PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2021 – 25

Happier Elderly Residents
The positive impact of physical activity on objective and subjective health condition of elderly people in nursing homes. Evidence from a multi-site randomized controlled trial

Claudia Senik
Guglielmo Zappalà
Carine Milcent
Chloé Gerves-Pinquié
Patricia Dargent-Molina

JEL Codes: I11, I12
Keywords: Physical activity, Retirement homes, Impact study, falls, Subjective health

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

Happier Elderly Residents

The positive impact of physical activity on objective and subjective health condition of elderly people in nursing homes. Evidence from a multi-site randomized controlled trial.

Claudia Senik[1], Guglielmo Zappalà*[2], Carine Milcent[2], Chloé Gerves-Pinquié[3], and Patricia Dargent-Molina[4]

[1]Sorbonne University and Paris School of Economics, Paris, France

[2]Paris School of Economics, Paris, France

[3]Institut de Recherche en Sante Respiratoire des Pays de la Loire, Beaucouzé, France

[4]INSERM, Paris, France

Abstract

We explore the effects of adapted physical exercise programs in nursing homes, in which some residents suffer from dementia and/or physical limitations and other do not. We use data from 452 participants followed over 12 months in 32 retirement homes in four European countries. Using a difference-in-difference with individual random effects model, we show that the program has exerted a significant impact on the number of falls and the self-declared health and health-related quality of life of residents (EQ-5D). The wide scope of this study, in terms of sites, countries, and measured outcomes, brings generality to previously existing evidence. A simple computation, in the case of France, suggests that such programs are highly cost-efficient.

JEL codes: I11, I12

Keywords: physical activity, retirement homes, impact study, falls, subjective health.

*Corresponding author: Paris School of Economics, 48 bd Jourdan, 75014 Paris, France. E-mail: guglielmo.zappala@psemail.eu

1 Introduction

With the increase in life expectancy and the ensuing aging of the population, the share of elderly people in developed countries has risen considerably (Rechel et al., 2013; Ezeh et al., 2012; Terroso et al., 2014; OECD, 2020), and so has the demand for long term care (Pickard et al,. 2007), including residential care (Cremer et al., 2012). In light of this new phenomenon, the quality of life of older people living in institutions, and the cost of their care, have become urgent questions. This has triggered the development of non-pharmacological interventions meant to improve the health-related quality of life of elderly people. This study investigates the benefits of one of the most popular among these interventions, adapted physical activity, on the general state of mental and physical health of senior people in residency homes.

The quality of life of elderly people has become a major concern for families and society as a whole. In particular, the question of how to take care of the elderly when they loose autonomy has attracted mounting attention. Essentially, the choice is to let the burden of dependent persons being borne by family caregivers - which turn out to be mostly females (Checkovich and Stern, 2002), or to recourse to residency homes. When considering the possibility of a placement in a residency home, the wellbeing of residents is naturally a major element of choice (Hiedemann et al. 2017; Rapp et al, 2018). Accordingly, this study is based on the EuroQol questionnaire, which provides a health-related quality of life measure. It also uses the Geriatric Depression Scale. Our second major outcome is falls, which represent the most frequent, serious, and costly health problem of elderly people (Sattin, 1992; Dieleman et al., 2016; Burns et al., 2016; Burns and Kakara, 2018). Reduced balance increases the risk of falling, and falls and fractures are common among nursing home residents, especially with dementia (Thomas et al., 2002; van Doorn et al., 2003; Tinetti, 1995; Jensen, 2003).

Previous randomized control trials have assessed the efficacy of programs that were either targeted to a specific pathology (Hauer et al., 2012; Deschamps et al., 2010; Williams and Tappen, 2007; Rolland et al., 2007), such as Alzheimer, or run in a single country (Alessi et al., 2005; Kerse et al., 2008), or receiving only one type of physical training (Baum et al., 2003). Studies have demonstrated that exercising can improve the muscle strength, balance and autonomy of cognitively healthy elderly people in nursing home (Grönstedt et al., 2011; DeChamps et al., 2010; Taguchi et al., 2010). For residents with dementia, the literature is scarce (see Pitkälä (2013) or Chan et al. (2015) for a systematic review). Rolland et al. (2007) showed that walking combined with strength and flexibility exercises reduced the decline in ADL-function. Telenius et al. (2015) show that a high intensity functional exercise program improved balance and muscle strength of patients with dementia placed in nursing home patients.

Systematic reviews of the literature dedicated to interventions and randomized controlled trials (RCT) consisting in training activity aimed at improving health-related quality of life, physical fitness, and balance of institutionalized older people (Weening-Dijksterhuis et al., 2011; Van Malderen et al., 2013; Chang et al., 2004; Rimland et al. 2016) show that studies almost never include both objective measures such as falls (Campbell et al. 1997; Lord et al. 2003; Province et al. 1995; Alvarez et al., 2015; Jill,

2018; Cameron et al., 2012; Silva et al., 2013), and self-assessed quality of life measures (Chin et al., 2004; Schoenfelder and Rubenstein, 2004; Grönstedt et al., 2013; Park et al., 2014). Experiments rarely include residents with cognitive impairment, which limits the extrapolation of their results. This study adds generality to these findings by its wide scope: it is run in four countries, on groups of residents suffering both from physical and/or mental limitations; it measures the impact of the program both on objective outcomes, such as the number of falls, and on self-assessed scores of physical and mental health, and depression.

Note that, although a large number of studies have been made on elderly people living in their home (e.g. El-Khoury et al., 2015), their results cannot be readily extrapolated to residents in nursing homes, as the latter may differ due to self-selection, and because their mental and physical health may be impacted by the transition to nursing homes itself, in case of loss of routine mobility, or feelings of estrangement for instance (MacRae et al., 1996; Drageset et al., 2008; Foster et al., 2009; Cress et al., 2011).

We analyze the results of a randomized control trial run by INSERM[1] and a French non-for-profit organization Siel Bleu[2], with support from the European Union[3]. The program, called Healthy Activity and Physical Program Innovations in Elderly Residences (HAPPIER), was run in 32 retirement homes, located in four countries (Belgium, Spain, France and Ireland), on groups of residents suffering from physical and/or mental limitations. It offered four types of physical activity, each corresponding to a certain level of physical and mental autonomy. It took place over 12 months, starting in January 2013.

It turns out that the program improved residents' self-assessed health (EuroQol-5) and reduced the prevalence of the feeling that "life is empty" (Geriatric Depression Scale). It improved residents' balance test scores, performance scores in everyday-life activities, and self-assessed autonomy in daily activities. Moreover, it reduced by 13% the number of persons who refused care (Aggressive Behavior Scale). The most impressive impact of the program is the prevention of falls. Estimates with full controls show that the program halved the number of falls, i.e. led on average to one fall less every year per person. We quantify the efficiency of the program using the number of falls, the only outcome measure that is readily objectively quantifiable. The cost-benefit analysis of the fall avoidance allowed by the program, if it were generalized in case of France, showed a net benefit per year in the range of €421 million to €771 millions.

The remainder of the paper is organized as follows. Sections 2 and 3 describe the Materials and methods and Results. Section 4 concludes.

[1] VIMA-UMR-S U1168 (INSERM/UVSQ)
[2] www.sielbleu.org
[3] Convention VS/2011/0059

2 Materials and Methods

Design

This study is based on a randomized controlled trial, implemented in parallel in 32 retirement homes located in four countries (Belgium, Spain, France and Ireland) [4]. The total sample of residents comprised 452 persons. The experiment was led from January 2013 until January 2014. Measures were taken in January 2013, June 2013 and January 2014. The control groups received the treatment after the end of the trial, in January 2014 (the treatment group did not anymore, after the end of the trial). During the experiment, the control group received care-as-usual.

Participants

The sample is composed of 452 participants (respectively, 118 in Belgium, 111 in Spain, 107 in France and 116 in Ireland, who were recruited in December 2013 (Table 1). Outcome measures were taken before the start of the program, in January 2013, and during the program, in June 2013 and January 2014. This generated a total of 1220 observations. Table 2 provides a description of the regression sample at baseline T_0. The average age was 86 (sd=8.1); men accounted for 22% of the sample; in terms of physical and mental impairment: 29% of the sample are with no pathology (H1), 24.5% with locomotion problems (H2), 26.7% with cognitive problems (H3), and 20% with both locomotion and cognition problems (H4).

Each selected retirement home then constituted a sample comprising of all of the willing and eligible residents over 75 years old. Participants gave their written informed consent. For those suffering from mental dementia, consent was given by their legal representative. Eligibility criteria were assessed by the residency home's doctor: a medical certificate of fitness and no counter-indication for participation, and a life expectancy of more than six months. Within each retirement home, the entire group of eligible participants was then randomly divided into a treatment group and a control group by a data manager located outside the retirement home, at Siel Bleu's head office. The time line of the selection process in presented in Figure A1 in the Appendix. Hence, in each country, each of the four programs was proposed to the treatment group in two residency homes.

The main estimates are run in a balanced panel, i.e. on residents who were included in the sample in all three points in time when the measures were taken (January 2013, June 2013 and January 2014). We also checked that our results are robust to using the larger entire unbalanced panel.

[4] The ethics approval was obtained in CCPg Ile de France center (n° ID RCB: 2012-A00646-37). It is an open-label experiment. The time-line of the trial is presented in the Appendix (Figure A1). The protocol was registered under the acronym HAPPIER (Healthy Activity and Physical Program Innovations in Elderly Residences). There was no data monitoring committee other than the Institute for Public Policy that oversaw the study. The study protocol can be fund online at http://ec.europa.eu/employment/defis/publigrant/public/publications/266/frame?publicationLanguage=fr

Outcome measures

Outcome measures are of two types: objective and subjective, i.e. self-declared. In this article, we report the effects of the intervention on all of the outcomes of the RCT. Our primary outcome concerned self-assessed health-related quality of life measures based on EuroQol questionnaire (EQ-5D, see https://euroqol.org and Table A10 in the Appendix). Therefore the trial was initially set in order to be able to display a relative difference on the EuroQol score between the control group and the intervention group. Our other self-assessed measures are the Geriatric Depression Scale (GDS-4, 4 modalities, see Yesavage et al., 1982 and Table A11 in the Appendix), and the self-assessed Performance in daily activities.

Objective outcome measures consist primarily in the number of falls per resident (recorded by the staff as they happened); they also include a series of balance and locomotion tests, such as Timed Get-Up and Go, Unipodal station, Keep walking when talking, as well as Capacity of performing daily activities.

Concerning the Capacity and Performance measures, these address the ability of residents to perform daily activities without help, such as: eating a meal, washing oneself, getting dressed, putting one's shoes on, going to the toilets, standing up, and moving from one point to another (either by walking or on a wheelchair). For each action, the evaluator first asked the resident whether she was capable of performing the activity (Performance score) then asked the resident to perform an action similar to the one evoked, for instance, washing oneself alone and moving a towel on one's body (Capacity score). GDS measures were collected by nurses; all of the other tests were carried out by an evaluator from Siel Bleu, independent of the person who was teaching the gym class. Finally, nurses filled the Aggressive Behavior Scale survey (Perlman and Hirdes, 2008) that captures their perception of patients' behavior.

All of the tests had been previously validated by the scientific literature, except for the performance tests in daily activities, which were elaborated by Siel Bleu. A detailed presentation of the questionnaires is included in the Appendix (Tables A10 to A12).

Intervention

The program was designed and implemented by Siel Bleu on the basis of their 17-year experience in supplying adapted physical activity for elderly people living in institutions. Sessions were organised on a weekly basis in small groups of fewer than ten people, under the direction of a physical educator from Siel Bleu, and the supervision of a nurse of the centre. Four types of adapted physical activity programs were proposed, depending on the physical and mental conditions of residents. Group H1 was composed of residents with no physical or mental impairment. They were offered a Prevention of falls program. Group H2 comprised residents with walking problems but no cognitive problems. They were offered a Gym on a chair program. Group H3's residents suffered from senility but no walking difficulty, and received Alzheimer's gymnastics. Finally, Group H4's residents suffered from both mental and physical impairment and practiced Gymnastics around the table. A detailed description of the physical activity offered to each group is provided in the Appendix.

The full description of the recruitment and randomization for each group is presented in the Appendix Figure A1.

Procedure

Retirement homes were selected via a call for participation. The conditions for participation were the following: at least two years of existence; no program of physical activity ongoing or programmed; location in France, Belgium, Ireland or Spain; a room for physical exercise of at least 30 square meters, a sufficient number of residents (from eight to 20) with the same degree of mobility and dependence, a staff capable of measuring the outcomes on residents. Each retirement home chose one type of adapted physical activity program among four.

A total of 59 homes were candidates: 15 in Belgium, 22 in Spain, 14 in France and 8 in Ireland. Retirement homes were selected by a jury composed of members of the administration of three French Ministries (Health, Labor and Sports) and of Siel Bleu, a non-for-profit association specialized in the delivery of adapted physical activity for the elderly that conceived and implemented the intervention. When possible, efforts were made in order to constitute a diversified sample in terms of location (rural versus urban), status (private versus public) and size (number of residents). A short list of retirement homes was established on the basis of these criteria; it was then sent to the selection committee who chose the selected homes. Selected homes were offered the program for free.

This makes eight retirement homes per country, for a total of 32 retirement homes. The total sample of residents comprised 452 people. In order to avoid frustration, members of the control group were informed that they would be offered a similar training after the termination of the experiment. There was no masking of the experiment.

The trial was initially set to include at least 448 residents in the tests, and monitor them over one year, in order to be able to display a relative difference of 32.25% on the EuroQol score between the control group (224 people) and the intervention group (224 persons), with a statistical power greater than 80% and a p-value of 5%.

Retirement homes were recruited between February 2012 and March 2012. The experiment was led from January 2013 until January 2014. Measures were taken in January 2013, June 2013 and January 2014. The control groups received the treatment in January 2014 (the treatment group did not anymore, after the end of the trial). Figure A2 in the Appendix illustrates the Trial Profile.

Data analyses

The analysis was conducted on an intention-to-treat basis. The impact of the program is evaluated in a difference-in-difference setting (Angrist and Pischke, 2008), using an OLS model and individual controls. The main specification is run on a balanced sample, i.e. including only participants for which we have observations in all of our three points in time (January 2013, June 2013 and January 2014). Using a

balanced panel allows evaluating the impact of the program on all those, and only those, who received it. The drawback is that this sample may be different from the initial one, hence less representative of the entire population of residents. Hence, for robustness, we also ran all our estimates on the entire unbalanced sample of whoever residents took the tests at each period of time.

Our main specification includes individual random effects. This is because this model allows the inclusion of a treatment dummy (which is time invariant, hence cannot be included in a model with individual fixed-effects). However, for robustness, we ran alternative specifications that include individual fixed-effects (without the main effect of the treatment dummy), as well as estimates in first-differences. All these models were run both on balanced and unbalanced samples. We tested the validity of the random-effects model, with a Hausman test whereby we ran the model sequentially with individual fixed-effects and individual random-effects (Wooldridge, 2016): the coefficients on the independent variables included in both models turned out to be the same, showing that the results could not be attributed to any invariant characteristics of the residents.

Estimates were clustered at the level of retirement homes. These statistical specifications are typically used to analyze the results of such Randomized Control Trials.

Although outcomes were measured three times for each individual, we identify the impact of the program by contrasting January 2013 (T_0) versus June 2013 and January 2014 pooled together, hereafter T_1. We thus compare the intervention period to the pre-intervention period. During the intervention, measures were taken in twice (in June 2013 and January 2014) for the sake of robustness. De facto, as one can see from Table 1, some residents participated in the measurement in June 2013, but not January 2014, and conversely (we have not information about the reason for this, as discussed in the Limitations section of the Discussion). We do not expect a dynamic improvement of the measures during the experiment because, due to the age of the subjects, their health tends to deteriorate over time; hence the program is expected to provide improvement and protection for residents' current health, as opposed to remote future benefits.

The impact of the program is measured by the coefficient on the interaction term between T_1 and a categorical variable coding 1 if the individual belonged in the treatment group and 0 in the control group. Hence, the estimates capture the differential evolution in the treatment group as opposed to the control group (difference in differences).

Other controls include the main effects (time and treatment dummy variables), type of mental or physical limitation (H1: no pathology, H2: locomotion problems, H3: mental senility, H4: both locomotion and cognition problems), gender, age, age squared, and country of residence dummies. These measures control for potential differences between groups or individuals, that may arise by chance in spite of the random allocation method, and may have an impact on the effect of program. This potential unexpected heterogeneity is also the reason for running double-differences estimates rather than simple difference estimates, in spite of the fact that the data is generated by a Randomized Controlled Trial.

We thus estimate equations of the type:

$$Y_{it} = \alpha_1 E_i + \alpha_2 E_i * T_1 + \alpha_3 T_1 + \alpha_4 X_{it} + \alpha_5 C_j + \mu_i + \varepsilon_{it} \quad (1)$$

Where Y_{it} is the outcome of interest, E_i is a dummy coding 1 if individual i is in the treatment group and 0 otherwise, T_1 stands for June 2013 or January 2014, as opposed to January 2013, which is the reference category (T_0). X_{it} is a vector of individual covariates, C_j is a country fixed effect, μ_i is an individual random or fixed effect, and ε_{it} is the error term. Our coefficient of interest is α_2, which will measure the impact of the program as a difference-in-differences, i.e. the impact of belonging in the treatment group, as opposed to the control group, during the treatment period.

Alternatively, for robustness, we measure the impact of the treatment separately in June 2013 (now t_1) and January 2014 (now t_2), as opposed to January 2013 (now t_0). We thus estimate equations of the type:

$$Y_{it} = \beta_1 E_i + \beta_2 t_1 + \beta_3 t_2 + \beta_4 E_i * t_1 + \beta_5 E_i * t_2 + \beta_6 X_{it} + \beta_7 C_j + \mu_i + \varepsilon_{it} \quad (2)$$

where our coefficients of interest are β_4 and β_5, which measure the impact of belonging in the treatment group, as opposed to the control group, during the treatment period.

Building a Summary Index Measure

Due to the fact that the ten outcomes we consider are not independent from one another, we create a summary index measure following Kling, Liebman and Katz (2007). This allows identifying the effects that are consistent across outcomes, when these present idiosyncratic variations. This summary index aggregates information and averages together the scores corresponding to ten different outcomes: Total Falls, EuroQol, Capacity score, Performance score, ABS, Timed Up & Go, GDS, Talking while walking, Standing on the right foot, Standing on the left foot.

Each outcome has been normalized in standardized units, so that it is possible to study mean effect sizes of the indices relative to the standard deviation of the control group. The normalization works as follows: for each outcome Y_j of the ten different scores, we build a normalized $Y_j^* = (Y_j - \mu_j)/\sigma_j$ where μ_j is the control group mean for outcome j and σ_j is the standard deviation of the control group. Therefore, each score has mean 0 and standard deviation 1 for the control group. As in Kling, Liebman and Katz (2007), we change the sign for negative outcomes (i.e. total falls, ABS, GDS and Timed Up) so that, for all normalized scores, a higher value represents a *healthier* outcome.

The summary index is the equally weighted average of ten specific scores, i.e. $Y^* = \sum_j Y_j^*/J$. We also follow the same missing value imputation methodology as in Kling, Liebman and Katz (2007), where if an individual has a score for at least one outcome, then the missing value for the other scores are imputed at the mean of the group she belongs to. Following this procedure, the index can be interpreted

as the average of results for different scores scaled to standard deviation units.

Surveys of the Staff

Two questionnaires were passed among the staff of nursing homes. Firstly, a specific questionnaire (quality of life at work) was passed among supervisors (nurses and staff) about the behavior of residents. Secondly, at the start and at the end of the program, a questionnaire was passed among consulting doctors (PRAPIS). The answers cannot be treated as part of the control trial, as the supervisors were not randomized nor blinded. They took care of the members of both the control and the treatment groups, and were aware of who was in which group. However, it is of interest to learn about their perception of the program.

Preliminary validity checks

Before analyzing the results, we checked that the composition of the control and treatment groups were generally the same in T_0, before the experiment began. In principle, this should be the case since all the groups were chosen randomly. De facto, there were no statistically significant differences across the two groups in T_0 (see Table 2).

We also checked that there was no endogenous attrition. In the particular case of this protocol, it is necessary to ensure that the residents who exited the sample were not discouraged or affected in a specific way by the treatment. Attrition from the sample can be due to mortality. Residents in poorer health at the beginning could also be more likely to drop out. Non-participation on the day of the exercise session could also be due to temporary mental or physical health condition. In such cases, the impact of the treatment would be over-evaluated because of the elimination of the weakest persons from the treatment group. To enquire, we checked that attrition was not correlated with participation in the treatment group. Appendix Table A1 shows, in the case of the EuroQol score, that among people who were present before the trial started (in T_0) belonging in the treatment group did not affect the probability to be absent at the time the measures were taken during the treatment (first column). We also verified that the probability of death was not higher in the treatment group than in the control group (Table A1, second column). We also followed Contoyannis et al. (2004) to assess a potential "survivor bias": we compared the coefficients measuring the impact of the program obtained with the balanced sample (residents who are present during the entire period of study as compared with the total unbalanced sample (allowing for attrition). No statistical difference was found.

3 Results

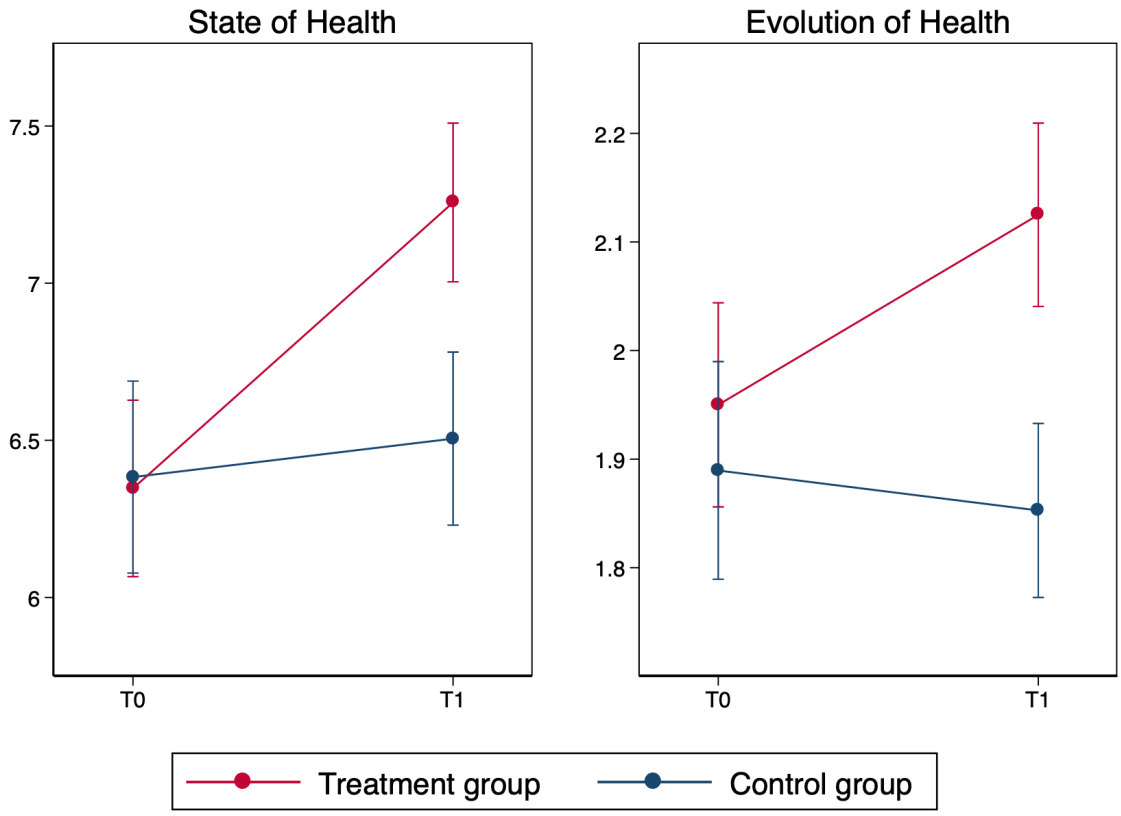
We now turn to the estimates of the program's impact on residents' self-assessed conditions and quality of life, as well as objective outcomes, using the difference-in-difference method.

Table 3 displays the partial correlation coefficients of the multiple linear regression of the scores averaged within each type. Gender, age, and age square are not statistically associated with the outcomes, whereas there are some differences across countries. Unsurprisingly, outcomes are often poorer in groups H2-H4 than in group H1, where people do not suffer from severe problems of locomotion or cognition.

Impact on Residents' Subjective Mental Wellbeing

As previously highlighted, the primary outcome of the study is EuroQol. For this reason, the trial was designed in order to have enough statistical power to detect an impact on the EuroQol score. As reported in Table 3 (column (2)), it is possible to detect the treatment effect on the average score of the EQ-5D: the program improves the average EuroQol score by 0.17 points (sd=0.05) and the effect is statistically significant at the 1% level. The raw score decreases from 2.12 to 2.07 in the control group, and increases from 2.18 to 2.28 in the treatment group (on a 0-3 scale). Table 4 displays the effect of the program on each component of the EuroQol measure. All components are affected in a favorable way by the treatment, however the effect is not always statistically significant. The scores on Mobility, Pain, Health evolution and State of Health are statistically significantly improved (respectively, the former two at the 5% level and the latter two at the 1% level). On the other hand, the effect of the program on Autonomy, Daily activities and Anxiety and Depression is not statistically significant at any conventional level. The measures most affected by the program are Health evolution and State of health, which respectively answer to the following questions: (i) "Compared with my general level of health in the past 12 months, my state of health today is: 1 (worse); 2 (about the same); 3 (better)", and (ii) "Give a score between 0 and 10 for your state of health – with 0 being the worst state of health imaginable and 10 the best". It turns out that the program raises the first score by 0.62 (on a 0-10 scale) and the second score by 0.22 (on a 1-3 scale). This last result shows that the majority of residents who participate in the programme feel that their state of health has improved. Figure 1 illustrates this evolution.

Figure 1: Impact of the program on residents' subjective health.



Note: This graph shows the estimated average change over time in the subjective health of residents based on EuroQol, in the treatment and control groups of the HAPPIER program. T_0 stands for January 2013, before the start of the program. T_1 stands for the average value of the outcome in June 2013 and January 2014. Panel A - State of Health - plots the answers to the question: "Give a score between 0 and 10 for your state of health". Panel B - Evolution of Health - uses the answer to the question: "Compared with my general level of health in the past 12 months, my state of health today is: 1 (worse); 2 (about the same); 3 (better)". Dots are estimates of the mean of reported subjective health for the two groups and bars indicate 95% confidence intervals for population means.

Dropping people who suffer from cognitive limitations (groups H3 and H4) and running the same estimates on the sub-sample of groups H1 and H2, we obtained the same order of magnitude of the impact of the treatment on the general EuroQol score, with the effect being statistically significant at the 5% level (coefficient of 0.112; sd=0.05, see Appendix Table A3, last column).

Finally, concerning the Geriatric Depression Scale (GDS), although the average score was not statistically significantly impacted by the program (Table 3, last column), a detailed analysis shows that a semester of training reduces by around 11 percentage points the prevalence among residents of the "feeling that life is empty" with the coefficient of interest being statistically significant at the 5% level. In the treatment group, the proportion of residents who feel that their life is empty is 29% before treatment

and 18% afterwards, hence a reduction by one third of the prevalence of this sad feeling. By contrast, this proportion remains stable is the control group. (Table A4 in the Appendix, column (2)).

Objective Impact Measures

Among all outcomes, the result with the greatest magnitude is the reduction in the number of falls, a particularly high risk for elderly people. Estimates with full controls show that the program leads on average to 0.38 less fall every semester per person (first column). The raw number of falls (last rows of the table) is up from 0.66 to 0.91 per semester in the control group, but down from 0.82 to 0.65 in the treatment group. Hence, it is safe to consider that the treatment reduces the number of falls by half, on average. One could be surprised that the number of falls is initially higher in the treatment group in spite of the random allocation of the residents into the two groups. However, the difference is not statistically significant.

One could wonder whether the impact of the program on residents depends on their locomotion problems. This does not mean that the program is not useful or tailored for those who do suffer from such limitations: on the contrary some specific exercises are designed for them (see Section 2). De facto, when we ran the estimate of equation (2) on the sub-sample of residents of groups H1 and H3 without locomotion problems, we obtained a higher coefficient of 0.56 (0.25) for the number of falls (Table A2 in the Appendix). The raw number of falls increase from 0.66 to 0.91 in the control group, and decreases from 1.11 to 0.74 in the treatment group. Hence, again, a conservative statement about the program is that it lowers the number of falls by about one half.

As it can be seen in Table 3, all of the other objective physical outcomes (Capacity, Performance tests, Timed Get Up and Go, unipodal station, etc.) are also impacted in the expected direction, however the statistical power of the tests is too low to be able to reject the null hypothesis of no impact, therefore the coefficients of interest are not statistically significant at any conventional level.

Finally, the self-declared performance scores in daily activities were always improved by the program, although not always in a statistically significant way. Out of the 17 types of daily activities, seven were significantly improved by the program (not shown), but the average impact was not statistically significant (Table 3, columns 3 and 4). Similarly, on average, the capacity tests in daily activities run by the nurses were not statistically significantly improved by the program.

Alternative models and samples

For robustness, we replicated all of the statistical estimates using alternative statistical models and samples, always clustered at the level of nursing homes, i.e.: the same individual random effect model on an unbalanced sample (Appendix Table A5); an individual fixed-effects OLS model, ran both on the balanced sample (Appendix Table A6) and unbalanced (Appendix Table A7) sample. The results remain essentially unchanged.

We also estimated the impact of the program on a synthetic index of outcomes, as explained in Section 2. The results are shown in Table 5, which displays all the specifications with individual random effects and fixed-effects on the unbalanced and balanced samples, as well as using a first-difference approach (column (5)). As can be seen, this synthetic measure of the impact of the program turns out to be statistically significant in all of the specifications.

As explained in the Data Analysis Section, for robustness, we measure the impact of the treatment separately in June 2013 (now t_1) and January 2014 (now t_2), as opposed to January 2013 (now t_0), as described by equation (2). Tables A14 and A15 in the Appendix show that considering the two measurement points separately leads to the same results. Table A14 replicates the results of Table 3: the coefficients of interest are negative and statistically significant in the estimates of the total number of falls, and positive and statistically significant in the estimates of EuroQol indices. Table 15 replicates Table 5 and displays the estimates of the synthetic index of outcomes: the results are also robust to this specification with two different points time during the treatment period. We choose to keep, as our main specification, the estimates that pool the two measurement points that were taken during the treatment.

Cost-Efficiency

In order to understand the cost-efficiency of the program, we focus on the most easily quantifiable outcome: the number of falls. We propose a back-of-the-envelope calculation in the case of France.

Falls and their complications represent the most serious and most frequent health problems for the elderly: they cause half of the injuries suffered by people over 65 years, one-third of the reasons for hospitalizations and hospital mortality, and half of the cost of their health care expenditure. Note that in France, people over 75 years represent one-fifth of health care expenditure, although they account for only one-tenth of the population.

A cost-benefit analysis for the sole fall avoidance allowed by the program, in the case of France, assuming the program was extended to all nursing homes of the country, shows a net benefit per year, with an order of magnitude between 421 million and 771 million euros. This is because the program reduces the number of falls per year by about 0.5 per resident per year (adopting a conservative point of view). According to SielBleu, the cost of the program is €158 per resident per year, and the average cost of a fall varies between 2,000 to 3,400 euros according to *la Haute Autorité de la Santé*. Therefore, the savings per resident by year ranges from €842 at the low end of the scale (0.5*2000 - 158 = 842) to €1542 , at the high end (0.5*3400 - 158 = 1542). With half a million people in residency homes, this creates a potential net benefit in the aforementioned range.

How supervisors judged the program

More than 80 per cent of employees in residences reported that their work conditions have been improved by the program, especially daily life within the establishment and the meaning of their work. A majority

of the staff saw an improved good mood and a lesser degree of aggression among the residents (as measured by the Aggressive Behavior Scale described in Appendix Table A13).

Secondly, at the start and at the end of the program, a questionnaire was passed among consulting doctors (PRAPIS). The evolution of their answers is striking. Doctors changed from a highly skeptical view of adapted gym to an almost unanimous praise of the program. As shown by Appendix Table A9, they agree unanimously that the adapted physical activity program is a preventive health strategy. They support its wider implementation for any degree of dementia or physical disorder of the residents. They regret the lack of funding for this kind of non- pharmaceutical care.

4 Discussion

Unlike previous programs, this intervention is a comprehensive program of adapted physical activity for nursing home residents. It is intended for all residents, regardless of their physical and cognitive abilities; and is adapted to each resident's condition, as the latter receive a combination of different exercises according to their physical and cognitive abilities. For this purpose, residents are assessed individually and classified into four subgroups; with each subgroup receiving a specific physical activity program. This intervention differs from previously tested physical activity programs that either consisted of a single exercise program, proposed to all participants; or offered physical exercises combined with other preventive measures in so-called multifactorial programs. The program exerts a sizeable impact on the subjective quality of life and objective outcomes of residents, whether they suffer from cognitive or mobility limitations or not. Run in 32 nursing homes located in four different countries, it thus adds generality to previous empirical evidence in favor of adapted physical activity for the elderly.

Interpretation

The theoretical implications of the findings confirm the Latin motto *mens sana in corpore sano*, "healthy mind in a healthy body", according to which physical exercise is an important part of mental and psychological well-being. There are three possible underlying mechanisms for our findings. The first one is the physical channel from physical activity to physical functioning and balance, and the positive mental impact of feeling more fit and autonomous. The second one is the direct impact of physical activity on the brain (Martinsen 1990; Duman et al. 2008; Greenwood et al. 2003). A third possible channel is the "intervention effect", i.e. the possibility that the perspective of participating in a group activity has, in itself, a positive impact on residents' physical and mental wellbeing. The fact that the intervention significantly reduced the prevalence among residents of the "feeling that life is empty" talks to this interpretation. In the same vein, interacting with a younger and dynamic trainer might have lifted residents' mood, hence their emotional wellbeing. In order to identify with more precision the specific channel through which physical activity exerts its impact, future research direction would consist

in running a program with two treatments, for example physical activity and nutrition, and assessing whether it makes a difference when they are provided in a separate or combined way.

For some of the outcomes, such as the Capacity and Performance tests, Timed Get Up and Go, and unipodal station, the coefficients of interest were not statistically significant, although most have the expected sign. However, the experiment was designed in order to have enough statistical power to detect an impact on the EuroQol index, but not on these secondary measures.

Limitations

The main potential limitations to the results provided in this study concerns attrition. Some patients were unable to answer or take the tests on certain days, but we have no further information about their precise condition at that time. In addition, the fact that our sample is composed of voluntary residents could induce a self-selection bias. We have no way of dealing with this issue, as we have no information about residents who did not volunteer or were not eligible. However, it is unlikely that adapted physical activity would ever be forced on residents who would not be willing to participate.

Future research

It would have been useful to know how much adapted physical activity was used as a substitute for medication, but the data did not allow to infer this. This question is left open for future research.

Conclusions

This study shows that adapted physical activity hinders the otherwise rapid degradation of the physical and mental health of elderly residents in retirement homes. In particular, it reduces balance problems that result in costly falls. We show that the program is very cost-efficient, even considering only the reduction of falls. However, our cost-benefit calculation is limited to falls and does not take into account the positive impact of the program on residents' quality of life, as the latter is difficult to quantify in monetary terms. Nevertheless, we provide empirical evidence of the benefits of the program on residents' subjective evaluation of their quality of life. Our estimates thus constitute a lower bound of the potential net benefit - both economic and intangible - of extending adapted physical activity programs in retirement homes. Overall, our results plead in favor of the diffusion of such programs. This experiment is also an illustration of the usefulness of collecting self-assessed measures of well- being in the evaluation of public policy.

References

Alessi, C. A., Martin, J. L., Webber, A. P., Cynthia Kim, E., Harker, J. O., & Josephson, K. R. (2005). Randomized, controlled trial of a nonpharmacological intervention to improve abnormal sleep/wake patterns in nursing home residents. Journal of the American Geriatrics Society, 53(5), 803-810.

Alvarez, K. J., Kirchner, S., Chu, S., Smith, S., Winnick-Baskin, W., & Mielenz, T. J. (2015). Falls reduction and exercise training in an assisted living population. Journal of aging research, 2015.

Angrist, J. D., & Pischke, J. S. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Baum, E. E., Jarjoura, D., Polen, A. E., Faur, D., & Rutecki, G. (2003). Effectiveness of a group exercise program in a long-term care facility: a randomized pilot trial. Journal of the American Medical Directors Association, 4(2), 74-80.

Burns, E., & Kakara, R. (2018). Deaths from falls among persons aged 65 years—United States, 2007–2016. Morbidity and Mortality Weekly Report, 67(18), 509.

Burns, E. R., Stevens, J. A., & Lee, R. (2016). The direct costs of fatal and non-fatal falls among older adults—United States. Journal of safety research, 58, 99-103.

Cameron, I. D., Dyer, S. M., Panagoda, C. E., Murray, G. R., Hill, K. D., Cumming, R. G., & Kerse, N. (2018). Interventions for preventing falls in older people in care facilities and hospitals. Cochrane database of systematic reviews, (9).

Campbell, A. J., Robertson, M. C., Gardner, M. M., Norton, R. N., Tilyard, M. W., & Buchner, D. M. (1997). Randomised controlled trial of a general practice programme of home based exercise to prevent falls in elderly women. Bmj, 315(7115), 1065-1069.

Chan, W. C., Yeung, J. W. F., Wong, C. S. M., Lam, L. C. W., Chung, K. F., Luk, J. K. H., ... & Law, A. C. K. (2015). Efficacy of physical exercise in preventing falls in older adults with cognitive impairment: a systematic review and meta-analysis. Journal of the American Medical Directors Association, 16(2), 149-154.

Chang, J. T., Morton, S. C., Rubenstein, L. Z., Mojica, W. A., Maglione, M., Suttorp, M. J., ... & Shekelle, P. G. (2004). Interventions for the prevention of falls in older adults: systematic review and

meta-analysis of randomised clinical trials. Bmj, 328(7441), 680.

Checkovich, T. J., & Stern, S. (2002). Shared caregiving responsibilities of adult siblings with elderly parents. Journal of Human Resources, 441-478.

Chin A. ,Paw M.J., van Poppel M.N., Twisk J.W., van Mechelen W. (2004). Effects of resistance and all-round, functional training on quality of life, vitality and depression of older adults living in long-term care facilities: a "randomized" controlled trial [ISRCTN87177281]. BMC Geriatr 4:5. doi:10.1186/1471-2318-4-5.

Choi, M., & Hector, M. (2012). Effectiveness of intervention programs in preventing falls: a systematic review of recent 10 years and meta-analysis. Journal of the American Medical Directors Association, 13(2), 188-e13.

Contoyannis, P., Jones, A. M., & Rice, N. (2004). The dynamics of health in the British Household Panel Survey. Journal of Applied Econometrics, 19(4), 473-503.

Cremer, H., Pestieau, P., & Ponthière, G. (2012). The economics of long-term care: a survey. Nordic economic policy review, (2), 108-148.

Cress, M. E., Orini, S., & Kinsler, L. (2011). Living environment and mobility of older adults. Gerontology, 57(3), 287-294.

Dechamps, A., Diolez, P., Thiaudière, E., Tulon, A., Onifade, C., Vuong, T., ... & Bourdel-Marchasson, I. (2010). Effects of exercise programs to prevent decline in health-related quality of life in highly deconditioned institutionalized elderly persons: a randomized controlled trial. Archives of internal medicine, 170(2), 162-169.

Dieleman, J. L., Baral, R., Birger, M., Bui, A. L., Bulchis, A., Chapin, A., ... & Murray, C. J. (2016). US spending on personal health care and public health, 1996-2013. Jama, 316(24), 2627-2646.

Drageset, J., Natvig, G. K., Eide, G. E., Clipp, E. C., Bondevik, M., Nortvedt, M. W., & Nygaard, H. A. (2008). Differences in health-related quality of life between older nursing home residents without cognitive impairment and the general population of Norway. Journal of clinical nursing, 17(9), 1227-1236.

Duman, C. H., Schlesinger, L., Russell, D. S., & Duman, R. S. (2008). Voluntary exercise produces

antidepressant and anxiolytic behavioral effects in mice. Brain research, 1199, 148-158.

El-Khoury, F., Cassou, B., Latouche, A., Aegerter, P., Charles, M. A., & Dargent-Molina, P. (2015). Effectiveness of two year balance training programme on prevention of fall induced injuries in at risk women aged 75-85 living in community: Ossébo randomised controlled trial. Bmj, 351.

Ezeh, A. C., Bongaarts, J., & Mberu, B. (2012). Global population trends and policy options. The Lancet, 380(9837), 142-148.

Crocker, T., Forster, A., Young, J., Brown, L., Ozer, S., Smith, J., ... & Greenwood, D. C. (2013). Physical rehabilitation for older people in long-term care. Cochrane database of systematic reviews, (2).

Greenwood, B. N., Foley, T. E., Day, H. E., Campisi, J., Hammack, S. H., Campeau, S., ... & Fleshner, M. (2003). Freewheel running prevents learned helplessness/behavioral depression: role of dorsal raphe serotonergic neurons. Journal of Neuroscience, 23(7), 2889-2898.

Grönstedt, H., Frändin, K., Bergland, A., Helbostad, J. L., Granbo, R., Puggaard, L., ... & Hellström, K. (2013). Effects of individually tailored physical and daily activities in nursing home residents on activities of daily living, physical performance and physical activity level: a randomized controlled trial. Gerontology, 59(3), 220-229.

Grönstedt, H., Hellström, K., Bergland, A., Helbostad, J. L., Puggaard, L., Andresen, M., ... & Frändin, K. (2011). Functional level, physical activity and wellbeing in nursing home residents in three Nordic countries. Aging clinical and experimental research, 23(5), 413-420.

Hauer, K., Schwenk, M., Zieschang, T., Essig, M., Becker, C., & Oster, P. (2012). Physical training improves motor performance in people with dementia: a randomized controlled trial. Journal of the American Geriatrics Society, 60(1), 8-15.

Hiedemann, B., Sovinsky, M., & Stern, S. (2018). Will you still want me tomorrow? The dynamics of families' long-term care arrangements. Journal of Human Resources, 53(3), 663-716.

Jin, J. (2018). Prevention of falls in older adults. Jama, 319(16), 1734-1734.

Kerse, N., Peri, K., Robinson, E., Wilkinson, T., von Randow, M., Kiata, L., ... & Arroll, B. (2008). Does a functional activity programme improve function, quality of life, and falls for residents in long term

care? Cluster randomised controlled trial. BMj, 337.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. Econometrica, 75(1), 83-119.

Lord, S. R., Castell, S., Corcoran, J., Dayhew, J., Matters, B., Shan, A., & Williams, P. (2003). The effect of group exercise on physical functioning and falls in frail older people living in retirement villages: a randomized, controlled trial. Journal of the American Geriatrics Society, 51(12), 1685-1692.

MacRae, P. G., Schnelle, J. F., Simmons, S. F., & Ouslander, J. G. (1996). Physical activity levels of ambulatory nursing home residents. Journal of Aging and Physical Activity, 4(3), 264-278.

Martinsen, E. W. (1990). Benefits of exercise for the treatment of depression. Sports Medicine, 9(6), 380-389.

OECD (2020), *OECD Labour Force Statistics 2020*, OECD Publishing, Paris, https://doi.org/10.1787/5842cc7f-en.

Pahor, M. (2019). Falls in older adults: prevention, mortality, and costs. Jama, 321(21), 2080-2081.

Park, S. H., Han, K. S., & Kang, C. B. (2014). Effects of exercise programs on depressive symptoms, quality of life, and self-esteem in older people: a systematic review of randomized controlled trials. Applied Nursing Research, 27(4), 219-226.

Perlman, C. M., & Hirdes, J. P. (2008). The aggressive behavior scale: a new scale to measure aggression based on the minimum data set. Journal of the American Geriatrics Society, 56(12), 2298-2303.

Pickard, L., Comas-Herrera, A., Costa-Font, J., Gori, C., di Maio, A., Patxot, C., ... & Wittenberg, R. (2007). Modelling an entitlement to long-term care services for older people in Europe: projections for long-term care expenditure to 2050. Journal of European social policy, 17(1), 33-48.

Pitkälä, K., Savikko, N., Poysti, M., Strandberg, T., & Laakkonen, M. L. (2013). Efficacy of physical exercise intervention on mobility and physical functioning in older people with dementia: a systematic review. Experimental Gerontology, 48(1), 85-93.

Province, M. A., Hadley, E. C., Hornbrook, M. C., Lipsitz, L. A., Miller, J. P., Mulrow, C. D., ...

& Weiss, S. (1995). The effects of exercise on falls in elderly patients: a preplanned meta-analysis of the FICSIT trials. Jama, 273(17), 1341-1347.

Rechel, B., Grundy, E., Robine, J. M., Cylus, J., Mackenbach, J. P., Knai, C., & McKee, M. (2013). Ageing in the European union. The Lancet, 381(9874), 1312-1322.

Rimland, J. M., Abraha, I., Dell'Aquila, G., Cruz-Jentoft, A., Soiza, R., Gudmusson, A., ... & Cherubini, A. (2016). Effectiveness of non-pharmacological interventions to prevent falls in older people: a systematic overview. The SENATOR Project ONTOP Series. PloS one, 11(8), e0161579.

Rolland, Y., Pillard, F., Klapouszczak, A., Reynish, E., Thomas, D., Andrieu, S., ... & Vellas, B. (2007). Exercise program for nursing home residents with Alzheimer's disease: A 1-year randomized, controlled trial. Journal of the American Geriatrics Society, 55(2), 158-165.

Sattin, R. W. (1992). Falls among older persons: a public health perspective. Annual review of public health, 13, 489-508.

Schoenfelder, D. P., & Rubenstein, L. M. (2004). An exercise program to improve fall-related outcomes in elderly nursing home residents. Applied Nursing Research, 17(1), 21-31.

Silva, R. B., Eslick, G. D., & Duque, G. (2013). Exercise for falls and fracture prevention in long term care facilities: a systematic review and meta-analysis. Journal of the American Medical Directors Association, 14(9), 685-689.

Taguchi, N., Higaki, Y., Inoue, S., Kimura, H., & Tanaka, K. (2010). Effects of a 12-month multicomponent exercise program on physical performance, daily physical activity, and quality of life in very elderly people with minor disabilities: an intervention study. Journal of epidemiology, 0911020115-0911020115.

Telenius, E. W., Engedal, K., & Bergland, A. (2015). Effect of a high-intensity exercise program on physical function and mental health in nursing home residents with dementia: an assessor blinded randomized controlled trial. PloS one, 10(5), e0126102.

Terroso, M., Rosa, N., Marques, A. T., & Simoes, R. (2014). Physical consequences of falls in the elderly: a literature review from 1995 to 2010. European Review of Aging and Physical Activity, 11(1), 51-59.

Van Malderen, L., Mets, T., & Gorus, E. (2013). Interventions to enhance the Quality of Life of older people in residential long-term care: a systematic review. Ageing research reviews, 12(1), 141-150.

Weening-Dijksterhuis, E., de Greef, M. H., Scherder, E. J., Slaets, J. P., & van der Schans, C. P. (2011). Frail institutionalized older persons: A comprehensive review on physical exercise, physical fitness, activities of daily living, and quality-of-life. American journal of physical medicine & rehabilitation, 90(2), 156-168.

Williams, C. L., & Tappen, R. M. (2007). Effect of exercise on mood in nursing home residents with Alzheimer's disease. American Journal of Alzheimer's Disease & Other Dementias, 22(5), 389-397.

Wooldridge, J. M. (2016). *Introductory Econometrics: A modern approach*. Nelson Education.

Declarations

Funding

The experiment was funded by the European Commission, Convention VS/2011/0059. It also benefited from the financial support of the French Agence nationale de la recherche (ANR-16-MYBL-0001-02), and the program Investissements d'Avenir (ANR-10-LABX-93-01).

Conflicts of interest/Competing interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Authors' Contributions

The protocol was designed by Siel Bleu with the contribution of Chloé Gerves, who also collected the data. Patricia Dargent contributed to the design of the protocol, in particular the outcome measures. Carine Milcent was initially the statistician for this study, and was later replaced by Guglielmo Zappalà as statistical analyst. Claudia Senik conducted the analysis of the data and wrote the report, which was reviewed, revised and subsequently approved by all authors.

Acknowledgments

Jean-Daniel Mueller, Guillaume Lefebvre and Jean-Michel Ricard (Siel Bleu) designed and implemented the protocol, with the contribution of Chloé Gerves, who also collected the data. We thank two anonymous referees for their insightful comments.

Abbreviations

ABS: Aggressive Behavior Scale
ADL: Activity of Daily Living
EuroQol: European Quality of Life Survey (https://euroqol.org/eq-5d-instruments/)
GDS: Geriatric Depression Scale
OLS: Ordinary Least Square
RCT: Randomized Controlled Trial

Compliance with Ethical Standards

The protocol was registered under the acronym HAPPIER (Healthy Activity and Physical Program Innovations in Elderly Residences). There was no data monitoring committee other than the Institute

for Public Policy that oversaw the study. The study protocol can be fund online at http://ec.europa.eu/employment/defis/publigrant/public/publications/266/frame?publicationLanguage=fr.

Ethics Approval

The ethics approval was obtained in CCPg Ile de France center (n° ID RCB: 2012-A00646-37).

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Disclosure of potential conflicts of interest

The authors have no relevant financial or non-financial interests to disclose.

Tables

Table 1: Trial Profile

	Belgium	Spain	France	Ireland	TOTAL
January (2013) (T_0)	118	111	107	116	452
June (2013) (T_1)	84	93	98	105	380
January (2014)(T_1)	73	97	94	124	388
Total number of observations (control and treatment)	275	301	299	345	1220

Note: This table indicates the total number of residents in each country and at each wave. Some residents did not participate in the program in all of the waves. For instance, in Belgium, 118 residents were initially selected (in T_0) to participate in the program; in June 2013, 84 of them were interviewed; in January 2014, the number of residents interviewed was 73. NB: Residents who died or whose status was unknown were excluded from the sample.

Table 2: Composition of the Treatment and Control Groups in T_0

$\mathbf{T_0}$	**Control Group** Average value	**Treatment Group** Average value	Testing equality Student/χ^2 p-values)
Socio-Demographic controls			
Age	86.05	86.53	$\Pr(\|T\| > \|t\|) = 0.557$
% Men	22.71	22.01	$\Pr(\|T\| > \|t\|) = 0.865$
% H1	28.44	29.07	$\Pr(\|T\| > \|t\|) = 0.882$
% H2	24.44	24.67	$\Pr(\|T\| > \|t\|) = 0.956$
% H3	26.67	26.87	$\Pr(\|T\| > \|t\|) = 0.961$
% H4	20.44	19.38	$\Pr(\|T\| > \|t\|) = 0.778$
Mean outcome variables			
Falls	0.63	0.77	$\Pr(\|T\| > \|t\|) = 0.318$
EuroQol	2.08	2.12	$\Pr(\|T\| > \|t\|) = 0.473$
Capacity Score	14.78	14.69	$\Pr(\|T\| > \|t\|) = 0.892$
Performance Score	33.42	34.14	$\Pr(\|T\| > \|t\|) = 0.520$
ABS	1.19	1.23	$\Pr(\|T\| > \|t\|) = 0.335$
Timed up & go	2.62	2.72	$\Pr(\|T\| > \|t\|) = 0.432$
GDS	0.28	0.28	$\Pr(\|T\| > \|t\|) = 0.857$

Table 3: Aggregate Impact Measures. Balanced sample. Individual Random-Effects Estimates

	Total Falls	EuroQol	Capacity Score	Performance Score	ABS	Timed Up & Go	Talking while walking	GDS
T_0	-0.279	0.0610	1.470	3.479*	-0.0199	0.0724	0.0259	-0.0481
	(0.176)	(0.0585)	(0.961)	(1.393)	(0.0404)	(0.117)	(0.0766)	(0.0267)
Treatment	0.0590	0.0133	-0.517	1.558	0.119*	-0.0846	-0.0423	0.00941
	(0.151)	(0.0637)	(0.768)	(1.047)	(0.0573)	(0.113)	(0.0756)	(0.0410)
Treatment*T_1	-0.385*	0.169**	0.484	1.292	-0.0847	-0.189	-0.0491	-0.0213
	(0.178)	(0.0534)	(0.994)	(1.465)	(0.0592)	(0.140)	(0.0694)	(0.0407)
H2	-0.223	-0.209**	-1.402	-3.434***	0.0178	0.760***	0.0456	-0.0147
	(0.255)	(0.0661)	(0.902)	(0.980)	(0.0317)	(0.229)	(0.0746)	(0.0250)
H3	-0.0231	-0.00230	0.143	-0.589	0.231***	-0.0714	0.0303	0.0205
	(0.245)	(0.0596)	(0.584)	(0.858)	(0.0499)	(0.152)	(0.0680)	(0.0293)
H4	-0.226	-0.575***	-3.413***	-10.56***	0.396***	0.849***	0.177*	-0.0129
	(0.233)	(0.166)	(0.963)	(1.759)	(0.102)	(0.176)	(0.0756)	(0.0286)
Male	-0.0839	0.0274	-0.909	-0.121	-0.00137	-0.0248	0.0753	-0.0541
	(0.124)	(0.0544)	(0.576)	(0.890)	(0.0310)	(0.112)	(0.0750)	(0.0286)
Age	0.119*	-0.0266	0.00454	-0.530	0.00583	-0.0851	-0.0231	0.0176
	(0.0555)	(0.0241)	(0.277)	(0.365)	(0.0163)	(0.0482)	(0.0177)	(0.0103)
Age^2	-0.0688*	0.0149	-0.0296	0.251	-0.00665	0.0609	0.0127	-0.0116
	(0.0346)	(0.0164)	(0.180)	(0.241)	(0.0105)	(0.0314)	(0.0118)	(0.00655)
Belgium	0.0941	-0.0606	1.567	1.533	0.0709	-0.0406	0.109	0.00424
	(0.158)	(0.105)	(1.163)	(1.355)	(0.0505)	(0.233)	(0.0856)	(0.0366)
France	0.120	0.0345	3.580***	1.868	0.128*	-0.792***	-0.0503	-0.0391
	(0.188)	(0.0861)	(0.926)	(1.225)	(0.0531)	(0.156)	(0.0686)	(0.0245)
Ireland	0.119	-0.0626	2.030	-0.494	-0.0366	0.303	-0.109	-0.164***
	(0.252)	(0.0805)	(1.051)	(1.003)	(0.0401)	(0.232)	(0.0808)	(0.0164)
Constant	-4.060	3.404***	14.57	59.54***	0.998	5.176**	1.611*	-0.253
	(2.152)	(0.884)	(10.60)	(14.22)	(0.624)	(1.852)	(0.673)	(0.403)
N	724	660	580	654	660	542	550	662
Overall R^2	0.0183	0.125	0.0985	0.157	0.155	0.233	0.0541	0.0939
Individuals	362	330	290	327	330	271	275	331
Control Group								
Mean Outcome in T_0	0.66	2.12	15.19	34.22	1.14	2.60	0.62	0.27
Mean Outcome in T_1	0.91	2.07	13.49	30.21	1.17	2.28	0.60	0.33
Treatment Group								
Mean Outcome in T_0	0.82	2.18	14.97	35.70	1.24	2.48	0.58	0.28
Mean Outcome in T_1	0.65	2.28	13.95	33.08	1.20	2.28	0.52	0.32

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References categories: H1, Spain

Table 4: EuroQol module. Balanced sample, individual random-effects estimates

	Mobility	Autonomy	Daily activities	Pain	Anxiety and depression	Health evolution	State of Health	Mean EuroQol Score
T_0	0.0617	-0.00949	-0.0300	-0.142*	-0.0774	0.0364	-0.1372	0.0610
	(0.0505)	(0.0651)	(0.0615)	(0.0675)	(0.0775)	(0.0551)	(0.2036)	(0.0585)
Treatment	0.0279	0.0737	0.0308	-0.000600	-0.0524	0.0378	-0.015	0.0133
	(0.0511)	(0.0756)	(0.0800)	(0.0782)	(0.0888)	(0.0778)	(0.2209)	(0.0637)
Treatment*T_1	0.0919*	0.0733	0.105	0.150*	0.0938	0.224**	0.6202**	0.169**
	(0.0416)	(0.0715)	(0.0595)	(0.0694)	(0.0804)	(0.0811)	(0.2299)	(0.0534)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	606	618	596	614	602	590	598	660
Overall R^2	0.135	0.106	0.121	0.109	0.0862	0.0853	0.0873	0.125
Individuals	303	309	298	307	301	295	268	330
Control Group								
Mean Outcome in T_0	2.34	2.29	2.29	2.24	2.39	1.89	6.38	2.12
Mean Outcome in T_1	2.28	2.30	2.30	2.41	2.48	1.85	6.51	2.07
Treatment Group								
Mean Outcome in T_0	2.40	2.39	2.35	2.29	2.34	1.95	6.35	2.18
Mean Outcome in T_1	2.41	2.46	2.45	2.54	2.53	2.12	7.26	2.28

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Control variables: H2, H3, H4, sex, Age, Age^2, country fixed-effects

Table 5: Synthetic Index à la Kling, Liebman and Katz (2007). Individual random effects estimates: Unbalanced sample in (1) and Balanced Sample in (2). Fixed effects estimates: Unbalanced sample in (3) and Balanced Sample in (4). First difference estimates in (5)

	(1)	(2)	(3)	(4)	(5)
T_0	0.0731*	0.0757*	0.0757*	0.0757*	
	(0.0332)	(0.0336)	(0.0334)	(0.0334)	
Treatment	-0.0396	-0.0369			
	(0.0364)	(0.0365)			
Treatment*T_0	0.106**	0.111**	0.111**	0.111**	0.115**
	(0.0372)	(0.0377)	(0.0374)	(0.0374)	(0.0363)
Controls	Yes	Yes	Yes	Yes	Yes
Individuals	389	380	389	380	380
N	769	760	769	760	380

The Index is built from the scores from Total Falls, EuroQol, Capacity Score, Performance Score, ABS, Timed Up Go, GDS, Talking while walking, Standing on the right foot, Standing on the left foot.

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: Sex, Age, Age^2, Country fixed-effects, H2, H3, H4

Online Appendix

Supplement to Intervention Section

Four types of adapted physical activity programs were proposed, depending on the physical and mental conditions of residents.

- Group H1. The Prevention of falls program was proposed for individuals still in possession of their mental faculties and capable of moving by themselves and without help, at least within their residence. These people nevertheless occasionally need assistance to go to the toilet or to get dressed. 25% of the program was dedicated to flexibility (upper limb, lower limb, stretching), 25% to strength, 45% to balance (unipodal station, tracking gait, clearing obstacles, double task performance), and 5% to relaxation (breathing, self-massage).

- Group H2. The Gym on a chair program was intended for individuals not suffering from cognitive trouble, but who have serious difficulties walking. These people have intact mental faculties; they maintain partial autonomy of movement, but may need a walking stick or frame to get around. Their risk of falling is very high (a priori, more than once a month). They need help every day, several times a day, in activities linked to their physical care. 40% of the program was dedicated to flexibility (upper limb, lower limb), 45% to strength (upper limb), 10% to playing game, 5% to relaxation.

- Group H3. Alzheimer's gymnastics: This activity is designed for residents suffering from senility but who have no difficulty walking. This group includes residents suffering from Alzheimer's disease, or another form of dementia; they still have their motor skills and can move themselves around and fulfil basic daily physical needs. 25% of the program was dedicated to socialization (communication, verbalization), 20% to strength and flexibility; 30% to balance (orientation in time and in space, obstacle clearing), 20% to memory (reflex games), and 5% to relaxation.

- Group H4. Gymnastics around the table: This programme is for those individuals suffering from both Alzheimer's disease or another form of dementia and locomotive difficulties. These people need help with most daily physical actions. 20% of the program was dedicated to memory (fine motricity tasks, spatial representation skills), 45% to upper limb flexibility; 30% to playing games (dexterity, skill, posture), and 5% to relaxation.

Figure A1: Time Line of the Selection of Residency Homes and Groups

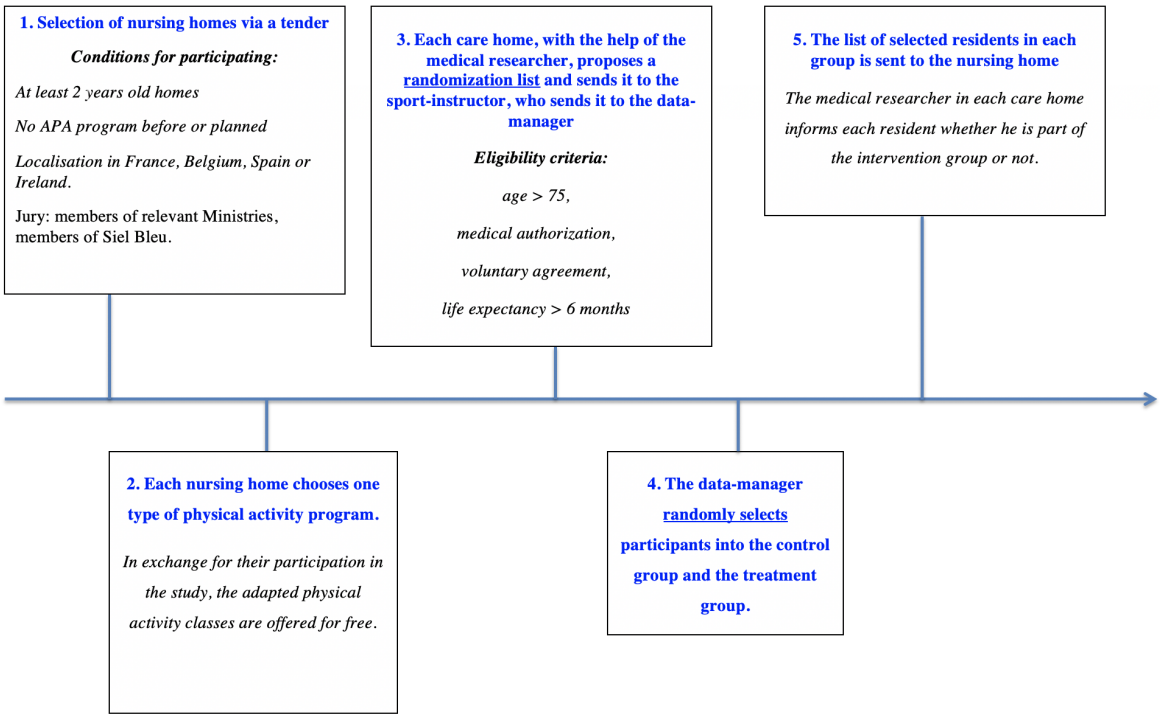


Figure A2: Trial Profile

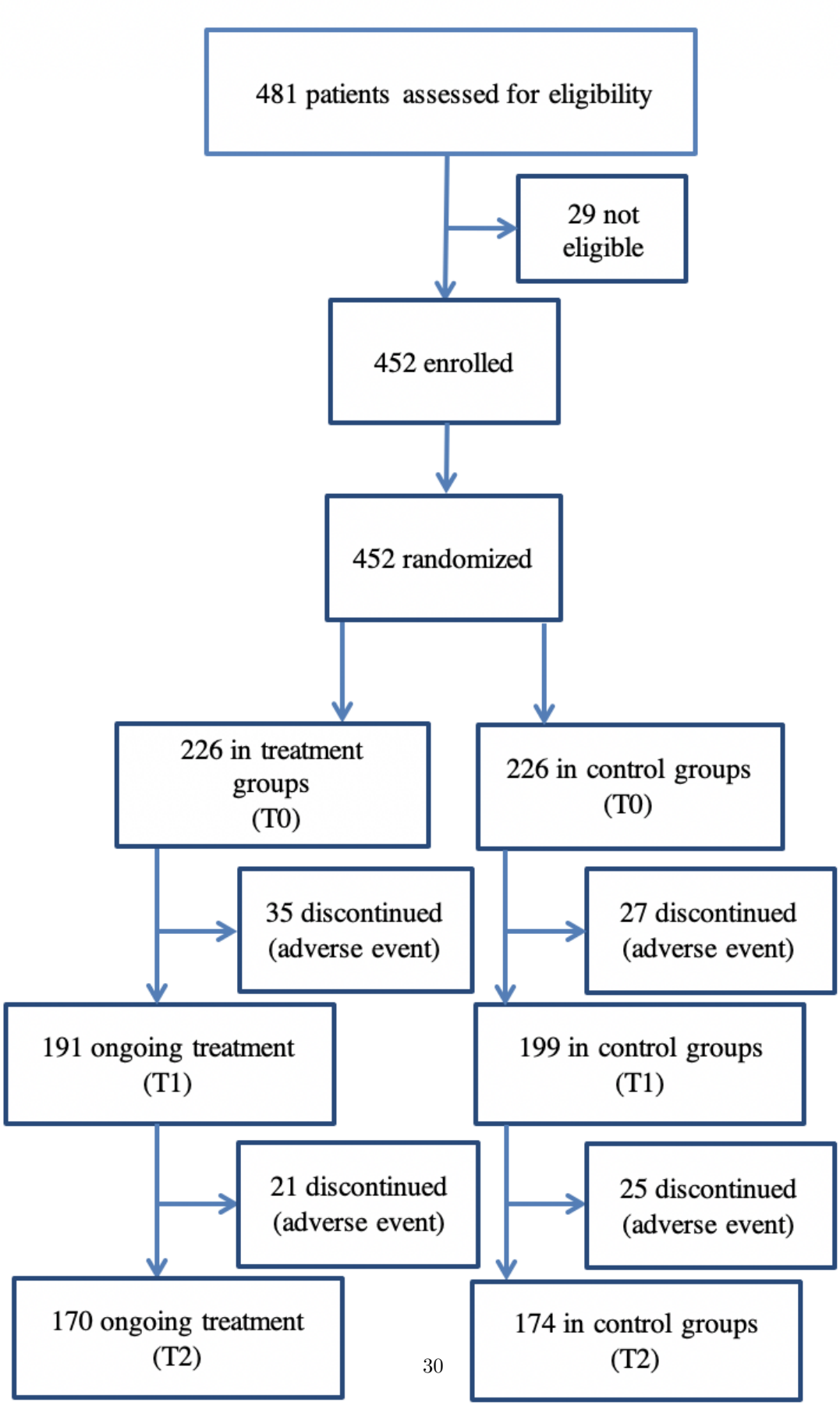


Table A1: Independence between Attrition and Treatment

	Present at the time of EuroQol module measures	Death during the trial
	OLS	OLS
Treatment	**0.0334**	**0.0510**
	(0.1164)	(0.1733)
Reference group: H1		
H2 (locomotion problems)	-0.1301	0.3130
	(0.2603)	(0.2992)
H3 (mental impairment)	0.1204	0.0937
	(0.2477)	(0.2169)
H4 (locomotion and mental impairment)	0.4693*	0.4898*
	(0.2284)	(0.2016)
Male	-0.1816	-0.0327
	(0.1460)	(0.1731)
Age	-0.0012	0.3443
	(0.0664)	(0.3049)
Age^2	0.0075	-0.1869
	(0.0410)	(0.1759)
Reference country: Spain		
Belgium	-1.9011***	-0.8744**
	(0.3729)	(0.3327)
France	-0.7012**	-0.1264
	(0.2295)	(0.2831)
Ireland	-0.1353	-0.1055
	(0.2370)	(0.2633)
Constant	-1.0945	-17.5087
	(2.7223)	(13.2236)
Observations	1167	1167
R^2	0.1385	0.0738
Clusters (retirement homes)	32	32

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The first column shows, in the case of the EuroQol score, that among people who were present before the trial started (in T_0) belonging in the treatment group does not affect the probability to be absent at the time the measures were taken during the treatment (T_1). The other columns show that the probability of death was not higher in the treatment group than in the control group.

Table A2: Total number of falls without people suffering from locomotion problems (H2 and H4). Balanced sample, individual random effects estimates.

	Total Falls
T_0	-0.3010
	(0.2795)
Treatment	0.3408
	(0.2224)
Treatment*T_1	-0.5558*
	(0.2544)
Controls	Yes
Constant	-1.9742
	(2.4270)
Observations	414
Overall R^2	0.0323
Individuals	207
Control Group	
Mean Outcome in T_0	0.66
Mean Outcome in T_1	0.91
Treatment Group	
Mean Outcome in T_0	1.11
Mean Outcome in T_1	0.74

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: Same as in Table A3

Table A3: EuroQol module without people suffering from cognitive problems (groups H3 and H4). Balanced sample, individual random-effects estimates.

	Mobility	Autonomy	Daily activities	Pain	Anxiety and depression	Health evolution	State of Health	Mean EuroQol Score
T_0	0.0761	-0.0591	-0.0111	-0.0824	-0.100	0.0714	-0.431	-0.0300
	(0.0705)	(0.0942)	(0.0862)	(0.0888)	(0.0763)	(0.0755)	(0.264)	(0.0647)
Treatment	-0.0457	0.0720	-0.0366	-0.0918	-0.0434	0.0731	-0.136	0.0114
	(0.0668)	(0.106)	(0.107)	(0.118)	(0.112)	(0.104)	(0.0802)	(0.0802)
Treatment*T_1	0.161**	0.0313	0.107	0.287***	0.0559	0.227*	0.395	0.112*
	(0.0518)	(0.0756)	(0.0799)	(0.0724)	(0.0917)	(0.0981)	(0.290)	(0.0538)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	372	374	366	366	366	368	330	380
Overall R^2	0.0803	0.0545	0.104	0.0817	0.0676	0.122	0.0941	0.0938
Individuals	186	187	183	183	183	184	165	190
Control Group								
Mean Outcome in T_0	2.27	2.22	2.30	2.14	2.31	1.87	6.26	2.14
Mean Outcome in T_1	2.22	2.28	2.30	2.26	2.41	1.80	6.65	2.17
Treatment Group								
Mean Outcome in T_0	2.28	2.31	2.29	2.13	2.27	1.96	6.05	2.18
Mean Outcome in T_1	2.34	2.43	2.41	2.47	2.45	2.10	7.07	2.33

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Control variables: H2, H3, H4, Male, Age, Age^2, country fixed-effects

H1, Spain: Reference category

Table A4: Geriatric Depression Scale module. Balanced sample, individual random-effects estimates

	(1) Discouraged or Sad	(2) Your life is empty	(3) Happy	(4) Your situation is desperate
T_0	-0.0414	0.00610	-0.101	-0.258***
	(0.0718)	(0.0340)	(0.0584)	(0.0490)
Treatment	0.0133	0.102	0.0730	0.00472
	(0.0617)	(0.0540)	(0.0706)	(0.0457)
Treatment*T_1	-0.0228	-0.113*	-0.0579	-0.0345
	(0.0705)	(0.0493)	(0.0752)	(0.0498)
Controls	Yes	Yes	Yes	Yes
Constant	-0.185	-0.0663	2.097***	0.427
	(0.747)	(0.461)	(0.610)	(0.439)
N	660	638	622	618
Overall R^2	0.0579	0.0527	0.0682	0.186
Individuals	330	319	311	309
Control Group				
Mean Outcome in T_0	0.41	0.20	0.66	0.14
Mean Outcome in T_1	0.48	0.19	0.78	0.42
Treatment Group				
Mean Outcome in T_0	0.43	0.29	0.74	0.15
Mean Outcome in T_1	0.47	0.18	0.80	0.40

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: Sex, Age, Age^2, Country fixed-effects, Physical and mental conditions of residents

H1, Spain: Reference category

Table A5: Aggregate Impact Measures. Unbalanced sample. Individual Random-Effects Estimates

	Total Falls	EuroQol	Capacity Score	Performance Score	ABS	Timed Up & Go	Talking while walking	GDS
T_0	-0.2225	0.0418	1.4243	3.1099*	-0.0047	0.0793	0.0307	-0.0433
	(0.1615)	(0.0583)	(0.9017)	(1.3594)	(0.0373)	(0.1100)	(0.0808)	(0.0260)
Treatment	0.0465	0.0400	-0.4800	1.4414	0.0898	0.0496	-0.0197	0.0065
	(0.1357)	(0.0536)	(0.6748)	(1.0497)	(0.0484)	(0.0869)	(0.0711)	(0.0392)
Treatment*T_1	-0.3403*	0.1430**	0.5799	1.3686	-0.0639	-0.2028	-0.0494	-0.0184
	(0.1610)	(0.0517)	(0.8931)	(1.5083)	(0.0524)	(0.1334)	(0.0697)	(0.0375)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-3.1812	3.3433***	19.0139	58.8221***	1.2620*	5.2403**	1.8391**	-0.2859
	(2.0501)	(0.8467)	(9.9767)	(15.0030)	(0.5859)	(1.7253)	(0.6803)	(0.3540)
Observations	769	706	666	703	711	616	614	709
Overall R^2	0.0149	0.1347	0.1044	0.1600	0.1422	0.2231	0.0416	0.0753
Individuals	389	376	376	376	381	345	339	378
Control Group								
Mean Outcome in T_0	0.63	2.08	14.78	33.42	1.19	2.62	0.60	0.28
Mean Outcome in T_1	0.84	2.08	13.16	29.98	1.18	2.56	0.58	0.33
Treatment Group								
Mean Outcome in T_0	0.77	2.12	14.69	34.14	1.23	2.72	0.58	0.28
Mean Outcome in T_1	0.60	2.27	13.55	33.01	1.20	2.36	0.51	0.32

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Control variables: H2, H3, H4, Male, Age, Age^2, Country fixed-effects

H1, Spain: Reference category

Table A6: Aggregate Impact Measures. Balanced Sample, Individual Fixed-Effects Estimates

	Total Falls	EuroQol	Capacity Score	Performance Score	ABS	Timed Up & Go
T_0	-0.2787	0.0610	1.4696	3.4790*	-0.0199	0.0724
	(0.1743)	(0.0581)	(0.9529)	(1.3819)	(0.0401)	(0.1154)
Treatment*T_1	-0.3848*	0.1694**	0.4837	1.2915	-0.0847	-0.1893
	(0.1771)	(0.0530)	(0.9857)	(1.4541)	(0.0588)	(0.1383)
Constant	1.0190***	2.0844***	13.5649***	31.5225***	1.2234*	2.4807**
	(0.1137)	(0.0358)	(0.6718)	(0.9959)	(0.0283)	(0.0738)
Observations	724	660	580	654	660	542
Overall R^2	0.0057	0.0124	0.0104	0.0244	0.0003	0.0116
Individuals	362	330	290	327	330	271
Control Group						
Mean Outcome in T_0	0.66	2.12	15.19	34.22	1.14	2.60
Mean Outcome in T_1	0.91	2.07	13.49	30.21	1.17	2.56
Treatment Group						
Mean Outcome in T_0	0.82	2.18	14.97	35.70	1.24	2.48
Mean Outcome in T_1	0.65	2.28	13.95	33.08	1.20	2.28

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Control variables: H2, H3, H4, Male, Age, Age^2, country fixed-effects

H1, Spain: Reference category

Table A7: Aggregate Impact Measures. Unbalanced Sample, Individual Fixed-Effects Estimates.

	Total Falls	EuroQol	Capacity Score	Performance Score	ABS	Timed Up & Go	Talking while walking	GDS
T_0	-0.2448	0.0610	1.4696	3.4790*	-0.0199	0.0724	0.0259	-0.0481
	(0.1613)	(0.0581)	(0.9505)	(1.3811)	(0.0401)	(0.1153)	(0.0759)	(0.0265)
Treatment*T_1	-0.3725*	0.1694**	0.4837	1.2915	-0.0847	-0.1893	-0.0491	-0.0213
	(0.1583)	(0.0530)	(0.9832)	(1.4532)	(0.0588)	(0.1382)	(0.0687)	(0.0404)
Constant	0.9792	2.0697***	13.3891***	30.9526***	1.2222***	2.5615**	0.5682**	0.3278***
	(0.1020)	(0.0367)	(0.6642)	(1.0121)	(0.0285)	(0.0762)	(0.0528)	(0.0199)
Observations	769	706	666	703	711	616	614	709
Overall R^2	0.0043	0.0110	0.0104	0.0136	0.0001	0.0120	0.0083	0.0041
Individuals	389	376	376	376	381	345	339	378
Control Group								
Mean Outcome in T_0	0.63	2.08	14.78	33.42	1.19	2.62	0.60	0.28
Mean Outcome in T_1	0.84	2.08	13.16	29.98	1.18	2.56	0.58	0.33
Treatment Group								
Mean Outcome in T_0	0.77	2.12	14.69	34.14	1.23	2.72	0.58	0.28
Mean Outcome in T_1	0.60	2.27	13.55	33.01	1.20	2.36	0.51	0.32

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Control variables: H2, H3, H4, Male, Age, Age^2, country fixed-effects

H1, Spain: Reference category

Table A8: First Difference Regressions

	(1) Total Falls	(2) EuroQol	(3) Capacity Score	(4) Performance Score	(5) ABS	(6) Timed Up & Go
Treatment	-0.387*	0.152**	0.642	1.346	0.0992	0.206
	(0.176)	(0.0487)	(0.978)	(1.427)	(0.0594)	(0.133)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	362	330	290	327	330	271

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: Sex, Age, Age^2, Country, Physical and mental conditions of residents

H1, Spain: Reference category

Table A9: PRAPIS Questionnaire

How do you perceive the physical activities offered to the patients in your facility ? Do you agree with the following statements :	T0 (% agreeing)	T1 (% agreeing)
It is a recreational activity	19%	73%
It is an aid for the personnel	54%	62%
It is a preventive care act	1%	96%
It is a sports activity	50%	58%
It is a socializing activity	4%	92%

Balanced Sample (26 doctors)

Table A10: Distribution of answers to the EuroQol questionnaire in T_0

	T_0(%)
Mobility	
A – I have no problem getting around on foot	3
B – I have some problems getting around on foot	52
C – I am confined to my bed	33
D – Cannot answer	12
Personal autonomy	
A – I have no problem taking care of myself	15
B – I have difficulties washing or dressing on my own	32
C – I cannot wash or dress on my own	42
D - Cannot answer	11
Daily activities (e.g. work, studies, housework family activities or leisure)	
A – I have no problem carrying out daily activities	12
B – I have some problem carrying out daily activities	37
C - I am incapable of carrying out daily activities	38
D - Cannot answer	14
Pain/discomfort	
A – I have neither pain nor discomfort	14
B – I have a little pain or moderate discomfort	36
C – I am in pain or in extreme discomfort	38
D – Cannot answer	12
Anxiety/Depression	
A – I am neither anxious nor depressed	9
B - I am moderately anxious or depressed	37
C - I am extremely anxious or depressed	40
D – Cannot answer	14
Compared to my general state of health over the last 12 months, my current state of health is:	
A - Better	20
B – About the same	53
C - Worse	12
D – Cannot answer	15

Table A11: GDS. "Circle the proposition that corresponds best to your situation (taking into account the last few weeks)":

Geriatric Depression Scale	Number of respondents	% Yes
Do you often feel sad or discouraged? (yes/no)	439	41
Do you feel that your life is empty? (yes/no)	427	25
Are you happy most of the time? (yes/no)	428	69
Do you think that your situation is hopeless? (yes/no)	431	14

Table A12: Balance Test: Unipodal station. Answers in T_0

	Number of residents	% Success
On right foot	281	26
On left foot	277	25

Table A13: Aggressive Behavior Scale. Statistics in T_0.

	N of residents	Average score	S.D.	Min	Max
Verbal aggression (threats, insults, cursing others)	441	1.3	0.7	1	4
Physical aggression (punching, scratching, sexual aggression towards others)	439	1.1	0.4	1	4
Socially inept/disturbing behavior (disturbing sounds, cries, noise, spitting, projecting food or stools, hoarding objects, violating the property of others)	441	1.2	0.7	1	4
Aggressively refusing care (refusing to take medicine, to receive injections, to accept ADI, or eating assistance)	436	1.2	0.6	1	4

1=not present over the last 7 days; 2=1 to 3 times over the last 7 days; 3=4 to 6 times over the last 7 days; 4=everyday over the last 7 days.

Table A14: Aggregate Impact Measures. Individual Random-Effects Estimates using three time periods

	Total Falls	EuroQol	Capacity Score	Performance Score	ABS	Timed Up & Go	Talking while walking	GDS
T1	0.225	-0.00395	-2.703*	-4.631**	0.0330	-0.0763	0.0305	0.0739*
	(0.157)	(0.0532)	(1.097)	(1.525)	(0.0527)	(0.128)	(0.0863)	(0.0366)
T2	0.282	-0.0963	0.322	-2.314	0.00649	-0.106	-0.104	0.0254
	(0.206)	(0.0732)	(0.951)	(1.499)	(0.0431)	(0.117)	(0.0842)	(0.0243)
Treatment	0.0445	0.0120	-0.482	1.561	0.117*	-0.0859	-0.0403	0.00921
	(0.137)	(0.0635)	(0.776)	(1.049)	(0.0576)	(0.113)	(0.0753)	(0.0406)
Treatment*T$_1$	-0.308	0.143**	0.484	1.918	-0.0964	-0.164	-0.0483	-0.0134
	(0.185)	(0.0441)	(1.070)	(1.374)	(0.0684)	(0.160)	(0.0951)	(0.0471)
Treatment*T$_2$	-0.442*	0.175*	0.602	1.256	-0.0754	-0.219	-0.0274	-0.0368
	(0.181)	(0.0859)	(1.125)	(2.091)	(0.0670)	(0.149)	(0.0813)	(0.0452)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.928	3.763***	22.24*	63.96***	0.528	4.934**	1.490*	-0.0940
	(2.188)	(0.966)	(9.843)	(15.26)	(0.605)	(1.830)	(0.667)	(0.343)
N	1126	941	804	929	942	754	762	945
Overall R^2	0.0138	0.123	0.124	0.137	0.124	0.235	0.0670	0.0962
Individuals	380	330	290	327	330	271	275	331

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: Sex, Age, Age^2, Country, Physical and mental conditions of residents

H1, Spain: Reference category

Table A15: Index from Kling, Liebman and Katz (2007) using three time periods.

	(1)	(2)	(3)	(4)
T1	-0.0776*	-0.0769	-0.0792	-0.0769
	(0.0394)	(0.0418)	(0.0399)	(0.0416)
T2	-0.0373	-0.0388	-0.0401	-0.0388
	(0.0269)	(0.0285)	(0.0271)	(0.0283)
Treatment	-0.0374	-0.0352		
	(0.0339)	(0.0351)		
Treatment*T$_1$	0.0927*	0.0932*	0.0946*	0.0932*
	(0.0407)	(0.0433)	(0.0409)	(0.0431)
Treatment*T$_2$	0.115**	0.123**	0.120**	0.123**
	(0.0407)	(0.0431)	(0.0403)	(0.0429)
Individuals	389	366	389	366
N	1144	1098	1144	1098

Robust standard errors adjusted for retirement homes clustering in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Individual random effects estimates: Unbalanced sample in (1) and Balanced Sample in (2).

Fixed effects estimates: Unbalanced sample in (3) and Balanced Sample in (4)

Controls in columns (1) and (2): Sex, Age, Age^2, Country fixed-effects, H2, H3, H4