.662	-31.349	31.769	-31.332	31.753	-8.067	8.059	-8.096	8.089	-8.092	8.084	-7.923	8.204	-7.954	8.231	-7.950	8.226
MTR	0.000	1.910	0.000	1.929	0.000	1.926	0.000	21.233	0.000	21.313	0.000	21.298	0.000	5.134	0.000	5.153	0.000	5.150	0.000	6.164	0.000	6.189	0.000	6.185
MTR + MTS	0.000	0.551	0.000	0.596	0.000	0.588	0.000	4.770	0.000	4.959	0.000	4.925	0.000	0.927	0.000	0.971	0.000	0.964	0.000	1.141	0.000	1.196	0.000	1.189
MTR + MTS + MIV (bias corr.)	0.000	0.494	0.000	0.624	0.000	0.596	0.000	4.463	0.000	4.755	0.000	4.690	0.000	0.910	0.000	0.965	0.000	0.953	0.000	1.042	0.000	1.128	0.000	1.109
MTR + MTS + MIV (no bias corr.)	0.000	0.424	0.000	0.553	0.000	0.525	0.000	4.282	0.000	4.572	0.000	4.508	0.000	0.863	0.000	0.918	0.000	0.906	0.000	0.996	0.000	1.082	0.000	1.063
Number of observations	66,872						173,898						173,898						173,898					

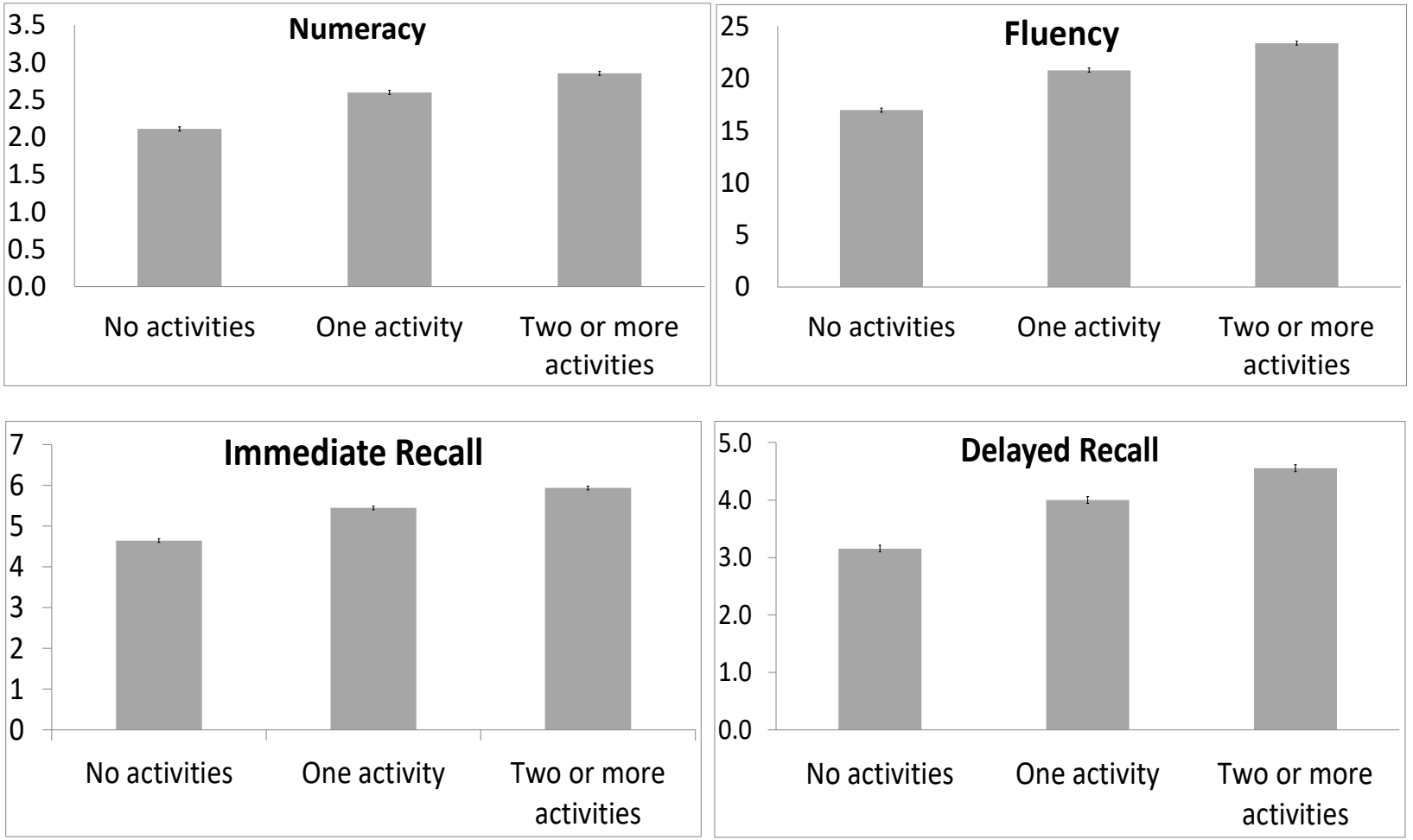
Note: See note to Table A.1.

Table A.7. Average treatment effect of social activities on cognitive test scores, non-imputed data

Assumptions	Numeracy						Fluency						Immediate recall						Delayed recall					
	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI	Lower bound	Upper bound	Low 95% CI	High 95% CI	Low 90% CI	High 90% CI
Panel A. ATE of a change from no activities to one activity																								
Exogenous treatment selection	0.483		0.454	0.506	0.460	0.503	3.861		3.724	4.002	3.744	3.976	0.838		0.801	0.871	0.807	0.865	0.954		0.914	0.994	0.922	0.988
No assumptions	-2.146	2.246	-2.159	2.259	-2.156	2.257	-19.825	23.778	-19.907	23.854	-19.890	23.837	-5.255	5.924	-5.276	5.943	-5.271	5.941	-4.621	6.563	-4.650	6.585	-4.646	6.581
MTR	0.000	2.134	0.000	2.146	0.000	2.143	0.000	21.921	0.000	21.980	0.000	21.974	0.000	5.448	0.000	5.464	0.000	5.461	0.000	5.940	0.000	5.961	0.000	5.957
MTR + MTS	0.000	0.508	0.000	0.529	0.000	0.526	0.000	4.154	0.000	4.291	0.000	4.262	0.000	0.893	0.000	0.925	0.000	0.919	0.000	1.021	0.000	1.058	0.000	1.053
MTR + MTS + MIV (bias corr.)	0.006	0.478	0.000	0.526	0.000	0.515	0.000	3.805	0.000	4.032	0.000	3.982	0.001	0.817	0.000	0.860	0.000	0.851	0.000	0.926	0.000	0.976	0.000	0.965
MTR + MTS + MIV (no bias corr.)	0.017	0.450	0.000	0.498	0.000	0.488	0.000	3.685	0.000	3.913	0.000	3.863	0.016	0.796	0.000	0.839	0.000	0.829	0.011	0.902	0.000	0.951	0.000	0.940
Panel B. ATE of a change from no activities to two or more activities																								
Exogenous treatment selection	0.733		0.701	0.767	0.706	0.761	6.344		6.166	6.528	6.190	6.499	1.300		1.259	1.341	1.264	1.334	1.521		1.471	1.573	1.480	1.567
No assumptions	-2.437	2.440	-2.451	2.453	-2.449	2.451	-21.867	26.118	-21.958	26.206	-21.947	26.191	-5.819	6.484	-5.844	6.507	-5.840	6.503	-5.022	7.287	-5.052	7.310	-5.047	7.305
MTR	0.000	2.440	0.000	2.453	0.000	2.451	0.000	26.118	0.000	26.206	0.000	26.191	0.000	6.484	0.000	6.507	0.000	6.503	0.000	7.287	0.000	7.310	0.000	7.305
MTR + MTS	0.000	0.733	0.000	0.767	0.000	0.761	0.000	6.344	0.000	6.528	0.000	6.499	0.000	1.300	0.000	1.341	0.000	1.334	0.000	1.521	0.000	1.573	0.000	1.567
MTR + MTS + MIV (bias corr.)	0.031	0.715	0.002	0.783	0.009	0.768	0.195	5.984	0.055	6.241	0.086	6.184	0.050	1.254	0.018	1.304	0.025	1.293	0.062	1.426	0.025	1.498	0.033	1.482
MTR + MTS + MIV (no bias corr.)	0.042	0.666	0.013	0.733	0.020	0.719	0.253	5.848	0.113	6.105	0.144	6.049	0.065	1.219	0.033	1.269	0.040	1.258	0.077	1.388	0.040	1.460	0.049	1.444
Panel C. ATE of a change from one activity to two or more activities																								
Exogenous treatment selection	0.250		0.213	0.286	0.219	0.281	2.483		2.293	2.680	2.325	2.651	0.462		0.420	0.507	0.425	0.499	0.566		0.508	0.619	0.519	0.610
No assumptions	-3.414	3.317	-3.425	3.329	-3.423	3.327	-32.057	32.355	-32.160	32.449	-32.144	32.434	-8.260	8.256	-8.284	8.279	-8.279	8.275	-8.093	8.414	-8.120	8.439	-8.116	8.435
MTR	0.000	1.869	0.000	1.882	0.000	1.880	0.000	21.331	0.000	21.403	0.000	21.391	0.000	5.217	0.000	5.233	0.000	5.230	0.000	6.325	0.000	6.345	0.000	6.342
MTR + MTS	0.000	0.580	0.000	0.614	0.000	0.608	0.000	4.999	0.000	5.184	0.000	5.152	0.000	1.008	0.000	1.045	0.000	1.039	0.000	1.187	0.000	1.241	0.000	1.231
MTR + MTS + MIV (bias corr.)	0.000	0.562	0.000	0.629	0.000	0.614	0.000	4.729	0.000	4.964	0.000	4.912	0.000	0.989	0.000	1.036	0.000	1.026	0.000	1.119	0.000	1.188	0.000	1.173
MTR + MTS + MIV (no bias corr.)	0.000	0.517	0.000	0.584	0.000	0.569	0.000	4.597	0.000	4.831	0.000	4.780	0.000	0.952	0.000	1.000	0.000	0.989	0.000	1.080	0.000	1.148	0.000	1.133
Number of observations	120,384						227,898						227,786						226,871					

Note: See note to Table A.1.

Fig. A.1. Means of the four cognitive test scores, by the level of social activity



Notes: The height of the histogram bars corresponds to the weighted average of the cognition score within the group of individuals who exhibit a given level of social activities. The vertical lines in the middle of the bars denote 95 percent confidence intervals.