Erasmus University Rotterdam

MSc Thesis Economics

# The Effect of Advance Care Planning on Dutch Healthcare Expenditures.

Prithviraj Basu Mallik

Student Number: 492948

Supervised by Dr. Bram Wouterse

26-07-2022

# Statement of Originality

This document is written by Prithviraj Basu Mallik who declares to take full responsibility for the contents of this document.

I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it.

The Erasmus School of Economics is responsible solely for the supervision of completion of the work, not for the contents.

**Abstract**

This research investigates the cost effects of an advance care planning intervention. It also inspects whether frail elderly are getting into advance care planning programs. Advance Care Planning is important because policy makers want to organise resource efficiently and society aims to help the frail elderly by minimising emotional stress. This research is relevant to economists and stakeholders in the health sector who are interested in payment systems, and the frail elderly. Using panel data from a large Dutch health insurer, Menzis, from 2015 to 2019 , I apply a difference-in differences with propensity score matching to identify the causal effect on costs of the frail elderly as a consequence of being in the program. Combining these approaches leads to accounting unobserved fixed effects and observable such that effects are minimised from bias in a doubly robust manner. I apply a logit model to asses the probability of being in the program conditional on their characteristics. I find evidence of frail elderly with multiple chronic illnesses being with GPs who are able to manage elderly. Further, I find that Advance Care Planning has limited cost decreases in terms of total cost and some indication uptake in costs related to advance care planning.

# Contents

# 1 Introduction

Families and friends of frail elderly aim to help them age at home rather than in an intensive care facility to prevent unnecessary costs and emotional stress. A possible way to achieve their goal is by Advance Care Planning (ACP). ACP refers to the planning between General Practitioners (GPs) and patients over 65 about end-of-life care. Planning often leads to higher-quality and goal-oriented patient care (Bischoff et al., 2013). ACP enables GPs to make prudent choices about old-age care, for example, by providing a trained nurse for lower-cost care at home in the comfort of their family rather than going to the hospital (Klingler et al., 2016 and Dixon et al., 2015).

The current issue is that frail older individuals end up in hospitals with expensive treatment towards the end of their life even if morbidity and mortality remain unaffected while more cost-efficient care could suffice. Since frail elderly are occupying healthcare resources unnecessarily, an opportunity cost arises because the resources could have been used by other patients who could have morbidity or mortality gains. The effects of unnecessary expensive treatment and its opportunity costs are of increasing concern because the health care needs of the frail elderly are expected to become more complex due to longevity and chronic illnesses, as seen in Lubitz et al. (2003) and de Meijer et al. (2013).

The higher costs incurred by patients and their opportunity costs can be linked to the fact that GPs lack the resources to implement ACP (Bekker et al., 2021, and de Vleminck et al., 2013). Hence, GPs require additional financing and resources to do ACP. Policymakers in the health sector need to establish how effective ACP and its reimbursement scheme can be. However, the existing literature does not answer to what extent ACP can be cost-saving using a methodologically sound quasi-experimental approach nor how payment systems interact with the resulting effects (Marckmann et al., 2013).

A candidate for an optimal payment system for ACP is a prospective reimbursement. In the context of ACP, there is an expectation that prospective reimbursement would help realise the value of ACP. Firstly, it would finance GPs to obtain resources to implement and realise the cost savings of ACP. Secondly, by design, prospective reimbursement provides an incentive to be prudent and cost-conscious (R. Ellis and McGuire, 1986). Hence, this paper aims to contribute to the literature by answering if ACP can be cost-saving using a quasi-experimental approach whilst interpreting effects in conjunction with the incentive structure of prospective reimbursement.

This research will investigate a prospective scheme in the form of a capitated fee that a Dutch health insurer introduced to encourage ACP for the frail elderly above the age of 65. The research will use a feature of the allocation mechanism in the compensation scheme for ACP where one group of GPs received the staggered annual roll-out of the reimbursement fee till 2019 to identify the causal impact of ACP on healthcare expenditures of the elderly.

The research aims to see if the introduction of a capitated fee for GPs to provide ACP leads to lower health care costs among frail elderly persons involved. A preliminary step of the research comes from the concern of families and friends of the frail elderly. I analyse if the frail elderly are actually in programs with GPs having the additional incentives for ACP. Then, I analyse if ACP affects costs because the additional financing and incentives will give GPs sufficient resources to organise lower-cost home care. This leads to the following research questions:

Are frail elderly enrolled in programs where the GP has additional finances for Advanced Care Planning? And is there an effect of a capitated fee for Advance Care Planning on the healthcare expenditures of the elderly aged above 65 from 2015 to 2019?

A logit model is used to estimate the probability of entering the treatment group conditional

on the characteristics of a frail elderly. For the causal effect on costs, I exploit the timing of the implementation where a certain group of GPs were selected to receive the treatment based on eligibility criteria in a specific year. Hence, I implement a difference in differences design along with propensity score matching, which enables closer comparison between treatment and control in a quasi-experimental manner.

The rest of the text is structured as follows. First, I give some background on the ACP implementation and then I contextualise it in the academic literature. Then, I explain the data and the empirical strategy used in the research. This is followed by the results that include summary statistics, time trend data, regression estimates, and sensitivity tests. Then, I discuss the assumptions' validity and the research's limitations. Finally, I discuss the implications of the estimates and answer the research question.

# 2 Background

This section will establish the background for the forthcoming empirical analysis. I start with a brief exposition of the Dutch healthcare system, followed by a deeper explanation of the elderly care module by the Dutch health insurer. Then, the empirical evidence on prospective reimbursement and advance care planning follows. The section ends by connecting the elements of the background, the hypotheses, and their implications for interpreting the results.

## 2.1 Elderly Care Module

First, it is good to give some background on the health care system in the Netherlands and then elaborate on the ACP implementation by the health insurer. In the Dutch healthcare system, GPs function as gatekeepers for specialist care. Hence, when frail elderly consult a GP, the GP can decide what care is necessary. Furthermore, GPs tend to work in groups with integration from primary care such as physiotherapists, mental health professionals, and nurses (Schäfer et al., 2010).

In 2015, the Dutch primary care system evolved to a three-segment payment model. In the first segment, GPs receive a base payment composed of a consultancy fee and capitation fee based on the number of patients enrolled, and in the second, they receive a bundled payment for chronic care. In the third segment, health insurers are encouraged to engage in different financial innovations that provide additional coverage that is missed in the first two segments, i.e. a segment for performance and innovation (Schut and Varkevisser, 2017). ACP is an example of an area that does not get sufficient focus under the first two segments (Bekker et al., 2021) and requires financial innovation to fund it.

Frail elderly are defined as individuals over the age of 65 with multiple chronic conditions according to the RIVM (2016). Menzis is one of the largest health insurers in the Netherlands. They have introduced an Elderly Care Module (ECM) in the third segment, which is a capitated fee per patient over the age of 65 for GPs so that they can organise multidisciplinary care for the frail elderly, including ACP.

GPs who fulfil specific criteria are eligible to choose the ECM option to bill from Menzis at the start of every calendar for the duration of that year. To be eligible for the scheme, a GP must be enrolled with Menzis, employ a nurse trained specially for elderly care, and be a part of a GP care group, i.e. not a standalone GP. Requiring to meet the criteria implicitly means that the scheme is allocated at the GP level because GPs can only take the program if the GPs have a sufficient number of frail older people in their practice. By construction, this means that the frail elderly in practices with many young, healthy people are less likely to be in the program. The eligibility criteria prevent malpractice, and the quality controls

by Menzis can be found in Menzis (2017). Some relevant quality controls include hiring an adequately qualified nurse and setting up an integrated care network such that a GP without a majority of frail elderly clientele would not take the program. de Vleminck et al. (2013) suggests that GPs in the Netherlands did not have sufficient resources to set up such networks, and hence ACP related activities are likely to be rudimentary before the ECM program. Additionally, there were no other changes to the innovation scheme that could affect the estimates in the forthcoming analysis. Also, there were no other shocks during this time period, which means that there are no other policy changes that are free-riding or diminishing the effects of the module.

The activities related to ACP that GPs must do as a part of the ECM entail multiple parts. First, a process of case findings occurs where a nurse and the GP determine frail elderly based on signals from the healthcare network and informal care. Then, a problem inventory is accounted for based on historical usage by the healthcare provider and the local care network. Finally, a case management process follows where a GP may assign a nurse, an elderly advisor, or the GP themselves. In consultation with the patient, the care manager makes an individual care plan for the end of life (advanced care planning) depending on the complexity of the illness and the existing local care network. However, it is an empirical question whether GPs in the ECM have the patient profile of a frail elderly. The patient profile of a treated GP is essential to know because for the program to show effects, the GPs need to have patients that could benefit from the program[1]. This leads to the following hypothesis,

*Hypothesis 1: Frail elderly are more likely to be with GPs that have the ECM.*

So far, I have introduced the Dutch health care system and how the ECM builds into the system by adding financial innovation to an area that is not sufficiently accommodated. Then, I argue that it is important to empirically test if the frail elderly are actually with GPs with the ECM. The arguments lead to the first hypothesis, which assesses if frail elderly are entering the program. There is also a second goal to the program related to allocative efficiency, which is the next topic of discussion.

## 2.2 Allocative Efficiency of Advanced Care Planning

The second goal of the program is to improve resource allocation within the healthcare system. The resource allocation problem occurs because frail elderly are often prescribed expensive hospital treatment without notable benefits to mortality or morbidity. The notion of limited additional benefit from treatment is challenging to quantify, but an impression from the literature can offer a better idea.

First, I am going to show that expenses are notably higher in the last year of life with limited scope of benefits. Older evidence from Polder et al. (2006) shows that total costs are approximately 13 times more at the end of life and projects these costs to fall by 10 percent over the forthcoming years. Stooker et al. (2001) suggest that total cost over a lifetime is approximately 10 percent in the last year of life. Melberg et al. (2013) show for Norway that 10.6 percent of lifetime costs are in the last year of life. So, this evidence suggests that costs tend to be skewed toward the end of life expenses. French et al. (2017) suggest that these end-of-life costs are driven by elderly with multiple chronic illnesses, i.e. frail elderly. The literature review Lehnert et al. (2011) show that higher costs at the end of life can be associated with higher resource use for those with multiple chronic conditions. So, it seems that the elderly are spending large amounts at the end of life and using resources; however, they are not reaping life-extending benefits.

[1] An extended discussion on the goals of the program can be found in the appendix.

If ACP is sufficiently done, patients may not end up in these expensive treatment groups, and instead, GPs can actively substitute hospital care for cheaper resource use and cost through palliative care, such as at-home care with a nurse. The intuition is that expensive hospital treatment can be avoided by organising and planning care for the elderly. So, from a theoretical standpoint, ACP is likely to reduce costs because it can initiate the substitution of unnecessary expensive end-of-life care among the frail elderly.

However, to what extent is ACP cost-saving? Does proactive planning for a patient's future lead to cost savings? Systematic literature reviews are appropriate to find to what extent ACP can be cost-reducing. Firstly, the literature review of Klingler et al. (2016) find a statistically significant association between ACP and healthcare savings. The effect is focused on people living in nursing homes with high support needs and low income. Second, the literature review of Dixon et al. (2015) find that savings of ACP are in groups with dementia, in nursing homes, and high expected end-of-life spending. They also note that studies do not offer mechanisms that explain the effect of ACP. Additionally, O'Sullivan et al. (2016) look at a new ACP policy introduction in Ireland. In terms of costs, they find approximately 4,000 euro decreases in cost per episode and 491 euros per day on average. Using a probabilistic sensitivity analysis, they extrapolate the findings to 17 to 42 million euro decreases in expenditure country-wide annually. The high savings is quite a notable decrease in costs; however, the methodological aspects are dubious because they do not account for existing cost trends or model any counterfactual. In terms of non-cost-related outcomes, O'Sullivan et al. (2016) also found uptake of ACP activities (similar to those in the Menzis program) by 50 percentage points, decreases in in-hospital deaths by approximately 10 percentage points, and increases in length of stay at hospitals by 7 to 9 days. A literature review by Khandelwal et al. (2015) corroborates evidence in O'Sullivan et al. (2016) concerning health care usage rates. K.-H. Nguyen et al. (2017) report similar findings for Australia but also lack methodological strength since they do not account for existing cost trends or model any counterfactual.

More studies consider the effectiveness of ACP with cost-related outcomes. Brinkman-Stoppelenburg et al. (2014) showed that ACP positively impacts the quality of end-of-life care and suggests that complex ACP interventions are effective. They also advise that ACP planning needs to be extensively done, as evidenced by the lack of documentation and preparation. Furthermore, more recent global systematic reviews by Jimenez et al. (2018) find evidence that institutional policies influence ACP development and that interventions that target multiple stakeholders are particularly effective. This is similar to the Menzis scheme, which seeks to involve multiple stakeholders( elaborated in section 2.4). A recent paper by Gabbard et al. (2021) studied a randomised trial of an ACP intervention and found higher uptake of documentation of ACP on a small-scale introduction and found limited evidence on usage of hospitalisation services. The critical point here is that if ACP is implemented in a focused way, the planning process takes place, which means that the supervisory conditions in the documentation can already drive some uptake of ACP by GPs.

Finally, in their literature review, Marckmann et al. (2013) note that they do not find methodologically sound analyses of ACP interventions. The literature reviews use inclusion criteria made by Higginson et al. (2002). For a paper to be in the inclusion criteria, costs must be a primary measure because utilisation rates as sole indicators for cost reductions fail because they do not provide an account of the net resource use resulting from ACP interventions. The reasoning serves as an important indication as to why cost should be measured as expenditure rather than only utilisation rates.

Overall, these systematic reviews show ACP can be cost decreasing, and in Dutch healthcare, it is insufficiently done. Furthermore, there are few studies with a methodologically sound quasi-experimental approach to identify the causal effects of an ACP intervention. The

possible cost reductions and inefficiencies in the current system imply possible marginal returns to investment and that financial incentives for GPs may be helpful. Hence, the following hypothesis can be formed.

*Hypothesis 2: Hospital costs are reduced by the introduction of the ECM.*

## 2.3 Funding and Incentives

So, we have established that allocative efficiency will improve because shifts from expensive to cheaper care will free up resources in intensive hospital care. However, why are GPs unable to do this already? Before the Menzis scheme, there was no funding or financial incentives for ACP, which could be associated with the low levels of ACP in the Netherlands Bekker et al. (2021). There are a few drivers with the current system that does not lead to ACP and hence an inefficient allocation of resources. de Vleminck et al. (2013) note two sets of issues. The first set is that GPs lack skills to deal with patients' vague requests and the frictions with defining the appropriate time to start ACP. The second set of issues - with stronger evidence - refers to the limited ability of a GP to predict health issues and respond to a patient's initiation of ACP.

The funding of the scheme can help resolve these issues. The first group of issues can be solved by hiring a nurse specialising in elderly care so that the doctor has additional support when making decisions. The second group of issues can be solved by setting up a geriatric assessment or an integrated intervention network so GPs can make a more informed decision. These solutions only become feasible with additional funding and address the resource allocation's sub-optimality. However, simply receiving funding does not lead to an effective scheme. The payment system is essential, which is why a prospective reimbursement system's implications are relevant. Currently, there is a gap in the literature that links ACP to costs whilst noting the implications of the funding scheme. So, it is crucial to explain how prospective reimbursement can uncover or reduce the underlying value of ACP.

The purpose of analysing the incentive structure is to understand its impact on the estimates in the results. It is hard to disentangle to what extent the effect is driven by additional funding versus having a particular incentive to do an activity. As explained above, funding will likely lead to an effect regardless of the incentive scheme because GPs will have additional resources. However, it is important to optimally fund this scheme to realise maximum returns. So, it is necessary to break down the incentive structure of the prospective reimbursement scheme to see how it will affect the expenditures.

Prospective reimbursement can drive the cost effects of the program both upwards and downwards. First, I start with a theoretical model to show how prospective reimbursement can be beneficial. The model will be used to illustrate the incentives and as a heuristic to interpret results. In a healthcare system, GPs behave as an agent for providing care for patients and hospitals. Sometimes GP may behave as an imperfect agent, i.e. the GP does not behave in the patient's best interest. The agency problem to consider is where the GP is the agent for hospitals and patients as done by R. Ellis and McGuire (1986). This is a theoretical model for understanding the mechanism behind the effect of a prospective scheme, and it has been empirically verified by Godager and Wiesen (2013).

Using utility optimisation, R. Ellis and McGuire (1986) propose a mixed payment system under the purview of imperfect agency of GPs. This is similar to the current Dutch healthcare system, where there is a mixed payment system and will provide a framework to interpret the results. They consider three actors in a health economic transaction. First, the patient's benefit function $B(q, s)$ is composed of the quantity of services $q$ from the hospital and inputs by the doctor $s$.

Second, a hospital that faces a mixed reimbursement system who have a profit function $\pi$ consisting of a fee for service and a prospective system.

$$\pi = a + (r-1)cq \tag{1}$$

where $a$ is a prospective amount, $r$ is the proportion of revenue shared between the physician and hospital, and $cq$ is the value of the fee for service.

Finally, the physician's utility function comprises of profit function of the hospitals and the benefits to the patients.

$$\max_q U(a + (r-1)cq, B(q,s)) \tag{2}$$

and the first order condition using implicit differentiation,

$$\frac{\partial U}{\partial \pi}(r-1)c + \frac{\partial U}{\partial B}b(q) = 0 \iff MRS_{\pi,B} = \frac{(1-r)c}{b(q)} \tag{3}$$

Equation 3 shows that the Marginal Rate of Substitution (MRS) is given by the ratio of the fraction of cost paid to marginal patient benefit. R. Ellis and McGuire (1986) argue that such a mixed payment scheme balances the strong incentives to reduce care under prospective payment and will lead to a more efficient quantity of services if the physician is not a perfect agent for patients and doctors.

This model applies to the Dutch healthcare system because it is a mixed reimbursement system. The ECM is a prospective system of reimbursement and thus makes the overall system more prospective for the GPs that are eligible for it, and following from the model above, it should also make GPs more cost-conscious. Specifically, it encourages GPs to do the necessary ACP to realise the benefits whilst not having multiple ACP meetings with insufficient marginal value. Hence, the key point to take away here is that the prospective system limits incentives for over-provision. By providing a prospective payment to focus on ACP, GPs are encouraged to be cost-conscious, leading to efficiency gains such as a single prudent meeting with patients.

The empirical literature is hard to apply since the context of the Menzis capitated fee is unique; however, some indications can still be observed. Prospective reimbursements tend not to have a large effect on costs relative to the estimated sizes seen in the empirical literature on ACP. Older evidence from the U.S.A shows a reduction by 2 to 6 percentage points per year in total costs on average (Coelen and Sullivan, 1981). Coburn et al. (1993) find an estimated 3.03 dollar decrease in total variable costs per person three years after the intervention relative to a fee-for-service system. Dobrez et al. (2010) found that introduction of prospective payment was associated with small, statistically significant reductions in measures related to cognitive functions of patients. Finally, Sood et al. (2008) apply an IV strategy to study a new prospective reimbursement on inpatient rehabilitation facilities and find around 7 percent reductions in cost in that care facility without changes to patient outcomes. In this case, the underlying value of inpatient rehabilitation facilities is better reflected under prospective reimbursement; however, a key limitation is that they only consider mortality as a patient outcome, and under-provision of care can manifest as morbidity harm. The main point is that an effective prospective scheme will amplify effects such that the value of the healthcare program will come through.

We can now consider the alternate side where prospective reimbursement is ineffective. Prospective reimbursement can also decrease the effect of a program due to the risk of

underprovision. How is under-provision manifested in the context of ACP? An important part of the ACP duties in the ECM is organising meetings with a patient about their end of life and setting up a care network for the individual. If the GP finds the fee insufficient, then these meetings are less likely to occur, and the intervention's impact is less likely to be significant. Hence, when observing effects, an explanation for no notable effect is that the GP was insufficiently compensated for doing the ACP in the ECM. An argument against this can be that the GPs who insufficiently do the activities of ECM may become ineligible in the forthcoming years. However, as mentioned in section 2.1, the module's conditions do not police the extent of activities but only if the set-up exists. Hence, it is still plausible for the GP to behave as an imperfect agent due to insufficient compensation; thus, the value of ACP may not be realised.

Once again, empirical literature can not be directly applied in the ACP context, but some indications of the ineffectiveness of prospective reimbursement exist. The famous paper of R. P. Ellis (1998) find that providers avoid care for high severity patients relative to the social optimum. Thomas and Davis (1987) surveyed GPs and found that they do not become cost-conscious after implementing a prospective reimbursement. Recent evidence by Reif et al. (2020) suggests that physicians and medical students consider the third party's payoffs, implying a possibility of under-provision of medical care.

So, there are two drivers, funding that makes advanced care planning possible because GPs can spend on services they require to improve their ACP capabilities. Conditional on sufficient policing, incentives can amplify the effects of ACP if the GP finds the compensation sufficient, but it can also reduce effects if the fee is insufficient. Hence, the analysed effect will be a mixture of ACP's efficiency gains and an ambiguous effect of prospective reimbursement. Consequently, if no effects are observed, a possible mechanism can be that the GPs found the ACP reimbursement insufficient and hence did not implement it adequately.

Now we know ACP is effective, and if the prospective reimbursement works well, the effect of ACP will shine through. However, it is not yet clear where the effects of ACP will be observed other than hospital costs. To understand this, I consider some literature on preferences for end of life in the Netherlands.

## 2.4 Last Days of Life in the Netherlands

What aspects does Dutch society find necessary in their last days of life? Specifically, where would people prefer to die or experience their last years of life? This is important to determine where effects will be expected. For example, if people prefer to die at home, then hospital costs are likely to be lower, and palliative care costs are likely to be higher; however, if people have a strong preference to be kept alive, then hospital costs are likely to be higher.

Survey literature can illuminate what the preferences are for members of Dutch society, which in turn will help indicate where to expect effects. Abarshi et al. (2009) find that 90 percent of people in the Netherlands preferred to die at home. A later study by de Roo et al. (2014) indicates that 60 percent of GPs in the Netherlands knew the preference of the place of death of their patients. This percentage would be likely to rise through ACP. A telephone-based population survey by Gomes et al. (2012) corroborates evidence; they find that 84 percent of people in the Netherlands prefer to die at home. The general indication is that Dutch people prefer to die at home, and thus, ex-ante, I expect palliative care costs to rise because Dutch people prefer to die at home in the company of their family on average. However, this literature is prone to poor responses from individuals because answers on phone calls or surveys on preferences on death are likely to be prejudiced.

An apt area that overcomes shortcomings of surveys for preferences of death is euthanasia

literature since patients tend to be meticulous in explaining their preferences. First, Parry and Munson (2015) explore a variety of case studies in the Netherlands where they document the need and concerns of terminally ill patients. They point out the particularity of each part of the dying process and that patient-family dynamics play a key role. Next, Norwood et al. (2009) note that a key component of the Dutch euthanasia policy is to affirm social bonds. So, there is an indication that the presence of family is important, which is often more possible in a home setting rather than an intensive care setting. So, both euthanasia and survey literature indicates that patients in the Netherlands prefer a specific process.

So, the literature seems to suggest that Dutch individuals prefer to die at home in the company of their family. Hence, there is an expectation that the costs of palliative care will be higher since people would request GPs to substitute away from expensive hospital care. In terms of observations, a fall in hospital costs should lead to more expenditure on palliative care such as nursing care at home, geriatric care, and GP care to identify evidence of ACP working. However, it must be noted that all palliative care does not need to have an uptake because of substitution effects. One kind of palliative care outcome might have a higher uptake, while other care is substituted preference of the patient and the advice from the case manager. This leads to the third hypothesis.

*Hypothesis 3: The ECM leads to higher expenditures in palliative care.*

Also, it can be checked if hospitals are changing the length of stay for those that have the ECM; this offers some indications in terms of non-cost measures. This could happen because patients in consultation with their case managers can identify that the patient could have less time spent in hospitals, and then the length of stay would decrease. Furthermore, the preferred place of death in the Netherlands is at home, which implies that length of stay is also unlikely to increase by people who prefer to pass at hospitals. With this information, the fourth hypothesis follows,

*Hypothesis 4: Length of stay is decreased by the introduction of the ECM.*

The final part of ACP's effectiveness is considering its implications as a complete system. It is good that substitution is happening, but total costs can illuminate if the net effect is tangible. A framework for a holistic analysis follows in the next subsection.

## 2.5 The Dynamics in the Health Care System

The final component that is necessary to explain is the interactions within the health care system resulting from the policy change. O'Hanlon et al. (2018) offers a framework for analysing how different stakeholders are affected by a change in policy in the context of ACP. He highlights that studies that evaluate an ACP intervention tend to focus excessively only on certain outcomes and not the whole healthcare system.

The final hypothesis follows from the holistic analysis of the healthcare system, which is about the total cost. If total costs are unchanged, it is an indication that hospital care was substituted for high amounts of alternate care and would, in turn, suggest that although resources may be better allocated, the overall costs do not decrease, which means cheaper less intensive care offsets the decreases in expensive hospital costs. However, if total costs decrease, it will signal that the decrease in hospital costs is not offset by an increase in alternate care, which would be in the interest of the policymakers since this was one of the program's goals. However, it is still an empirical question whether total costs change or not through substitution effects. So, the final hypothesis follows,

*Hypothesis 5: The ECM leads to a change in total expenditures.*

In this section, the background of the study was established. Menzis, a Dutch health insurer, has introduced a capitated fee for GPs with frail elderly patients. ACP is an effective way

to realise societal goals but also a better allocation of resources. However, the method of financing the scheme is also important, and so the prospective reimbursement literature says that the scheme will be successful if there is sufficient compensation. So, a possible driver of no effect could be that GPs are insufficiently compensated. Finally, the outcomes of total cost, palliative care, and length of stay will gauge to what extent the intervention was successful. With these ideas in mind, the empirical strategy follows.

# 3 Empirical Strategy

## 3.1 Data

In this section, I will describe how the data set is formed, the general structure of the data set, how variables are defined, and what outcomes are measured. I analyse Menzis' claims and demographic data for frail elderly patients. Frail vulnerable elderly are defined as those who are above the age of 65 and have multiple chronic conditions.(Menzis, 2017 and RIVM (2016)). The chronic conditions encompass areas of physical, social, cognitive, and psychological vulnerability. The sample is selected to patients who are frail elderly because this is where the effects are intended.

The panel data set runs from 2015 to 2019, and I analyse 2 intervention years, i.e. 2017 and 2018. An intervention year or cohort refers to the first period where a GP bills the ECM. The treatment group refers to the patients whose GP bills the ECM in the intervention year and is assumed to remain in the intervention group. This means GPs and the corresponding patients who first received the module in 2015, 2016, and 2019 are dropped from the sample. When the intervention year is 2017, the observations with the first occurrence of the ECM in 2018 are dropped and vice versa. This strategy is done because the timing of the treatment is important for identifying the effect. This is explained in detail in the next section. Next, the control group comprises frail elderly patients whose GPs were not eligible or did not choose the ECM. Thus, they can serve as an appropriate comparator since they are similar to patients in the treatment group but do not receive the fee because of exogenous factors, namely, the GP facilities and their peers' health.

The dependent variables in this study are annual measurements of health care expenditures in euros. They include expenditure from the GP, nursing home care and nurse visits, geriatric rehabilitation care, hospital, and the length of stay at the hospital. Additionally, total costs are in the data, which are the summation of the aforementioned costs.

Additional control variables include age, gender, number of chronic illnesses, and disease severity. The disease severity indicators need to be explained further. Disease severity is operationalised using expected cost indicators. The intuition is that if patients are in the high expected cost category then they are associated with more disabling illnesses since they would require more treatment. There are two measures of illness severity. The first is from diagnostic care groups (DKGs), which are formed based on historical hospital resource usage and insurance risk. DKG's dummies for the severity of illness are formed based on an ordinal list of indicators from 1 to 15 where a unit increase only states if expenditure was higher than the previous category. A patient was considered to have low severity if they had a dummy between 1 to 5, medium severity if 6 to 10, and high severity if 11 to 15. However, this indicator considers mainly hospital usage; hence, a more general indicator is necessary. So, the second is based on the quartiles of historical total costs. There are dummies, low severity for the first quartile, medium for the middle quartile, and high for the top quartile. The indicators are measured in the year before they are indexed. For example, the DKG and historical costs of 2017 are based on the consumption of 2016.

The chronic disease indicators are formed based on Zorginstituut-Nederland (2017). In-

surance companies collect data on the medicine consumption of the insured. The national insurance agency of the Netherlands has reference categories for medicinal consumption that are documented in Zorginstituut-Nederland (2017). If an individual consumes medicine beyond a specific reference range, they receive an indicator for it. Certain combinations of these reference indicators relate to chronic illnesses in the data set. A simple example, a patient has a reference indicator for inhaled corticosteroids and fluticasone, then they would get a dummy for asthma. Finally, all variables are measured at the end of a calendar year.

## 3.2 Methods

In this section, I will discuss the methodological basis for the empirical strategy. First, I will explain a binary outcome model and then a quasi-experimental research design to check for the cost outcomes will follow.

The binary outcome model will be used for propensity score matching and testing the first hypothesis, i.e. if frail elderly are with GPs in the ACP intervention group. The binary outcome of interest is having or not having the treatment. A logit model is an appropriate method to test binary outcome variables because the logit model will indicate the probability of being in the treatment group conditional on the characteristics of a frail elderly individual. Furthermore, it bounds fitted values between 0 and 1 which is suitable for interpreting probabilistic outcomes . I also add variables for the severity of illness based on historical expenditure and usage in the hospital. Dummies for gender are added to inspect any differences by gender which are likely to arise in this older age group (Waldron and Johnston, 1976). The specification of the model follows,

$$logit(Y_i) = \ln(\frac{\pi}{1-\pi}) = \alpha + \beta_1 Age_i + \beta_2 Age^2 + \beta_3 Chron_i + \beta_4 DKG_i + \beta_5 past_i + \beta_6 Male + \epsilon_i \tag{4}$$

where $\pi = Probability(Y = Treated|X_1 = x_1, ..., X_n = x_n)$ , $\alpha$ is the intercept, $\beta$ are regression coefficients, and $X$ refers to the set of predictors. The parameters $\beta$ and $\alpha$ are estimated using maximum likelihood. $Chron$ refers to the number of chronic diseases, $past$ refers to a categorical variable for disease severity based on past total costs, $Male$ is a dummy for gender, and $DKG$ is disease severity based on the $DKG$ categorical variable explained earlier. The time index is the same for all variables; hence the subscript for time is omitted; the model will be estimated for cross-section at $t = -1$.

The model will be estimated with respect to only the year prior to participating in the program for two reasons. Firstly, from a logical standpoint, GPs evaluate their patient group in the year prior to billing the module to the insurer. Hence, it is fair to use the year prior to check if frail elderly are with GPs with the module. Secondly, from an econometric standpoint, a logit model for short panel data can be inconsistent due to the incidental parameter problem (Stammann et al., 2016), which holds true for this data set since there is insufficient data on years prior to the treatment allocation. The estimates from the logit model in equation 4 will be useful for the quasi-experimental research design.

The second goal of the empirical part relating to the costs is to compare patients whose GPs billed the fee to patients whose GPs did not bill the fee. However, there is an identification issue because the fee was allocated based on eligibility criteria which means that there is a selection effect. As explained in section 2.1, the eligibility criteria are linked to the GP's ability to treat the elderly and the number of elderly in a GP practice. Hence, the selection leads to the treatment and control group not being equivalent a priori because eligible GPs are more suitable to treat the elderly already. Thus, a cross-sectional comparison is ineffective because there is a systematic difference between the treatment and control groups.

A before and after comparison is not complete either because there are factors that vary over time, such as patients becoming more ill or rising expenditure costs.

To solve the identification issue, a difference in differences (DiD) design as seen in Wooldridge (2015) is appropriate because it accounts for systematic time-invariant differences and the time-varying differences. Time-invariant factors can be modelled by using a dummy for the initial difference. The time varying differences between groups can be modelled if the identifying assumption is met, which makes the results closer to a causal interpretation. The identifying assumption of a DiD is that a control group and treatment follow a common trend such that the control group can be a valid counterfactual for the treatment group. So, time-varying parameters can be accounted for as long as they are the same in both the treatment and control groups. In the context of this study, it means that a priori, I expect the costs for patients in the treatment group to follow in parallel to the cost trend of the control group GPs. Some other factors need to be considered to ensure consistency and efficiency of the estimates. Firstly, time-fixed effects need to be included to account for time trends. Secondly, GP-clustered, heteroskedasticity-robust standard errors are needed to account for correlation in expenditures between patients of the same GP and serial correlation for the GP cluster similar to Hayen et al. (2021). With the information in mind, the estimation equation is in equation 5.

$$Y_{igt} = \delta_0 + \delta_1 ECM_g + \sum_{j \neq k} \beta_j Time_{t=j} + \sum_{j \neq k} \beta_j ECM_g \cdot Time_{t=j} + X'_{it}\gamma + \epsilon_{igt} \tag{5}$$

$Y_{igt}$ refer to the annual spending for individual $i$ assigned to GP $g$ in the year $t$. Costs are from multiple categories, and this is explained later. $ECM_g$ is a dummy variable which is 1 if the GP is in the Elderly Care Module (ECM) group. $Time_t$ are time dummies referring to each year, and the year prior to treatment serves as the reference; Time dummies model the time fixed effects which model aggregate shocks that belong to both groups. $X_{it}$ refers to a vector of patient-level control variates namely, $age$, $age^2$, gender, the severity of illness based on hospital expenditures, and severity of illness based on past expenditures. These are included to pick up any residual variation missed by propensity score matching (explained later) and improve the precision of estimates (T.-L. Nguyen et al., 2017).

The years $j \leq k$ are the pre-treatment periods. All periods $j \neq k$ are representative of post-treatment period dummies. The $\beta_j$'s are estimates of the treatment effect for each year. So, every $\beta_j$ is an estimate of the $j-$th year specific treatment effect. To get the overall average treatment effect in the post treatment years, the model can be specified with an interaction term between a dummy for the post treatment period and a dummy for the treatment group. The equation for this setup is below,

$$Y_{igt} = \beta_1 ECM_g + \beta_2 Post_t + \beta_3 ECM_g \times Post_t + X'_{it}\gamma + \epsilon_{igt} \tag{6}$$

*Post* refers to a dummy that identifies the years after the intervention $\beta_3$ will identify the average treatment effect on the treated (ATT). $ECM_g$ is a dummy variable which is 1 if the GP is in the Elderly Care Module (ECM) group. Once again, $X_{it}$ refers to a vector of patient-level control variates, namely, age, $age^2$, gender, the severity of illness based on hospital expenditures, the severity of illness based on past expenditures, and $\epsilon_{igt}$ is the error term.

It is important to explain how the different hypotheses from the theoretical framework will be tested using the DiD approach. Hypothesis 2 and 5 test if the hospital and total costs have decreased, respectively. Hence, for these hypotheses, $Y_{it}$ will be hospital and total costs, and the coefficient of interests and its corresponding test will be,

$H_0 : \beta_j = 0$

$H_1 : \beta_j < 0$

Hypothesis 3 considers the length of stay at the hospital. So,$Y_{it}$ will be the length of stay in the number of days and will have the following test based on the idea that length of stay is expected to be unaffected.

$H_0 : \beta_j = 0$

$H_1 : \beta_j < 0$

Hypothesis 4 tests if palliative care costs have increased. But it is important to make the test two-sided so that the substitution effect can be checked, i.e. one care expenditure might decrease whilst the other can increase. Palliative care consists of nursing care, geriatric care, and GP care. Hence, $Y_{it}$ will consider each of these outcomes of palliative separately, and it will be tested using the following,

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

If the null hypothesis is rejected at a 5 percent significance, it means that there is statistically significant evidence of cost reduction or increase (depending on which hypothesis is tested) in the treatment group of GPs with the ECM relative to the control group without the ECM.

Next, I will explain the doubly robust method for implementing a DiD design. The parallel trends assumption is crucial for the control group to be a valid counterfactual. Funk et al. (2011) explain the doubly robust estimation, which helps further validate the counterfactual. Doubly robust estimation involves propensity score matching (Heckman et al., 1997) and adding control variables in the outcome estimation to pick up notable differences between treatment and control. If either of the matching or the outcome estimation is correctly specified, then the results can be unbiased. The intuition is that by matching observations, a more valid counterfactual is formed because observations would have more similar characteristics, which in turn would make it more likely for them to follow the trend that the treatment group would follow in the absence of the treatment.

Now, I will explain Propensity Score Matching (PSM) and how it will be used to form a valid counterfactual. PSM involves 2 steps. First, a binary outcome model is estimated so that that the dimensions of the data can be reduced to one dimension i.e. the propensity score. For this study, the logit model in equation 4 will be used since the above specification structures the important variables that are relevant for the prediction of the probability of treatment. The second step involves matching observations on their propensity scores such that the differences between the covariates are minimised. There are a variety of algorithms that perform this matching, and this study will use a 1:1 nearest neighbour matching without replacement based on propensity scores from a logit model as defined in equation 4 (Caliendo and Kopeinig, 2008 and Rubin, 1973. If observations have identical propensity scores, one of them is chosen at random by the 'MatchIt' package in R (Ho et al., 2011), which was used to run the 1:1 nearest neighbour matching without replacement algorithm. The algorithm was chosen for computational speed. The algorithm has the advantage of being intuitively appealing because it finds the closest possible observation between treatment and control observation. However, this algorithm has the downside of discarding a large number of observations since it looks for a 1:1 match. However, the review of Stuart (2010) shows from multiple studies that the power concern is negligible if the treatment group does not change in size, making it an appropriate matching algorithm. Finally, the propensity score matching weights are implemented into a weighted regression of the DiD specification (Heckman et al., 1997).

In this section, the methodological basis for the empirical part has been established. First, the use of a logit model was explained to understand how the probability of treatment based on the characteristics of the frail elderly would be investigated. Then, the use of the doubly robust method of DiD, which involves propensity score matching and control variables, was explained. Finally, I described the exact model specification of the DiD and how the hypotheses from the theoretical framework would be tested. With these ideas in mind, the results can be analysed.

# 4 Results

## 4.1 Summary Statistics

An appropriate place to start is with summary statistics to see identify characteristics of the sample that is studied. Table 1 and table 2 show the descriptive statistics for the the intervention years 2017 and 2018 divided by treatment and control group. The values are the pooled means across all the years 2015 to 2019. The summary statistics do not change if a single cross-sectional is summarised because these variables are not affected by the treatment[2]. The greatest difference by magnitude is in age. The treated group is older on average in both intervention years. Another notable difference in the 2018 cohort is that the control group has 25 percent of its people in the high-cost category compared to 17 percent in the treated group. Since the cost category is supposed to represent the severity of illnesses, it can be argued that the control group is sicker on average. The difference in age and illness severity means the matching is a necessary step because significantly different characteristics can lead to different responses to changes over time. With these descriptive statistics in mind, the empirical evidence can be evaluated.

| **Variable** | **Treated** | **Control** | **SE** |
|---|---|---|---|
| Number of Observations | 2535 | 80580 | |
| Number of Patients | 507 | 16116 | |
| Number of GPs | 122 | 1803 | |
| Age | 80.97 (5.14) | 75.84 (6.72) | 0.135*** |
| Number of Chronic Illnesses | 10.17 (8.14) | 10.18 (8.12) | 0.020* |
| Male | 0.41 | 0.43 | 0.009* |
| Severity of illness based on DKG | | | |
| Low Cost | 0.87 | 0.86 | 0.004*** |
| Medium Cost | 0.12 | 0.11 | 0.004** |
| High Cost | 0.01 | 0.03 | 0.001*** |
| Severity of illness based on past Total Costs | | | |
| Low Cost | 0.28 | 0.25 | 0.009** |
| Medium Cost | 0.50 | 0.55 | 0.010 *** |
| High Cost | 0.22 | 0.25 | 0.009 *** |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Descriptive Statistics intervention year 2017. *SE* refers to the standard error of the mean difference. The stars refer to statistical significance of a t-test of the mean differences.

[2]Cross sections summary statistics can be inspected in the appendix.

| Variable | Treated | Control | SE |
|---|---|---|---|
| Number of Observations | 1845 | 80580 | |
| Number of Patients | 369 | 16116 | |
| Number of GPs | 187 | 1803 | |
| Age | 80.96 (5.14) | 75.84 (6.72) | 0.157 *** |
| Number of Chronic Illnesses | 10.26 (8.14) | 10.19 (8.12) | 0.191 |
| Male | 0.41 | 0.43 | 0.006 *** |
| Severity based on DKG | | | |
| Low Cost | 0.87 | 0.86 | 0.006 *** |
| Medium Cost | 0.12 | 0.13 | 0.007 *** |
| High Cost | 0.007 | 0.01 | .001 |
| Severity based on past Total Costs | | | |
| Low Cost | 0.28 | 0.25 | 0.006** |
| Medium Cost | 0.55 | 0.50 | .007 *** |
| High Cost | 0.17 | 0.25 | 0.006 *** |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Descriptive Statistics intervention year 2018. *SE* refers to the standard error of the mean difference. The stars refer to statistical significance of a t-test of the mean differences.

## 4.2 Treatment Probabilities

Hypothesis 1 checks if frail elderly are getting into the treatment group. Table 3 shows the estimates from the logit model, which will be used to provide evidence for hypothesis 1. Firstly, it can be seen that older people are more likely to get into the treatment in both intervention years. Secondly, in the 2017 intervention year, more chronic diseases make you more likely to be in the treatment group. However, in the 2018 intervention having more chronic diseases makes an individual less likely to be in the treatment group. It could be that the 2018 group has more severe illnesses than the 2017 group. In the 2018 intervention, high and medium severity of illness categories are more likely to be in the treatment group. Severity based on DKGs shows no significant differences. Overall, it seems that in the 2017 intervention year, there were a higher number but less severe chronic illnesses among the treated; in the 2018 intervention year, there were fewer but more severe chronic illnesses. Having fewer severe illnesses and higher low-severity illnesses are likely to be notably both impairing in terms of morbidity and quality of life. Hence, there is evidence for the first hypothesis because older frail elderly have a higher probability of getting into the program.

The fitted values from this logit model are used to form propensity scores like in Heckman et al. (1997). These propensity scores need to be matched using a matching algorithm; this is the next part's topic.

| | Logit Models | |
|---|---|---|
| | 2017 | 2018 |
| (Intercept) | $-95.42^{***}$ | $-78.54^{***}$ |
| | (3.75) | (6.32) |
| Age | $2.19^{***}$ | $1.76^{***}$ |
| | (0.09) | (0.15) |
| $Age^2$ | $-0.01^{***}$ | $-0.01^{***}$ |
| | (0.00) | (0.00) |
| Num. Chronic | $0.01^{*}$ | $-0.42^{***}$ |
| | (0.00) | (0.03) |
| $Medium_{DKG}$ | 0.02 | 0.19 |
| | (0.06) | (0.11) |
| $High_{DKG}$ | −0.12 | 0.04 |
| | (0.24) | (0.39) |
| Male | 0.02 | −0.02 |
| | (0.04) | (0.08) |
| $High_{past}$ | $-0.68^{***}$ | $0.87^{***}$ |
| | (0.06) | (0.12) |
| $Medium_{past}$ | −0.08 | $0.63^{***}$ |
| | (0.05) | (0.12) |
| AIC | 20472.58 | 6809.43 |
| BIC | 20556.54 | 6893.31 |
| Log Likelihood | −10227.29 | −3395.72 |
| Deviance | 20454.58 | 6791.43 |
| Num. obs. | 83115 | 82425 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Logit models for both intervention years. For the 2017 cohort, the model has been based on data from 2016. For 2018 cohort, the model has been based on data from 2017. The subscript $DKG$ refers to severity of illness based on insurance risk. The subscript *past* refers to severity of illness based on past consumption

## 4.3 Matching and Parallel Trends

So far, propensity scores have been collected based on the logit model from 4. These estimates are now matched using a 1:1 nearest neighbour matching without replacement algorithm based on the propensity scores. For the intervention year 2017, 2535 observations are matched from the possible 80580 original observations in the control group. Figure 7 shows the balance of the propensity scores and table 4 shows the balance on the covariates.

In the intervention year 2017, 2535 observations are matched from the possible 80580 original observations in the control group. No treated observations were dropped, meaning that the power concern from a 1:1 matching is less of a concern. The histograms in figure 7 show the effect of the matching for the 2017 intervention cohort. Firstly, post-matching, the propensity score distribution of the control closely resembles the shape of the distribution of the treated. It seems as though the smaller propensity scores are not similar between treatment and control. Checking the balance on covariates can help interpret how successful the matching is and which variables have been weighted the most. The balance between the covariates can be inspected in table 4.

Table 4 communicates two important points. First, the balance table indicates which pa-

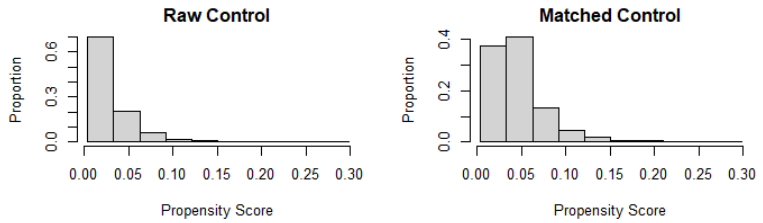


Figure 1: Histogram for the distribution of the propensity scores before and after a 1:1 nearest neighbour matching without replacement algorithm based on the propensity scores for intervention year 2017. The treated group's distribution does not change, this can be inspected in the appendix.

| **Variable** | **Means Treated** | **Means Control** | **Std. Mean Diff.** |
|---|---|---|---|
| Age | 80.97 | 80.98 | -0.0014 |
| Num. Illnesses | 10.1657 | 10.1700 | -0.0005 |
| $Medium_{DKG}$ | 0.1140 | 0.1132 | 0.0025 |
| $High_{DKG}$ | 0.0071 | 0.0039 | 0.0376 |
| Male | 0.4103 | 0.4107 | -0.0008 |
| $High_{past}$ | 0.1688 | 0.1708 | -0.0053 |
| $Medium_{past}$ | 0.5546 | 0.5519 | 0.0056 |

Table 4: Balance table for matched data of covariates for the year prior to intervention year 2018. The subscript 'past' refers to severity of illness based on past expenditure. The subscript 'DKG' refers to severity of illness based on DKG groups. Num. Illnesses refers to the number of chronic illnesses. The differences between the variables are not statistically significant.

rameters were changed notably after the matching once a comparison to table 1 is made. The greatest change was in the age category, where the control group now, on average, has older individuals. The remaining parameters had less notable changes. The covariates for each individual are similar on average after the matching, as indicated by the low magnitude of the standardised mean difference between the group. Therefore, the matching can be considered successful because the covariates and the propensity scores distribution are similar after matching.

For the intervention year 2018, 1845 observations are matched from the possible 80580 original observations in the control group. No treated observations were dropped, which means that the power concern from a 1:1 matching are less of a concern. The matching balance needs to be checked for the intervention year 2018 as well which can be found in figure 2 and table 5. The propensity score distribution after matching treatment and control are similar. Once again, the most notable change is in the age category after the weighting. This is not surprising given that similar groups are expected to enter the treatment group for both years. Finally, the magnitude of the standardised mean differences across the covariates is small. Overall, the matching has been successful for both intervention years because the distribution of the propensity scores is similar after the matching, and the standardised mean differences between the variables are negligible. Consequently, the weights from the PSM can be used for further analysis. The first use is for the parallel trends assumption, which will be analysed next.

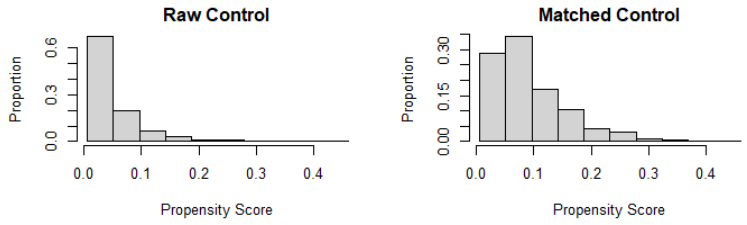


Figure 2: Histogram for the distribution of the propensity scores before and after a 1:1 nearest neighbour matching without replacement algorithm based on the propensity scores for intervention year 2018. The treated group's distribution does not change, this can be inspected in the appendix.

| **Variable** | **Means Treated** | **Means Control** | **Std. Mean Diff.** |
|---|---|---|---|
| Age | 81.4786 | 81.4846 | -0.0010 |
| Num. Illnesses | 10.2625 | 10.2805 | -0.0022 |
| $Medium_{DKG}$ | 0.1574 | 0.1599 | -0.0069 |
| $High_{DKG}$ | 0.0138 | 0.0102 | 0.0305 |
| Male | 0.3271 | 0.3246 | 0.0053 |
| $Medium_{past}$ | 0.4370 | 0.4382 | -0.0025 |
| $High_{past}$ | 0.4445 | 0.4431 | 0.0029 |

Table 5: Balance table for matched data of covariates for the year prior to intervention year 2018. The subscript 'past' refers to severity of illness based on past expenditure. The subscript 'DKG' refers to severity of illness based on DKGs. Num. Illnesses refers to the number of chronic illnesses. The differences between the variables are not statistically significant.

Now, the validity of common trends assumption will be assessed. This will be done by comparing the treatment and control group's raw and weighted time trends. Figure 3 shows the raw and weighted time trends for the intervention year 2017, and figure 4 shows the raw and weighted time trends for intervention year 2018. The graphs on the right-side column are interesting since they show the weighted values after the matching. The unweighted column on the left-hand side is for heuristic purposes since it shows the effect of the matching on the outcome variables when compared to the right-hand side.

First, total costs seem to follow a close trend. Both lines for treatment and control move in parallel and are closer to each other after matching, indicating that there could be fewer time-invariant mean differences between groups. The change in the trend after the intervention for 2017 seems to be lagged by a year. There does not seem to be a discernible change in the trends in 2018 intervention since they both follow the same path and do not diverge or change.

Second, GP expenditures in the treatment and control follow a close trend with a slight divergence prior to the treatment in 2017, but they were reasonably parallel in 2018. In 2018, the treated group breaks the trend relative to the control; this could be an anticipatory effect, but this can not be said with confidence. It is something worth noting for the regressions. Third, nursing expenditures seem to follow parallel trends for both interventions years. The divergence between treatment and control groups is notable for intervention year 2017 and less stark for 2018.

Next, hospital expenditures for 2017 and 2018 interventions shows the strongest visual evidence of parallel trends. Both groups move in parallel prior to the treatment years. It seems that in the treatment year 2017, the treatment group changed more than its control

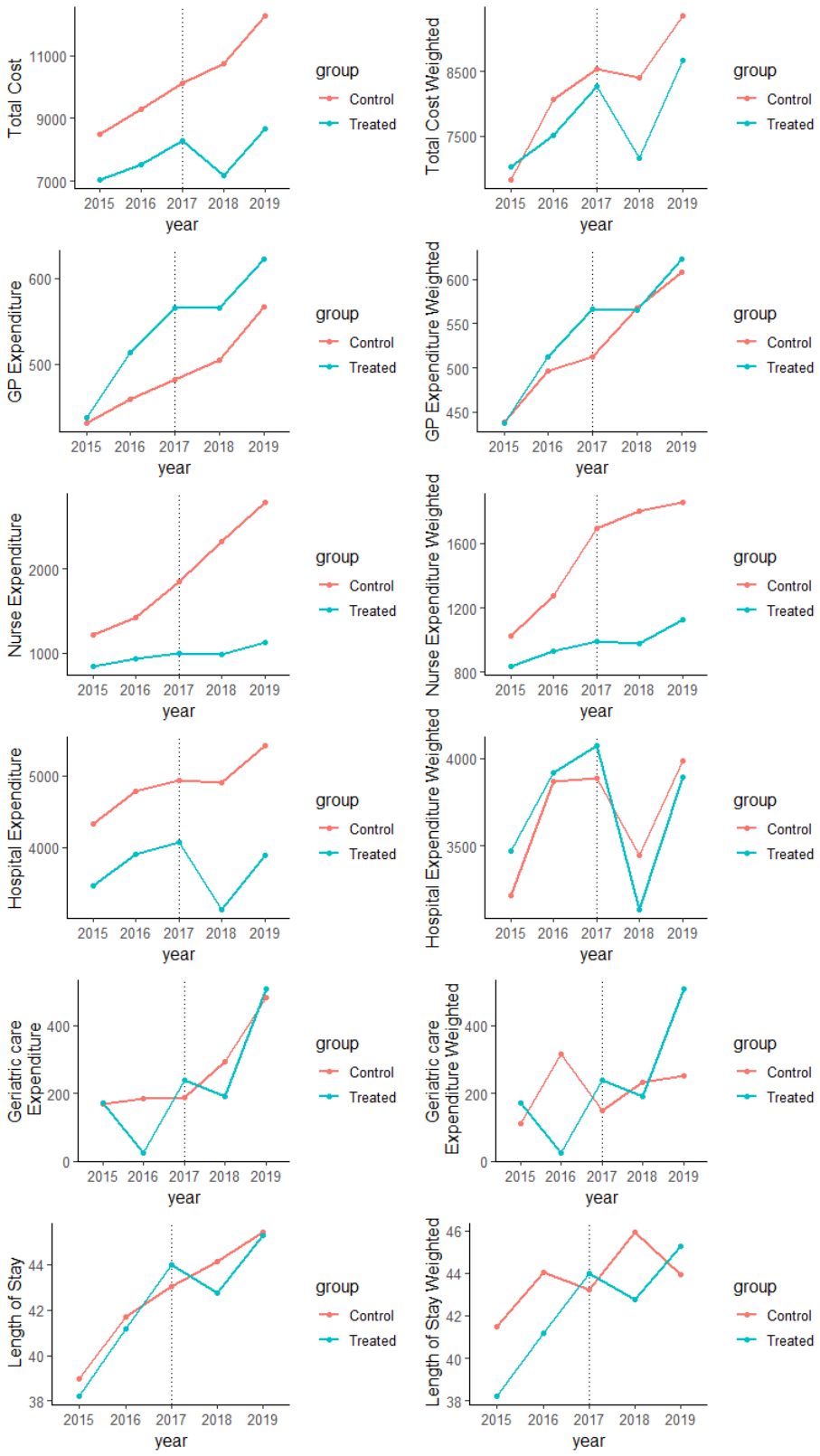


Figure 3: Time Trends between treatment and control for when the intervention took effect in 2017. Expenditures are measured as annual averages in euros per year. *Length of stay* is measured in days per patient per year. *Weighted* on the labels refers to weights assigned to values after the PSM. The vertical line is in the year of the treatment, so the first change at the intervention is more discernible.

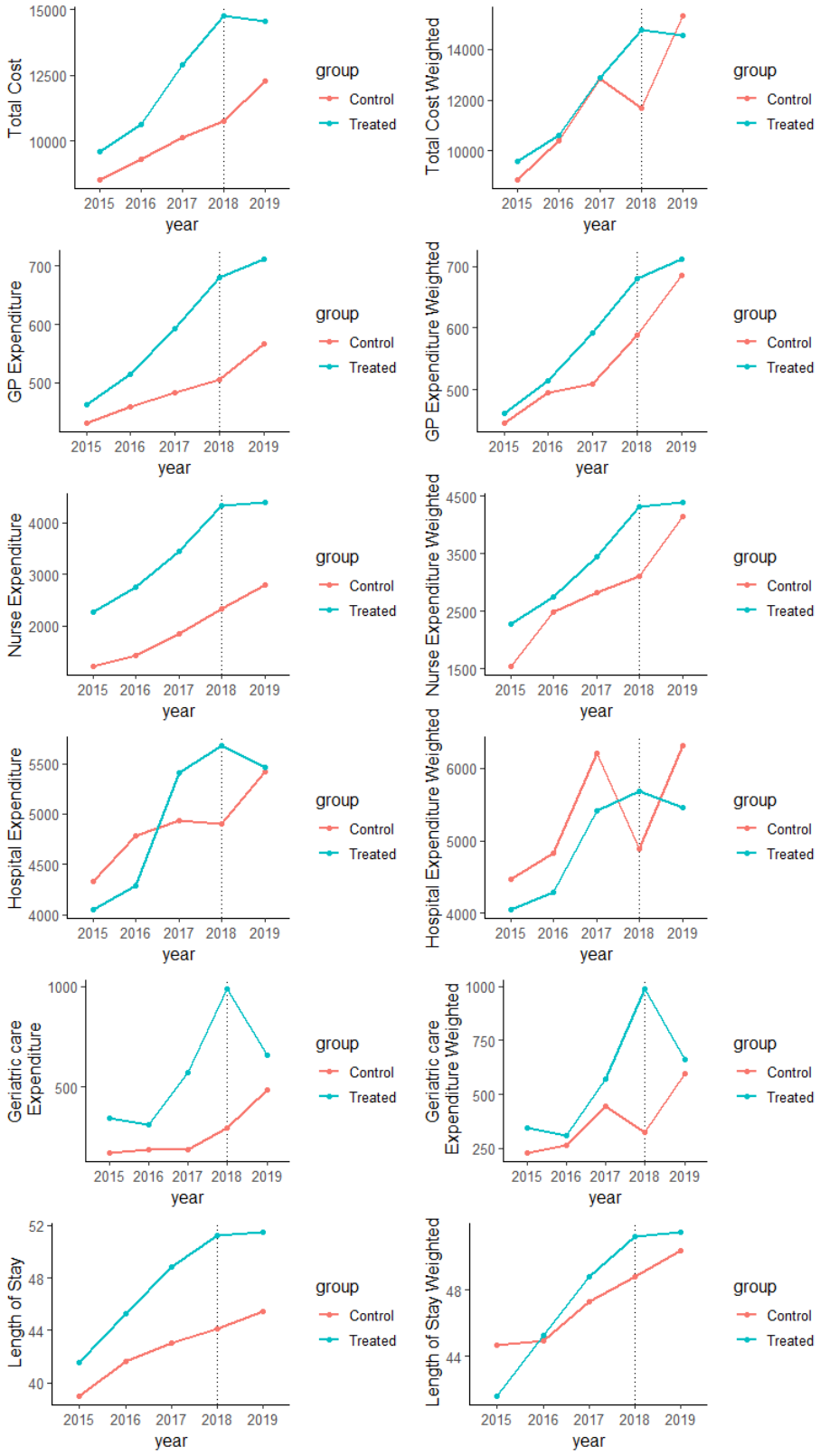


Figure 4: Time Trends between treatment and control for when the intervention takes effect in 2018. Expenditures are measured as annual averages in euros per year. *Length of stay* is measured in days per patient per year. *Weighted* on the labels refers to weights assigned to values after the PSM. The vertical line is in the year of the treatment so that the first change at the intervention is more discernible.

group compared to the treatment and control trends for 2018. This could indicate that the magnitude of the effect in hospital expenditures could be larger in 2017 than in 2018.

Furthermore, the time trend for geriatric care expenditure in both intervention years is not notably parallel in 2017. The expenditure pattern in 2017 seems to move in the opposite direction, and it could still be a valid counterfactual as long as one can assume that this pattern would remain. The time trends in 2018 seem to move more in parallel relative to 2018. So, it is fair to argue that parallel trends exist since the 2017 and 2018 groups are similar in composition. An explanation for the erratic behaviour of the time trend could be due to high variance in the sample of geriatric expenditure since there is generally less use of this type of care.

Finally, the length of stay in the intervention year 2017 moves in parallel for the treatment and control groups. It seems that treated groups tend to stay longer at the hospital than those in the control group. In the intervention year 2018, the trends are similar to those of geriatric expenditure in the 2017 intervention year. The trends move in opposite directions. Once again, it is fair to argue that parallel trends exist since the 2017 and 2018 groups are similar in composition.

In this part, the effectiveness of the PSM and time trends have been analysed. Evidence of an effective matching was evaluated based on histograms of the propensity scores and a balance table of covariates. All outcomes for both intervention years showed evidence of parallel trends. Some outcomes showed uncertainty on the validity as a counterfactual, hospital expenditures in the 2018 intervention year showed some evidence of an anticipation effect, geriatric care, and length of stay had erratic trends. After looking at the descriptive validity of the counterfactual, the DiD results can be discussed.

## 4.4 Difference in Differences

The results for the DiD has estimates for the cohort with the intervention 2017 in table 6 and 2018 is in table 7. All estimates in the forthcoming tables have been estimated using control variables described in the data section in line with equation 5 and 6. The 2017 effect period will have three post periods since the data runs till 2019, and then the 2018 effect period will have two post-treatment periods. The analysis considers both intervention years and compares the Average Treatment Effect on the Treated (ATT) as year-specific estimates obtained from equation 5 and those estimates averaged from equation 6. All estimates below are interpreted assuming ceteris paribus. Also, if estimates are not statistically significant, the magnitude and direction of the effect are still analysed. This is important to assess because there can still be a signal from the data that is impaired by low precision.

I first consider hypothesis 2, which tests whether hospitals' costs have decreased or not. Statistically, it seems that hospital expenditures are unaffected by the implementation of the ECM in both intervention years. In 2017, in the program's first year, the costs increased by 339.78; in the second year, 151.79, and 112.80 euros in the third. The average effect across these years is an increase of 173.70 euros per year. The trend in the estimates suggests the relative cost change reduces every year but the magnitude is not especially large. In the 2018 intervention, in the program's first year, hospital cost increased by 536.18 followed by 582.13 in 2019. The average effect across this years is an increase of 785.45 euros annually. So, for both intervention cohorts, hospitals costs have increased but not statistically significantly; an effect size 785.34 is notable and it seems the large standard errors are driving a lower t-statistic.

The next hypothesis to consider tests whether total costs have decreased or not. In terms of statistical significance, it seems that total costs are unaffected by the implementation of the ECM. In 2017, in the first year of the program, the total costs increased by 510.14; in the

second year, 333.31, and 57.48 euros per person per year. The trend is that the magnitude of the total cost effect reduces every year. In 2018, $ATT_0$ suggests a 1389.22 euro increase in the first year, followed by 364.90 in the second, with an average effect of 1202.34 euro increase. Once again, all estimates have a relatively large standard error, suggesting notable unexplained variation and possibly the driving factor for lack of statistical significance. Overall, it seems that there has been no net as a consequence of the intervention.

The subsequent outcome to analyse is the length of stay, which relates to hypothesis 4. In the intervention year 2017, the days spent at the hospital increased to a maximum of 3.72 days on average in the third period. In the intervention year 2018, the days spent at the hospital increased to a average effect size of 2.51 days in the second period. None of these estimates was statistically significant nor large in magnitude, which suggests that the length of stay at the hospital was unaffected by the ECM.

The final outcome to consider is palliative care, the topic of hypothesis 3. Palliative care outcomes are measured by GP care, nursing home care, and geriatric care. The hypothesis suggests that palliative care will have cost increases but cost reductions can also be plausible if there is a possible substitution effect within the types of care. For example, higher geriatric care and lower of nursing care. In the intervention year 2017, in the first period of the program, GP care increased by 34.84 euros, geriatric care by 380.24 euros, and nursing care decreased by 350.45 euros. The increases in GP and geriatric care were statistically significant, indicating the uptake of care associated with advance care planning. In the following two years, GP expenses decreased by 12.14 euros followed by 4.23 euros, and nursing expenses decrease by 192.29 and 494.87. However, geriatric care increases statistically significantly by 348.65 and 529.57. So, it can be said that there is uptake for palliative care and that there are observations consistent with the notion of substitution where there is substitution away from GP and nursing home care to uptake in geriatric care.

In the intervention year 2018, there is sparse evidence of the substitution effect in palliative care or any uptake in it. This time, nursing care has increased expenditure by 359.49 in the first period of effect and a decrease by 77 euros in the second period. The average effect over the program years also suggests a 302.64 euro per year per patient uptake. There has been a increase in geriatric care and negligible changes in GP care. Geriatric care expenditure on average by 278.01 euros. Hence, there is indication of treatment heterogeneity by cohort because nursing sees an uptake and then a decrease whereas in the 2017 cohort, there are only decreases. However, it is worth noting that none of the estimates are statistically significant.

The main message from the results is that the cost effects of ACP planning are sparse. Figure 5 shows a coefficient plot for the significant estimates of the 2017 intervention cohort. GP care initially increases significantly and falls over the next two periods, although the latter two estimates are not significant. The figure also indicates that geriatric care has evidence of pre-intervention effects. Recall that geriatric care had notably erratic pre-trends. So, the estimates of geriatric care carry some uncertainty given the significant pre-intervention $ATT$ coefficients and the erratic pre-trends. In the 2018 cohort, there were no statistically significant estimates in any of the care groups, which suggests that the group with the intervention did not benefit from the financial efficiency expectation.

| | Intervention year 2017 — DiD with PSM | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-2}$ | −24.55 | 52.87 | 0.07 | 333.05* | 0.15 | 367.33 |
| | (18.31) | (251.96) | (383.10) | (154.68) | (2.34) | (456.53) |
| $ATT_0$ | 34.84* | −350.45 | 339.78 | 380.24* | 3.31 | 510.14 |
| | (16.10) | (241.00) | (475.37) | (156.50) | (2.09) | (518.90) |
| $ATT_1$ | −12.14 | −192.29 | 151.79 | 348.65** | 0.41 | 333.31 |
| | (18.47) | (262.06) | (415.30) | (132.58) | (2.30) | (463.98) |
| $ATT_2$ | −4.23 | −494.87 | 112.80 | 529.57* | 3.72 | 57.48 |
| | (22.52) | (295.46) | (533.51) | (216.32) | (2.40) | (624.99) |
| $ATT_{average}$ | 16.61 | −386.72 | 173.70 | 234.86* | 2.82 | 44.56 |
| | (14.12) | (205.64) | (336.66) | (107.71) | (1.73) | (401.98) |
| Adj. $R^2$ | 0.14 | 0.27 | 0.35 | 0.06 | 0.07 | 0.57 |
| Num. obs. | 83115 | 83115 | 83115 | 83115 | 83115 | 83115 |
| RMSE | 64.85 | 911.22 | 1555.20 | 477.94 | 8.78 | 1789.79 |
| N Clusters | 1813 | 1813 | 1813 | 1813 | 1813 | 1813 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Effect of the intervention in 2017. All variables are in average expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation 6.

| | Intervention year 2018 — DiD with PSM | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-3}$ | −63.61** | −138.37 | −438.44 | −95.34 | −5.18 | −614.75 |
| | (22.61) | (486.20) | (868.44) | (238.69) | (3.10) | (981.05) |
| $ATT_{-2}$ | −51.72** | −385.51 | −255.21 | −126.51 | −0.59 | −458.37 |
| | (19.66) | (395.33) | (733.63) | (245.25) | (2.68) | (913.41) |
| $ATT_0$ | 11.72 | 359.49 | 536.18 | 436.61 | 0.11 | 1389.22 |
| | (22.41) | (441.59) | (931.51) | (266.11) | (3.00) | (947.58) |
| $ATT_1$ | −44.95 | −77.00 | 582.13 | −4.92 | 1.15 | 364.90 |
| | (34.37) | (525.72) | (1036.08) | (279.49) | (3.71) | (1210.17) |
| $ATT_{average}$ | 19.60 | 302.64 | 785.47 | 278.01 | 2.51 | 1202.34 |
| | (21.74) | (377.53) | (688.08) | (186.15) | (2.75) | (755.77) |
| Adj. $R^2$ | 0.15 | 0.36 | 0.26 | 0.07 | 0.09 | 0.54 |
| Num. obs. | 82425 | 82425 | 82425 | 82425 | 82425 | 82425 |
| RMSE | 73.31 | 1157.71 | 1859.33 | 543.65 | 8.29 | 2148.18 |
| N Clusters | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Effect of the intervention in 2018. All variables are in average expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation 6.

## 4.5 Sensitivity, Robustness, Specification

Before discussing the results, it is useful to consider the alternate model specification, relax assumptions, and assess if or to what extent results change. Hence, in this section, I focus on estimates that have a notable change in significance, magnitude, or direction.

First, I consider a log specification of the DiD model to account for a possibly non linear relationship with the outcome and that costs are bound to be positive. The estimates for the 2017 cohort can be found in table 8 and the 2018 cohort in table 9

| | Intervention year 2017 — Log Specification | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-2}$ | $-0.10^{**}$ | 0.47 | 0.02 | $0.38^{*}$ | 0.15 | 0.00 |
| | (0.04) | (0.38) | (0.08) | (0.18) | (2.34) | (0.03) |
| $ATT_0$ | $0.06^{*}$ | 0.07 | 0.10 | 0.41 | 3.31 | 0.01 |
| | (0.03) | (0.44) | (0.07) | (0.21) | (2.09) | (0.03) |
| $ATT_1$ | $-0.03$ | $-0.57$ | 0.15 | 0.37 | 0.41 | 0.01 |
| | (0.03) | (0.46) | (0.08) | (0.21) | (2.30) | (0.03) |
| $ATT_2$ | $-0.03$ | $-1.22^{*}$ | 0.09 | 0.38 | 3.72 | $-0.02$ |
| | (0.04) | (0.50) | (0.08) | (0.21) | (2.40) | (0.03) |
| $ATT_{average}$ | 0.05 | $-0.76^{*}$ | 0.10 | 0.16 | 2.82 | $-0.00$ |
| | (0.03) | (0.34) | (0.05) | (0.13) | (1.73) | (0.02) |
| Adj. $R^2$ | 0.15 | 0.29 | 0.48 | 0.09 | 0.07 | 0.81 |
| Num. obs. | 83115 | 83115 | 83115 | 83115 | 83115 | 83115 |
| RMSE | 0.12 | 1.44 | 0.26 | 0.58 | 8.78 | 0.10 |
| N Clusters | 1813 | 1813 | 1813 | 1813 | 1813 | 1813 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: Effect of the intervention in 2017 with dependent variables that are log-transformed. All variables are in percentage points expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation. 6

Total costs, length of stay, and hospital costs do not show any notable changes relative to the

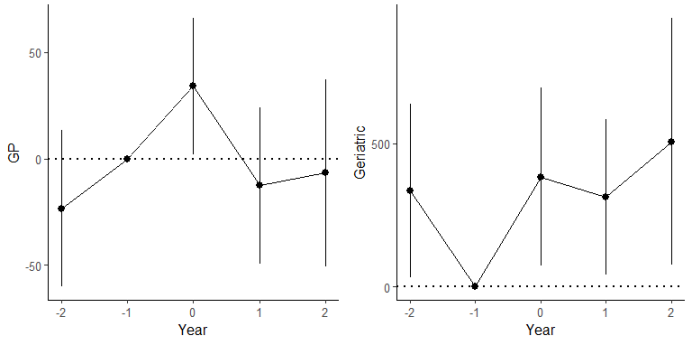


Figure 5: Coefficient plot for the 2017 cohort. The Y axis refers to the estimates in average euros spent per patient. Year -1 refers to the year prior to the intervention which is the reference period.

| | Intervention year 2018 — Log Specification | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-3}$ | $-0.11^{*}$ | $-0.17$ | 0.02 | $-0.01$ | $-5.18$ | 0.05 |
| | (0.05) | (0.62) | (0.12) | (0.33) | (3.10) | (0.05) |
| $ATT_{-2}$ | $-0.10^{**}$ | $-0.43$ | $-0.05$ | $-0.31$ | $-0.59$ | $-0.00$ |
| | (0.04) | (0.53) | (0.09) | (0.32) | (2.68) | (0.04) |
| $ATT_0$ | 0.01 | $4.11^{***}$ | 0.15 | 0.71 | 0.11 | $0.12^{**}$ |
| | (0.04) | (0.66) | (0.10) | (0.38) | (3.00) | (0.04) |
| $ATT_1$ | $-0.06$ | $-1.39^{*}$ | 0.03 | $-0.10$ | 1.15 | 0.00 |
| | (0.04) | (0.63) | (0.11) | (0.39) | (3.71) | (0.05) |
| $ATT_{average}$ | 0.04 | $1.43^{**}$ | 0.10 | 0.39 | 2.51 | 0.04 |
| | (0.03) | (0.48) | (0.08) | (0.25) | (2.75) | (0.03) |
| Adj. $R^2$ | 0.19 | 0.40 | 0.42 | 0.09 | 0.09 | 0.83 |
| Num. obs. | 82425 | 82425 | 82425 | 82425 | 82425 | 82425 |
| RMSE | 0.11 | 1.39 | 0.24 | 0.75 | 8.29 | 0.09 |
| N Clusters | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 9: Effect of the intervention in 2018 with dependent variables that are log transformed. All variables are in percentage point changes per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation 6.

original specification in the 2017 cohort. However, the log specification suggests an effect in the pre-treatment period for both the GP and geriatric care. The graphical evidence of the parallel trends along with the PSM weights make it surprising for the GP outcome. However, the graphs for geriatric care in the 2017 intervention year were erratic and less plausible for a valid counterfactual. This is corroborated by the fact that in the log specification, there is no significance. However, the magnitude is still notably large i.e. 41 percent in the first year 37 percent in the second and 38 percent in the third year. Next, nursing home costs decrease by a statistically significant 122 percent in the first period and 76 percent in the second period of the intervention. For the 2018 cohort, hospital costs, GP costs, and length of stay are not sensitive to the log specification is the same. Nursing costs' effects suggests that there has been an uptake in nursing care with average increasing in costs of 143 percent but the period specific effect show heterogeneity. $ATT_0$ suggests a 411 percent increase but $ATT_0$ suggests 139 percent decrease. Furthermore, total costs in the first period of effect show a 12 percent increase, suggesting that the module is more expensive.

Overall, the changing estimates imply that the estimates ought to be interpreted with uncertainty since they are sensitive to the log specification. This could imply a non-linear relationship between being in the treated and the resulting costs in the true data generating process since logs specification can model non-linearity better. However, the economic rationale for a log specification is unclear. A possible idea could be that the intervention affects the growth of medical expenditure but this argument can not explain the heterogeneity in the nursing cost effects. Hence, it remains a topic for future research to understand the heterogeneity behind care managers' or patients' choices.

Next, I estimate the model without weights. The weights were applied to make the control and treatment groups as similar as possible, making the counterfactual time trend more likely to be valid. However, King and Nielsen (2019) and Smith and Todd (2005) show that PSM in isolation can lead to less efficient estimates. The DiD analysis uses PSM as weights for the weighted least squares, but it is interesting to check to what extent this affected the estimates compared to an ordinary least squares estimation. The estimates are reported in table 10 for the 2017 cohort and table 11 for the the 2018 cohort.

| | Intervention year 2017 — No PSM | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-2}$ | $-46.43^{***}$ | 247.81 | 65.06 | 177.03 | $-0.18$ | 555.47 |
| | (9.12) | (148.86) | (267.41) | (93.13) | (1.29) | (313.71) |
| $ATT_0$ | $32.94^{***}$ | $-247.19$ | 467.08 | $240.93^{*}$ | 2.14 | 646.93 |
| | (9.27) | (124.32) | (329.13) | (95.95) | (1.32) | (377.07) |
| $ATT_1$ | 18.00 | $-449.69^{**}$ | 372.66 | 145.07 | 1.22 | 286.33 |
| | (10.81) | (166.43) | (323.46) | (74.62) | (1.45) | (338.31) |
| $ATT_2$ | 10.75 | $-989.79^{***}$ | 372.96 | 241.89 | 2.22 | $-271.17$ |
| | (13.98) | (189.26) | (384.35) | (172.09) | (1.53) | (436.55) |
| $ATT_{average}$ | 10.96 | $-615.18^{***}$ | 299.47 | $-90.25$ | $-4.22^{*}$ | $-326.98$ |
| | (9.83) | (141.15) | (201.55) | (46.54) | (1.64) | (185.97) |
| Adj. $R^2$ | 0.12 | 0.29 | 0.30 | 0.04 | 0.10 | 0.52 |
| Num. obs. | 83115 | 83115 | 83115 | 83115 | 83115 | 83115 |
| RMSE | 270.39 | 4665.30 | 8530.08 | 2166.65 | 36.26 | 9814.63 |
| N Clusters | 1813 | 1813 | 1813 | 1813 | 1813 | 1813 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 10: Effect of the intervention in 2017 that is estimated using ordinary least squares. All variables are expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation 6.

The general observation for these estimates is that absolute values for the estimates tend to increase, and the standard errors are smaller. I will highlight some of the notable changes. The pre-trends for the GP are significant for both cohorts, which gives some evidence that the matching was successful since this effect disappears with weights in the 2017 cohort in 6. Nursing care shows a larger value of decreasing and is statistically significant in period-specific and on average. Geriatric care is only statistically significant in the first period of the effect at $ATT_0$.

Furthermore, LOS becomes significant with an average decrease of 4.22 days. The rest of the estimates share a trend of larger absolute values and smaller estimates except for geriatric care. The consistently higher absolute values of the estimates and lower standard errors without the weighting could be a case of a bias-variance trade-off. The matching leads to less biased estimates since it enables a fairer comparison between treatment and control groups but at the cost of efficiency, which is partly in line with King and Nielsen (2019).

In the main models of the research, the treatment timing was considered important due to possible treatment heterogeneity by year and cohort. That was why the cohorts of 2017 and 2018 were analysed separately. However, it is also possible to pool these estimates into a single model in the generalised difference-in-differences design by standardising time. The functional form of this DiD specification is the same as in equation 5. Time is standardised

| | Intervention year 2018 — No PSM | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-3}$ | −66.23*** | −42.16 | −234.60 | −154.09 | −2.00 | −379.74 |
| | (14.60) | (328.82) | (643.11) | (185.39) | (1.76) | (735.98) |
| $ATT_{-2}$ | −44.96*** | −88.34 | −760.61 | −242.63 | −1.47 | −926.36 |
| | (12.48) | (239.83) | (497.29) | (197.66) | (1.88) | (685.56) |
| $ATT_0$ | 57.45*** | 110.68 | −129.13 | 273.15 | 0.21 | 299.28 |
| | (13.78) | (203.08) | (654.84) | (219.78) | (2.00) | (677.61) |
| $ATT_1$ | 37.02 | 33.51 | 175.74 | −177.58 | 0.82 | 237.10 |
| | (21.28) | (357.51) | (797.93) | (212.85) | (2.57) | (927.74) |
| $ATT_{average}$ | 84.28*** | 115.59 | 354.89 | 180.04 | 1.67 | 703.41 |
| | (13.45) | (260.46) | (544.30) | (157.59) | (2.14) | (597.26) |
| Adj. $R^2$ | 0.12 | 0.30 | 0.29 | 0.05 | 0.10 | 0.52 |
| Num. obs. | 82425 | 82425 | 82425 | 82425 | 82425 | 82425 |
| RMSE | 273.36 | 4714.61 | 8602.54 | 2181.29 | 36.38 | 9875.96 |
| N Clusters | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 11: Effect of the intervention in 2018 that is estimated using ordinary least squares. All variables are expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation 6.

by subtracting the treatment year from the time variables and then appending the two separate panel data. Consequently, there is additional statistical power since the number of observations is increasing.

Generalised difference-in-differences is a relatively new area of research (Callaway and Sant'Anna, 2021) and relies additionally on constant treatment effects, i.e. for each cohort, the treatment effect must have the same magnitude. The estimates at the ends of the observation window consist of one less cohort of observations. The 2017 cohort would have an additional year where effects are assessed, whereas the 2018 cohort would have an additional year of pre-trend. Also, it is unclear whether propensity score matching is plausible in the generalised setup because it is still being researched Callaway and Li (2021) and Athey and Imbens (2022). Hence, the propensity score matching and weighting is left out for simplicity. The estimates for Standardised Time Generalised DiD can be found in table 12.

| | Standardised Time Generalised DiD | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-3}$ | $-47.78^{**}$ | $624.50^{*}$ | $-450.16$ | 80.27 | $-0.05$ | 329.06 |
| | (14.16) | (245.74) | (451.13) | (134.78) | (1.78) | (545.31) |
| $ATT_{-2}$ | $-44.98^{***}$ | 122.16 | $-316.22$ | $-5.38$ | $-0.71$ | $-72.90$ |
| | (7.61) | (138.93) | (244.40) | (97.10) | (1.11) | (338.35) |
| $ATT_0$ | $42.42^{***}$ | $-100.14$ | 213.77 | $251.54^{*}$ | 1.16 | 487.93 |
| | (7.77) | (113.81) | (320.93) | (107.14) | (1.10) | (346.42) |
| $ATT_1$ | $24.29^{*}$ | $-230.06$ | 248.63 | $-4.87$ | 1.18 | 234.42 |
| | (11.62) | (181.45) | (371.34) | (100.40) | (1.29) | (441.38) |
| $ATT_2$ | $-3.06$ | $-1488.06^{***}$ | 559.26 | 70.20 | 0.60 | $-773.37$ |
| | (15.30) | (235.90) | (395.78) | (180.42) | (1.76) | (501.14) |
| $ATT_{average}$ | $51.70^{***}$ | $-587.25^{***}$ | 468.47 | 85.84 | 1.51 | 74.36 |
| | (8.36) | (139.14) | (260.90) | (81.31) | (1.12) | (300.42) |
| Adj. $R^2$ | 0.11 | 0.29 | 0.29 | 0.05 | 0.10 | 0.52 |
| Num. obs. | 165540 | 165540 | 165540 | 165540 | 165540 | 165540 |
| RMSE | 273.13 | 4700.57 | 8571.88 | 2174.29 | 36.39 | 9852.13 |
| N Clusters | 1820 | 1820 | 1820 | 1820 | 1820 | 1820 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 12: Generalised difference in differences in a standardised time framework estimated using ordinary least squares. Effect of the intervention in 2017 that is estimated using ordinary least squares. All variables are expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period. $ATT_{average}$ is estimated using equation 6.

There are heterogeneous treatment effects by year. In the first and second years of effect, GP and Geriatric care have a statistically significant change. Geriatric care sees an increase in the first year, whereas GP care has an uptake at $ATT_0$ and $ATT_1$. At $ATT_2$, GP care decreases, nursing care decreases significantly, whereas hospital consumption continues to increase. However, both GP and nursing care show evidence of pre-intervention effects. Also, nursing care has decreases in costs by 587.25 euros per year. Overall, the generalised DiD shows more evidence of various heterogeneous effects across the outcomes by year after assuming the treatments occur at the same time.

Some other more nuanced details of the analysis need to be noted. The first point relates to the treatment of missing values in the length of stay variable. A missing value in length of stay could either be an individual who has not been to a hospital or a truly missing value. Changes in this variable do not make a difference since there are few missing values. This is to be expected since the data set comprises frail elderly who are likely to visit the hospital. Also, the first intervention year of the ECM was 2016; however, this year is not analysed because insufficient pre-trend data is available to make any meaningful analysis. Bertrand et al. (2004) note that block bootstrapped stand errors are likely to be necessary if there are an insufficient number of clusters due to the risk of serial correlation. No exact number has been noted as sufficient (Angrist and Pischke, 2009). However, the DiD standard errors consist of more than 1800 clusters, a large number of clusters sufficient to manage serial correlation.

# 5 Assumptions and Limitations

In this section, I will describe the assumptions as well as the limitations of the study. I consider the assumptions critical to identifying the estimates from Lechner et al. (2011). First, I will justify the Stable Unit Value Treatment Assumption (SUTVA). The assumption states that only the treatment group is affected by the intervention, and consequently, there are no spillover effects. The assumption is fairly satisfied by the design of the program. The benefits of enrolling with a GP are individual-specific, and the resulting cost benefits are unlikely to spill over to individuals without it. Next, the exogeniety assumption states that the onset of the treatment does not influence the covariates. Exogeniety is critical to ensure that the estimates are not biased, for example, including variable as control that is also simultaneously an outcome. Age and gender are not influenced by the treatment and can be assumed to be exogenous. The severity of illness dummies based on the past year's total cost can be endogenous. In the years after the intervention, an individual could shift the severity of illness categories because of being with a treated GP. However, I expect this variation to be negligible since total costs were not affected on average. Furthermore, ACP is not expected to improve patients' health but rather allocate resources more effectively and satisfy non-financial goals so the severity of illness is not a confounder in this context.

Next, the analysis assumes that there are no effects pre-treatment (NEPT). In this case, GPs would be conducting ACP activities even before receiving the funding. This is not plausible because GPs require the funding and incentive to do the funding (Bekker et al. (2021) and Marckmann et al. (2013)) and there is no rationale for pre-treatment activities. Hence, it would be fair to conclude that NEPT holds. The final assumption is parallel trends. This has been elaborated on extensively in the methods section and the results. It would be fair to say this assumption holds because the comparable units have been matched using PSM, and any remaining observable differences have been modelled in the regression such that the assumption should hold conditional on the covariates.

Now, the limitations of the study need to be discussed. The first and most important limitation follows from the definition of frail elderly. To understand this, one must consider the GP's perspective. A GP must try and identify frail elderly within his care group, and this is done based on the GPs intuition and past medicinal consumption of the patient. These aspects are not fully observed in the data, so some individuals cannot be considered frail elderly even if they have multiple illnesses, and this can be worrisome if GP care groups are composed of such clientele. This is because there is no guarantee that GP will provide the ACP protocol to these people, but according to the data, they should be getting the treatment. This means that frail elderly in my treatment group might not even be exposed to the treatment if they do not correspond with their GP's subjective view. The treated group not getting the treatment is paradoxical and is likely to bias the absolute values of the estimates downwards since they do not receive the treatment despite being in 'treatment' group.

A second limitation relates to unobservable effects as a consequence of ACP. The review of Bom et al. (2019) find evidence of a negative impact on the mental and physical health of the informal caregiver. ACP is likely to put an individual at home, and despite planning, there is still likely some level of informal care, and this effect can not be observed. Health effects are difficult to quantify in terms of costs, but it would reduce the benefits obtained in a non-monetary sense.

The third limitations refer to how well cost can serve as a measure of allocative efficiency. Costs are composed of price and volume. Hence, costs can be driven down by moving individuals to lower-priced care or giving less treatment. Hence, it is important to disentangle costs into a price and volume effect by imputing prices of services at their median level (Hayen, 2018). This data was not accessible; this can serve as an addition to future research.

# 6 Discussion and Conclusion

Family and friends of frail elderly aim to help them age at home rather than in a nursing home to prevent unnecessary costs and emotional stress. A possible way to achieve society's goal is by Advance Care Planning. Insurers, patients, and providers want to ensure that that the frail elderly have ACP and also like to identify the financial efficiency of the module. But do frail elderly enter a ACP program? and is there a causal effect of a capitated fee for Advance Care Planning on the healthcare expenditures of the elderly aged above 65 from 2015 to 2019? This research elucidates these questions using a binary outcome model and a quasi-experimental approach.

A logit model was used to estimate the probability of entering the treatment group conditional on the characteristics of a frail elderly. For the causal effect on costs, I exploit the timing of the ECM implementation where a certain group of GPs were selected to receive the treatment based on eligibility criteria. Hence, I implement a difference in differences design along with propensity score matching, which enables closer comparison between treatment and control in a quasi-experimental manner.

The evidence suggests that the 2017 treated cohort has more chronic illnesses, but illnesses are less disabling, whereas the 2018 cohort has fewer illnesses but more disabling. It can be said that few severe diseases and many diseases with less severity may have similar impacts on the overall quality of life. Hence, there is evidence for frail elderly being more likely to be in treated group, which implies that there is evidence of the familial goal of the ECM being satisfied and the first hypothesis.

The causal effect has multiple aspects to the evidence. There is mixed evidence of hospital costs increasing. In 2017, hospitals costs are up and in 2018, they are down, but neither is statistically significant. This is plausible because patients can still need hospital care if agreed upon by the care manager, although it is unexpected. Next, there is stronger evidence of palliative care taking place, and there is evidence of substitution within services along with heterogeneity in service choice by cohort. For the 2017 cohort, there is an uptake in geriatric care and a decrease in GP and nursing care. In the 2018 cohort, there is some uptake in nursing care but a decrease in geriatric care and a slight decrease in GP care. However, none of these estimates is statistically significant nor large in magnitude. It is worth noting that these substitution effects are implied and not directly observed. However, from a logical standpoint, it does make sense that the care manager would prioritise a certain type of care conditional on the properties of the care service and the characteristics of the patients. This in in turn would lead to substitution within the umbrella of palliative care service. A more concerning explanation for the lack of effects in the 2018 cohort could be that the GPs found the fee insufficient and applied the finances from the ECM in other parts of their practice.

Also, length of stay at the hospital is not affected significantly by participating in the ECM in either cohort. This implies, on average, that care managers do not plan for patients to have shorter stays at hospitals but rather allow the required amount of treatment that could be necessary. Finally, total costs are unaffected in both cohorts. Since there is no impact on total cost and there is evidence of substitution, it could be that there is a high volume of cheaper care that is occurring (mainly in the 2017 cohort), and this is beneficial from a patient's perspective because palliative care is generally less invasive than hospital care. The result of total cost also suggests that care providers and payers are not worse off and the patients could be better off.

The main contribution of this study is that ACP has much less of a cost-benefit compared to other studies. Studies such as those of O'Sullivan et al. (2016), and K.-H. Nguyen et al. (2017) find large savings of ACP programs. These studies do not account for any counterfactual, so its estimates can be dubious. This paper accounts for the counterfactual in a DiD framework with PSM to find sparse cost benefits of the ECM and its associated ACP activities. However, there is evidence that the frail elderly are with GPs who have the necessary funding to conduct ACP-related activities. For future research, I suggest exploring this topic with a similar methodological framework with additional data on quality of life and the price of care providers. I also recommend researching GPs' or care managers' treatment preferences based on patient information so that the estimated effects are more intuitive to interpret and that policymakers can identify where resources are being consumed.

Overall to answer the research question, the evidence suggests that frail elderly are likely to be with GPs who have the ECM. However, the evidence is limited on the financial effectiveness of the module. Hence, the recommendation is that the module is reevaluated regarding GP activities related to ACP. As of now, the GPs can identify frail elderly to a certain extent; however, the cost effects of the ECM are not entirely clear, and it would be fruitful to identify what activities are being undertaken and where or if the capitated fee is being used.

# References

Abarshi, E., Onwuteaka-Philipsen, B., Donker, G., Echteld, M., Van den Block, L., & Deliens, L. (2009). General practitioner awareness of preferred place of death and correlates of dying in a preferred place: A nationwide mortality follow-back study in the netherlands. *Journal of pain and symptom management*, *38*(4), 568–577.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, *226*(1), 62–79.

Bekker, Y. A., Suntjens, A., Engels, Y., Schers, H., Westert, G. P., & Groenewoud, A. S. (2021). Advance care planning in primary care. a retrospective medical record study among patients with different illness trajectories.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Bischoff, K. E., Sudore, R., Miao, Y., Boscardin, W. J., & Smith, A. K. (2013). Advance care planning and the quality of end-of-life care in older adults. *Journal of the American Geriatrics Society*, *61*(2), 209–214.

Bom, J., Bakx, P., Schut, F., & Van Doorslaer, E. (2019). The impact of informal caregiving for older adults on the health of various types of caregivers: A systematic review. *The Gerontologist*, *59*(5), e629–e642.

Brinkman-Stoppelenburg, A., Rietjens, J. A., & Van der Heide, A. (2014). The effects of advance care planning on end-of-life care: A systematic review. *Palliative medicine*, *28*(8), 1000–1025.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31–72.

Callaway, B., & Li, T. (2021). Supplementary appendix: Policy evaluation during a pandemic.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Campbell, S. M. (2020). Well-being and the good death. *Ethical Theory and Moral Practice*, *23*(3), 607–623.

Coburn, A. F., Fortinsky, R., McGuire, C., & McDonald, T. P. (1993). Effect of prospective reimbursement on nursing home costs. *Health Services Research*, *28*(1), 45.

Coelen, C., & Sullivan, D. (1981). An analysis of the effects of prospective reimbursement programs on hospital expenditures. *Health Care Financing Review*, *2*(3), 1.

de Meijer, C., Wouterse, B., Polder, J., & Koopmanschap, M. (2013). The effect of population aging on health expenditure growth: A critical review. *European journal of ageing*, *10*(4), 353–361.

de Roo, M. L., Miccinesi, G., Onwuteaka-Philipsen, B. D., Van Den Noortgate, N., Van den Block, L., Bonacchi, A., Donker, G. A., Lozano Alonso, J. E., Moreels, S., Deliens, L., et al. (2014). Actual and preferred place of death of home-dwelling patients in four european countries: Making sense of quality indicators. *PloS one*, *9*(4), e93762.

de Vleminck, A., Houttekier, D., Pardon, K., Deschepper, R., Van Audenhove, C., Vander Stichele, R., & Deliens, L. (2013). Barriers and facilitators for general practitioners to engage in advance care planning: A systematic review. *Scandinavian journal of primary health care*, *31*(4), 215–226.

Dixon, J., Matosevic, T., & Knapp, M. (2015). The economic evidence for advance care planning: Systematic review of evidence. *Palliative medicine*, *29*(10), 869–884.

Dobrez, D., Heinemann, A. W., Deutsch, A., Manheim, L., & Mallinson, T. (2010). Impact of medicare's prospective payment system for inpatient rehabilitation facilities on stroke patient outcomes. *American journal of physical medicine & rehabilitation*, *89*(3), 198–204.

Ellis, R., & McGuire, T. G. (1986). Provider behaviour under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, *5*(2), 129–151. https://EconPapers.repec.org/RePEc:eee:jhecon:v:5:y:1986:i:2:p:129-151

Ellis, R. P. (1998). Creaming, skimping and dumping: Provider competition on the intensive and extensive margins. *Journal of health economics*, *17*(5), 537–555.

French, E. B., McCauley, J., Aragon, M., Bakx, P., Chalkley, M., Chen, S. H., Christensen, B. J., Chuang, H., Côté-Sergent, A., de Nardi, M., et al. (2017). End-of-life medical spending in last twelve months of life is lower than previously reported. *Health Affairs*, *36*(7), 1211–1217.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American journal of epidemiology*, *173*(7), 761–767.

Gabbard, J., Pajewski, N. M., Callahan, K. E., Dharod, A., Foley, K. L., Ferris, K., Moses, A., Willard, J., & Williamson, J. D. (2021). Effectiveness of a nurse-led multidisciplinary intervention vs usual care on advance care planning for vulnerable older adults in an accountable care organization: A randomized clinical trial. *JAMA Internal Medicine*, *181*(3), 361–369.

Godager, G., & Wiesen, D. (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of health economics*, *32*(6), 1105–1116.

Gomes, B., Higginson, I. J., Calanzani, N., Cohen, J., Deliens, L., Daveson, B., Bechinger-English, D., Bausewein, C., Ferreira, P., Toscani, F., et al. (2012). Preferences for place of death if faced with advanced cancer: A population survey in england, flanders, germany, italy, the netherlands, portugal and spain. *Annals of oncology*, *23*(8), 2006–2015.

Graafland, J. J. (2009). Utilitarianism. *Handbook of economics and ethics*. Edward Elgar Publishing.

Hayen, A. (2018). *Shared savings and patient cost sharing in the dutch health care system* (Doctoral dissertation). Ipskamp.

Hayen, A., van den Berg, M. J., Struijs, J. N., & Westert, G. P. (2021). Dutch shared savings program targeted at primary care: Reduced expenditures in its first year. *Health Policy*, *125*(4), 489–494.

Heckman, J. J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator. *The review of economic studies*, *65*(2), 261–294.

Higginson, I. J., Finlay, I., Goodwin, D. M., Cook, A. M., Hood, K., Edwards, A. G., Douglas, H.-R., & Norman, C. E. (2002). Do hospital-based palliative teams improve care for patients or families at the end of life? *Journal of pain and symptom management*, *23*(2), 96–106.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. https://doi.org/10.18637/jss.v042.i08

Jimenez, G., Tan, W. S., Virk, A. K., Low, C. K., Car, J., & Ho, A. H. Y. (2018). Overview of systematic reviews of advance care planning: Summary of evidence and global lessons. *Journal of Pain and Symptom Management*, *56*(3), 436–459.

Khandelwal, N., Kross, E. K., Engelberg, R. A., Coe, N. B., Long, A. C., & Curtis, J. R. (2015). Estimating the effect of palliative care interventions and advance care planning on icu utilization: A systematic review. *Critical care medicine*, *43*(5), 1102.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454.

Klingler, C., in der Schmitten, J., & Marckmann, G. (2016). Does facilitated advance care planning reduce the costs of care near the end of life? systematic review and ethical considerations. *Palliative Medicine*, *30*(5), 423–433.

Lechner, M. et al. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, *4*(3), 165–224.

Lehnert, T., Heider, D., Leicht, H., Heinrich, S., Corrieri, S., Luppa, M., Riedel-Heller, S., & König, H.-H. (2011). Health care utilization and costs of elderly persons with multiple chronic conditions. *Medical Care Research and Review*, *68*(4), 387–420.

Lubitz, J., Cai, L., Kramarow, E., & Lentzner, H. (2003). Health, life expectancy, and health care spending among the elderly. *New England Journal of Medicine*, *349*(11), 1048–1055.

Marckmann, G. et al. (2013). The economics of advance care planning: Empirical data and ethical implications. *BMJ Supportive & Palliative Care*, *3*(2), 241–241.

Melberg, H. O., Godager, G., & Gregersen, F. A. (2013). Hospital expenses towards the end of life. *Tidsskrift for Den norske legeforening*.

Menzis. (2017). Bijlage 9 zorginkoopbeleid 2018-2019 huisartsenzorg en multidisciplinaire zorg.

Nguyen, K.-H., Sellars, M., Agar, M., Kurrle, S., Kelly, A., & Comans, T. (2017). An economic model of advance care planning in australia: A cost-effective way to respect patient choice. *BMC health services research*, *17*(1), 1–8.

Nguyen, T.-L., Collins, G. S., Spence, J., Daurès, J.-P., Devereaux, P., Landais, P., & Le Manach, Y. (2017). Double-adjustment in propensity score matching analysis: Choosing a threshold for considering residual imbalance. *BMC medical research methodology*, *17*(1), 1–8.

Norwood, F., Kimsma, G., & Battin, M. P. (2009). Vulnerability and the 'slippery slope'at the end-of-life: A qualitative study of euthanasia, general practice and home death in the netherlands. *Family practice*, *26*(6), 472–480.

O'Hanlon, C. E., Walling, A. M., Okeke, E., Stevenson, S., & Wenger, N. S. (2018). A framework to guide economic analysis of advance care planning. *Journal of Palliative Medicine*, *21*(10), 1480–1485.

O'Sullivan, R., Murphy, A., O'Caoimh, R., Cornally, N., Svendrovski, A., Daly, B., Fizgerald, C., Twomey, C., McGlade, C., & Molloy, D. W. (2016). Economic (gross cost) analysis of systematically implementing a programme of advance care planning in three irish nursing homes. *BMC research notes*, *9*(1), 1–10.

Parry, J. K., & Munson, C. (2015). *Negotiating a good death: Euthanasia in the netherlands*. Routledge.

Polder, J. J., Barendregt, J. J., & van Oers, H. (2006). Health care costs in the last year of life—the dutch experience. *Social science & medicine*, *63*(7), 1720–1731.

Reif, S., Hafner, L., & Seebauer, M. (2020). Physician behavior under prospective payment schemes—evidence from artefactual field and lab experiments. *International Journal of Environmental Research and Public Health*, *17*(15), 5540.

RIVM. (2016). Ouderen van nu en straks: Zijn er verschillen in kwetsbaarheid?

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 159–183.

Schäfer, W., Kroneman, M., Boerma, W., Van den Berg, M., Westert, G., Devillé, W., Van Ginneken, E., Organization, W. H., et al. (2010). The netherlands: Health system review.

Schut, F. T., & Varkevisser, M. (2017). Competition policy for health care provision in the netherlands. *Health Policy*, *121*(2), 126–133.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of econometrics*, *125*(1-2), 305–353.

Sood, N., Buntin, M. B., & Escarce, J. J. (2008). Does how much and how you pay matter? evidence from the inpatient rehabilitation care prospective payment system. *Journal of Health Economics*, *27*(4), 1046–1059.

Stammann, A., Heiß, F., & McFadden, D. (2016). Estimating fixed effects logit models with large panel data. (G01-V3). http://hdl.handle.net/10419/145837

Stooker, T., Van Acht, J. W., Van Barneveld, E. M., Van Vliet, R. C., van Hout, B. A., Hessing, D. J., & Busschbach, J. J. (2001). Costs in the last year of life in the netherlands. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *38*(1), 73–80.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Thomas, D. R., & Davis, K. M. (1987). Physician awareness of cost under prospective reimbursement systems. *Medical Care*, *25*(3), 181–184. Retrieved June 27, 2022, from http://www.jstor.org/stable/3765247

Waldron, I., & Johnston, S. (1976). Why do women live longer than men? *Journal of human stress*, *2*(2), 19–30.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Zorginstituut-Nederland. (2017). *Fkg's vereveningsmodel*. https://www.zorginstituutnederland.nl/publicaties/publicatie/2017/10/09/referentiebestand-fkg%E2%80%99s-2018

# 7 Appendix

## 7.1 Goals of the Program

The first goal is that the frail elderly should have access to the activities in the elderly care module. The justification of this goal can be understood in two ways: constrained utility optimisation from economics and moral philosophy. Constrained utility optimisation shows that welfare can be increased by aligning patients' preferences with their medical consumption. Suppose ACP allows patients' preferences to be adequately accounted for. In that case, this leads to welfare benefits to society since the utilitarian view of economics implies that it is necessary to maximise the sum of individual welfare to make society better off (Graafland, 2009).

The utilitarian theory of morals is often criticised for not being able to capture more 'human' aspects of a policy (Graafland, 2009). The notion of a good death from ethical theory gives another justification for the program's first goal. In ethical theory literature, there is a notion of the 'good death' where a good death constitutes a preexisting normative view of one own death (Campbell, 2020). From the perspective of a deontologist - the idea that actions are good or bad according to a clear set of rules - Dutch society has a duty to enable people to die at home (or place of preference) to experience a good death. The ECM can allow a patient's death to be aligned with their preferences and, in turn, allow a good death. Overall, the frail elderly need to be in the ECM to maximise societal welfare and enable society's duty to allow an individual to experience a good death.

## 7.2 Histograms for Propensity Score Matching

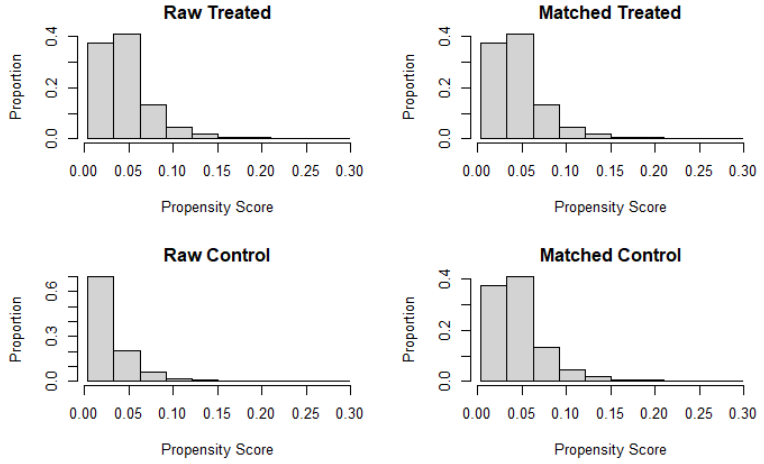


Figure 6: Histogram for the distribution of the propensity scores before and after a 1:1 nearest neighbour matching without replacement algorithm based on the propensity scores for intervention year 2017.

The treated group has some observations in the region .3 and greater, which is not present in the control group. This means that those with a high probability of treatment conditional on their characteristics will not have the strongest match. It is hard to interpret this because a propensity score is a reduced dimension of multiple covariates.

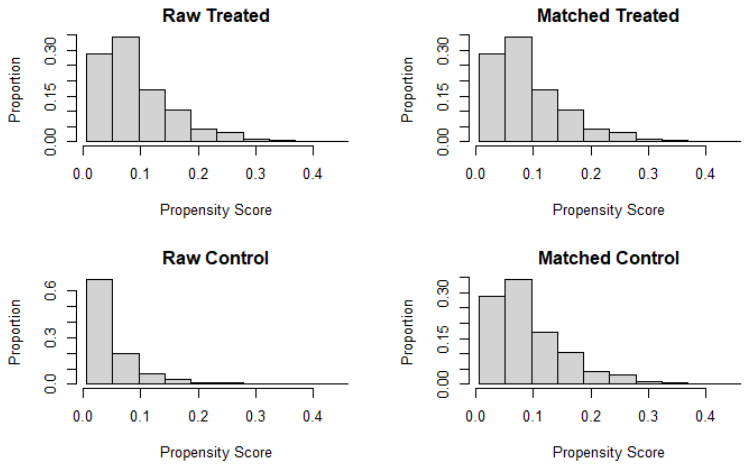


Figure 7: Histogram for the distribution of the propensity scores before and after a 1:1 nearest neighbour matching without replacement algorithm based on the propensity scores for intervention year 2018.

## 7.3 Year Specific Summary Statistics

| Variable | Treated | Control | T-statistic |
|---|---|---|---|
| Number of Observations | 2535 | 80580 | |
| Number of Patients | 507 | 16116 | |
| Number of GPs | 122 | 1803 | |
| Age | 80.97 (5.14) | 75.84 (6.72) | 38.1*** |
| Number of Chronic Illnesses | 10.17 (8.14) | 10.18 (8.12) | 2.03* |
| Male | 0.41 | 0.43 | 2.2* |
| Severity of illness based on DKG | | | |
| Low Cost | 0.87 | 0.86 | 4.4*** |
| Medium Cost | 0.12 | 0.11 | 2.8** |
| High Cost | 0.01 | 0.03 | 4.2 *** |
| Severity of illness based on past Total Costs | | | |
| Low Cost | 0.28 | 0.25 | 3.1** |
| Medium Cost | 0.50 | 0.55 | 5.6 *** |
| High Cost | 0.22 | 0.25 | 9.59 *** |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 13: Descriptive Statistics intervention year 2017. The stars refer to statistical significance of a t-test of the mean differences.

| **Variable** | **Treated** | **Control** | **T-statistic** |
|---|---|---|---|
| Number of Observations | 1845 | 80580 | |
| Number of Patients | 369 | 16116 | |
| Number of GPs | 187 | 1803 | |
| Age | 80.51 | 75.84 | 29.59 *** |
| Number of Chronic Illnesses | 10.19 | 10.18 (8.12) | 0.49 |
| Male | 0.39 | 0.43 | 3.3 *** |
| Severity based on DKG | | | |
| Low Cost | 0.85 | 0.86 | 3.1 *** |
| Medium Cost | 0.14 | 0.13 | 2.9 *** |
| High Cost | 0.01 | 0.01 | .004 |
| Severity based on past Total Costs | | | |
| Low Cost | 0.19 | 0.25 | 6.5** |
| Medium Cost | 0.48 | 0.50 | 0.96 *** |
| High Cost | 0.33 | 0.25 | 7.65 *** |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 14: Descriptive Statistics intervention year 2018. *SE* refers to the standard error of the mean difference. The stars refer to statistical significance of a t-test of the mean differences. Absolute values of the t statistic are reported.

## 7.4 Full Regression Tables

| | Intervention year 2017 — Did with PSM | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-2}$ | −24.55 | 52.87 | 0.07 | 333.05* | 0.15 | 367.33 |
| | (18.31) | (251.96) | (383.10) | (154.68) | (2.34) | (456.53) |
| $ATT_0$ | 34.84* | −350.45 | 339.78 | 380.24* | 3.31 | 510.14 |
| | (16.10) | (241.00) | (475.37) | (156.50) | (2.09) | (518.90) |
| $ATT_1$ | −12.14 | −192.29 | 151.79 | 348.65** | 0.41 | 333.31 |
| | (18.47) | (262.06) | (415.30) | (132.58) | (2.30) | (463.98) |
| $ATT_2$ | −4.23 | −494.87 | 112.80 | 529.57* | 3.72 | 57.48 |
| | (22.52) | (295.46) | (533.51) | (216.32) | (2.40) | (624.99) |
| $Group_{treated}$ | 18.83 | −346.85 | −50.84 | −293.40** | −2.62 | −661.54* |
| | (18.21) | (213.76) | (310.34) | (109.04) | (2.38) | (307.86) |
| $Year_{-2}$ | 368.17*** | 2386.03 | −4197.83 | −912.03 | 39.30*** | 3558.80 |
| | (86.48) | (1210.38) | (2219.00) | (586.27) | (10.90) | (2823.50) |
| $Year_0$ | 13.69 | 348.29 | −8.98 | −183.53 | −1.13 | 351.31 |
| | (13.05) | (206.06) | (367.07) | (126.01) | (1.59) | (386.70) |
| $Year_1$ | 551.43*** | 3037.00* | −5057.66 | −1143.16 | 50.28*** | 4670.87 |
| | (100.36) | (1461.35) | (2769.49) | (707.39) | (12.85) | (3541.56) |
| $Year_2$ | 580.98*** | 3047.80* | −4739.90 | −1091.53 | 47.06*** | 5209.37 |
| | (98.16) | (1448.78) | (2893.11) | (732.32) | (12.67) | (3664.74) |
| (Intercept) | 1405.67 | 1127.35 | 17119.42 | 6421.82 | 61.04 | 37117.29 |
| | (1185.05) | (16808.18) | (17933.81) | (6272.40) | (139.37) | (23411.78) |
| Age | −42.90 | −182.54 | −132.64 | −150.61 | −2.71 | −953.76 |
| | (28.91) | (397.47) | (449.72) | (153.80) | (3.48) | (587.97) |
| $Age^2$ | 0.29 | 1.65 | −0.12 | 1.07 | 0.02 | 5.57 |
| | (0.18) | (2.44) | (2.69) | (0.95) | (0.02) | (3.55) |
| Male | −14.72 | −768.33*** | 1201.27*** | −94.37 | 0.71 | 232.93 |
| | (11.74) | (139.26) | (234.82) | (56.12) | (1.77) | (240.81) |
| $Medium_{DKG}$ | −3.98 | 395.71 | 329.66 | −107.87 | 4.45* | 755.58 |
| | (13.25) | (242.62) | (399.42) | (93.15) | (1.95) | (456.69) |
| $High_{DKG}$ | −17.45 | −782.18 | 1011.87 | 350.34 | −12.92* | 534.00 |
| | (65.51) | (1732.30) | (1013.56) | (628.99) | (5.61) | (3332.27) |
| Num. Chronic | 28.09*** | 156.58 | −279.74 | −55.07 | 2.84*** | 252.67 |
| | (5.71) | (81.09) | (155.86) | (38.88) | (0.73) | (196.36) |
| $Medium_{past}$ | 122.90*** | 363.24*** | 1747.23*** | 14.95 | 10.07*** | 3682.40*** |
| | (9.13) | (60.82) | (91.06) | (15.09) | (1.48) | (96.47) |
| $High_{past}$ | 232.70*** | 5905.49*** | 12911.49*** | 1249.40*** | 25.18*** | 23893.00*** |
| | (16.60) | (459.02) | (578.96) | (161.16) | (2.36) | (709.34) |
| Adj. $R^2$ | 0.14 | 0.27 | 0.35 | 0.06 | 0.07 | 0.57 |
| Num. obs. | 83115 | 83115 | 83115 | 83115 | 83115 | 83115 |
| RMSE | 64.85 | 911.22 | 1555.20 | 477.94 | 8.78 | 1789.79 |
| N Clusters | 1813 | 1813 | 1813 | 1813 | 1813 | 1813 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 15: Effect of the intervention in 2017. All variables are in average expenditures in euro per year except LOS which is length of stay in days. The subscripts for *ATT* and *Year* refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period.

| | Intervention year 2017 — Log Specification | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| $ATT_{-2}$ | $-0.10^{**}$ | 0.47 | 0.02 | $0.38^{*}$ | 0.15 | 0.00 |
| | (0.04) | (0.38) | (0.08) | (0.18) | (2.34) | (0.03) |
| $ATT_0$ | $0.06^{*}$ | 0.07 | 0.10 | 0.41 | 3.31 | 0.01 |
| | (0.03) | (0.44) | (0.07) | (0.21) | (2.09) | (0.03) |
| $ATT_1$ | −0.03 | −0.57 | 0.15 | 0.37 | 0.41 | 0.01 |
| | (0.03) | (0.46) | (0.08) | (0.21) | (2.30) | (0.03) |
| $ATT_2$ | −0.03 | $-1.22^{*}$ | 0.09 | 0.38 | 3.72 | −0.02 |
| | (0.04) | (0.50) | (0.08) | (0.21) | (2.40) | (0.03) |
| $Group_{treated}$ | $0.09^{*}$ | $-0.97^{*}$ | −0.04 | $-0.35^{**}$ | −2.62 | −0.02 |
| | (0.04) | (0.38) | (0.07) | (0.13) | (2.38) | (0.02) |
| $Year_{-2}$ | $0.66^{***}$ | −2.59 | $-0.76^{*}$ | −1.13 | $39.30^{***}$ | $0.60^{***}$ |
| | (0.16) | (1.88) | (0.31) | (0.65) | (10.90) | (0.13) |
| $Year_0$ | 0.04 | 0.32 | −0.07 | −0.10 | −1.13 | 0.04 |
| | (0.02) | (0.29) | (0.05) | (0.17) | (1.59) | (0.02) |
| $Year_1$ | $1.03^{***}$ | −3.01 | $-0.91^{*}$ | −1.28 | $50.28^{***}$ | $0.78^{***}$ |
| | (0.19) | (2.26) | (0.36) | (0.76) | (12.85) | (0.15) |
| $Year_2$ | $1.08^{***}$ | −2.68 | $-0.85^{*}$ | −1.35 | $47.06^{***}$ | $0.80^{***}$ |
| | (0.18) | (2.22) | (0.36) | (0.77) | (12.67) | (0.15) |
| (Intercept) | $6.91^{***}$ | −45.40 | 0.64 | −10.29 | 61.04 | $9.40^{***}$ |
| | (1.84) | (23.65) | (3.75) | (7.51) | (139.37) | (1.43) |
| Age | −0.06 | 0.82 | $0.19^{*}$ | 0.04 | −2.71 | −0.07 |
| | (0.04) | (0.56) | (0.09) | (0.18) | (3.48) | (0.03) |
| $Age^2$ | 0.00 | −0.00 | $-0.00^{*}$ | −0.00 | 0.02 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.02) | (0.00) |
| Male | −0.03 | $-1.18^{***}$ | $0.21^{***}$ | $-0.25^{***}$ | 0.71 | −0.00 |
| | (0.02) | (0.22) | (0.04) | (0.06) | (1.77) | (0.02) |
| $Medium_{DKG}$ | −0.02 | 0.49 | $0.24^{***}$ | −0.17 | $4.45^{*}$ | $0.08^{***}$ |
| | (0.03) | (0.31) | (0.04) | (0.11) | (1.95) | (0.02) |
| $High_{DKG}$ | −0.09 | −3.51 | $0.47^{**}$ | 0.65 | $-12.92^{*}$ | −0.03 |
| | (0.11) | (1.83) | (0.15) | (1.03) | (5.61) | (0.09) |
| Num. Chronic | $0.05^{***}$ | −0.18 | $-0.04^{*}$ | −0.07 | $2.84^{***}$ | $0.04^{***}$ |
| | (0.01) | (0.13) | (0.02) | (0.04) | (0.73) | (0.01) |
| $High_{past}$ | $0.42^{***}$ | $10.37^{***}$ | $2.99^{***}$ | $1.93^{***}$ | $25.18^{***}$ | $2.60^{***}$ |
| | (0.03) | (0.43) | (0.07) | (0.19) | (2.36) | (0.02) |
| $Medium_{past}$ | $0.26^{***}$ | $2.46^{***}$ | $1.41^{***}$ | $0.06^{*}$ | $10.07^{***}$ | $1.12^{***}$ |
| | (0.02) | (0.21) | (0.04) | (0.02) | (1.48) | (0.02) |
| Adj. $R^2$ | 0.15 | 0.29 | 0.48 | 0.09 | 0.07 | 0.81 |
| Num. obs. | 83115 | 83115 | 83115 | 83115 | 83115 | 83115 |
| RMSE | 0.12 | 1.44 | 0.26 | 0.58 | 8.78 | 0.10 |
| N Clusters | 1813 | 1813 | 1813 | 1813 | 1813 | 1813 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 16: Effect of the intervention in 2017. All variables are in percentage points expenditures in euro per year except LOS which is length of stay in days. The subscripts for *ATT* and *Year* refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period.

| | Intervention year 2018 — DiD with PSM | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| (Intercept) | −644.02 | 11188.96 | 36550.94 | 4651.95 | −370.04* | 47303.02 |
| | (1577.88) | (23800.10) | (26352.66) | (10355.03) | (152.97) | (32252.67) |
| $Group_{treated}$ | 76.94** | 649.20 | −551.06 | 150.83 | 1.29 | 395.88 |
| | (25.33) | (449.74) | (718.07) | (181.93) | (2.99) | (810.83) |
| $Year_{-3}$ | 779.33*** | 12815.95*** | −5751.66 | −1015.02 | 42.73* | 13033.07** |
| | (166.14) | (2803.97) | (4476.15) | (950.56) | (17.10) | (4961.01) |
| $Year_{-2}$ | 8.89 | 1062.41 | −131.97 | 3.25 | −8.38* | 695.00 |
| | (41.61) | (655.62) | (848.48) | (293.70) | (4.22) | (986.29) |
| $Year_0$ | 981.18*** | 12251.87*** | −8530.61 | −1596.79 | 75.29*** | 10659.33* |
| | (186.98) | (3255.74) | (4742.45) | (1100.90) | (18.58) | (5368.54) |
| $Year_1$ | 1043.36*** | 11806.47*** | −8390.83 | −1486.00 | 81.56*** | 10856.91* |
| | (202.51) | (3427.53) | (4487.67) | (1185.62) | (19.21) | (5258.61) |
| Age | −10.28 | −830.19 | −427.28 | −76.65 | 7.58* | −1406.60 |
| | (38.86) | (581.63) | (666.05) | (248.22) | (3.78) | (791.77) |
| $Age^2$ | 0.12 | 6.27 | 1.02 | 0.48 | −0.04 | 8.11 |
| | (0.24) | (3.62) | (4.03) | (1.52) | (0.02) | (4.79) |
| Male | 4.96 | −1195.09*** | 1737.87*** | −178.35* | −2.05 | 447.63 |
| | (19.20) | (296.83) | (421.94) | (88.23) | (2.17) | (416.69) |
| $Medium_{DKG}$ | −16.59 | −122.90 | 1111.63 | 112.57 | 3.89 | 1876.01** |
| | (16.58) | (328.69) | (564.72) | (152.04) | (2.44) | (652.38) |
| $High_{DKG}$ | −38.46 | −2096.91** | 4638.07 | −277.77 | 2.38 | 2642.01 |
| | (54.54) | (717.37) | (2597.73) | (374.42) | (7.94) | (2935.11) |
| Num. Chronic | 53.77*** | 756.46*** | −426.53 | −81.45 | 3.87*** | 740.68* |
| | (10.35) | (179.65) | (273.15) | (59.66) | (1.03) | (308.99) |
| $High_{past}$ | 256.17*** | 8205.37*** | 11604.36*** | 1423.56*** | 28.26*** | 25174.56*** |
| | (23.87) | (436.32) | (507.74) | (135.60) | (2.58) | (584.38) |
| $Medium_{past}$ | 118.91*** | 441.69*** | 1733.13*** | 6.42 | 10.35*** | 3684.58*** |
| | (13.65) | (113.21) | (138.70) | (25.91) | (1.63) | (136.85) |
| $ATT_{-3}$ | −63.61** | −138.37 | −438.44 | −95.34 | −5.18 | −614.75 |
| | (22.61) | (486.20) | (868.44) | (238.69) | (3.10) | (981.05) |
| $ATT_{-2}$ | −51.72** | −385.51 | −255.21 | −126.51 | −0.59 | −458.37 |
| | (19.66) | (395.33) | (733.63) | (245.25) | (2.68) | (913.41) |
| $ATT_0$ | 11.72 | 359.49 | 536.18 | 436.61 | 0.11 | 1389.22 |
| | (22.41) | (441.59) | (931.51) | (266.11) | (3.00) | (947.58) |
| $ATT_1$ | −44.95 | −77.00 | 582.13 | −4.92 | 1.15 | 364.90 |
| | (34.37) | (525.72) | (1036.08) | (279.49) | (3.71) | (1210.17) |
| $ATT_{average}$ | 19.60 | 302.64 | 785.47 | 278.01 | 2.51 | 1202.34 |
| | (21.74) | (377.53) | (688.08) | (186.15) | (2.75) | (755.77) |
| Adj. $R^2$ | 0.15 | 0.36 | 0.26 | 0.07 | 0.09 | 0.54 |
| Num. obs. | 82425 | 82425 | 82425 | 82425 | 82425 | 82425 |
| RMSE | 73.31 | 1157.71 | 1859.33 | 543.65 | 8.29 | 2148.18 |
| N Clusters | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 17: Effect of the intervention in 2018. All variables are in average expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ and $Year$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period.

| | Intervention year 2018 — Log Specification | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| (Intercept) | 1.31 | −58.10* | 2.51 | −14.29 | −370.04* | 9.10*** |
| | (2.17) | (24.65) | (3.80) | (12.15) | (152.97) | (1.41) |
| $Group_{treated}$ | 0.13** | 1.27* | −0.09 | 0.18 | 1.29 | 0.01 |
| | (0.04) | (0.64) | (0.09) | (0.25) | (2.99) | (0.03) |
| $Year_{-3}$ | 0.84*** | 1.39 | −1.30* | −1.90 | 42.73* | 0.86*** |
| | (0.21) | (2.59) | (0.50) | (1.43) | (17.10) | (0.16) |
| $Year_{-2}$ | −0.07 | −1.08 | −0.05 | −0.26 | −8.38* | 0.05 |
| | (0.06) | (0.71) | (0.11) | (0.38) | (4.22) | (0.05) |
| $Year_0$ | 1.42*** | 4.33 | −1.03 | −1.83 | 75.29*** | 0.85*** |
| | (0.24) | (2.79) | (0.56) | (1.58) | (18.58) | (0.17) |
| $Year_1$ | 1.57*** | 6.10* | −0.80 | −1.43 | 81.56*** | 0.87*** |
| | (0.25) | (2.92) | (0.56) | (1.65) | (19.21) | (0.17) |
| Age | 0.06 | 0.89 | 0.16 | 0.16 | 7.58* | −0.07* |
| | (0.05) | (0.60) | (0.09) | (0.29) | (3.78) | (0.03) |
| $Age^2$ | −0.00 | −0.00 | −0.00* | −0.00 | −0.04 | 0.00* |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.02) | (0.00) |
| Male | −0.01 | −2.28*** | 0.26*** | −0.24* | −2.05 | 0.01 |
| | (0.03) | (0.29) | (0.05) | (0.12) | (2.17) | (0.02) |
| $Medium_{DKG}$ | −0.03 | 0.05 | 0.25*** | 0.14 | 3.89 | 0.09*** |
| | (0.03) | (0.36) | (0.05) | (0.20) | (2.44) | (0.02) |
| $High_{DKG}$ | −0.06 | −1.69 | 0.56* | −0.53 | 2.38 | 0.15 |
| | (0.10) | (1.17) | (0.21) | (0.54) | (7.94) | (0.08) |
| Num. Chronic | 0.07*** | 0.22 | −0.06* | −0.11 | 3.87*** | 0.06*** |
| | (0.01) | (0.15) | (0.03) | (0.09) | (1.03) | (0.01) |
| $High_{past}$ | 0.45*** | 12.58*** | 2.64*** | 2.26*** | 28.26*** | 2.62*** |
| | (0.04) | (0.35) | (0.07) | (0.18) | (2.58) | (0.03) |
| $Medium_{past}$ | 0.27*** | 4.04*** | 1.25*** | 0.03 | 10.35*** | 1.13*** |
| | (0.03) | (0.29) | (0.05) | (0.04) | (1.63) | (0.02) |
| $ATT_{-3}$ | −0.11* | −0.17 | 0.02 | −0.01 | −5.18 | 0.05 |
| | (0.05) | (0.62) | (0.12) | (0.33) | (3.10) | (0.05) |
| $ATT_{-2}$ | −0.10** | −0.43 | −0.05 | −0.31 | −0.59 | −0.00 |
| | (0.04) | (0.53) | (0.09) | (0.32) | (2.68) | (0.04) |
| $ATT_0$ | 0.01 | 4.11*** | 0.15 | 0.71 | 0.11 | 0.12** |
| | (0.04) | (0.66) | (0.10) | (0.38) | (3.00) | (0.04) |
| $ATT_1$ | −0.06 | −1.39* | 0.03 | −0.10 | 1.15 | 0.00 |
| | (0.04) | (0.63) | (0.11) | (0.39) | (3.71) | (0.05) |
| $ATT_{average}$ | 0.04 | 1.43** | 0.10 | 0.39 | 2.51 | 0.04 |
| | (0.03) | (0.48) | (0.08) | (0.25) | (2.75) | (0.03) |
| Adj. $R^2$ | 0.19 | 0.40 | 0.42 | 0.09 | 0.09 | 0.83 |
| Num. obs. | 82425 | 82425 | 82425 | 82425 | 82425 | 82425 |
| RMSE | 0.11 | 1.39 | 0.24 | 0.75 | 8.29 | 0.09 |
| N Clusters | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 18: Effect of the intervention in 2018. All variables are in average expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ and *Year* refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period.

| | Standardised Time Generalised DiD | | | | | |
|---|---|---|---|---|---|---|
| | GP | Nurse | Hospital | Geriatric | LOS | Total |
| (Intercept) | 287.62 | 34492.56*** | −40326.80*** | 242.57 | −89.82** | −2280.33 |
| | (230.56) | (3906.66) | (4819.31) | (1046.80) | (27.41) | (5127.14) |
| $Group_{treated}$ | 45.96*** | −255.57 | 107.34 | −3.50 | −2.61 | −103.49 |
| | (11.33) | (175.96) | (248.20) | (61.12) | (1.35) | (279.02) |
| $year_{-3}$ | −11.65*** | −55.41* | −185.40** | −13.56 | 0.42* | −278.45*** |
| | (1.53) | (24.84) | (57.33) | (12.65) | (0.21) | (62.95) |
| $year_{-2}$ | −9.24*** | −66.85*** | −66.38 | 2.31 | 0.27 | −184.71*** |
| | (1.03) | (15.37) | (35.79) | (8.13) | (0.14) | (40.44) |
| $year_0$ | 18.42*** | 184.21*** | −78.78* | −13.74 | 0.76*** | 277.89*** |
| | (1.07) | (17.10) | (32.08) | (7.63) | (0.14) | (39.85) |
| $year_1$ | 55.27*** | 366.56*** | 20.13 | 64.48*** | 1.43*** | 867.62*** |
| | (1.92) | (29.57) | (51.99) | (14.89) | (0.21) | (64.70) |
| $year_2$ | 80.06*** | 430.31*** | 118.06 | 136.32*** | 1.15*** | 1219.24*** |
| | (2.79) | (39.30) | (76.08) | (26.37) | (0.28) | (95.41) |
| Age | −3.17 | −1012.61*** | 1176.29*** | −17.07 | 2.22** | 111.98 |
| | (6.02) | (102.52) | (122.87) | (27.57) | (0.71) | (131.13) |
| $Age^2$ | 0.05 | 7.35*** | −8.46*** | 0.19 | −0.01 | −0.89 |
| | (0.04) | (0.67) | (0.78) | (0.18) | (0.00) | (0.83) |
| Male | −22.72*** | −728.53*** | 1183.89*** | −84.95*** | 0.88 | 432.64*** |
| | (3.53) | (59.60) | (99.77) | (15.98) | (0.49) | (105.33) |
| $Medium_{DKG}$ | −2.94 | 70.90 | 644.52*** | 75.45* | 4.37*** | 1331.37*** |
| | (3.52) | (72.84) | (132.46) | (31.02) | (0.54) | (154.54) |
| $High_{DKG}$ | −44.75*** | 38.92 | 6390.63*** | −233.79* | 5.63** | 8212.76*** |
| | (12.89) | (467.35) | (1264.45) | (99.48) | (2.10) | (1394.78) |
| Num. Chronic | 0.55*** | −1.22 | 7.48* | −3.68*** | 0.11*** | 12.75** |
| | (0.10) | (1.63) | (3.27) | (0.92) | (0.01) | (4.04) |
| $High_{past}$ | 215.95*** | 6486.04*** | 13488.13*** | 1016.17*** | 27.49*** | 25369.62*** |
| | (5.61) | (135.72) | (177.16) | (32.64) | (0.62) | (186.41) |
| $Medium_{past}$ | 109.02*** | 333.69*** | 1901.10*** | −12.53*** | 11.46*** | 3845.30*** |
| | (2.86) | (18.98) | (26.96) | (3.29) | (0.34) | (26.53) |
| $ATT_{-3}$ | −47.78** | 624.50* | −450.16 | 80.27 | −0.05 | 329.06 |
| | (14.16) | (245.74) | (451.13) | (134.78) | (1.78) | (545.31) |
| $ATT_{-2}$ | −44.98*** | 122.16 | −316.22 | −5.38 | −0.71 | −72.90 |
| | (7.61) | (138.93) | (244.40) | (97.10) | (1.11) | (338.35) |
| $ATT_0$ | 42.42*** | −100.14 | 213.77 | 251.54* | 1.16 | 487.93 |
| | (7.77) | (113.81) | (320.93) | (107.14) | (1.10) | (346.42) |
| $ATT_1$ | 24.29* | −230.06 | 248.63 | −4.87 | 1.18 | 234.42 |
| | (11.62) | (181.45) | (371.34) | (100.40) | (1.29) | (441.38) |
| $ATT_2$ | −3.06 | −1488.06*** | 559.26 | 70.20 | 0.60 | −773.37 |
| | (15.30) | (235.90) | (395.78) | (180.42) | (1.76) | (501.14) |
| $ATT_{average}$ | 51.70*** | −587.25*** | 468.47 | 85.84 | 1.51 | 74.36 |
| | (8.36) | (139.14) | (260.90) | (81.31) | (1.12) | (300.42) |
| Adj. $R^2$ | 0.11 | 0.29 | 0.29 | 0.05 | 0.10 | 0.52 |
| Num. obs. | 165540 | 165540 | 165540 | 165540 | 165540 | 165540 |
| RMSE | 273.13 | 4700.57 | 8571.88 | 2174.29 | 36.39 | 9852.13 |
| N Clusters | 1820 | 1820 | 1820 | 1820 | 1820 | 1820 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 19: Generalised difference in differences in a standardised time framework estimated using ordinary least squares. Effect of the intervention in 2017 that is estimated using ordinary least squares. All variables are expenditures in euro per year except LOS which is length of stay in days. The subscripts for $ATT$ refer to years before and after the intervention, where -1 is the year before the intervention that serves as the reference period.