

Quality of life and resource use in nursing homes

The effect of the COSMOS-intervention

Lillian Ovedal Myrtveit

Master thesis

The thesis is submitted to complete the degree

Master in economics

University of Bergen, Department of Economics

July 2018



UNIVERSITETET I BERGEN

Acknowledgements

The journey of writing this thesis has had its ups and downs, but has in total been a very rich and positively challenging learning process. I would like to thank my supervisor, Julie Riise, for valuable inputs and discussions throughout the process. I would also like to thank SEFAS for giving me access to the data, and co-supervisor Christine Gulla for helpful thoughts and comments. All help and feedback are sincerely appreciated.

A big “thank you” goes to my amazing husband, Magnar. I could not have done this without your support, academic discussions, ready-to-eat dinners, and uplifting words. I also want to thank Magnar, Irene Aasmul, and Makalani for constructive feedback.

Bergen, July 2018

Lillian Ovedal Myrtveit

Abstract

Dementia is a costly disease. Future projections show a higher share of elderly, which will increase prevalence of dementia and thus health care expenditures consumed by people with dementia. In addition, comes the costs of informal care. Caring for a loved one can make informal caregivers experience a high burden of care and high opportunity costs, even after the patient is moved to a nursing home. Dementia can also negatively affect the quality of life for the diseased person. Due to these factors, interventions that can decrease costs and positively impact the affected people are of high value to society.

In 2015-2016, a randomized controlled trial, the COSMOS-intervention, was implemented in various nursing homes across Norway. The main objective was to improve quality of life in nursing home patients, with secondary objectives including improving cost-effectiveness and reducing unnecessary medication and hospital admissions. This thesis evaluates how the COSMOS-intervention affects quality of life, formal- and informal time use, and use of psychotropic medication in patients with dementia. To do so, the average treatment effect on the treated is estimated using regressions with individual fixed effects. Further, the thesis aims to investigate the economic implications of the intervention. The results show no change in quality of life, but a decreased use of psychotropic medication and informal care. Formal care has increased considerably. The causality of the results is challenged due to various factors, including a small sample size. Whether the changes can be considered economically beneficial, depends on whose perspective is taken, since costs and benefits are valued differently by the patient, nursing home management, and society. All calculations and estimations are performed using Stata IC 14.2.

Contents

<u>ACKNOWLEDGEMENTS.....</u>	<u>II</u>
<u>ABSTRACT</u>	<u>III</u>
<u>FIGURES.....</u>	<u>VIII</u>
<u>TABLES.....</u>	<u>VIII</u>
<u>1. INTRODUCTION.....</u>	<u>1</u>
<u>2. CONTEXT</u>	<u>4</u>
2.1 PUBLIC HEALTH EXPENDITURES	4
2.2 COST OF DEMENTIA IN NORWAY	5
2.3 THE DEMENTIA PLAN 2020.....	6
<u>3. EMPIRICAL LITERATURE.....</u>	<u>7</u>
3.1 QUALITY OF LIFE.....	7
3.2 RESOURCE USE IN NURSING HOMES	9
3.2.1 FORMAL CARE	9
3.2.2 INFORMAL CARE.....	11
3.2.3 PSYCHOTROPIC DRUGS	12
3.2.4 DRUG OR TIME USE.....	13
<u>4. DATA</u>	<u>15</u>
4.1 THE COSMOS-STUDY.....	15
2.1 METHOD AND DESIGN	16
2.2 INTERVENTION AND IMPLEMENTATION	17

2.3 OUTCOMES.....	17
4.2 THE SAMPLE DATA.....	17
4.2.1 ASSESSMENT OF THE VARIABLES.....	17
4.2.2 WEAKNESSES.....	19
4.2.3 STRENGTHS.....	19
4.2.4 MISSING VALUES.....	20
<u>5. METHOD.....</u>	<u>21</u>
5.1 RANDOMIZED CONTROLLED TRIALS.....	21
5.1.1 IS RANDOMIZATION ENOUGH?.....	22
5.2 DESCRIPTIVE STATISTICS.....	25
5.2.1 BALANCE TESTS.....	25
5.2.2 DISTRIBUTIONS.....	29
5.3 THE ROAD TO CAUSALITY.....	30
5.3.1 IMPACT EVALUATION.....	30
5.3.2 COSMOS: POSSIBLE FACTORS CHALLENGING THE CAUSALITY.....	31
5.4 DATA ANALYSES.....	32
5.4.1 PANEL DATA.....	32
5.4.2 FIXED EFFECTS.....	32
5.4.3 AVERAGE TREATMENT EFFECT ON THE TREATED.....	34
5.4.4 STANDARD ERROR ADJUSTMENT.....	36
5.4.5 MODEL.....	38
<u>6. RESULTS.....</u>	<u>38</u>
6.1 QoL.....	39
6.2 FORMAL CARE.....	41
6.3 INFORMAL CARE.....	44
6.4 PSYCHOTROPIC MEDICATION.....	46
6.4 POSSIBLE CHALLENGES AND ROBUSTNESS TESTS.....	47

6.4.1 POSSIBLE CHALLENGES	47
6.4.2 ROBUSTNESS TESTS	48
6.4.3 MEASUREMENT ERRORS.....	51
<u>7. DISCUSSION.....</u>	<u>52</u>
7.1 QoL	52
7.2 ECONOMIC IMPLICATIONS	53
7.2.1 FORMAL CARE	54
7.2.2 INFORMAL CARE	55
7.2.3 PSYCHOTROPIC MEDICATION	57
7.2.4 PSYCHOTROPIC DRUGS OR TIME USE	58
7.3 TIME HORIZON.....	59
<u>8. CONCLUDING REMARKS</u>	<u>60</u>
8.1 FUTURE RESEARCH	62
<u>BIBLIOGRAPHY</u>	<u>63</u>
<u>APPENDIX 1 ACCESS TO VARIOUS PERSONNEL: DESCRIPTIVE STATISTICS</u> <u>AT BASELINE.....</u>	<u>67</u>
<u>APPENDIX 2 DISTRIBUTIONS</u>	<u>68</u>
<u>APPENDIX 3 ATET OF COSMOS ON DEPRESSION</u>	<u>70</u>
<u>APPENDIX 4 EXCLUDED PATIENTS: DESCRIPTIVE STATISTICS AND</u> <u>REGRESSIONS.....</u>	<u>71</u>
<u>APPENDIX 5 HOW PRECISION IS AFFECTED: CLUSTERING AND</u> <u>BOOTSTRAPPING.....</u>	<u>75</u>

APPENDIX 6 EXCLUSION OF PATIENTS WITH LARGE CHANGE IN FORMAL CARE76

APPENDIX 7 REGRESSIONS WITHOUT FIXED EFFECTS77

APPENDIX 8 CHANGE IN QOL FOR HOSPITALIZED PATIENTS81

APPENDIX 9 CHANGE IN HOSPITAL ADMISSIONS.....82

Figures

Figure 1 Forecasted global costs of dementia (Source: World Alzheimer Report 2015)	1
Figure 2 Change in QoL	39
Figure 3 Change in formal care.....	41
Figure 4 Change in informal care.....	44
Figure 5 Change in psychotropic medication	46

Tables

Table 1 Health care expenditures (NOK million) measured in constant 2010-prices ...	4
Table 2 Patient: Descriptive Statistics at Baseline	26
Table 3 Ward: Descriptive Statistics at Baseline	28
Table 4 ATET of COSMOS on QoL.....	39
Table 5 ATET of COSMOS on hours of formal care per month.....	42
Table 6 ATET of COSMOS on components of formal care	43
Table 7 ATET of COSMOS on components of informal care	45
Table 8 ATET of COSMOS on use of psychotropic medication	47
Table 9 Costs and benefits in different perspectives	54

1. Introduction

There are almost 50 million people living with dementia worldwide (The Norwegian Health Association, 2017), and this number is expected to double within the next three decades (Husebo et al., 2015). With an estimated cost of \$818 billion, global dementia care would be the 18th largest economy, if it were a country. This enormous sum exceeds the market value of large companies such as Apple and Google and following the future projections, is expected to grow massively in the coming years (Prince et al., 2015). Figure 1 shows that the costs worldwide are expected to rise to above \$2000 billion¹ in 2030. These numbers do not include informal care, which would increase the numbers considerably.

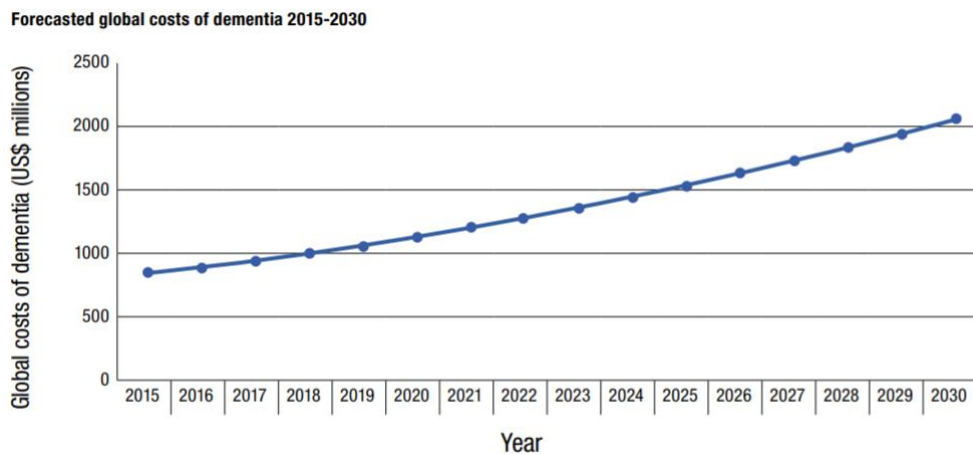


Figure 1 Forecasted global costs of dementia (Source: World Alzheimer Report 2015)

In Norway, it has been predicted that the number of people with dementia will increase from 78,000 today to around 112,000 in 2030, and to 200,000 in 2060². The cost of dementia is found to be 360,000 Norwegian kroner (NOK) per person per year, giving a staggering yearly cost of around 28 billion NOK (Vossius et al., 2015). The future projections indicate that these costs, and thus share of the health budget, will

¹ From my understanding, it must be a mistake in the figure from the original source. On the y-axis in Figure 1, it should say 'billions' and not 'millions'.

² This number is uncertain because data from Norway doesn't exist. For that reason, the prediction is based on data from other countries. The numbers for Norway will be available in 2020.

increase. It is therefore of major importance to utilize resources spent on people with dementia in an efficient way; and it will be even more so in the future.

Much of the costs related to dementia is incurred through government expenditure on resource use. But the dementia disease has more than just a direct economic impact. Firstly, the person with dementia can experience reduced quality of life (QoL), higher mortality, ill health and disability. Secondly, family and friends are often the cornerstone for care and support, and thus time and attention that could have been spent working or on leisure activities, are devoted to caring for the diseased person. This is in the literature called *informal care*. Informal care is associated with opportunity costs of lost productivity (Prince et al., 2015) and health deterioration for the caregiver, and often constitutes a high care burden. A high care burden for both informal and formal caregivers, is often related to the fact that the majority of people with dementia develop stressful behavioral and psychological symptoms, also called neuropsychiatric symptoms, such as depression, agitation, hallucinations, and aggression.

In the nursing homes (NHs) such care-related challenges raise the question of having labor-intensive or material-intensive manners of care. For instance, to control and monitor a resident's behavioral problems, the decision of constant monitoring by staff versus use of chemical or physical restraints must be made (Cawley et al., 2006). A chemical restraint often used to prevent these behaviors are psychotropic drugs, like antipsychotics (Quinn, 2013). Psychotropic drugs, however, have been linked to negative health outcomes, such as increased risk of mortality and seizures. In a recent article, the Norwegian Broadcasting Corporation (NRK) shed light on today's problematic use of psychotropic medication in elderly in Norwegian NHs (Sørbø and Moen Nilsen, 2018). Very large amounts of antipsychotics were given to an elderly patient as a mean to control the patient's behavioral problem. Not long after, the patient was hospitalized, and she died a few weeks later. Though the reason for death is not clear, this shows that use of psychotropic medication in elderly is a relevant topic in Norwegian NHs today. Elderly NH patients often use many drugs, including

multiple psychotropic drugs, and the Norwegian Ministry of Health and Care Services states that one out of three NH patients use one or more superfluous drugs, and that approximately 10% of all hospital admissions are due to drug-related problems (Ministry of Health and Care Services, 2015).

Since there is no current cure or treatment that can slow or stop the progression of dementia (Alzheimer's Association, 2017), trying to maximize the patients' QoL can be the most important thing to do on a day-to-day basis. At the same time, we must take into consideration that resources are scarce and should be used efficiently. An intervention that can reduce costs and heighten QoL for people with dementia is valuable to society, when considering today's situation and future projections of cost and prevalence. In my thesis, I will focus on NH patients with mild to severe dementia. More specifically, I will compare QoL, time use and use of psychotropic medication of patients with dementia in randomly assigned control and intervention groups which is a part of the cluster randomized clinical hybrid trial, COSMOS. The main purpose of the COSMOS-study is to measure the effect of a randomized controlled intervention on QoL. In addition to investigate the effect on QoL, I will exploit the randomization to further evaluate the effect of the intervention on resource use. Due to data availability, resource use is restricted to mean formal- and informal time use and use of psychotropic medication.

The two main purposes of the thesis are: (1) Evaluate how the COSMOS-intervention affects the QoL in people with dementia, and their use of psychotropic medication, formal care, and informal care. (2) Investigate the economic implications of the intervention.

The structure is as follows: Chapter 2 puts dementia, and especially the cost of dementia, in the context of Norway today. Chapter 3 examines the empirical literature. Chapter 4 describes the COSMOS-study and the sample data. Chapter 5 details the methodology and estimation strategy. Chapter 6 presents the results, while Chapter 7 discusses the findings in light of the literature and specifically explore the economic implications of the intervention. Chapter 8 concludes.

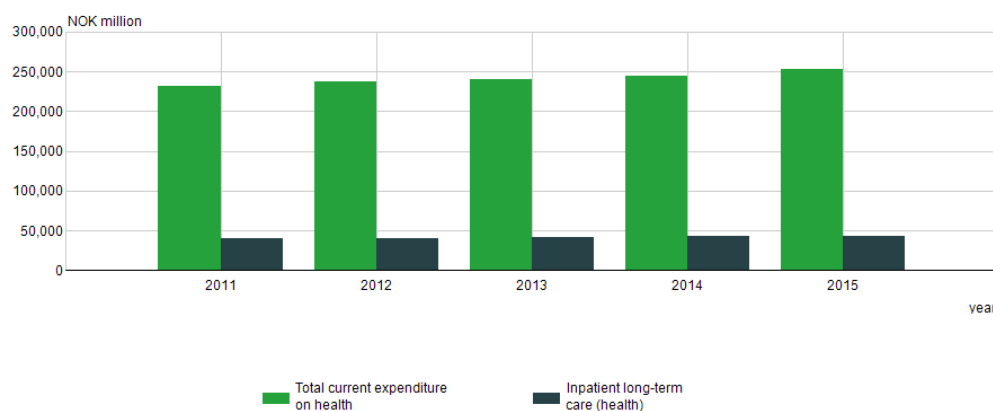
2. Context

This chapter gives a picture of the economic motivation of studying the population of people with dementia. It specifically examines the cost of dementia in more detail, and also describes Norway's expenditures on health and the elderly; a population in which people with dementia will be a larger share of in the coming years. Further, a current governmental plan concerning dementia is presented, showing that interventions focusing on dementia are of public interest in Norway today.

2.1 Public health expenditures

The demand for health care is increasing, and we can expect to spend more on health in the future. The contribution is especially large from the growing share of elderly, following the changing demographics that is evident all over the world. Today, 875,000 people in Norway, or approximately 17 % of the population, are 65 years old or older, and Norway has never before had a proportion of elderly this large (Statistics Norway, 2017). This group is expected to increase to 21% of the population by 2050. Since the greatest risk factor of dementia is increasing age (Alzheimer's Association, 2018), the future projections of growing elderly leads to a higher incidence of dementia, and thus costs related to this disease.

Table 1 Health care expenditures (NOK million) measured in constant 2010-prices



Source: Statistics Norway

From 2011 to 2015, the total expenditures on health care increased by around 20,000 million NOK. The spending on NHs (inpatient long-term care) has also increased, with around 2,500 million NOK. The age group of patients >67 years old, stands for most of the resources used on NHs and similar long-term residential facilities, where nine of ten NOK are used on the elderly in these facilities (Statistics Norway, 2013). As 80% of NH patients are found to have dementia (Selbæk et al., 2007), the majority of NH costs are related to people with dementia.

2.2 Cost of dementia in Norway

Vossius et al. (2015) have tried to map the total resource use and cost of dementia-related illness in Norway in the nation-wide report *Resource Use and Disease Course in Dementia* (REDIC). The following cost estimations of direct, indirect and intangible costs are found in this report.

It is calculated that one patient with dementia will use health and care services for approximately 2.9 million Norwegian kroner (NOK), where NH costs counts for almost two-thirds of the total costs. The total cost of people with dementia is calculated to 28 billion NOK per year. These numbers do not include informal care. During the course of the disease, staffing-resources equal to 3.37 man-years are used. In total, 32,451 man-years in health and care services are spent on people with dementia per year, while it in NH locations are spent 1.06 man-years per patient per year.

There are approximately 28,091 people with dementia in Norwegian NHs. About half of the NH patients receive informal care from loved ones, with an average of 6.7 hours per month. This is valued between 0 and 2466 NOK, depending on the assumed pricing, e.g. no wage or a nurse's wage.

When it comes to QoL, Vossius et al. (2015) find that it seems like people with dementia perceive their QoL equally good as their peers without dementia. At the same time, they find consistent reporting of health care personnel and next of kin valuing QoL lower than the patients themselves. Thus, they conclude that society

value the illness of dementia with lower health related QoL, and that the personal experience of higher QoL is probably not actual, but merely a symptom of the disease. If this valuation is used, each person with dementia will lose 1 quality-adjusted life year (QALY) because of reduced health related QoL. When higher mortality is considered, dementia stands for a loss of 4.2 QALYs per patient.

Vossius et al. (2015) also evaluate how much the costs of psychotropic medication amounts to in the dementia care. They find that people with dementia have 4.3 times higher costs of psychotropic medication than people without dementia. The monthly cost is 111 NOK per patient and amounts to less than 1% of the total costs per month. These estimates are based on price and quantity and does not take into consideration how QoL can be affected by use of these drugs.

2.3 The Dementia Plan 2020

The Government of Norway (hereafter the Government) recognizes dementia of current and future importance and has dementia care and research on its agenda. As the first country in the world, Norway presented an individual plan for the services that affects people with dementia. Dementia was put on the Government's agenda with the Dementia Plan 2015, and has stated that "One of the biggest care challenges we face [...] is the expectation of twice as many people in the next 35 years with various dementia disorders (Ministry of Health and Care Services, 2008, p.7). The Dementia Plan 2020 is built on experiences from the Dementia Plan 2015 and was published by the Norwegian Ministry of Health and Care Services in 2015 as a guidance to dementia care for the years 2016-2020. It strives to improve the services offered to people with dementia and their next of kin. The strategies and measures in the plan are

"intended to help put dementia and its challenges on the agenda of ordinary municipal planning. The aim is to ensure long-term, integrated planning for the local community and the physical environment, and high-quality services

designed to meet the needs of the growing number of people with dementia and their families.” (Ministry of Health and Care Services, 2015, p.9).

The plan includes many strategies and measures to reach these aims, such as providing support and help for informal caregivers by establishing an informal care program to respite the caregiver, and securing “better procedures to ensure correct medication management for older people in general and people with dementia in particular” (Ministry of Health and Care Services, 2015, p.32). A part of the long-term strategy is combating the extensive use of psychotropic drugs such as antipsychotics and other psychiatric drugs.

3. Empirical literature

Together, dementia, QoL, psychotropic drugs, and formal- and informal care have wide economic implications, concerning different types of costs. There exists literature to various degrees regarding these topics and the association between them, though the causal relationships are explored to a smaller degree. This chapter presents a selection of empirical literature attempting to shed light on informal care, QoL and its relationship with resource use, and the relationship between psychotropic drugs and formal care in a NH setting for people with dementia.

3.1 Quality of Life

QoL is defined by WHO as “an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns” (World Health Organization, 2017). For a person with dementia, QoL is affected by the consequences of the disease, including present and future forecast, for example, on cognitive abilities. Factors that can further influence QoL in people with dementia is loss of independence, side-effect of drugs, incontinence, social isolation, paranoid reactions, sleeplessness, hallucinations and delusions, and depression (Korczyn and Davidson, 1999).

In studies connecting QoL and dementia, the findings vary. QoL in people with dementia has been associated negatively with impairment in activities of daily living (ADL), severity of dementia and cognitive impairment³, neuropsychiatric symptoms, depression and pain (Mjørud et al., 2014, Barca et al., 2011, Rostad et al., 2017, Sloane et al., 2005). Sloane et al. (2005) found that ADL-function was the strongest predictor of QoL scores. Depressive symptoms and agitation was also associated negatively with QoL, although these factors explained less of the variance. Pain, on the other hand, showed little association with QoL in this study. Due to the cross-sectional designs of these studies and for the lack of randomization, one cannot make any causal conclusions, only imply associations. Cross-sectional studies do not follow the same individual over time but observe the sample at a specific point in time. We thus do not know if something underlying are affecting the results, how they would change for the same person over time, or if one factor is causing the other.

In other words, QoL in dementia seems to be complex and the empirical findings are inconsistent, which can make it difficult to identify and quantify how and why dementia affects QoL. Additionally, due to the disease's impact on the brain, especially in cases of severe dementia, many patients find it hard, or even impossible, to communicate how he or she feels. This leads to the need of a proxy assessment, performed by the people that are closest to the patient.

An issue is that proxy-rating can be biased by the proxy-rater, as people are not neutral beings. In a recent systematic review, Robertson et al. (2017) found that proxy-rated QoL were rated lower where cognitive impairment was higher for care home residents with dementia. This may reflect how the observer's view of the patient's health status has changed, or how the observer would feel if in a similar situation (Moyle and Murfield, 2013). Specifically, proxy-rated QoL in long-term care residents with dementia are found to be different from patient-rated QoL (Sloane et al., 2005). In the Norwegian context, Vossius et al. (2015) found that next of kin

³ Cognitive impairment is in the literature used as a synonym for dementia. Higher cognitive impairment is equivalent to more severe dementia.

considered NH patients' QoL as lower than that experienced by the patients themselves. On the other hand, it has also been found reasonable compliance between proxy rating between staff and relatives (Robertson et al., 2017) and between proxy- and patient-rating (Essen, 2004), even though the latter found that moderate cognitive impairment was associated with lower compliance. So, QoL can be underestimated and the validity questioned. Still, proxy rating is important for individuals that cannot express themselves – secondhand information is better than none information at all.

3.2 Resource use in nursing homes

3.2.1 Formal care

Dementia includes a wide range of symptoms related to decline in memory or other thinking abilities that have an impact on a person's everyday activities (Alzheimer's Association, 2017). People with dementia often end up in a NH, either because they can no longer take care of themselves, or because family and relatives can no longer look after them. In Norway, after the diagnosis has been given, half of the people with dementia move to a NH within three years (Vossius et al., 2015). 80 % of patients in Norwegian NHs are found to have dementia (Selbæk et al., 2007, Bergh et al., 2012). They require excessive care, which increases along with the progression of the disease. They need help in ADL and other tasks, and NH staff provide extensive assistance with dressing, personal hygiene, transfers, toileting, communication, bed mobility, locomotion, and eating (Carpenter et al., 2006, Matsubasa et al., 2015).

Formal care and Dementia

Some research has tried to identify how dementia affects the staff's time use. In a Swedish longitudinal population-based study on aging and dementia, Nordberg et al. (2007) found that the dependency of formal care in ADL was higher in people with dementia, than of people without dementia. Having dementia and dependency of care in ADL was significantly associated with use of care time. Including age, gender, comorbidity and informal care into the model did not change the results. In the

Norwegian context, Vossius et al. (2015) found that functional disability, cognitive impairment, neuropsychiatric symptoms, and generally bad health were associated with increased resource use on an individual level for people with dementia.

Functional disability had the strongest association. In a NH context, they found that most residents needed help in tasks related to ADL and instrumental ADL (IADL), while only around half needed supervision. ADL-tasks had the highest use of care-hours on average. Overall, it seems like dementia is associated with higher resource use, where especially impairment in ADL, or functional disability, contributes to an increased use of formal care.

Formal care and QoL

Elderly NH patients require attention and assistance from their caregivers on a daily basis. An integrative⁴ review from 2012, reviewed the literature on nursing staff, quality of care and QoL. They found that higher nursing staff affected the residents' outcomes in a positive way and concluded that more nursing staff is required to reach optimal QoL for NH residents (Shin and Bae, 2012). Many of the studies in the review were observational, so a causal relationship cannot be stated. Another finding in the literature is that formal caregivers' time with NH residents are associated with better health outcomes, like fewer pressure ulcers (Horn et al., 2005). Since negative health outcomes are associated with poorer QoL, more time use can thus increase patients' QoL. To my knowledge, there is little research investigating the causal relationship between time use and QoL in NH patients with dementia, and the causal association is yet not clear.

⁴ The main difference between a systematic review and an integrative review is the types of studies that are included in the review. Integrative reviews include both quantitative and qualitative studies, while systematic review only include quantitative studies. Source:

<http://academicanswers.waldenu.edu/faq/145251>

3.2.2 Informal Care

Patients suffering from dementia are often taken care of by their family, friends and relatives. This is called *informal care* and is most commonly referred to when the caretaker is not residing in a long-term care institution. Even so, a Dutch cross-sectional study by Metzelthin et al. (2017) found that informal caregiving does not stop with the admission to a NH. Informal caregivers provide an important role in the NH environment, and often perform tasks similar to those performed when the patients is residing at home, including feeding, grooming, and providing transportation (Schulz et al., 2014). Informal caregivers have also been found to have an active role in caring activities, such as social group activities (e.g. bridge-club), providing coffee and tea, meal activities, cleaning and reacting to incidents (e.g. aggressive behavior towards other residents or staff) (Paulus et al., 2005).

Opportunity cost

Informal care often comes with an opportunity cost, which can be defined as the caregiver's foregone benefits due to time spent on providing informal care (Berg et al., 2004). There is an opportunity cost also for informal caregivers with a relative in the NH. A study by (Gautun and Bratt, 2017), investigated how public health services (NHs) affect the work attendance of adults with parents in need in Norway. Only adults with one parent alive were included, because a spouse could take on much of the care responsibility. The findings suggest that when the parent is residing in a NH, absence from work is lower, compared to if the parent is not in a NH. This further indicates that although less than if the parent is not residing in a NH, there still exists an opportunity cost for adults with parents living in NHs. The data are cross-sectional. Thus, the causal relationship is not identified in this study.

Another source of opportunity costs of informal care can be related to that the care-tasks lead to negative consequences for the caregivers. Caring for a loved one can make caregivers in both home-caring and institutions experience a high burden of care (Metzelthin et al., 2017). In this cross-sectional study, the researchers considered both objective burden (provided caregiving hours) and subjective burden (self-rated

burden scale). They found that even though the informal caregivers in institutionalized long-term care provided significantly less caregiving hours, they experienced a comparable level of subjective burden to caregivers living with the patient at home. This implies that the subjective burden the next of kin can experience, is not lifted when the patient is admitted to a NH. In their sample, most of the informal caregivers in institutionalized long-term care were adult children caring for their parents. Due to the study design, we cannot interpret this as a causal conclusion.

3.2.3 Psychotropic drugs

Neuropsychiatric symptoms are commonly seen in people with dementia and include behavioral and psychological symptoms such as agitation, irritability, aggression, apathy, and hallucinations (Quinn, 2013). In a Norwegian cross-sectional study, Selbæk et al. (2007) found that 89 % of individuals diagnosed with dementia exhibited a neuropsychiatric symptom the past week.

Psychotropic Drugs and Dementia

Neuropsychiatric symptoms in NH patients with dementia are often treated with psychotropic drugs, like antipsychotics and antidepressants. This is also the case in Norwegian NHs, where around 75 % of the residents receive one or more psychotropic drugs (Selbæk et al., 2007). It is found that one of three NH residents uses at least one unnecessary drug, and that many elderly use psychotropic medication that is unrecommended for use in the elderly (Ministry of Health and Care Services, 2015). Despite the widespread use, the effect of these drugs is debatable, where some find modest effects, and yet others find no effect (Sink et al., 2005, Ford, 2015). This excess use can lead to higher costs due to the patient's increased risk of hospital admission when on psychotropic medication. A large nationwide observational study based on register data of the general population of elderly in Sweden, found that psychotropic drug use, particularly when combined, is associated with higher risk of hospitalizations, with dementia as a predictor (Johnell et al., 2017).

Psychotropic Drugs and QoL

The use of psychotropic drugs in elderly NH patients with dementia is not unproblematic, however. According to Gurwitz et al. (2017, p.118), the “excessive use of antipsychotic drugs among long-term nursing home residents with dementia has been among the most challenging issues in the care of this vulnerable population”. Psychotropic drugs have been linked to various negative effects in elderly people, such as increased risk of mortality (Gill et al., 2007), increased risk of seizures (Bloechliger et al., 2015) and increased risk of falling (Cox et al., 2016). It has been found that NH residents with dementia tolerates discontinuation (Helvik et al., 2017), which indicates that patients can safely decrease the risk of side-effects, by taking less of these drugs. From this, one can conclude that psychotropic medication can affect elderly people’s QoL. For instance, consequences from a fall can be a bone fracture or head injury, which can drastically reduce the patient’s QoL. Some studies suggests that psychotropic medication negatively impacts QoL (Mjørud et al., 2014), while others find no such evidence (van de Ven-Vakhteeva et al., 2013).

3.2.4 Drug or time use

How to best use the resources in a NH is a continuous consideration for the decision-makers and NH management. As seen, people with dementia often have neuropsychiatric symptoms. Managing such behaviors can be very challenging for NH staff, especially those providing direct care (Galik et al., 2014). A common such symptom is *agitation*. Agitation includes restlessness, aggression, and excess motor activity, and is associated with excess staff burden (Rinaldi et al., 2005, Aronson et al., 1993, Sourial et al., 2001). How to best handle agitated patients can lead to the question if one should medicate or hire an extra caregiver. Should medication be used as a means of behavior control, or should a caregiver try to calm the patient with individualized care? Both alternatives can lead to the same goal, which can be to calm the patient and make sure he is not negatively interacting with other patients. Which input is chosen will affect the costs differently, depending on if the perspective of the patient, society or NH management is taken. Medicine (capital) is cheaper

than labor, and if the former is chosen the cost will be lower for the NH and society. On the other hand, it can also result in costs for the patient and society, in form of negative side-effects and increased hospitalization. A longitudinal study from 2017 by Helvik et al. (2017), shows high prevalence and persistence of psychotropic drugs in Norwegian NHs. This can indicate that the cheapest alternative, namely drugs, are often chosen in order to address neuropsychiatric symptoms often seen in people with dementia. This can be due to the economic gain, convenience, lack of staffing or knowledge, or that medication use is unsystematically reviewed, or not reviewed at all.

In the query of using drugs or time, to for instance calm the patient, a question is how these two resources work together or depend on each other. Specifically, there is the question if they work as substitutes in a NH setting. To the best of my knowledge, there is no literature investigating whether psychotropic drugs work as a substitute for formal care, or vice versa, in the NH. Still, some studies indicate that this can be the case, or that they at least depend on each other to some degree. Perhaps the most common hypothesis on the relationship between formal care and psychotropic medication, is that better staffing will lead to lower rates of drug use (Schmidt et al., 1998). Using an instrumental variables method on longitudinal data, Cawley et al. (2006) found that higher wages for NH workers increased the use of psychotropic medication, where higher wage was driven by reduced labor supply. Further, a study investigating the relationship of facility characteristics and use of psychotropic medication, found that the only significant factor correlated with excessive use was the level of staffing: Facilities with better staffing had less excessive use of psychotropic medication (Schmidt et al., 1998). A systematic review that investigated how staffing influences QoL and quality of care for long-term dementia care residents, found that more staff and more qualified staff, lead to less use of psychotropic medication (Anderson et al., 2016). Staff distress and less attention and care provided for the residents, were associated with greater reliance on psychotropic drugs. The studies investigating this are of observational design, and no causal

relationship can be concluded on. These findings indicate that when more staff is available, they cover the care needs to a higher degree than if less staff is available. When there are not enough staff, other non-staffing means, such as medication, must “step in” to cover the care needs. But we do not know if psychotropic drugs actually work as a substitute for formal care, and if so, to what degree.

4. Data

The COSMOS-data was collected in 2015-2016 on behalf of Centre for Elderly and Nursing Home Medicine (SEFAS), University of Bergen (UiB). COSMOS is a longitudinal study with panel data on NH patients for three points in time, stretching for nine months. This chapter explains the COSMOS-study and the sample data in more detail.

4.1 The COSMOS-study

COSMOS is an acronym for Communication, Systematic assessment and treatment of pain, Medication review, Occupational therapy, and Safety (COSMOS). The overall objective is to improve QoL in NH patients. The COSMOS-program is based on evidence from the literature, and the research groups’ own scientific and clinical experiences. The intervention aims to improve the physical and mental health of the patients, increase safety and cost-effectiveness, and reduce unnecessary hospital admissions and medication. The researchers want to achieve their goal by educating NH staff in:

- i. enhanced communication and advance care planning
- ii. organizing meaningful activities
- iii. medication reviews
- iv. systematic assessment and treatment of pain (Husebo et al., 2015)

This thesis uses a selection of the COSMOS-sample, because only people with dementia are considered. In the COSMOS-study they also used multiple instruments to measure QoL, while in this thesis only one is considered. Therefore, my data is

different from the COSMOS-data and my findings will not apply to the COSMOS-study as a whole.

2.1 Method and Design

The COSMOS-study includes a two-month pilot study, with 128 NH patients distributed among nine Norwegian NHs, and a four-month multicenter, cluster randomized effectiveness-implementation clinical hybrid trial, with follow-up at month nine (Husebo et al., 2015). The COSMOS-intervention involves 765 patients from 71 NH wards (Aasmul et al., 2018). The inclusion criteria were (i) NH patients with and without dementia (ii) 65 years or older. Patients were excluded if they were dying (life expectancy of 6 months or less) or had schizophrenia. In order to get a more representative sample, NH wards in various parts of Norway were included, where one NH ward is defined as one cluster⁵. The clusters were randomized to either the control or intervention group, where the control groups continued with current best practice. The cluster design will take account of spill-over effects and the fact that people living and working in the same unit are more similar in the first place (e.g. they have the same health personnel/colleagues and environment) (Husebo et al., 2015).

The COSMOS-intervention is complex and comprehensive and includes many NHs and different health personnel. Complex interventions can lead to ineffective implementation, and it was thus important to test and improve the implementation during the investigation. According to Craig et al. (2008), a lack of impact can be due to implementation failure, rather than ineffectiveness, which makes it necessary to evaluate the process to identify possible problems connected to the implementation. In other words, poor implementation could be the reason for a possible unobserved effect. There was continuous surveillance of the implementation process with telephone calls, visits, and more. The COSMOS-study used a mixed method study

⁵ The definition of clusters was later changed. This is further elaborated in section 5.1.1.

design to cover the whole process of implementation and assessment of the intervention (Husebo et al., 2015).

2.2 Intervention and Implementation

The implementation happened between baseline and 4-month follow-up and had two stages. First, health personnel such as nurses, NH managers, licensed practical nurses or physicians from each NH unit participated in a two-day seminar, where they were educated on COSMOS implementation domains (activity, pain assessment, communication, medication review and safety). The seminar included written guidelines, case discussions, repeated theoretical and practical training, and role play. Nurses attending the seminar became the COSMOS-ambassadors (Aasmul et al., 2018). Second, the ambassadors implemented the intervention in the NH wards by educating and informing their colleagues, if possible multiple times a week. Each week the focus was on one of the implementation domains mentioned above. Additionally, a midway one-day evaluation, regular telephone calls, and a telephone hotline where NH staff could reach out if they had a concrete question related to data collection or internal education was offered to the intervention group. Medication reviews were completed in collaboration between the NH physician, NH staff and two COSMOS-researchers. The control group received care as usual and was monitored by telephone contact on a monthly basis. To motivate the control group to continue participating, they received the COSMOS-intervention after month 9 (Husebo et al., 2015).

2.3 Outcomes

The primary outcome in COSMOS is QoL in NH patients and secondary outcome measures include neuropsychiatric symptoms (NPS), pain, activities of daily living (ADL), sleep, depression, hospital admission, medication, communication, staff distress and mortality measured at baseline, and 4- and 9-month follow-up (Husebo et al., 2015).

4.2 The sample data

4.2.1 Assessment of the variables

Dementia was assessed with the Mini Mental Status Examination (MMSE) questionnaire, a tool to assess cognitive status. It contains 20 questions, including the categories orientation in time, orientation in place, immediate recalling/recognition, repetition, apprehension, reading, and writing/generating sentences. Low score indicates low cognitive status, while a high score indicates high cognitive status (Strobel and Engedal, 2008). In my analysis, the cut-of score for identifying NH patients with dementia was a MMSE total score < 25 , which identifies mild to severely cognitively impaired patients. This complies with the Norwegian MMSE manual (Strobel and Engedal, 2008) and the COSMOS-study (Husebo et al., 2015).

The instrument Quality of Life in late-stage Dementia (QUALID) was developed by Weiner et al. (2000) to assess QoL in people with late-stage dementia residing in long-term care facilities. As late-stage dementia patients have difficulty expressing themselves, a proxy rate is necessary. With QUALID, QoL is assessed by next of kin or health personnel that deal with the patient on a daily basis, and should be considered on the basis of concrete, observable behavior and emotional expressions of the patient. It is an 11-item, 5-point scale, with a total score range of 11-55, where lower score indicates higher QoL. The Norwegian version has been validated and is perceived as reliable (Røen et al., 2015).

The form Resource Utilization in Dementia – Formal Care (RUD-FOCA) was developed to serve as a standardized tool to measure the direct care time required in the nursing home. It measures direct care time spent the last four weeks, in the three categories activities of daily living (ADL), instrumental ADL (IADL), and supervision, for both formal- and informal caregivers. ADL includes care tasks like toileting, bathing, hygiene, and eating, which I have chosen to call “general care”. IADL includes care tasks like taking medicine, go to the hairdresser, and doing administrative tasks, which I call “everyday tasks”. Direct care time includes for example the time a health personnel spends on helping a wandering patient back to his room, but not time spent in a quiet ward (Luttenberger and Graessel, 2010). According to Luttenberger and Graessel (2010), RUD-FOCA is validated as a

suitable instrument for recording the direct care time in NHs, even though IADL needs improvement. In my thesis, time use will be the direct care time for each patient in the last month, measured in hours.

There are different ways of defining psychotropic drugs. I have defined psychotropic drugs as the psychoactive drugs: antipsychotics, antidepressants, anxiolytics, hypnotic/sedatives, and anti-dementia, which is in accordance with earlier studies (Røen et al., 2017, Helvik et al., 2017). My thesis does not take dosage or length of prescription into account, because it is outside the scope of this text.

4.2.2 Weaknesses

The sample are people with dementia. This leads to much of the data collected being secondhand-information, because many of the forms are not filled out by the patients themselves, but the responsible caregiver. This can be problematic because it has been found differences between what the patient reports and what the next-of-kin or health personnel reports when filling out forms regarding QoL (Sloane et al., 2005).

The intervention is measured only over nine months. If the patients were followed for a longer period of time, more information could be collected, and more precise estimates could be found. To follow the individuals until death, would be ideal. However, there are resource constraints that must be considered.

Further, there was no information collected regarding how much the health personnel learned throughout the intervention. So, there was no quantifying measure of how successful the implementation of the intervention was. Advantageously, this could be collected shortly after the intervention, to get information on to what degree the health personnel's knowledge increased. If the health personnel did not increase their knowledge, or only slightly, one could maybe not expect an effect of the intervention either.

4.2.3 Strengths

The data are longitudinal, and not observational. This means that we follow the same individual over time, which enables us to uncover causal relationships, and not just

associations. This is important because we want to know what the intervention *causes*, and not just what it is associated with. In an observational study, it is possible that other factors not accounted for causes the outcome. This is what we are trying to get around by using longitudinal data. The randomized control and intervention groups also makes it possible to identify the causal effect of the intervention.

When studying elderly people with dementia, many patients often die during the study period because of their frail health condition and old age, or they move. In my sample, no patient died or moved during the three data-collection periods in neither of the groups. That the sample is made up of the same patients during the whole study is a strength, since we know that the same people are compared before and after the intervention is implemented. We also know that a change in one of the outcome measures cannot be due to many people dying; for example, if QoL is increasing from month four to month nine, we know that it is not because patients with poor health and QoL died in-between data collection in months 4 and 9.

4.2.4 Missing values

In a study with many questionnaires, missing values is difficult to avoid. Proxy-raters can have trouble knowing what to answer or can be occupied with care duties in the middle of the procedure, and thus forget to fill in the appropriate answer. How to treat missing values, is a large and complicated field. My approach is to exclude all patients with missing values in variables used in the sample selection process, e.g. the measure for QoL and cognition. This was chosen because inclusion of patients with missing values could have led to incorrect inclusion of patients. An example can be derived from the process of identifying people with dementia. If patients with missing values were to be included while calculating total MMSE-scores, this could have overestimated the number of people with lower cognitive status: if a patient with a missing value had a score of 24, this missing value could give a point if filled out properly and thus make this patient actually go over the cut-off score. If this is the case, the patient is falsely included in the study. To ensure that only correctly

specified patients are included, all patients with an MMSE missing value are excluded from the study. It was only necessary to exclude patients with missing MMSE values in the baseline data, since the baseline score was used to identify the level of cognition. Also, patients with missing values in outcome variables were excluded. This was chosen so that the same patients were compared in all model specifications. There are missing data in the other measurement variables as well, but these can be believed to be missing at random: If the proxy-rater was interrupted, e.g. due to sudden care duties, in the middle of the procedure of filling out the questionnaire which then led to missing values, it is likely that the interruptions happened at random. Regarding how raters can have difficulties answering certain questions, this is more likely non-random and could lead to selection bias. To exemplify, take for instance the measurement of QoL. If the raters consistently find it difficult to answer if the patient has low QoL, but not high QoL, there would be much missing data when a patient has low QoL. This can further lead to little information about patients with low QoL, and the estimation of change of QoL over time can be biased.

5. Method

This chapter describes the method used in detail. The advantages of randomized controlled trials and randomization are discussed, but it is also questioned whether randomization is enough for drawing causal conclusions. Further, the descriptive statistics at baseline gives an insight into the successfulness of the randomization. To estimate the average treatment effect on the treated on the outcome variables, the chosen estimation strategy is a fixed effects regression model.

5.1 Randomized controlled trials

A randomized trial studies a sample of people from the same underlying population. A certain treatment is given to a randomly chosen subset of these people, and the rest carry on without (or with only partly) treatment. The key is the random assignment, making the comparison of groups *ceteris paribus*: the only difference

between them is the treatment status (Angrist and Pischke, 2015). In the COSMOS-study, the treatment was the COSMOS-intervention, and the control group carried on with care as usual. Wards, and not individuals, were randomized due to a high probability of spill-over effects if randomization was performed on an individual level. The NH wards in each municipality were randomized after they were recruited. NHs in the municipalities Sund, Kvam, Fjell and Øygarden were recruited and then the NH wards here were randomized. The same process was then repeated in Bærum, Sarpsborg, and Bergen. This ensured that the intervention and control group contained an equal number of wards that are considered rich/poor and rural/urban. This makes the two groups more similar to each other before the experiment begins and can minimize the random error due to systematic differences, and increase the internal validity. To ensure randomness, ideally, all NHs should be recruited and then randomized, but due to budgetary reasons and time-constraints the process described above was chosen.

Two randomly chosen groups are similar, as long as the sample is large enough. This is due to the powerful statistical property Law of Large Numbers (LLN), stating that a sample average can be brought close to the population average, simply by enlarging the sample. This ensures that systematic differences due to, for instance, individual characteristics like gender, will wash out (Angrist and Pischke, 2015). If a variable Y_i is generated through a random process, like the throw of a die, the population average $E[Y_i]$ is the average if this process is repeated infinitely. $E[Y_i]$ is called the mathematical expectation of Y_i . If Y_i comes from a sample survey, $E[Y_i]$ represents the average obtained if every person in the population from which the sample was drawn were included (Angrist and Pischke, 2015). In my sample, there are 154 people in the treatment group and 120 people in the control group. This is not a large sample size, and it can be questioned whether the property of LLN will be put into effect.

5.1.1 Is randomization enough?

Perfect randomization is difficult to achieve. Random samples are not perfect samples, and there will always be sample errors (Mitchell and Jolley, 2012).

Randomization done on an entire population, would be very reliable, as the group studied would both be very large and representative. Unfortunately, this is not possible, and randomization must be done on a smaller sample of the population. The sample might not always be a good representation of the population, especially so for smaller samples. For example, it can by chance be older or generally sicker than the population. This leads to systematic bias despite randomization.

This is further elaborated with inspiration from *Mastering Metrics* by Angrist and Pischke (2015). Let Y represent the outcome. To distinguish between the individuals in the control and intervention group, let Y_{0i} denote the outcome if individual i is in the control group, and Y_{1i} if he is in the intervention group. The difference in the outcome averages without randomization can be written as:

$$\textit{Difference in group means} = \textit{Average causal effect} + \textit{Selection bias}$$

Selection bias is here the difference in average Y_{0i} between the intervention and control group. In other words, the difference between the two groups in the outcome variables that would have been there even without treatment, because the difference was present before treatment. With random assignment on the other hand, we expect the two groups to be equal in both what can be observed, and what cannot be observed due to the random assignment process. Let D_i denote treatment. If treatment is randomly assigned, $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, and the expected outcome for individual i when not treated would be the same regardless of which group the individual belongs to. Then the differences in expectations by treatment status captures the causal treatment effect κ :

$$\begin{aligned} & E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \end{aligned}$$

$$\begin{aligned}
&= E[Y_{0i} + \kappa | D_i = 1]^6 - E[Y_{0i} | D_i = 0] \\
&= \kappa + E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0] \\
&= \kappa
\end{aligned}$$

Thus, a successful random assignment can eliminate selection bias, and uncover the causal effect – as long as the sample is large enough for the LLN to kick in. This process does not eliminate individual differences, but makes sure that the mix of individuals are equal in both groups (Angrist and Pischke, 2015). In my thesis, wards have been randomized, not individuals. Therefore, we expect the individuals to be equal on average across the wards in the two groups. Even though random assignment can eliminate selection bias, it does not ensure that all selection bias is eliminated. Though intuitive in theory, in practice it can be very difficult to ensure that the control and intervention group are very similar, that the observed effects are generalizable and that the effects are a function of only the intervention itself (Khandker et al., 2010). In my sample, there are some problems challenging the randomization. One problem, already mentioned, is that my sample is small, and hence that it is questionable if the properties of the LLN kicks in.

Another challenge is that the control group received information about the content and aim of the COSMOS-intervention. This was part of each NH’s process of deciding whether they wanted to participate or not. And after randomization, as part of the data collection, the control group staff also filled out and was trained in the various questionnaires used in the study (Husebo et al., 2015). Therefore, the control group can have gotten the equivalent of “partial treatment”, even though they should have continued with care as usual. The researchers in the COSMOS-study are aware of this, and state that “the control group may also derive a considerable learning effect” (Husebo et al., 2015, p.5).

⁶ $Y_{1i} = Y_{0i} + \kappa$, or ⁶ $Y_{1i} - Y_{0i} = \kappa$ because the difference between the outcome when treated (Y_{1i}) and the outcome when not treated (Y_{0i}) is the constant effect of the treatment κ . This is called the constant-effect assumption. If treatment is randomized, κ is the causal effect of treatment.

The control and intervention groups can be located in the same NH, since the randomization is performed on a NH ward level. Further, the same staff could work in both intervention and control wards. This is related to the definition of a ward being different between the NHs; in one NH a ward could include three groups or corridors, while this would be three separate wards at another NH. Therefore, a unifying definition of a ward was made, stating that a ward consists of patients sharing kitchen and living room. These factors can have led to spill-over effects on the control group.

Lastly, the researchers cannot force anyone to take the treatment or not, which basically means that the researcher have a lack of control. Randomizing a sample into treatment and control groups, does not ensure that everybody in the treatment group gets treatment, nor that nobody is treated in the control group. After being assigned to one group, the participant can refuse to participate, or try to get the other treatment some other way. In the COSMOS-study, the control group was offered the intervention nine months after they were invited to participate, since the control condition was “care as usual”. This can have led more NH wards assigned to control to withdraw from participation, than NH wards assigned to intervention. On the other hand, it could also ensure and encourage the NH wards to join the intervention, since they were ensured to get the intervention later. Further, there is no way of making sure that the people in the intervention wards actually did what they were supposed to when the researchers or COSMOS-ambassadors were not there.

So, random samples are not perfect, and validation of the data and randomization must be performed. The first and most important thing to do, is to perform a balance test on the baseline data.

5.2 Descriptive statistics

5.2.1 Balance tests

It is reasonable to believe that the outcome variables are affected both by *who* you are as a person and *which* ward you are placed in. There can be systematic differences between the wards and individuals in the control and intervention group, which could be present before the intervention. This implies that a potential difference observed after the intervention would have been observed also without the intervention. Hence, it is important to detect differences between the control and intervention group before treatment on both the individual- and ward level. Thus, we must perform a balance test on the baseline data. The balance checks should show non-significant results to support a successful randomization. This further supports that the patients in the two groups are similar in both observed and unobserved differences. For this cause, the t-test has been calculated for the various variables. In Table 2 the patient descriptive statistics are displayed, and in Table 3 the ward descriptive statistics can be seen.

Table 2 Patient: Descriptive Statistics at Baseline

	Scale	Intervention		Control		Difference	
		Mean	sd	Mean	sd	p-value	t-value
Demographics and health							
Age	year	86.64	(7.11)	86.55	(7.28)	0.916	-0.11
Men	proportion	0.26	(0.44)	0.23	(0.42)	0.509	-0.66
Married	proportion	0.27	(0.44)	0.19	(0.40)	0.149	-1.45
Weight	kg	64.80	(14.24)	63.89	(13.41)	0.603	-0.52
Height	m	1.64	(0.10)	1.64	(0.09)	0.832	0.21
Blood pressure		129.55	(21.09)	125.05	(19.89)	0.085	-1.73
Blood pressure		72.14	(11.24)	71.15	(11.81)	0.496	-0.68
Pulse		71.91	(12.23)	72.84	(13.23)	0.565	0.58
Scores from COSMOS-study							
Formal care (RUD-FOCA)	hrs/month	49.32	(48.88)	52.64	(46.83)	0.571	0.57
ADL		32.50	(25.67)	33.39	(27.48)	0.783	0.28
IADL		9.16	(14.79)	10.72	(12.40)	0.356	0.93
Supervision		7.66	(25.87)	8.53	(16.44)	0.748	0.32
Informal care (RUD-FOCA)	hrs/month	4.16	(14.01)	2.21	(6.66)	0.162	-1.40
ADL		1.48	(6.10)	0.64	(2.80)	0.160	-1.41
IADL		1.82	(7.69)	1.17	(3.93)	0.397	-0.85
Supervision		0.85	(3.79)	0.41	(2.36)	0.262	-1.12

QoL-score (QUALID)	11-55	20.99	(7.39)	21.13	(7.46)	0.884	0.15
Cognitive status (MMSE)	0-30	10.32	(6.67)	11.45	(7.19)	0.182	1.34
Pain-score (MOBID)	0-10	2.32	(2.47)	2.55	(2.76)	0.499	0.68
Agitation-score (CMAI)	29-203	41.77	(16.15)	42.03	(14.82)	0.893	0.14
Activities-of-daily-living-score	0-30	20.00	(11.04)	20.12	(12.00)	0.933	0.08
Depression-score	0-38	6.74	(5.96)	6.96	(5.74)	0.779	0.28
Neuropsychiatric symptoms-score	0-120	14.62	(17.26)	14.05	(17.95)	0.798	-0.26
Caregiver distress-score	0-50	7.77	(7.12)	7.24	(8.28)	0.606	-0.52
Psychotropic drug use							
Antipsychotics	n	0.18	(0.45)	0.14	(0.42)	0.449	-0.76
Anxiolytics	n	0.18	(0.39)	0.26	(0.54)	0.174	1.36
Hypnotics and sedatives	n	0.27	(0.46)	0.40	(0.59)	0.045*	2.01
Antidepressants	n	0.44	(0.60)	0.48	(0.56)	0.500	0.67
Anti-dementia drugs	n	0.16	(0.39)	0.13	(0.34)	0.518	-0.65
Total drug use	n	1.23	(1.12)	1.42	(1.19)	0.193	1.30
<i>N</i>		154		120		274	

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

Notes: The table shows the t-statistics of mean differences in the intervention and control group for different patient characteristics and scores at baseline. Standard deviations in parentheses.

As we can see in Table 2, patients in the control and intervention groups are quite similar, and there are mostly non-significant differences. Blood pressure and use of hypnotics and sedatives are significantly different on a 0.05 % and 0.10 % significance level, respectively. But in a table with many comparisons, a few isolated differences that are statistically significant are often also likely to chance (Angrist and Pischke, 2015). The patients are on average 86 years old, in other words a relatively old sample, and most of the patients are female (approximately 75 %) and not married or widowers (approximately 75 %). The average MMSE-score is around 11, which tells us that many of the patients in the sample has a far-advanced type of dementia. Most of the patients do not give the formal caregivers a large care burden, but are highly dependent in ADL-tasks as most of the formal time use is spent here. On average, the patients receive little informal care. Something worth noticing is that informal care, and to a smaller degree, formal care, psychotropic medication and some other variables, had different baseline levels. Even though the differences were

non-statistical, a problem could be that if the sample size increased, the differences could diverge and become significantly different.

Table 3 Ward: Descriptive Statistics at Baseline

	Scale	Intervention		Control		Difference	
		Mean	sd	Mean	sd	p-value	t-value
Patients	n	15.59	(6.88)	18.05	(8.42)	0.267	1.12
Staff, daytime	n	5.07	(2.21)	5.91	(2.41)	0.212	1.26
Staff, evening	n	3.63	(2.06)	3.66	(0.97)	0.951	0.06
Staff, nighttime	n	1.40	(0.65)	1.47	(0.66)	0.717	0.36
Weekend: Staff, daytime	n	4.21	(2.17)	4.55	(1.44)	0.537	0.62
Weekend: Staff, evening	n	3.62	(2.07)	3.61	(0.98)	0.986	-0.02
Weekend: Staff, nighttime	n	1.36	(0.62)	1.47	(0.66)	0.549	0.60
Doctoral visits, per week	hours	6.63	(6.39)	6.43	(7.80)	0.926	-0.09
Total staff	n	22.35	(10.94)	24.48	(9.19)	0.478	0.71
Registered nurses	n	5.31	(2.82)	5.77	(2.18)	0.535	0.62
Assistant nurses	n	10.93	(6.33)	11.33	(4.76)	0.807	0.25
Unskilled nurses	n	5.85	(4.81)	7.00	(4.63)	0.402	0.85
<i>N</i>		27		22		49	

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

Notes: This table shows the t-statistics of the mean differences between NH ward characteristics in the control and intervention group. Standard deviations in parentheses.

The ward descriptive statistics are non-significant, and we can see that the two groups are quite similar on the ward level. In general, there are more intervention wards than control wards. The wards in the intervention group are on average smaller, which is probably why there are systematically less staff here compared to the control group. Six of the wards are not included in the descriptive statistics due to missing values, where most misses all information.

There was also information on the wards' access to different personnel. The differences are mostly non-significant, though one (access to social workers) came up as statistically significant. The table can be found in Appendix 1. An interesting finding is that around 60 % of the wards have access to volunteers, while none of the wards in the intervention group, and only 15 % of the wards in the control group,

have access to a social worker. This indicates that social and other activities are mostly driven by the staff or volunteers like next-of-kin or volunteering organizations. Despite the small sample size, the intervention and control groups are quite similar, which supports that randomization was successful.

5.2.2 Distributions

Even though the average values in the groups are similar in most variables, the underlying distributions can be different. By looking at the distributions, one can inspect for worrisome differences, and other potential problems, like outliers, and evaluate if this needs to be considered in the analyses or discussion. This section discusses the dependent variables' distributions, which can be found in Appendix 2.

For formal care, the distributions are largely skewed to the left, with few relatively high time-consuming patients. In the intervention group there are a few more highly time-consuming individuals. In the control group, the distribution peaks at around 50 hours per month, while the peak in the intervention group are a little less than 50 hours. In general, the two groups have a relatively equal spread of observations. The distributions for informal care are also relatively equal. There are some patients in the intervention group that receive a very high amount of informal care compared to the others (between 50 and 90 hours per month), which can explain why the average of informal care is higher in the intervention group compared to the control group. But most patients receive little or no informal care. Also for QoL, the distributions are relatively similar and skewed to the left. Since the QoL instrument, QUALID, uses a scale where a low score means higher QoL, this means that the majority of patients have a relatively high QoL. The distributions of psychotropic medication are fairly similar for both intervention and control group. They are skewed to the left, with between 20 % and 30 % of the patients taking no psychotropic medication in both groups. The highest percentage (approximately 35 %) of patients in both groups use one psychotropic drug. Less than 15 % in the control group and 10 % in the intervention group takes three or more different psychotropic medications.

5.3 The road to causality

5.3.1 Impact evaluation

This section is largely inspired by Khandker et al. (2010). In order to say something about how successful an intervention is, we want to uncover causal effects. This means to isolate the effect that is caused by the intervention, which would not have been there without it. Ideally, we want individual i to both get the intervention, and not get the intervention. Afterwards, we would subtract the differences in outcome and get the causal effect of treatment. Of course, this is not possible, since at a particular point in time, a person cannot both get treated or not treated. To get an approximation of the unobserved counterfactual outcome, we can estimate it. One way to do this by replacing the counterfactual outcome with a control group that has not been given the treatment.

We want to estimate the *average treatment effect on the treated* (ATET). Let there be two periods, one period before treatment (0) and one after (1). All individuals either get treated ($D=1$) or not treated ($D=0$). Outcome when treated is Y_1 , while outcome when not treated is Y_0 . Then we want to estimate

$$\alpha_{ATET} = E[Y_1(1) - Y_0(1)|D = 1]$$

This is an estimate for the difference in the outcome if treated and the outcome if not treated for the individuals in the intervention group. $Y_0(1)|D = 1$ is the counterfactual and unobservable outcome, and what we want to estimate. When individuals are successfully randomized to control and intervention groups, they are expected to be equal in every way, both in observed and unobserved factors. If this assumption holds, the outcome for the randomized non-treated patients is a good measure of the counterfactual outcome for the treated, and $Y_0(1)|D = 0$ is a valid counterfactual estimate of the outcome $Y_0(1)|D = 1$. A successful randomization leads to the observed outcomes in the control group being a counterfactual outcome for the intervention group if they did not get the COSMOS-intervention. When the groups are similar before treatment, if not for the treatment, they should also be similar in

the post-treatment period. A possible difference after treatment, will thus be due to the intervention.

5.3.2 COSMOS: Possible factors challenging the causality

As we have seen, a successful randomization of treatment to NH wards says that any differences in the average treatment effect between the treatment and control group are likely due to the COSMOS-intervention. This is further supported when the balance checks on both individual- and ward-level at baseline passed as mainly non-significant. However, perfect randomization and zero selection bias is difficult to achieve. Presented below are some factors in the COSMOS-study challenging the causality.

Firstly, the intervention was always implemented in the NH wards by different COSMOS-ambassadors. Therefore, it was not possible to control exactly how the implementation was given in each ward, and whether it was identically implemented. For example, there could be differences in how many days a week education in the COSMOS components (pain, medication review etc.) would be given to the colleagues by the ambassadors, and thereby differences in education thoroughness or focus. These implementation differences can impact which focus a ward has, and thus explain some of the differences we see between the intervention wards. Secondly, the participants couldn't be completely blinded regarding which group they were allocated to because of the hybrid trial design, which allows for optimization during the process of intervention, education, and follow-up (Husebo et al., 2015). This can produce bias in the effect of the intervention, since people can respond (consciously or subconsciously) differently to the treatment if they think or know they are getting treated versus not getting treated. This can be due to e.g. preexisting expectations on the treatment's effect (Schulz and Grimes, 2002). Another issue is the Hawthorne effect; that the health personnel change their behavior, either consciously or subconsciously, *because* they are being observed. People want to be perceived in the best way possible, and not seen as someone who is caring too little for the patients or do not spend enough time with them. Regardless, it was not possible to fully blind

the participants in this study. Thirdly, when participating in a study, people are not unbiased. The health care personnel probably want to see an effect and can more and less consciously try to affect the outcome in the desired direction. They may both want their patients to have a better QoL and also be part of something that made the patients feel better. So, during questionnaires in month 9, if they wish to see an effect, that is what they see, and record down.

All these factors can challenge the causality and is something to bear in mind when interpreting the results. The next section describes how α_{ATE} is estimated.

5.4 Data analyses

5.4.1 Panel Data

The COSMOS-data are panel, or longitudinal, data. Each patient is measured at three different points in time, making it possible to follow each individual over time. The ability to observe the same patients over time is an advantage, since it permits controlling for certain unobserved characteristics of the individual. This makes causality more believable than if two cross-sectional units were compared (Wooldridge, 2014).

5.4.2 Fixed Effects

If the intervention and control group are successfully randomized, we can expect both observed and unobserved differences to be equal in the two groups on average. In addition to the factors challenging the randomization and causality, there are many factors I do not have data on in my sample, like education, diseases, or former hospital admissions. Thus, pre-treatment levels of these variables, and others, could not be checked for balance in the two groups. Therefore, fixed effects were included to further control for unobserved heterogeneity.

The data are extracted from many different NHs and NH wards and places in Norway, and we can expect there to be fixed differences between the entities. These differences can explain some of the variance in the data collected. Differences can be due to the local socio-economic environment, geographical location, or the wellbeing

of staff and working environment on the various wards and NHs. Differences on an individual level can also be expected, such as education, income, pain tolerance, or general attitude towards life. The differences can be both observable and non-observable, be constant or vary over time, and can affect our dependent variables (Wooldridge, 2014), independent of treatment status, and must be taken into account when performing regression analysis. The problem of just running a pooled ordinary least squared (OLS) regression on panel data, is that the results would be biased if a time-invariant factor a_i (located in the error term u_i) is correlated with one or more of the independent variables. This would be a violation of the consistency of the OLS-estimators, so that $cov(u_i, x_{it}) \neq 0$, which is problematic. Examples of a_i in the context of COSMOS can be chronic diseases or living location of closest relatives. To illustrate the fixed effect model, a model from Wooldridge (2014) is presented. Consider a model with a single explanatory variable:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2, \dots, T \quad [1]$$

The notation i denotes the person, and t denotes the time period. The error term u_{it} , is often called the idiosyncratic or time-varying error, because it embodies the unobserved factors that are time-variant and affects the dependent variable (Wooldridge, 2014).

If this equation is averaged over time within each individual, we get

$$\bar{y}_{it} = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i \quad [2]$$

where $\bar{y}_i = \sum_{t=1}^T \frac{y_{it}}{T}$, and so on. The variable a_i captures all time-constant, unobserved factors that affects our dependent variable y_{it} , and is often referred to as a fixed effect, or individual effect. Because it is time-constant, a_i will be the same in both equations. If we subtract [2] from [1], we get

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{u}_{it} \quad [3]$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_{it}$, and so on. [3] is called the time-demeaned data on y , x and u . The method is called the *fixed effects transformation* or the *within transformation*.

The important thing to notice here, is that the unobserved effect α_i has disappeared when the data was time-demeaned (Wooldridge, 2014). This means that we have gotten rid of all unobservable factors that do not change over time and could have been difficult to control for. An estimation of [3] will use the time variation in y and x *within* each individual and thus not take into account the between-variation (Wooldridge, 2014). This is what we want, because differences between individuals are often not due to the treatment we are studying but are constant and present independent of treatment or not.

In my analysis, patient fixed effects are used to control for fixed differences between patients over time. This means that we look at changes within each person, and thus exploit the within variation each person “creates” as the variables measured change over time. As patients do not change ward or NH in the study period, we thus also control for ward and NH fixed effects. To control for time-invariant differences are important because differences between patients can affect the dependent variable and lead to omitted variable bias. By using fixed effect regression, at least some of this bias is reduced, although one can never be sure that it is completely “taken care of”.

In the fixed-effects regression, a critical assumption is that the unobserved heterogeneity is constant over time. In the context of the COSMOS study, the unobserved heterogeneity, such as chronic diseases or budget, are likely constant during the study period as our observation period is only 9 months. Even so, we cannot be sure that there has not been a change in for example chronic diseases during this period. If the unobserved heterogeneity is constant over time, the observed effect is more likely due to the intervention.

5.4.3 Average treatment effect on the treated

Since the effect of the intervention may be found by comparing the intervention and control group, I rely on a method that is methodologically very similar to a difference-in-difference (DiD) approach. It can formally be written as

$$y_{it} = \alpha + \beta_1 TREAT_i + \beta_2 POST_t + \alpha_{ATET} (TREAT_i \times POST_t) + \varepsilon_{it} \quad [4]$$

The dummy for treatment group is denoted $TREAT_i$. Since no patient changes treatment status during the intervention period, this variable is constant over time and will be omitted because of the fixed effect regression. The dummy for post-treatment period, $POST_t$, varies over time and controls for the fact that, whether treated or not, conditions change. The interaction term $TREAT_i \times POST_t$ is the ATET (Angrist and Pischke, 2015). Mathematically it can be expressed by

$$\alpha_{ATET} = (\bar{Y}_{intervention,post} - \bar{Y}_{control,post}) - (\bar{Y}_{intervention,pre} - \bar{Y}_{control,pre})$$

Though the method is the same, a DiD approach and causality interpretation will not be correct or reliable for this data set. This is because I cannot verify the identifying assumption for DiD; the common trends assumption, which tells us that in the absence of treatment, the two group-outcomes would move in parallel (Angrist and Pischke, 2015). Thus, prior to the introduction of a treatment the control and intervention group should follow a common trend in the variable inspected. To rely on DiD estimates, data on the outcome variables would have to be collected for multiple points in time before COSMOS was introduced for both groups. If they did not follow a common trend, we would not know if possible changes after treatment would be due to treatment, or something else that happened before treatment was introduced. For instance, if the intervention group was steadily upwards-trending in all time-periods before treatment, while the control group had a flat trend, some of the difference seen after treatment would most likely also have been observed without treatment. Consequently, a DiD approach has not been used, even though the methodology is the same. I will look at ATET over time, comparing the control and intervention groups in the post-treatment period. Even though I do not have information about the pre-treatment trends, patients are randomized to control and intervention groups. As mentioned, if randomization was successful, we can expect the two groups to be similar before treatment. Then observable differences after treatment should be the ATET due to the intervention. The balance checks show that only a few of the variables are statistically significant and the groups seem quite similar. This makes a successful randomization more believable and it is more likely

that the ATET is causal, especially after also controlling for fixed effects. Still, we have to remember the challenges discussed earlier.

5.4.4 Standard error adjustment

Clustered standard errors

In panel data, the standard errors can suffer from serial-correlation and intraclass correlation. The problems arise when following individuals over time, and if groups of people within the sample are similar to each other. As pointed out by Moulton (1990), non-clustering the standard errors on these groups can lead to a down-wards bias in the OLS standard errors, which should be dealt with and adjusted for.

Serial correlation is the tendency for one observation to be correlated with the observations in the periods before (Angrist and Pischke, 2009). If you observe a patient with depression and anxiety, it is more likely that you will also observe depression and anxiety in this patient in the next observation period. According to Angrist and Pischke (2009), the easiest and mostly used approach to this problem is to cluster on a higher level. In my case this means to scale up the clustering level from patient to NH wards, and from NH wards to NHs. The authors go on by saying that the reduced number of clusters is problematic. They claim that fewer than 42 clusters could be worrisome, though this number is very uncertain. In my sample, the number of clusters are reduced from 66 to 31 when changing cluster levels from NH wards to NHs. This is a significant decrease, which can lead to problems such as biased standard errors and misleading inferences (Angrist and Pischke, 2009). In statistics, it is generally preferred to cluster on the highest level possible where we believe there is in-group correlation. For this reason, I have chosen to cluster on NH level. To correct for bias in the standard errors due to too few clusters, I have used bootstrapped standard errors, described below.

In my study, patients from different NH are compared to each other. Observations within NHs are likely to be correlated with each other, and the standard errors should be adjusted to correct for this. The appropriate formula to use for this intraclass correlation is, again, the clustered standard errors, which allows for

correlated data within clusters defined by the researcher (Angrist and Pischke, 2009). This is sensible since we can assume that patients living in the same NH are more similar compared to patients living in other NHs, because they have equal conditions regarding NH economy, and socio-economic and geographical environment. Also, staff often interact or are involved in all or multiple wards in a NH.

Bootstrapped standard errors

The number of clusters needed to get unbiased standard errors is disputed, but my number of 31 is questionable. For this reason, a bootstrapping procedure can be used to obtain more accurate cluster-robust inference. Angrist and Pischke (2009) explains that Stata's "cluster" option relies on the asymptotic approximation of many clusters or time series observations. As the number of clusters are not asymptotically large, bootstrapped standard errors can be a better approximation to unbiased standard errors. Based on asymptotic formulas, bootstrapping is a resampling scheme and offers an alternative to inference. A bootstrap sample is a sample randomly drawn from our own data. We treat our sample N as the population and draw repeatedly from it, with replacement, and thus construct a sampling distribution for the "true" population N . When we treat N as the population, we know the "true" errors and can approximate the standard error of the true distribution with the errors of the resampled data. The bootstrapped standard error is the standard deviation of the resampled data (Angrist and Pischke, 2009).

More specifically, I have used block bootstrapping. Angrist and Pischke (2009) says that to preserve the clustered data's dependence structure, the block bootstrapping method can be used. This means to resample the data by randomly drawing blocks of data defined by the cluster group. In my data, it is not individuals that are resampled, but NHs. The ideal number of repetitions is disputed, but the main concern is to run enough repetitions for the asymptotic properties to be realized. I have thus used 1,500 repetitions to get an asymptotic approximation of the standard errors.

5.4.5 Model

The model used in the analysis is presented below. I estimate a fixed effects regression trying to identify the ATET of the COSMOS-intervention from baseline to the 9-month follow-up. This can be formalized in the following way:

$$\tilde{y}_{it} = \tilde{\beta}_{it}TREAT_i + \tilde{\gamma}POST_t + \tilde{\alpha}_{ATET}(TREAT_i \times POST_t) + controls + \tilde{u}_{it}, \quad t = 1,2,3 \quad [5]$$
$$i = 1, \dots, 275$$

The dependent variable \tilde{y}_{it} is QoL, number of psychotropic medication, and formal- and informal time use. Cognition score was the only time-variant control variable that could not be suspected to be an outcome of the intervention and were therefore the only control variable included in the full model.

Additionally, I look at the different components of formal and informal care separately. A sub analysis on the change in the care components ADL, IADL and supervision has been performed to better understand how the amount of formal- and informal care has changed. This can enlighten where the COSMOS-intervention has had the greatest effect. For instance, since ADL and IADL is where people with dementia are most dependent, an effectivization of ADL and IADL can be more important than that of supervision. Further, since the intervention was implemented between baseline and the 4-month follow-up, we can expect a larger effect between these points in time, than between 4- and 9-month follow-up. Therefore, separate interaction terms for 4- and 9-month are included.

6. Results

In this chapter the results from the analyses are presented. Each outcome gets a visual introduction, before the fixed-effects regression results are presented. For the first two outcomes, four estimation strategies leading to the full model are presented for the sake of comparisons. Issues that challenge the causality are described, and various robustness tests are performed in an attempt to identify sources of potential bias.

6.1 QoL

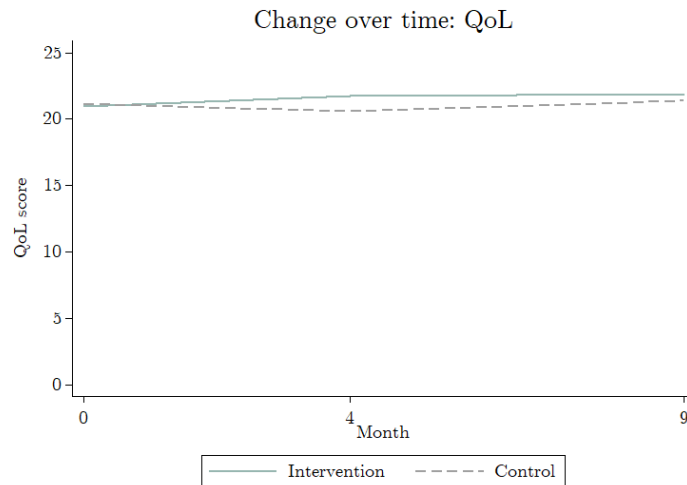


Figure 2 Change in QoL

When visually inspecting the change in QoL in the two groups (Figure 2), it is apparent that the QoL-score only has a marginal change from baseline to month 9. In the second follow-up, the average QoL score slightly increases (QoL slightly worse) compared to baseline. Nevertheless, over time the QoL-score has been almost constant in both groups, and there are no apparent differences on average.

Figure 2 can only tell us the descriptive association since it only displays the change in raw data. In Table 4, the estimation strategies attempting to find the causal effect of treatment are presented. To visualize how controls and fixed effects affect my results, different estimation strategies of how COSMOS has affected the patient's QoL is presented. The estimation strategy in column (4) represents the full model and is preferred because fixed effects controls for time-invariant unobserved heterogeneity and including the time-variant control variable 'development in dementia' controls for a possible endogenous variable affecting our outcomes.

Table 4 ATET of COSMOS on QoL

	(1)	(2)	(3)	(4)
After	-0.154 (0.622)	0.677 (1.103)	-0.274 (0.607)	
Interaction	0.969 (0.862)	-0.234 (1.233)	0.920 (0.855)	

Intervention	-0.131 (1.009)	-0.0965 (1.204)		
Month 4-Interaction			1.266 (0.909)	
Month 9-Interaction			0.578 (0.892)	
After: Month 4			-0.642 (0.674)	
After: Month 9			0.105 (0.641)	
Constant	21.13*** (0.746)	13.36** (6.524)	22.11*** (0.990)	22.06*** (0.950)
Observations	822	411	822	822
Controls	No	Yes	Yes	Yes
FE	No	No	Yes	Yes

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

Note: Cluster bootstrapped standard errors in parentheses.

This table presents various model specifications of mean difference over time on QoL. The first three columns include an interaction term where we look at the post-treatment period as one period, while in column (4) we look at each post-treatment period separately. Column (1) shows the results with no control variables or fixed effects. In column (2), time-invariant and time-variant control variables are included, and in column (3) and (4) time-variant control variables and fixed effects are included. Time-invariant control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, marital status, age and sex. The time-variant control is change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

When testing for differences in the intervention and control groups, no significant change in ATET on QoL can be seen from baseline to the 9-month follow-up; not with either estimation strategy. Including control variables and fixed effects do not increase precision. Since measuring QoL in people with dementia can be problematic in several ways, I have done a separate analysis on depression score, which can represent an objective measure of a person's QoL. This association has also been found in the literature (Barca et al., 2011). The QoL-score and depression score were strongly correlated (0.6417). I also considered pain score, but the correlation between QoL and pain score was poor (0.2913). The regression results of ATET of depression score also shows non-significant results and can be found in the Appendix 3. This supports that there has been no a change in the measured QoL of the patients after the intervention.

6.2 Formal care

Change over time

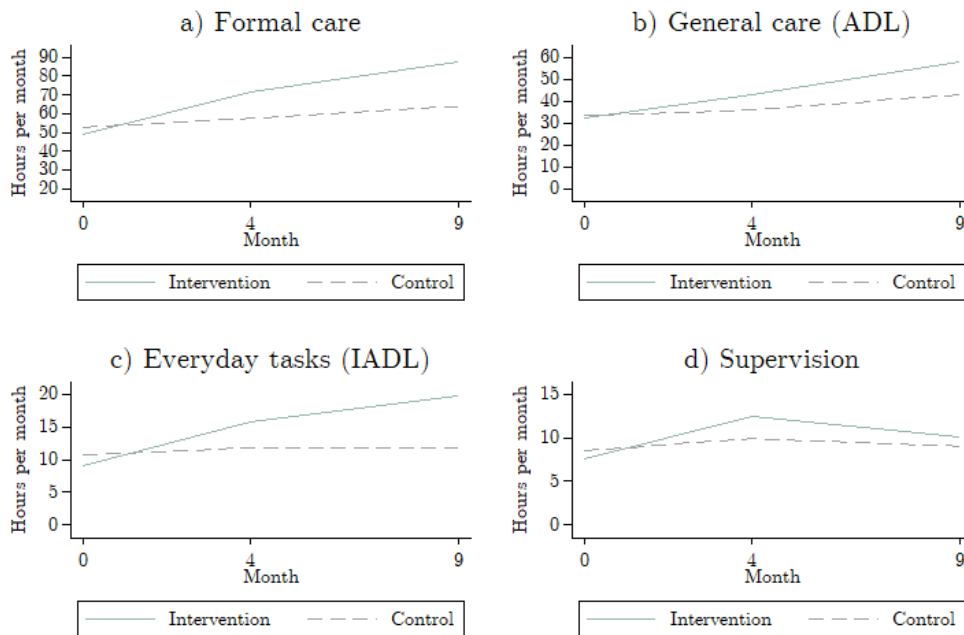


Figure 3 Change in formal care

In Figure 3, the development in formal care and the different components of formal care can be seen. a) displays the formal care in total, while b)-d) divides the total time use into the different components. A general tendency is that time use in the intervention group is increasing from baseline to the 9-month follow-up, while it in the control group is more constant. a) shows that the change in use of formal care in the intervention group is drastic; from baseline to the second follow-up, the mean increases from around 49 to 88 hours per month. The control group also has an increase in mean formal care, although much smaller, going from around 53 to 64 hours per month. As suspected, general care is the component taking most of the formal caregivers' time. In d) it is evident that supervision is affected the least in change over time and takes the least of the formal caregivers' time. In the intervention group, formal care used on IADL has increased a great deal, as is evident in c), while it has stayed almost constant in the control group.

Table 5 ATET of COSMOS on hours of formal care per month

	(1)	(2)	(3)	(4)
After	8.128 (6.476)	7.447* (4.167)	5.145 (5.939)	
Interaction	22.27* (13.03)	16.73** (8.317)	21.04* (11.34)	
Intervention	-3.314 (7.381)	1.667 (5.832)		
Month 4-Interaction				15.95 (11.73)
Month 9-Interaction				26.25* (14.27)
After: Month 4				2.725 (7.385)
After: Month 9				7.847 (6.439)
Constant	52.64*** (5.357)	31.65 (40.96)	77.11*** (9.989)	75.87*** (10.10)
Observations	822	411	822	822
Controls	No	Yes	Yes	Yes
FE	No	No	Yes	Yes

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

Clustered bootstrapped standard errors in parentheses.

Note: This table presents various model specifications of the ATET on formal care. The first three columns include an interaction term where we look at the post-treatment period as one period, while in column (4) we look at each post-treatment period separately. Column (1) shows the results with no control variables or fixed effects. In column (2), time-invariant and time-variant control variables are included, and in column (3) and (4) time-variant control variables and fixed effects are included. Time-invariant control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, marital status, age and sex. The time-variant control is change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

In Table 5, the ATET on formal care is presented. After including control variables, the number of observations drop substantially, due to missing data. The interaction terms are significant on either a 0.10 or 0.05 level using the different estimation strategies, except for the 4-month interaction in column (4). The full model shown in column (3) tells us that formal care per month has on average increased by over 20 hours per patient in the intervention group compared to the control group. This translates into a 43 % increase in time use. In column (4), it is evident that only in the 9-month follow-up a significant change is found, with an increase in time use

equivalent to 53 %. All the following regressions include fixed effects and time-variant control variables.

Table 6 ATET of COSMOS on components of formal care

	(1) General care (ADL)	(2) General care (ADL)	(3) Everyday tasks (IADL)	(4) Everyday tasks (IADL)	(5) Supervision	(6) Supervision
After	4.097 (2.893)		0.587 (1.509)		0.461 (2.098)	
Interaction	11.14* (6.763)		7.357** (2.983)		2.543 (3.587)	
Month 4- Interaction		7.278 (6.699)		5.393** (2.540)		3.277 (4.405)
Month 9- Interaction		15.12* (8.420)		9.342** (4.587)		1.789 (4.052)
After: Month 4		1.050 (3.535)		0.704 (1.867)		0.971 (2.750)
After: Month 9		7.426** (3.334)		0.521 (1.533)		-0.0997 (3.229)
Constant	50.96*** (7.625)	49.72*** (6.837)	14.07*** (2.038)	13.85*** (2.186)	12.08*** (2.998)	12.30*** (3.062)
Observations	822	822	822	822	822	822

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

Clustered bootstrapped standard errors in parentheses.

Note: This table presents the ATET on the three components of formal care. Columns (1), (3), and (5) include an interaction term where we look at the post-treatment periods as one period, while in columns (2), (4), and (6) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

The analysis of formal care components shows some interesting results and can be seen in Table 6. In line with the literature, the patients are most dependent in care-tasks related to ADL, or general care, with about four times as much time use compared to IADL and supervision. The results indicate that there is an increase of around 11 and 7,5 hours per month in formal care used on ADL-related and IADL-related tasks, respectively, in the intervention group compared to the control group. The results further suggest that there has been a change in IADL-tasks throughout the intervention, while in ADL-related tasks, only after the 4-month follow-up. There is no evidence of a significant change in time devoted to supervision. This suggests that changes in formal care is mostly driven by changes in IADL and ADL. Due to

higher precision, we can be more certain about the effect on COSMOS on changes in IADL, than ADL.

6.3 Informal care

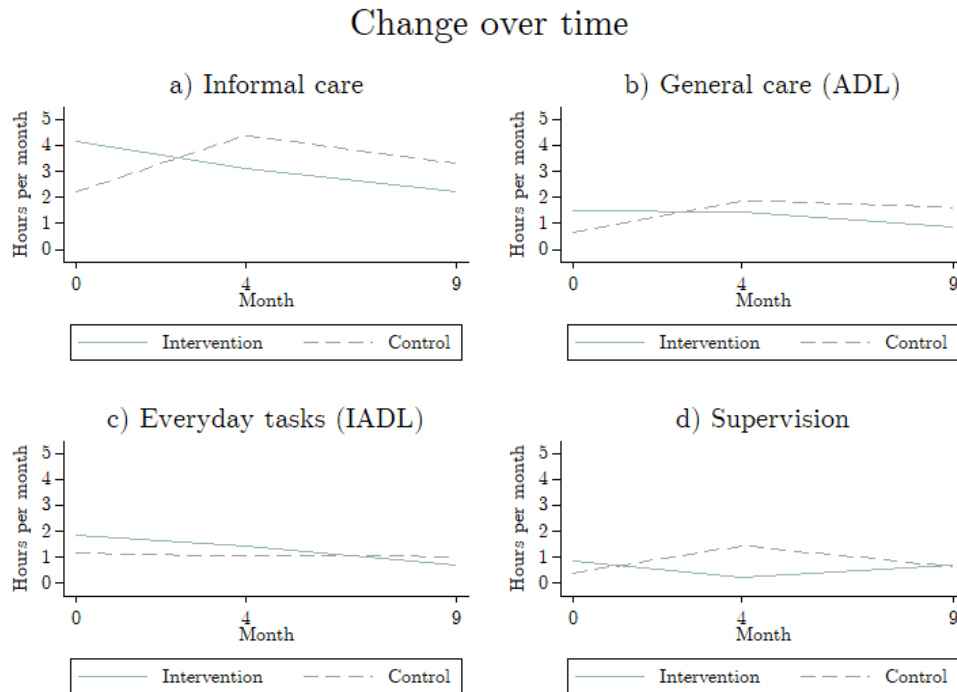


Figure 4 Change in informal care

Looking at Figure 4, it is apparent that informal care has changed over time. In a), we can see that the intervention group has a steady downwards sloping trend after the COSMOS-intervention. On the other hand, the control group has a mean increase of around 1 hour per month from before the intervention at baseline to the second follow-up. In b) and c) we can see that the amount of general care and everyday tasks performed in the intervention group has a downwards sloping trend, and has on average decreased by about 0.75 hours and 1 hour per month per patient, respectively. The development of time spent on supervision, which can be seen in d), is almost constant from baseline to 9-month follow-up, with a slight decrease in month 4. For the control group, informal care spent on everyday tasks, are constant over time. On the other hand, for general care and supervision, time used on informal care has increased somewhat on average. It is important to notice that hours of informal care have different pre-treatment levels in the two groups. In order to

identify a causal effect in this case, the estimation strategy used might not be sufficient. Additionally, information about pre-treatment trends are needed: We do not know if hours of informal care in the intervention group was already on a downwards sloping trend before treatment, and the opposite for the control group. Thus, the causality of the ATET should be interpreted with caution.

Table 7 ATET of COSMOS on components of informal care

	(1) Total informal care	(2) Total informal care	(3) General care (ADL)	(4) General care (ADL)	(5) Everyday tasks (IADL)	(6) Supervision	(7) Supervision
After	1.496** (0.674)		1.016** (0.428)		-0.172 (0.352)	0.652** (0.274)	
Interaction	-3.196*** (1.201)		-1.501** (0.654)		-0.669 (0.640)	-1.018** (0.443)	
Month 4-Interaction		-3.332** (1.574)		-1.366 (0.940)			-1.655** (0.686)
Month 9-Interaction		-3.072** (1.476)		-1.640** (0.648)			-0.383 (0.671)
After: Month 4		2.073* (1.225)		1.181* (0.710)			1.052* (0.602)
After: Month 9		0.890 (0.798)		0.838** (0.376)			0.249 (0.383)
Constant	4.701** (1.947)	4.825** (1.893)	2.041** (0.943)	2.099** (0.967)	2.038*** (0.587)	0.618 (0.775)	0.635 (0.789)
Observations	822	822	822	822	822	822	822

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

Clustered and bootstrapped standard errors in parentheses.

Note: This table presents the ATET on informal care and components of informal care. Columns (1), (3), (5), and (6) include an interaction term where we look at the post-treatment periods as one period, while in columns (2), (4) and (7) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

The regression of ATET in Table 7 shows that there with high precision is a decrease in hours of informal care of over 3 hours per patient per month on average in the intervention group compared to the control group. This amounts to an 80 % decrease from baseline levels. There is a significant change in the amount of informal care spent on general care and supervision: After the intervention, family and friends use about 1.5 hours less on help with general care-activities such as toileting, feeding and bathing, and around 1 hour less on supervision per patient per month. As for formal

care, informal care tasks related to ADL do not have a significant change before after the 4-month follow-up. For supervision on the other hand, the only significant finding is in month 4. More than 1,5 hours less informal care per month has been spent on ADL in month 9 and supervision in month 4. No significant change is seen in IADL-related tasks. The coefficient “after” shows that the control group has a significant increase of informal care going from baseline to post-intervention, except for in tasks related to IADL.

6.4 Psychotropic medication

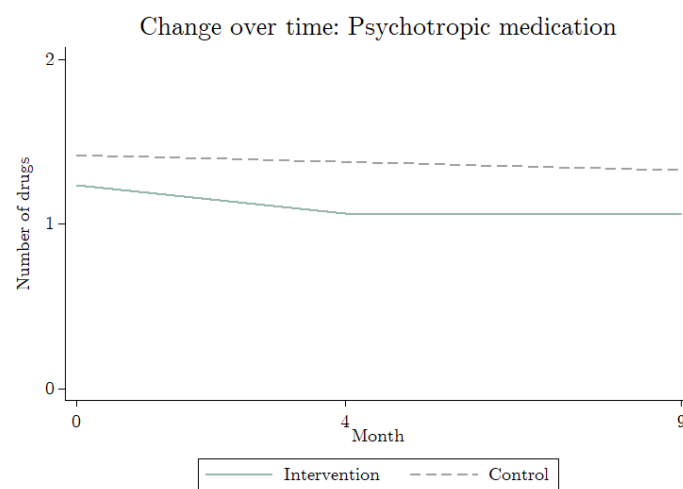


Figure 5 Change in psychotropic medication

The first impression of change in psychotropic medication when examining Figure 5, is that use of medication decreases slightly in the intervention group, while staying relatively constant in the control group. The biggest impact in the intervention group happens from baseline to the first follow-up around month 4, while it from then to the 9-month follow-up almost does not change at all.

Also here the pre-treatment levels are different, although not to a great extent. Even so, information of pre-treatment trends would be useful, and the causal effect must be interpreted with caution. The ATET of COSMOS on use of psychotropic medication can be seen in Table 8. The results showed a statistical significant decrease in use of psychotropic medication from baseline to month 4 in the intervention group compared to the control group. This indicates that the intervention has decreased the

use of psychotropic medication in the first follow-up month, but not further after month 4. The average use in the intervention group in month 4 decreased by 0.126, giving a decrease in use of psychotropic medication of over 10 % from baseline to 4-month follow-up.

Table 8 ATET of COSMOS on use of psychotropic medication

	(1)	(2)
After	-0.0638 (0.0710)	
Interaction	-0.104 (0.0788)	
Month 4-Interaction		-0.126** (0.0609)
Month 9-Interaction		-0.0827 (0.105)
After: Month 4		-0.0397 (0.0438)
After: Month 9		-0.0887 (0.105)
Constant	1.289*** (0.145)	1.292*** (0.137)
Observations	822	822

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

Clustered bootstrapped standard errors in parentheses.

Note: This table presents the ATET on use of psychotropic drugs. Column (1) include an interaction term where we look at the post-treatment periods as one period, while in column (2) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

6.4 Possible challenges and robustness tests

6.4.1 Possible challenges

It is challenging to control for everything that can affect the results, and there is probably omitted variable bias (OVB) in this model. Although the fixed effects regression controls for time-invariant omitted variables, like discussed in section 5.4.2, there may still be time-variant variables that are not controlled for. If the time-variant omitted variables affect the intervention and control group differently, the causal effect of the intervention is biased. One example is how the general health

changed after the intervention period, where information was available, though only after the intervention. It turns out that the patients in the control group had a higher health deterioration than the patients in the intervention group during the intervention period. If the patient had a worsening or improvement in their health, this could affect their QoL or time use regardless of the intervention. When health is declining, the patient becomes frailer, needs more medication, or more supervision, and we can expect the patient to become more dependent on his or her caregivers. Therefore, it is possible that some of the difference observed between the groups, is in fact due to differences in general health. Since the patients in the control group had a bigger health deterioration, the QoL-estimation can be down-wards biased and the formal care-estimation can be upwards-biased in the control group, and then also the causal effect of the intervention can be biased. There are various other control variables that would have been beneficial to have, reducing the possibility of OVB. For instance, the presence of chronic diseases and other comorbidities would be useful to control for. A change of these factors may have affected formal time use, for instance if the patients become bedridden. Also, access to staffing can affect available formal care if someone gets long-term illness, or if someone quit.

Another issue is the small sample size. A bigger sample could show different results – for both the coefficient and the significance level. The LLN works when $n \rightarrow \infty$, and my sample may be too small for its properties to properly kick in. As mentioned in section 5.2.1, a problem can be that if the sample size increase, differences observed at baseline can diverge and become statistically significant.

As mentioned by the COSMOS-researchers themselves, we expect the control group to also have an effect of the intervention. If this is the case, the effect of the intervention is underestimated and could also be the reason for not finding a significant difference between the two groups in some of the outcome variables.

6.4.2 Robustness tests

When including patients in the sample, I performed a listwise deletion of patients with missing values in the variables determining cognitive level at baseline and QoL.

When a listwise deletion is performed, all patients with a missing value in the relevant variable are excluded. This was, as mentioned before, because the inclusion criteria could have been violated if these individuals were included. Even so, this can still be problematic as important variation may have been deleted. Also, patients with missing values in the dependent variables were excluded. In total, 206 patients were excluded, which is a considerable amount, and a lot of variation was lost due to this. As a robustness test, I wanted to inspect if there was a systematic difference in the excluded and included patients. In Appendix 4, the descriptive statistics from baseline of the excluded patients can be seen. In general, they were relatively similar to the patients included in the sample. On average, they were a little older and more demented. They also used a little more formal care and a little less informal care. Many of these patients died during the study period, and it thus makes sense that the excluded patients are older, more demented and contributed a higher care-burden. It is common that patients of older age, have more advanced dementia and demonstrates a higher care-burden towards the end of life. Including the 206 excluded patients, I also performed the fixed effect regressions to investigate how the excluded patients affect the results. The results can be found in Appendix 4. Compared to the full model, the regressions show a significant decrease in QoL in the control group after the intervention, but there is still no difference between the groups. The change in formal care is almost identical, while there are some differences in the change in the components of formal care with generally smaller estimates. For informal care the results are also similar to the full model, with somewhat smaller estimates. For psychotropic medication the interaction term is now significant, and the estimated decrease in use of psychotropic drugs are higher.

There can be problems regarding the level of clustering. I chose to cluster on NH level to control for correlation between patients residing in the same NH. This resulted in 31 clusters. Another possibility was to cluster on ward level, which would give 66 clusters. 31 clusters can be too few, and as a result the standard errors were bootstrapped. A possibility is that clustering on NH wards would have been a

sufficiently “high” level where there is suspected correlation, and that this would be a better choice. To check whether choice of clustering level and bootstrapped standard errors affected my estimates to a high degree, I compared the estimates when the different model specifications were applied. They can be found in Appendix 5 and show that the standard errors are relatively similar. Formal care is only significant when clustering at NH level, but is very close to becoming significant (p -value=0.108) when clustering on ward level. Changing clustering levels, both increased and decreased precision in some outcome variables in the different model specifications. Bootstrapping increased precision in one of the outcomes (informal care), but the standard errors did not change much. This can indicate that the non-bootstrapped standard error is a good approximation of the population.

Since we are looking at the average treatment effect over time, a source of concern is that the results are driven by a few patients that have become more time-consuming due to something unrelated to the intervention, e.g. due to an infection, pressure ulcer or other diseases, or because they are towards the end of their lives. To approach this challenge, I excluded all patients that had a change of formal care use > 200 , which resulted in the deletion of 10 patients. I ran the same fixed effects regressions, and the results can be found in Appendix 6. They show that there is no longer a significant change of formal care hours in the intervention group compared to the control group. This may indicate that the results are driven by a few demanding individuals, and that the final results can be biased. Due to this, quantifying the median change instead of the average could be preferred. However, all patients but one, were in the intervention group, and the increased time use might likely be a result of the intervention.

Even though perfect randomization is difficult to achieve, the randomization seems quite successful. Thus, the fixed effect model may be redundant, and a model without fixed effects could have been more appropriate. Therefore, I checked how the results changed when not including fixed-effects, and with baseline control variables. The results, which can be found in Appendix 7, still show no significant change in QoL.

There is still a significant change in formal care and in the components general care and everyday tasks, though the coefficients are generally smaller. Also, informal care and the general care-component are significant, though also here the coefficients are generally smaller. No significant difference in supervision is found. Difference in psychotropic drug use is no longer statistically significant at month 4 (p-value=0.12). Without fixed effects, the results are still somewhat similar, and can indicate that time-invariant unobserved heterogeneity is present but not a big problem in this sample.

6.4.3 Measurement errors

There can be some challenges connected with the questionnaires used. One is if the sensitivity of the QoL-instrument is good enough. To get an impression of the sensitivity of the QoL-instrument, I inspected how the measured QoL changed for the patients that was admitted to a hospital during the intervention period. We would expect the negative side-effects due to the incidence leading to hospital admission, like falling, to be picked up by the QoL-instrument. But, visually inspecting the change in both groups show an almost constant development (Appendix 8).

There are also some drawbacks concerning how formal time use is measured, which reduces the reliability of the findings of time use. If the time use is not recorded correctly, it will not be a representation of reality. In COSMOS there was no time schedule where all caregivers recorded their time spent on direct care after their shift was over, like suggested in a study by Luttenberger and Graessel (2010). A caregiver who worked day shift would fill in the form, with only information about the daytime direct care. This is a weakness on how the resource use data is collect, as the daytime health care personnel does not know how much time personnel on night or evening shift spent on direct care. On the other hand, most direct care time will take place during the day, as the majority will sleep during the night shift. But even so, the direct care time performed during the evenings and night will either not have been included in the total care time, or it has been estimated by someone who was not there. This can to some degree make the recording of care time incorrect. Another

drawback of the RUD-FOCA is that it does not take group activities into account. Thus, we do not know if some of the recorded time was used in group activities, and that we have some double-counting in the measurements. The drawbacks of how time use is measured can create measurement errors and bias the true formal- and informal time use. Another challenge is that this measurement error can be different in the control and intervention group, which would bias our causal effect of the intervention. However, we cannot know how much, if any, of the difference is due to these setbacks.

7. Discussion

The results show no significant change in QoL for people with dementia after COSMOS was introduced. Regarding the measures for resource use, formal care has increased drastically, while informal care and psychotropic medication has decreased, but to a lesser extent. This chapter discusses the results in view of the literature and tries to shed light on the economic implications of the intervention. The economic implications are discussed from the perspectives of the patient, the NH management and society.

7.1 QoL

Earlier, the challenges with measuring QoL in a population of people with dementia and the dispersed finding in research was discussed. QoL is a subjective construct, and thus difficult to measure truthfully in this population, because the disease prevents them from being able to communicate and comprehend their QoL, and one has to rely on proxy-rated QoL. Proxy-rating of QoL has been found to be both in compliance and non-compliance with self-reports, but are often underestimated (Robertson et al., 2017). If the patients were reliably able to fill out the questionnaires in the COSMOS-study themselves, the results may have been different. QoL can have been affected by the intervention, maybe on a level that is difficult to measure in a questionnaire like QUALID. When measuring QoL with

QUALID, QoL is transformed into an objective thing by measuring how often the patients smiles, cries, seems sad and so forth. But in reality, QoL incorporates much more than outward expressions and behaviors. Another potential issue is that the proxy-rated QoL is biased by the proxy-raters perception of the disease and the impact it has on a persons' life (Moyle and Murfield, 2013). These challenges can also be present in the COSMOS-study and can be a reason for finding non-significant results. The subjective part of QoL is not measured in this study because it is not possible, but subjectively the patients may have experienced a changed QoL.

7.2 Economic implications

The society's resources should be used *efficiently*. Resources spent on COSMOS should be the best possible utilization, and the opportunity cost of the resources should be low. Ergo, the value foregone by not using the resources on the best possible alternative to COSMOS ought to be low. In any efficiency analysis, like a cost-benefit analysis (CBA) or a cost-utility analysis (CUA), one must always weigh the costs up against the benefits. What is viewed as the optimal utilization of the resources will be contingent on whether you ask the patients themselves, the NH management, or the Government, because they value costs and benefits differently. When taking the societal perspective, the society's welfare as a whole must be taken into account. When taking the patient's or NH's perspective, the welfare of society in total is not considered, rather the relevant costs and benefits for either the patient himself or the NH. According to the Norwegian Directorate for Health's guidelines for economic evaluation in the health sector, it is recommended for analyses in the health care sector to be conducted in a societal perspective (The Norwegian Directorate of Health, 2012b). A key question in any decision-making process is what the intervention is compared to. The most common reference case is "care as usual" (The Norwegian Directorate of Health, 2012b). Compared to "care as usual", implementing the COSMOS-intervention increases formal care, decreases informal care and use of psychotropic medication, and has no effect on QoL, if the results are interpreted as causal. If the intervention is beneficial, the sum of benefits exceeds the sum of costs.

Since the various effects of the intervention are not valued, it is difficult to know whether the sum of benefits exceeds the sum of costs. To do a proper CBA or CUA, all factors should be taken into account and valued. Since this is outside the scope of this text, I will discuss the observed benefits and costs related to the intervention, and what consequences it can lead to from the various perspectives; societal, patient and NH management. In table 9, an overview of what is considered benefits and costs in the various perspectives in my discussion can be seen.

Table 9 Costs and benefits in different perspectives

Increase in	Perspective		
	Societal	NH	Patient
Formal care time	cost/benefit	cost/benefit	benefit
Informal care time	cost/benefit	benefit	benefit
Psychotropic medication	cost	cost	cost
QoL	benefit	benefit	benefit

7.2.1 Formal care

The amount of formal care has increased tremendously. An increase of approximately 20 hours of formal care per patient per month is equivalent to around 45 minutes per patient per day, on average. For this to be possible without hiring extra staff in the intervention period, it seems unlikely that the staff have used their time very efficiently before the intervention. Alternatively, they must have spent less time on administrative tasks, talking with colleagues, or other activities that are not measured, and more time on the patient, after the intervention. There is no information about the change in staffing, but as the NHs received no extra resources in connection with the intervention, it is likely that the staffing stayed constant. On the other hand, they can have hired more staff⁷ to be able to successfully carry out

⁷ Due to the interventions short time-span, calling part-time workers in for more extra shifts than usual is the most likely scenario

the intervention. This would be important to have information on, as costs would then be largely affected. If the staff was very inefficient before the intervention (for instance used much of their time on non-caring activities), it caused efficiency gains if no extra staffing was required. Then the increased time use can be viewed as a benefit in both the NH management and societal perspective. This possible gain is not reflected in the development in QoL, however. On the other hand, if more staff would have to be hired, this would lead to increased costs. Another possible cost is if the staff already was efficient before the intervention. Then it can be argued that if the same staff use much more time, they spend their time more inefficiently. For the patient, increased time use is most likely a pure benefit, if he gets increased attention and more thorough care. A setback is that we do not actually know if the staff has increased their efficiency or not after the intervention, as this is not measured.

7.2.2 Informal care

After the COSMOS-intervention the amount of informal care has decreased with over 3 hours per patient per month. Since Vossius et al. (2015) found that informal care in Norwegian NHs amounts to 6.7 hours per month on average, a decrease of over 3 hours is a considerable amount. A reason for the need for informal care, can be due to understaffing. Informal caregiving measured with the RUD-FOCA questionnaire can be a sign of the formal caregivers not doing their job properly – at least in the eyes of family members or friends. For instance, if a patient needs to use the restroom while a family member is visiting, the next of kin might feel obliged to perform the task if the personnel is not immediately available. Understaffing can be reflected in the personnel per patient ratio. At baseline, the ratio during daytime is on average 0.34 in the NHs. In the evenings the ratio drops to 0.24. In Norway there is no mandated personnel-patient ratio, and one cannot say that this is too low according to a national standard. But it is plausible that the ratio is low enough for the patients having to wait for care, so that the informal caregivers feel the need to step in – especially in the evenings. In Norway, NHs are funded by the municipalities. To a large extent, each municipality's economy and political prioritizations will thus decide

the level of care and staffing. The amount of staffing varies greatly across the municipalities, unrelated to patient care-burden (Dokument 8:155 S (2010–2011), 2011). From a societal and NH management perspective, fewer informal care hours can lead to increased costs. This is especially true for municipalities with lower resource allocation to NHs. If the informal caregivers stop doing the care-tasks, someone else will have to fill the “void”. Informal care can be a huge resource to society, and with the increasing demand of health care due to the aging population it will likely become more and more important, as future projections show a shortage of health care personnel (The Norwegian Directorate of Health, 2012a). If the resources available are too scarce to properly fill the care-need (e.g. understaffing), the unpaid informal carers could cover care-needs otherwise not fulfilled. Therefore, a decrease in informal care can lead to higher costs for the society and NH management.

When an informal caregiver spends his or her time doing care chores, it is at the expense of something else. There is an opportunity cost when performing informal care, which can incur a cost in a societal perspective. The caregiver could be spending quality time with the patient, working, using their free-time doing leisure activities or voluntary work, or just relaxing after a tough week at work. Instead he or she can feel obliged to spend their time caring for a relative, maybe due to a sense of commitment or bad conscience. The literature shows that also in the NH setting, informal care can be stressful and be experienced as a burden for the informal carer (Gautun and Bratt, 2017). Health decline and stress can be a source of opportunity costs, as the care-tasks can be exhausting and, especially for the elderly caregivers, “steal” time that could otherwise be spent on, for instance, needed rest. Thus, a decrease of informal care can also be a benefit in a societal perspective. There is not much information on who the informal caregivers are, or why they participate in care tasks even after a relative or friend is admitted to a NH. Most of them are probably wives or husbands, adult children and friends. Many do most likely have a job, while some are retired. In fact, the descriptive statistics shows that only about 25 % of the patients were married, and we can thus expect many of the informal caregivers to be

adult children caring for their parents, and many of them having a job. There are different ways to value informal care, and one widely used method in economic evaluations is the opportunity cost method (Berg et al., 2004). For the working caregivers, the hours spent on informal care can be valued as their hourly salary. For retired spouses, the opportunity cost of informal care can be linked to foregone leisure time to spend with friends and family, needed rest or time to perform house chores. Normally one would value leisure time according to the hourly wage, but retirees do not have a job. A more appropriate valuation to the opportunity cost of wage can in this case be the replacement cost approach. This approach uses the market price of an equivalent service; in other words the price of a formal caregiver (Escribano-Sotos and Pardo-García, 2015). This can be a better valuation approach because the informal carer are doing work that a formal carer should be doing.

7.2.3 Psychotropic medication

There are two possible channels of efficiency in the reduction of psychotropic medication – reduced direct costs and less negative side-effects. The use of psychotropic drugs has decreased with over 10 % compared to the baseline levels and represents reduced direct medical costs for the society and NH. As the literature describes, psychotropic medication is related to negative side-effects. This can be regarded as intangible costs and are costs accounted for in the patient and societal perspective. As there is no proof of change in the QoL-measurement, the intangible costs seem to be unchanged.

Superfluous psychotropic medication has also been related to increased hospital admissions (Johnell et al., 2017), because overuse of medications can lead to serious side-effects and mistreatment. A decrease in psychotropic medication can thus be cost-saving in a societal perspective. Performing medication reviews were a part of the COSMOS-intervention, and my results indicate that the medication reviews have successfully reduced the use of psychotropic medication. In the *White Paper on Medical Products – Correct use - better health* the Norwegian Ministry of Health and Care Services suggests that a medication review for NH patients should be

mandatory. Their motivation is that a medication review will increase the quality of care and the safety of the patient, lead to better health, and can result in fewer hospital admissions (Meld. St. 28 (2014–2015), 2015). When checking how hospitalization has changed during the intervention period, I found no increase in hospitalization in the control group, while ten more patients are hospitalized going from month 4 to month 9 in the intervention group. The descriptive statistics can be found in Appendix 9. This indicates that reduced use of psychotropic medication does not lead to fewer hospital admissions in the short run. Since we do not have data on the patients' hospital use before the intervention, we do not know the change before and after, and thus not the causal relationship between hospital use and the COSMOS-intervention. But this finding can indicate that COSMOS has increased hospital admissions.

7.2.4 Psychotropic drugs or time use

The literature shows some evidence which points to that use of psychotropic medication depends on the use of formal care, and that they may work as substitutes (Cawley et al., 2006, Schmidt et al., 1998, Anderson et al., 2016). To claim that these resources are substitutes on the background of my data and results can be going too far, as the divergency could be unrelated to each other. Even so, my results can indicate that there is an association between them, though a causal relationship cannot be stated. If there is shortage of staff, psychotropic medication can be a mean to fill this shortage. My findings show a negative correlation between formal care and psychotropic medication. This can point to that when formal care goes up, the care-needs are filled to a larger degree by the caregivers and that the use of psychotropic medication then goes down. To the best of my knowledge, the estimated elasticity of substitution in this setting is not known. Thus, we don't know how big of a change one would expect in psychotropic medication if formal care change, or vice versa.

As labor is much more expensive, changing to more material-intensive methods can be a way of saving resources for society and the NH management. Using psychotropic medication as a substitute for labor, at least to a certain degree, can be especially

useful due to future projections of under-staffing in the health care sector and the growing number of elderly. It can be critical to rely on medication to some degree; not to make the patient dozey, but to avoid very burdensome behaviors, like aggression or agitation. My findings show that decreased medication and a substantial increase of labor does not affect the patients' QoL, neither in the proxy-rated or objective measure. If QoL is not negatively affected, using material-intensive methods up to a certain degree can be economically efficient.

The literature also shows that more use of psychotropic medication is related to less staff, especially less educated staff, and staff distress (Anderson et al., 2016).

Educated personnel will to a higher degree know how to handle, for instance, agitated patients and handle a stressing situation more efficiently because they know exactly what to do. Even though we know that the number of personnel and educated personnel were similar in the two groups at baseline, educated personnel in one group can by chance have been reduced, due to for instance absence. Having information about the change in staffing and of potential differences between the intervention and control group, could give a better understanding of psychotropic medication potentially working as a substitute for either (the lack of) staff in general or educated staff in specific.

7.3 Time horizon

The time horizon can affect the cost-effectiveness between the intervention and the reference case (care as usual), and should therefore be long enough for costs and health effects between the alternatives to be identified (The Norwegian Directorate of Health, 2012b). In this case, we only know the short-term effects of the intervention. The long-term effects can be different, and all costs and health effects are thus not identified. Especially, the substitution of formal care and psychotropic medication can be thought about as a process that is ongoing for a longer period of time, maybe after longer periods with a lot of pressure on the staff due to understaffing or long-term sickness. The empirical evidence linking QoL and psychotropic medication is as discussed earlier not clear, though psychotropic drugs are associated with an

increased risk of for instance seizures and falling, which we can expect to have a negative impact on QoL. But, I have found that even though there has been a reduction of these drugs, QoL has not changed significantly. A reason for not finding a significant change, can be that the decrease of psychotropic drugs is not large enough for an effect to be visible on QoL on the short-term. Side-effects such as increased risk of seizures are factors typically identified on the long-term. Thus, an intervention stretching for nine months may have a timespan too short for capturing such effects and how they could affect QoL.

Also, it would be interesting to know to what degree the knowledge obtained during the intervention has been internalized by the staff. When the intervention started, the health personnel might be very motivated, and feel more responsible as they got continuous follow-ups throughout the study period. After the intervention was over, they might lose motivation or feel less obliged to perform the changes introduced by the COSMOS-intervention and more or less go back to the routines they had before. Alternatively, they have internalized the new knowledge and will continue with what they have learned. Since the intervention was aiming at increasing knowledge and focusing on training staff, a longer time-span or yearly follow-up would be interesting and beneficial. If the intervention had a longer time-span, better statistical power could have been reached and the credibility of the results increased. A problem is that elderly people with dementia doesn't have a long life-expectancy, and we could risk losing a large part of the sample towards the end of the study.

8. Concluding remarks

Considering the large costs of dementia, both direct and intangible, interventions that increase efficiency in dementia care is valuable. Also, since the dementia disease is non-reversible and still untreatable, assuring the highest possible QoL can be the most important mission in health care regarding elderly people with dementia. The COSMOS-intervention intended to improve both cost-effectiveness and QoL. This thesis shows that after the COSMOS-intervention, the patients with dementia had no

significant change in QoL, formal care increased substantially, and informal care and use of psychotropic medication decreased. A possible explanation for finding no significant effects on QoL, can be related to the challenges of measuring QoL in people with dementia.

Whether the intervention seems economically desirable and improved cost-effectiveness, depends on the chosen perspective. There are indications of the intervention making it possible for informal carers to perform fewer care-tasks. This can lead to an increase in both benefits and costs for society, but since they are unvalued we cannot know which one exceeds the other. Additionally, QoL is unaffected and there is no evidence of decrease in hospitalization, while use of formal care increased. Increased costs may thus exceed benefits in a societal perspective. From the NH managements perspective, except maybe if the increased formal time use is due to more efficient time use, the benefits do not seem to exceed the costs. In the perspective of the patient, there is no apparent increase in costs. The patient experiences lower long-term risks of negative side-effects due to decreased use of psychotropic drugs, gets more formal care hours and QoL seems to be unaffected by less time spent with informal caregivers. This indicates that the benefits outweigh the costs for the patient. There are some indications of psychotropic drugs and formal care depending on each other, though the correlation is not very strong. To a small degree, increased formal care may have led to reduced dependency on psychotropic drugs. The causality of the results should be interpreted with caution due to various factor, such as a small sample size and the fact that also the control group were expected to have a learning-effect. Since none of the outcomes are valued, determining whether the intervention is efficient or not is challenging. These reasonings do not take into consideration if and how the other outcomes of the COSMOS-intervention changed (for instance change in activities, other drugs etc.), which may lead to a different conclusion.

Regardless of whether the intervention is efficient or not, it can still give us valuable information about the relationship between the outcomes investigated. It can also tell

us something about how the population of elderly people with dementia responds to changes in these areas. This information can be used when optimizing care and it can give us knowledge on how resources can be used efficiently in the future.

8.1 Future research

In the process of writing this thesis, several questions have emerged that would be interesting to pursue. Due to the limited scope of this text, I could not go further with these questions, but recognize that they can be of value for future research.

Regarding the relationship between staffing, psychotropic medication and hospitalization, some questions arises: Does understaffing lead to more frequent use of psychotropic medication in Norwegian NHs? Can more staff lead to less costs in the long-run due to fewer hospitalizations and consequences of side-effects? Are psychotropic medication and formal care substitutes? These are questions that future research should pursue, as a part of securing efficient use of the resources in the future.

It would be interesting to know to what degree informal- and formal care in a NH can be substituted. Will increased informal care lead to decreased formal care? If so, this can be an important resource to utilize in the future, due to future projections of shortage of health personnel. Future research should investigate the causal relationship, or rate of elasticity, between formal- and informal care in a NH setting.

Bibliography

- AASMUL, I., HUSEBO, B. S. & FLO, E. 2018. Description of an advance care planning intervention in nursing homes: outcomes of the process evaluation. *BMC geriatrics*, 18, 26.
- ALZHEIMER'S ASSOCIATION. 2017. *What is Dementia?* [Online]. Available: <http://www.alz.org/what-is-dementia.asp> [Accessed 07.09. 2017].
- ALZHEIMER'S ASSOCIATION. 2018. *Risk Factors* [Online]. Available: https://www.alz.org/alzheimers_disease_causes_risk_factors.asp [Accessed 11.06. 2018].
- ANDERSON, K., BIRD, M., MACPHERSON, S. & BLAIR, A. 2016. How do staff influence the quality of long-term dementia care and the lives of residents? A systematic review of the evidence. *International Psychogeriatrics*, 28, 1263-1281.
- ANGRIST, J. D. & PISCHKE, J.-S. 2015. *Mastering 'metrics : the path from cause to effect*, Princeton, N.J, Princeton University Press.
- ANGRIST, J. D. & PISCHKE, J. R.-S. 2009. *Mostly harmless econometrics : an empiricist's companion*. Princeton: Princeton University Press.
- ARONSON, M. K., POST, D. C. & GUASTADISEGNI, P. 1993. Dementia, agitation, and care in the nursing home. *Journal of the American Geriatrics Society*, 41, 507-512.
- BARCA, M. L., ENGEDAL, K., LAKS, J. & SELBÆK, G. 2011. Quality of Life among Elderly Patients with Dementia in Institutions. *Dementia and Geriatric Cognitive Disorders*, 31, 435-442.
- BERG, B., BROUWER, W. & KOOPMANSCHAP, M. 2004. Economic valuation of informal care. *Health Economics in Prevention and Care*, 5, 36-45.
- BERGH, S., HOLMEN, J., SALTVEDT, I., TAMBS, K. & SELBÆK, G. 2012. Demens og nevropsykiatriske symptomer hos sykehjemspasienter i Nord-Trøndelag. *Tidsskrift for den Norske Lægeforening*, 132, 1956.
- BLOEHLIGER, M., RÜEGG, S., JICK, S., MEIER, C. & BODMER, M. 2015. Antipsychotic Drug Use and the Risk of Seizures: Follow-up Study with a Nested Case–Control Analysis. *CNS Drugs*, 29, 591-603.
- CARPENTER, G. I., HASTIE, C. L., MORRIS, J. N., FRIES, B. E. & ANKRI, J. 2006. Measuring change in activities of daily living in nursing home residents with moderate to severe cognitive impairment. *BMC geriatrics*, 6, 7.
- CAWLEY, J., GRABOWSKI, D. C. & HIRTH, R. A. 2006. Factor substitution in nursing homes. *Journal of Health Economics*, 25, 234-247.
- COX, C. A., VAN JAARSVELD, H. J., HOUTERMAN, S., VAN DER STEGEN, J. C. G. H., WASYLEWICZ, A. T. M., GROULS, R. J. E. & VAN DER LINDEN, C. M. J. 2016. Psychotropic Drug Prescription and the Risk of Falls in Nursing Home Residents. *Journal of the American Medical Directors Association*, 17, 1089-1093.
- CRAIG, P., DIEPPE, P., MACINTYRE, S., MICHIE, S., NAZARETH, I. & PETTICREW, M. 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *British Medical Journal*, 337, 5.
- DOKUMENT 8:155 S (2010–2011) 2011. Representantforslag 155 S.
- ESCRIBANO-SOTOS, F. & PARDO-GARCÍA, I. 2015. Analyzing the Costs of Informal Care for Persons with Dementia in Spain. *Journal of Promotion Management*, 21, 459-474.
- ESSEN, L. V. 2004. Proxy ratings of patient quality of life Factors related to patient–proxy agreement. *Acta Oncologica*, 43, 229-234.
- FORD, A. H. 2015. Neuropsychiatric symptoms of dementia: are pharmacological treatments effective and safe? *Journal of Neurology, Neurosurgery & Psychiatry*, 86, 4.
- GALIK, E., RESNICK, B., HAMMERSLA, M. & BRIGHTWATER, J. 2014. Optimizing Function and Physical Activity Among Nursing Home Residents With Dementia: Testing the Impact of Function-Focused Care. *The Gerontologist*, 54, 930-943.

- GAUTUN, H. & BRATT, C. 2017. Caring too much? Lack of public services to older people reduces attendance at work among their children. *European journal of ageing*, 14, 155-166.
- GILL, S. S., BRONSKILL, S. E., NORMAND, S.-L. T., ANDERSON, G. M., SYKORA, K., LAM, K., BELL, C. M., LEE, P. E., FISCHER, H. D., HERRMANN, N., GURWITZ, J. H. & ROCHON, P. A. 2007. Antipsychotic drug use and mortality in older adults with dementia. *Annals of internal medicine*, 146, 775.
- GURWITZ, J. H., BONNER, A. & BERWICK, D. M. 2017. Reducing excessive use of antipsychotic agents in nursing homes. *JAMA*, 318, 118-119.
- HELVIK, A.-S., ŠALTYTĖ BENTH, J., WU, B., ENGEDAL, K. & SELBÆK, G. 2017. Persistent use of psychotropic drugs in nursing home residents in Norway. *BMC Geriatrics*, 17, 52.
- HORN, D. S., BUERHAUS, J. P., BERGSTROM, J. N. & SMOUT, J. R. 2005. RN Staffing Time and Outcomes of Long-Stay Nursing Home Residents: Pressure ulcers and other adverse outcomes are less likely as RNs spend more time on direct patient care. *AJN, American Journal of Nursing*, 105, 58-70.
- HUSEBO, B. S., FLO, E., AARSLAND, D., SELBAEK, G., TESTAD, I., GULLA, C., AASMUL, I. & BALLARD, C. 2015. COSMOS—improving the quality of life in nursing home patients: protocol for an effectiveness-implementation cluster randomized clinical hybrid trial. *Implementation Science*, 10, 131.
- JOHNELL, K., JONASDOTTIR BERGMAN, G., FASTBOM, J., DANIELSSON, B., BORG, N. & SALMI, P. 2017. Psychotropic drugs and the risk of fall injuries, hospitalisations and mortality among older adults. *International Journal of Geriatric Psychiatry*, 32, 414-420.
- KHANDKER, S. R., KOOLWAL, G. B. & SAMAD, H. A. 2010. Handbook on impact evaluation : quantitative methods and practices. Washington, DC: World Bank.
- KORCZYN, A. D. & DAVIDSON, M. 1999. Quality of life in Alzheimer's disease. *European Journal of Neurology*, 6, 487-489.
- LUTTENBERGER, K. & GRAESSEL, E. 2010. Recording care time in nursing homes: development and validation of the "RUD-FOCA" (Resource Utilization in Dementia -Formal Care). *Int Psychogeriatr*, 22.
- MATSUBASA, T., KIMURA, A., SHINOHARA, M. & ENDO, F. 2015. Time study of staff members in an institution for severe motor and intellectual disabilities. 57, 1154.
- MELD. ST. 28 (2014–2015) 2015. Legemiddelmeldingen — Riktig bruk – bedre helse. Oslo: Helse- og omsorgsdepartementet.
- METZELTHIN, S. F., VERBAKEL, E., VEENSTRA, M. Y., VAN EXEL, J., AMBERGEN, A. W., KEMPEN, G., ERASMUS SCHOOL OF HEALTH, P. & MANAGEMENT 2017. Positive and negative outcomes of informal caregiving at home and in institutionalised long-term care: A cross-sectional study. *BMC Geriatrics*, 17.
- MINISTRY OF HEALTH AND CARE SERVICES. 2008. *Dementia Plan 2015 «Den gode dagen»* [Online]. Available: https://www.regjeringen.no/globalassets/upload/hod/vedlegg/omsorgsplan_2015/demensplan2015.pdf [Accessed 12.05. 2018].
- MINISTRY OF HEALTH AND CARE SERVICES. 2015. *Dementia Plan 2020* [Online]. Available: https://www.regjeringen.no/contentassets/3bbec72c19a04af88fa78ffb02a203da/dementia-plan_2020_long.pdf [Accessed 10.05. 2018].
- MITCHELL, M. L. & JOLLEY, J. M. 2012. *Research design explained*, Belmont, Calif, Wadsworth Cengage Learning.
- MJØRUD, M., KIRKEVOLD, M., RØSVIK, J., SELBÆK, G. & ENGEDAL, K. 2014. Variables associated to quality of life among nursing home patients with dementia. *Aging & Mental Health*, 18, 1013-1021.
- MOULTON, B. R. 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. (Notes). *Review of Economics and Statistics*, 72, 334.

- MOYLE, W. & MURFIELD, J. E. 2013. Health-related quality of life in older people with severe dementia: challenges for measurement and management. *Expert review of pharmacoeconomics & outcomes research*, 13, 109.
- NORDBERG, G., WIMO, A., JÖNSSON, L., KÅREHOLT, I., SJÖLUND, B. M., LAGERGREN, M. & STRAUSS, E. V. 2007. Time use and costs of institutionalised elderly persons with or without dementia: results from the Nordanstig cohort in the Kungsholmen Project—a population based study in Sweden. *International Journal of Geriatric Psychiatry*, 22, 639-648.
- PAULUS, A. T. G., VAN RAAK, A. & KEIJZER, F. 2005. Informal and formal caregivers' involvement in nursing home care activities: impact of integrated care. *Journal of Advanced Nursing*, 49, 354-366.
- PRINCE, M., WIMO, A., GUERCHET, M., ALI, G.-C., WU, Y.-T. & PRINA, M. 2015. *World Alzheimer Report 2015. The Global Impact of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends*.
- QUINN, J. 2013. *Dementia*, Hoboken, UNITED KINGDOM, John Wiley & Sons, Incorporated.
- RINALDI, P., SPAZZAFUMO, L., MASTRIFORTI, R., MATTIOLI, P., MARVARDI, M., POLIDORI, M. C., CHERUBINI, A., ABATE, G., BARTORELLI, L., BONAIUTO, S., CAPURSO, A., CUCINOTTA, D., GALLUCCI, M., GIORDANO, M., MARTORELLI, M., MASARAKI, G., NIEDDU, A., PETTENATI, C., PUTZU, P., TAMMARO, V. A., TOMASSINI, P. F., VERGANI, C., SENIN, U. & MECOCCI, P. 2005. Predictors of high level of burden and distress in caregivers of demented patients: results of an Italian multicenter study. *International Journal of Geriatric Psychiatry*, 20, 168-174.
- ROBERTSON, S., COOPER, C., HOE, J., HAMILTON, O., STRINGER, A. & LIVINGSTON, G. 2017. Proxy rated quality of life of care home residents with dementia: a systematic review. *International psychogeriatrics*, 29, 569-581.
- ROSTAD, H. M., PUTS, M., T. E., SMÅSTUEN, M. C., GROV, E. K., UTNE, I. & HALVORSRUD, L. 2017. Associations between Pain and Quality of Life in Severe Dementia: A Norwegian Cross-Sectional Study. *Dementia and Geriatric Cognitive Disorders Extra*, 7, 109-121.
- RØEN, I., SELBÆK, G., KIRKEVOLD, Ø., ENGEDAL, K., LERDAL, A. & BERGH, S. 2015. The Reliability and Validity of the Norwegian Version of the Quality of Life in Late-Stage Dementia Scale. *Dementia and Geriatric Cognitive Disorders*, 40, 233-242.
- RØEN, I., SELBÆK, G., KIRKEVOLD, Ø., ENGEDAL, K., TESTAD, I. & BERGH, S. 2017. Resource Use and Disease Course in dementia - Nursing Home (REDIC-NH), a longitudinal cohort study; design and patient characteristics at admission to Norwegian nursing homes. *BMC Health Services Research*, 17, 365.
- SCHMIDT, I., CLAESSION, C. B., WESTERHOLM, B. & SVARSTAD, B. L. 1998. Resident characteristics and organizational factors influencing the quality of drug use in Swedish nursing homes. *Social Science & Medicine*, 47, 961-971.
- SCHULZ, K. F. & GRIMES, D. A. 2002. Blinding in randomised trials: hiding who got what. *The Lancet*, 359, 696-700.
- SCHULZ, R., ROSEN, J., KLINGER, J., MUSA, D., CASTLE, N. G., KANE, A. L. & LUSTIG, A. 2014. Effects of a Psychosocial Intervention on Caregivers of Recently Placed Nursing Home Residents: A Randomized Controlled Trial. *Clinical Gerontologist*, 37, 347-367.
- SELBÆK, G., KIRKEVOLD, Ø. & ENGEDAL, K. 2007. The prevalence of psychiatric symptoms and behavioural disturbances and the use of psychotropic drugs in Norwegian nursing homes. *International Journal of Geriatric Psychiatry*, 22, 843-849.
- SHIN, J. H. & BAE, S. H. 2012. Nurse staffing, quality of care, and quality of life in US nursing homes, 1996–2011: an integrative review. *Journal of gerontological nursing*, 38, 46-53.
- SINK, K. M., HOLDEN, K. F. & YAFFE, K. 2005. Pharmacological Treatment of Neuropsychiatric Symptoms of Dementia: A Review of the Evidence. *JAMA*, 293, 596-608.
- SLOANE, P. D., ZIMMERMAN, S., WILLIAMS, C. S., REED, P. S., GILL, K. S. & PREISSER, J. S. 2005. Evaluating the Quality of Life of Long-Term Care Residents With Dementia. *The Gerontologist*, 45, 37-49.

- SOURIAL, R., MCCUSKER, J., COLE, M. & ABRAHAMOWICZ, M. 2001. Agitation in Demented Patients in an Acute Care Hospital: Prevalence, Disruptiveness, and Staff Burden. *Int. Psychogeriatr.*, 13, 183-197.
- STATISTICS NORWAY. 2013. *Eldres bruk av helse- og omsorgstjenester* [Online]. Available: <https://www.ssb.no/helse/artikler-og-publikasjoner/attachment/125965?ts=13f8b5b6898> [Accessed 12.05. 2018].
- STATISTICS NORWAY. 2017. *Framtidens eldre i by og bygd* [Online]. Available: <https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/attachment/328026?ts=15f811146d0> [Accessed 11.05. 2018].
- STROBEL, C. & ENGEDAL, K. 2008. MMSE-NR. Norwegian revised mini mental status evaluation. Revised and expanded manual. *National Center for Aging and Health, Oslo*.
- SØRBØ, K. & MOEN NILSEN, P. 2018. Sjukeheim ville ha slutt på ropinga til Torbjørg (92) - seksdobla dosen med psykosemedisin. *Norwegian Broadcasting Corporation*, 14.06.2018.
- THE NORWEGIAN DIRECTORATE OF HEALTH 2012a. Norway and Health : an introduction. 2. utg. ed. Oslo: Norwegian Directorate of Health.
- THE NORWEGIAN DIRECTORATE OF HEALTH 2012b. Økonomisk evaluering av helsetiltak : en veileder. Oslo: Helsedirektoratet.
- THE NORWEGIAN HEALTH ASSOCIATION. 2017. *Demensforskning* [Online]. Available: <http://nasjonalforeningen.no/forskning/demensforskning/> [Accessed 07.09. 2017].
- VAN DE VEN-VAKHTEEVA, J., BOR, H., WETZELS, R. B., KOOPMANS, R. T. C. M. & ZUIDEMA, S. U. 2013. The impact of antipsychotics and neuropsychiatric symptoms on the quality of life of people with dementia living in nursing homes. *International Journal of Geriatric Psychiatry*, 28, 530-538.
- VOSSIUS, C., SELBÆK, G., YDSTEBO, A., BENTH, J. S., GODAGER, G. & LURAAS, H. 2015. Ressursbruk og sykdomsforløp ved demens (REDIC). *Sykehuset Innlandet, Report*, 1-155.
- WEINER, M. F., MARTIN-COOK, K., SVETLIK, D. A., SAINÉ, K., FOSTER, B. & FONTAINE, C. 2000. The quality of life in late-stage dementia (QUALID) scale. *J Am Med Dir Assoc*, 1, 114-6.
- WOOLDRIDGE, J. M. 2014. *Introduction to econometrics*, Andover, Cengage Learning.
- WORLD HEALTH ORGANIZATION. 2017. *WHOQOL: Measuring Quality of Life* [Online]. Available: <http://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/> [Accessed 20.09. 2017].

Appendix 1 Access to various personnel: Descriptive Statistics at Baseline

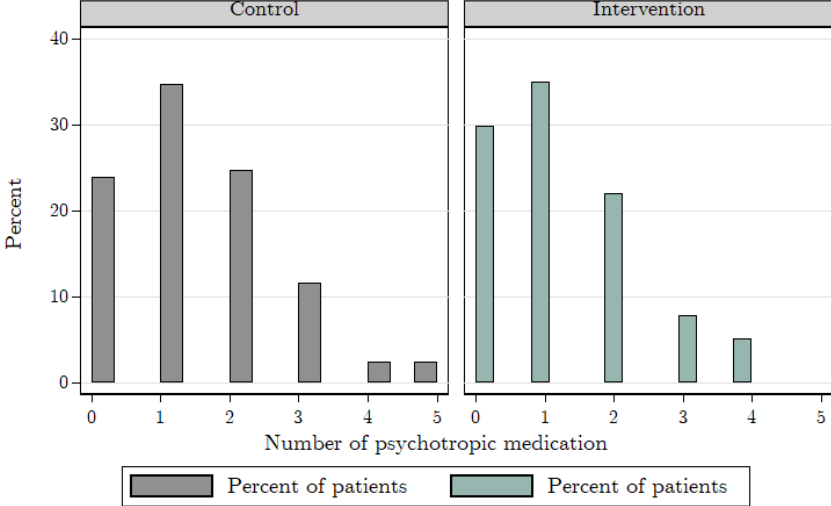
	Proportion of wards		Difference	p-value	t-value
	Intervention	Control			
Physical therapist	0.48	0.50	-0.02	0.898	0.13
Priest	0.66	0.69	-0.03	0.769	0.29
Dentist	0.62	0.69	-0.07	0.577	0.56
Hair dresser	0.90	0.81	0.09	0.351	-0.93
Volunteers	0.59	0.65	-0.06	0.606	0.52
Occupational therapist	0.21	0.23	-0.02	0.831	0.21
Social worker	0.00	0.15	-0.15	0.028*	2.17*
Chiropodist	0.86	0.81	0.05	0.586	-0.54
Activities	0.59	0.54	0.05	0.721	-0.36
<i>N</i>	29	26			

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

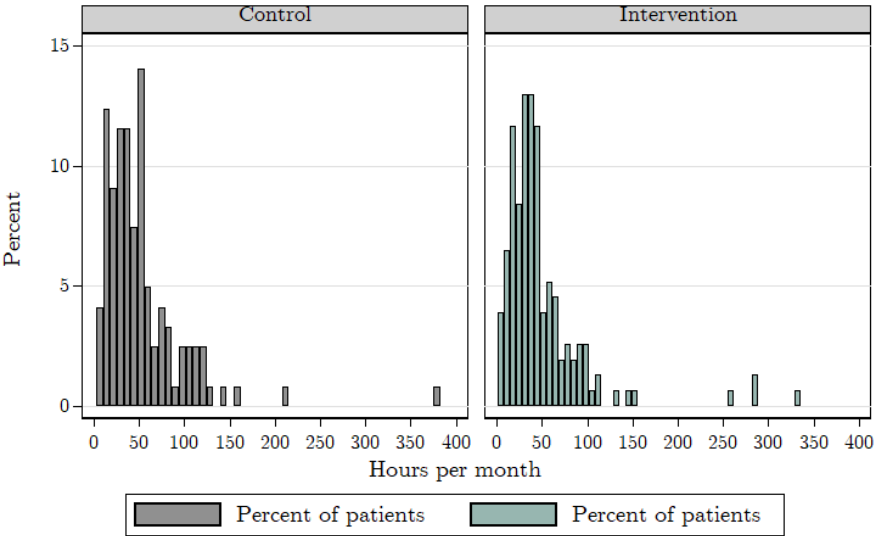
Notes: This table shows the t-statistics for the mean differences in access to resources in the control and intervention group.

Appendix 2 Distributions

Distribution of psychotropic medication
Baseline

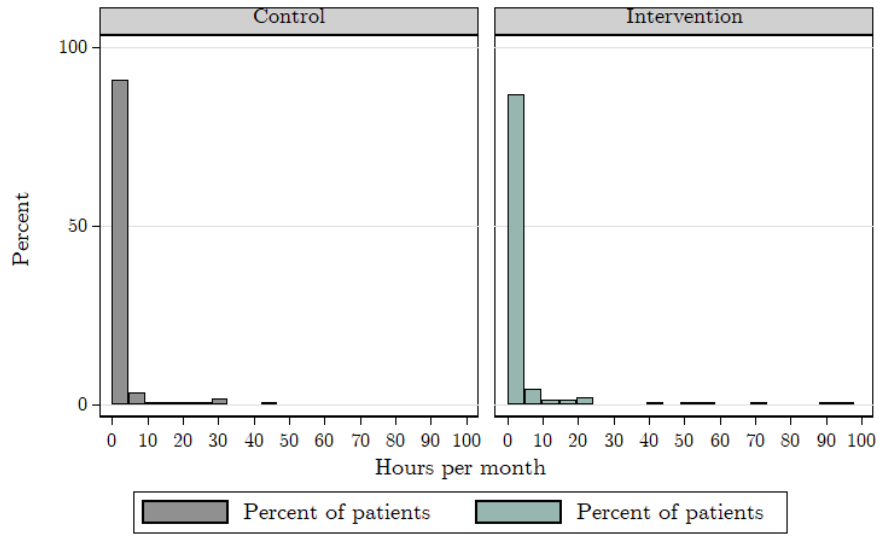


Distribution of formal care
Baseline



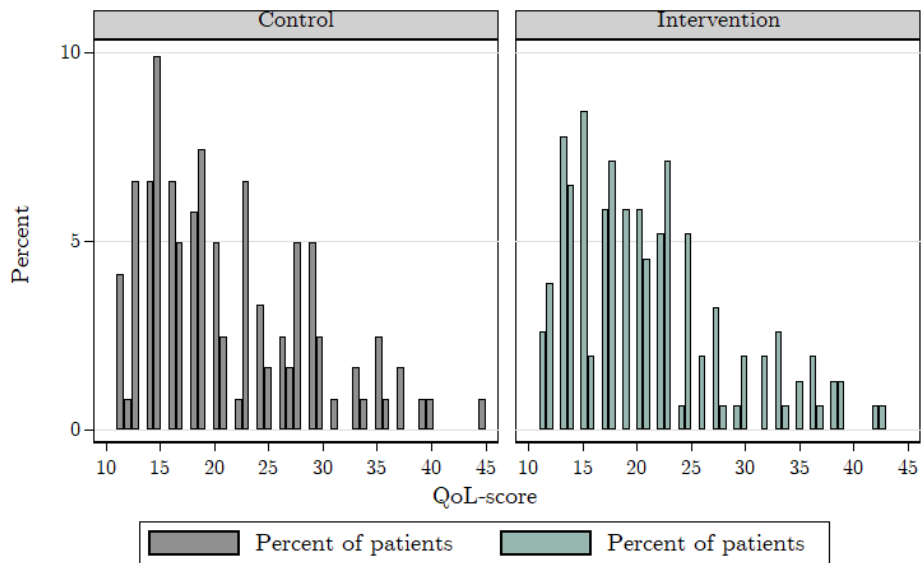
Distribution of informal care

Baseline



Distribution of QoL

Baseline



Appendix 3 ATET of COSMOS on depression

ATET of COSMOS on depression		
	(1)	(2)
After	0.0111 (0.518)	
Interaction	0.163 (0.875)	
Month 4-Interaction		0.395 (0.952)
Month 9-Interaction		-0.0510 (0.844)
After: Month 4		-0.403 (0.493)
After: Month 9		0.410 (0.612)
Constant	8.701*** (1.110)	8.659*** (1.085)
Observations	632	632

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

Clustered standard errors in parentheses.

Note: This table presents the mean difference over time for use of psychotropic drugs. Column (1) include an interaction term where we look at the post-treatment period as one period, while in column (2) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

Appendix 4 Excluded patients: Descriptive statistics and regressions

	Scale	Intervention Mean	se	Control Mean	se	Difference p-value	t-value
Demographics and Health							
Age	year	86.59	(7.96)	88.16	(7.38)	0.148	1.45
Men (proportion)	share	0.31	(0.46)	0.28	(0.45)	0.664	-0.43
Weight (kg)	kg	64.92	(14.34)	61.21	(13.73)	0.074	-1.80
Height (m)	m	1.63	(0.09)	1.64	(0.08)	0.346	0.95
Blod pressure		128.40	(21.17)	128.01	(23.54)	0.907	-0.12
Blod pressure		70.69	(12.69)	71.36	(13.80)	0.734	0.34
Pulse		73.59	(12.32)	71.00	(11.68)	0.152	-1.44
Scores from COSMOS-study							
Formal care (RUD-FOCA)	hrs/month	65.56	(51.76)	57.76	(51.94)	0.300	-1.04
Informal care (RUD-FOCA)	hrs/month	2.10	(5.91)	2.98	(10.01)	0.454	0.75
QoL-score (QUALID)	11-55	20.86	(7.75)	21.65	(9.24)	0.509	0.66
Cognitive status (MMSE)	0-30	8.25	(7.43)	7.38	(7.35)	0.395	-0.85
Pain-score (MOBID)	0-10	2.27	(2.37)	2.92	(2.66)	0.094	1.68
Agitation-score (CMAI)	29-203	41.58	(12.89)	42.70	(14.83)	0.578	0.56
Activities-of-daily-living-score	0-30	23.62	(13.38)	20.50	(12.36)	0.085	-1.73
Psychiatric-symptoms-score	0-120	15.00	(17.19)	16.20	(17.49)	0.636	0.47
Caregiver-distress-score	0-50	7.50	(6.89)	6.39	(6.14)	0.271	-1.11
Depression-score	0-38	6.14	(5.19)	7.88	(7.19)	0.094	1.69
Psychotropic drug use							
Use of antipsychotics	n	0.15	(0.39)	0.16	(0.39)	0.975	0.03
Use of anxiolytics	n	0.21	(0.43)	0.24	(0.43)	0.612	0.51
Use of hypnotics and sedatives	n	0.30	(0.50)	0.27	(0.45)	0.661	-0.44
Use of antidepressants	n	0.51	(0.67)	0.38	(0.57)	0.127	-1.53
Use of anti-dementia drugs	n	0.16	(0.37)	0.17	(0.37)	0.954	0.06
Use of total psychotropic drugs	n	1.34	(1.22)	1.21	(1.21)	0.452	-0.75
<i>N</i>		110		96		206	

Notes: This table shows the t-statistics of mean differences at baseline in the intervention and control group for the patients that was excluded due to missing values. Standard deviations in parentheses.

Standard deviations in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ATET of COSMOS on QoL

	(1)	(2)	(3)	(4)
After	-4.785*** (0.601)	-3.963*** (0.868)	-3.795*** (0.618)	
Interaction	0.711 (0.914)	0.900 (1.154)	1.230 (0.958)	
Month 4-Interaction				1.661 (1.137)
Month 9-Interaction				0.758 (0.969)
After: Month 4				-3.310*** (0.690)
After: Month 9				-4.358*** (0.705)
Constant	21.71*** (0.548)	3.235 (6.697)	16.02*** (0.697)	16.22*** (0.716)
Observations	1566	744	1566	1566

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

Note: Cluster bootstrapped standard errors in parentheses.

This table presents various model specifications of ATET for formal care. The first three columns include an interaction term where we look at the post-treatment period as one period, while in column (4) we look at each post-treatment period separately. Column (1) shows the results with no control variables or fixed effects. In column (2), time-invariant and time-variant control variables are included, and in column (3) and (4) time-variant control variables and fixed effects are included. Time-invariant control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, marital status, age and sex. The time-variant control is change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on hours of formal care per month

	(1)	(2)	(3)	(4)
After	6.799 (5.082)	9.188** (4.120)	7.075 (4.970)	
Interaction	19.99* (10.67)	10.86 (6.928)	21.30** (9.993)	
Intervention	3.327 (6.931)	4.981 (7.015)	0 (0)	0 (0)
Month 4-Interaction				15.97 (10.46)
Month 9-Interaction				27.92* (14.36)
After: Month 4				2.831 (5.564)
After: Month 9				12.30* (6.559)
Constant	54.12*** (5.530)	84.68*** (31.91)	69.19*** (7.530)	67.50*** (7.038)
Observations	1302	624	1302	1302
Controls	No	Yes	Yes	Yes
FE				

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

Clustered bootstrapped standard errors in parentheses.

Note: This table presents various model specifications of the ATET for formal care. The first three columns include an interaction term where we look at the post-treatment period as one period, while in column (4) we look at each post-treatment period separately. Column (1) shows the results with no control variables or fixed effects. In column (2), time-invariant and time-variant control variables are included, and in column (3) and (4) time-variant control variables and fixed effects are included. Time-invariant control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, marital status, age and sex. The time-variant control is change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on categories of formal care

	(1)	(2)	(3)	(4)	(5)	(6)
	General care	General care	Everyday tasks	Everyday tasks	Supervision	Supervision
After	-0.595 (2.958)		-0.801 (1.148)		-3.111** (1.523)	
Interaction	6.288 (5.226)		5.409** (2.520)		2.591 (3.371)	
Month 4- Interaction		5.854 (5.249)		3.544* (2.021)		3.416 (3.670)
Month 9- Interaction		6.797 (6.935)		7.272* (4.150)		1.742 (3.622)
After: Month 4		-2.653 (2.257)		-0.548 (1.217)		-3.001* (1.733)
After: Month 9		1.689 (4.378)		-0.962 (1.308)		-3.281 (2.327)
Constant	28.30*** (3.028)	27.65*** (3.018)	7.825*** (0.918)	7.600*** (1.075)	9.353*** (2.330)	9.517*** (2.164)
Observations	1555	1555	1552	1552	1561	1561

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

Clustered bootstrapped standard errors in parentheses.

Note: This table presents the ATET for three components of formal care. Columns (1), (3), and (5) include an interaction term where we look at the post-treatment periods as one period, while in columns (2), (4), and (6) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on informal care

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total informal care	Total informal care	General care	General care	Everyday tasks	Supervision	Supervision
After	1.514** (0.679)		0.398 (0.316)		-0.180 (0.326)	0.399** (0.161)	
Interaction	-2.193** (1.000)		-0.589 (0.481)		-0.172 (0.475)	-0.774*** (0.271)	
Month 4- Interaction		-1.884 (1.196)		-0.514 (0.608)			-1.085*** (0.301)
Month 9- Interaction		-2.593** (1.282)		-0.680 (0.445)			-0.464 (0.449)
After: Month 4		1.692** (0.744)		0.628* (0.339)			0.582** (0.270)
After: Month 9		1.287 (0.897)		0.143 (0.348)			0.215 (0.294)

Constant	3.825*** (1.139)	3.928*** (1.148)	0.942*** (0.366)	1.019*** (0.393)	1.160*** (0.291)	0.597** (0.250)	0.599** (0.251)
Observations	1297	1297	1561	1561	1555	1564	1564

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

Clustered and bootstrapped standard errors in parentheses.

Note: This table presents the ATET for informal care and components of informal care. Columns (1), (3), (5), and (6) include an interaction term where we look at the post-treatment periods as one period, while in columns (2), (4) and (7) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on use of psychotropic medication

	(1)	(2)
After	0.00983 (0.0630)	
Interaction	-0.192** (0.0753)	
Month 4-Interaction		-0.230*** (0.0646)
Month 9-Interaction		-0.147 (0.103)
After: Month 4		0.0393 (0.0478)
After: Month 9		-0.0253 (0.0883)
Constant	1.175*** (0.0911)	1.177*** (0.0912)
Observations	1336	1336

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

Clustered bootstrapped standard errors in parentheses.

Note: This table presents the ATET for use of psychotropic drugs. Column (1) include an interaction term where we look at the post-treatment periods as one period, while in column (2) we look at each post-treatment period separately. Control variables include change of cognition score over time. Standard errors are bootstrapped and clustered on NH level.

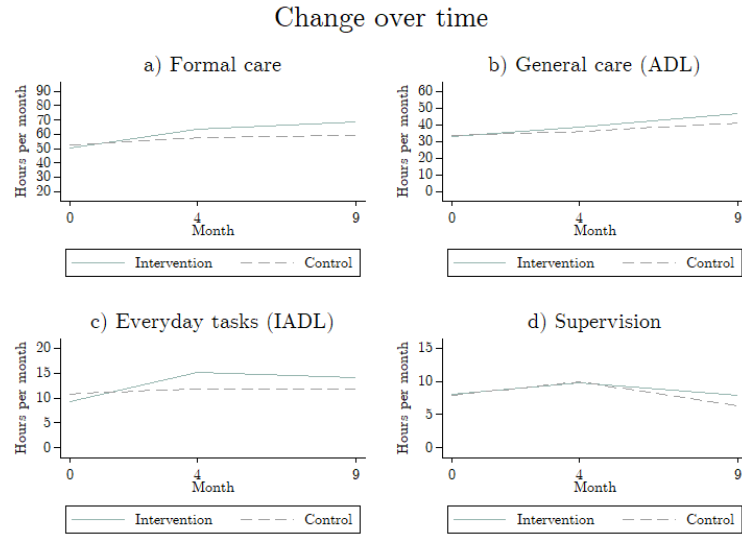
Appendix 5 How precision is affected: Clustering and bootstrapping

	(1) Cluster: ward	(2) Cluster: NH	(3) Cluster: NH & bootstrapped
<u>QoL</u>			
Interaction	0.920 (0.903)	0.920 (0.847)	0.920 (0.871)
<u>Formal care</u>			
Interaction	21.04 (12.89)	21.04* (11.29)	21.04* (11.35)
<u>Informal care</u>			
Interaction	-3.196** (1.296)	-3.196** (1.160)	-3.196*** (1.169)
<u>Psychotropic medication</u>			
Interaction (4-month)	-0.126** (0.0589)	-0.126* (0.0630)	-0.126* (0.0645)
Observations	822	822	822

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

Note: This table shows how the precision of estimates of the dependent variables change, depending on cluster level and if the standard errors are bootstrapped.

Appendix 6 Exclusion of patients with large change in formal care



ATET of COSMOS on hours of formal care per month

	(1)	(2)	(3)	(4)
After	6.080 (5.489)	6.233 (4.462)	4.139 (5.756)	
Interaction	9.636 (6.457)	9.587 (6.695)	8.809 (6.682)	
Intervention	-2.139 (8.387)	1.051 (5.336)	0 (0)	0 (0)
Month 4-Interaction				7.149 (8.063)
Month 9-Interaction				10.50 (7.895)
After: Month 4				3.596 (7.383)
After: Month 9				4.747 (5.299)
Constant	52.40*** (6.059)	-3.115 (29.14)	68.87*** (9.419)	68.58*** (9.059)
Observations	792	399	792	792
Controls	No	Yes	Yes	Yes
FE	No	No	Yes	Yes

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

Note: This table presents various model specifications of ATET for formal care, excluding patients with change in formal care use >200 hours from baseline to 9-month follow-up. The first three columns include an interaction term where we look at the post-treatment period as one period, while in column (4) we look at each post-treatment period separately. Column (1) shows the results with no control variables or fixed effects. In column (2), time-invariant and time-variant control variables are included, and in column (3) and (4) time-variant control variables and fixed effects are included. Time-invariant control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, marital status, age and sex. The time-variant control is change of cognition score over time. Standard errors are clustered on NH level.

Appendix 7 Regressions without fixed effects

ATET of COSMOS on QoL

	(1)	(2)
After	0.677 (1.039)	
Interaction	-0.234 (1.165)	
Pain score	0.314 (0.242)	0.314 (0.251)
Agitation score	0.00685 (0.0459)	0.00676 (0.0475)
Depression score	0.302*** (0.110)	0.303*** (0.110)
Neuropsychiatric symptoms score	0.0159 (0.0643)	0.0159 (0.0649)
ADL score	0.0537 (0.0467)	0.0536 (0.0477)
Care distress score	0.218 (0.158)	0.218 (0.155)
Age	0.0330 (0.0766)	0.0327 (0.0764)
Sex	-0.156 (1.062)	-0.157 (1.068)
Married	-0.0290 (0.851)	-0.0294 (0.817)
Month 4-Interaction		0.542 (1.232)
Month 9-Interaction		-1.013 (1.451)
After: Month 4		-0.0523 (1.071)
After: Month 9		1.404 (1.308)
Constant	13.36** (6.298)	13.40** (6.287)
Observations	411	411

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

Cluster bootstrapped standard errors in parentheses. Note: This table presents the ATET for use of psychotropic drugs. Column (1) include an interaction term where we look at the post-treatment period as one period, while in column (2) we look at each post-treatment period separately. Control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, change of cognition score over time, age, marital status and sex. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on categories of formal care

	(1)	(2)	(3)	(4)	(5)	(6)
	General care	General care	Everyday tasks	Everyday tasks	Supervision	Supervision
After	6.241** (3.093)		0.146 (1.141)		1.071 (2.484)	
Interaction	9.138* (5.539)		7.673*** (1.904)		-0.0710 (3.712)	
Pain score	0.981 (0.816)	0.981 (0.820)	0.212 (0.225)	0.213 (0.223)	-0.179 (0.432)	-0.179 (0.445)
Agitation score	0.293 (0.261)	0.292 (0.252)	0.0133 (0.0972)	0.0134 (0.102)	0.429*** (0.159)	0.429*** (0.157)
Depression score	-0.431 (0.423)	-0.429 (0.433)	-0.0283 (0.139)	-0.0285 (0.141)	-0.496* (0.274)	-0.496* (0.267)
Neuropsychiatric symptoms score	-0.854** (0.384)	-0.854** (0.376)	-0.0783 (0.0894)	-0.0783 (0.0908)	-0.0555 (0.172)	-0.0555 (0.174)
ADL score	0.915*** (0.169)	0.915*** (0.161)	0.171 (0.113)	0.171 (0.113)	0.0827 (0.186)	0.0827 (0.176)
Care distress score	1.987** (0.860)	1.986** (0.851)	0.263 (0.210)	0.263 (0.219)	0.153 (0.375)	0.153 (0.371)
Age	-0.243 (0.339)	-0.244 (0.341)	0.0147 (0.109)	0.0148 (0.110)	-0.0738 (0.193)	-0.0738 (0.190)
Sex	-4.529 (5.838)	-4.533 (5.693)	-2.322 (2.097)	-2.321 (2.025)	0.401 (2.346)	0.401 (2.284)
Married	1.367 (3.514)	1.365 (3.478)	-0.252 (1.007)	-0.252 (1.025)	0.211 (1.479)	0.211 (1.500)
Month 4-Interaction		4.584 (4.365)		6.463*** (2.457)		0.274 (3.589)
Month 9-Interaction		13.68 (8.912)		8.883*** (2.466)		-0.416 (4.802)
After: Month 4		1.539 (3.674)		0.428 (1.122)		0.422 (2.820)
After: Month 9		10.93** (4.602)		-0.134 (1.377)		1.720 (2.808)
Constant	25.38 (26.26)	25.61 (26.25)	6.203 (8.186)	6.182 (8.347)	-0.333 (16.86)	-0.324 (16.58)
Observations	411	411	411	411	411	411

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

Cluster bootstrapped standard errors in parentheses.

Note: This table presents the ATET for three categories of formal care. Columns (1), (3), and (5) include an interaction term where we look at the post-treatment period as one period, while in column (4) we look at each post-treatment period separately. Control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, change of cognition score over time, age, marital status and sex. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on informal care

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total informal care	Total informal care	General care	General care	Everyday tasks	Supervision	Supervision
After	1.471 (1.021)		0.914* (0.492)		0.367 (0.596)	0.225 (0.292)	
Interaction	-2.719** (1.379)		-1.587** (0.621)		-0.666 (0.664)	-0.426 (0.527)	
Pain score	0.163 (0.216)	0.163 (0.215)	0.0914 (0.0739)	0.0913 (0.0746)	-0.0383 (0.104)	0.111 (0.0861)	0.111 (0.0857)
Agitation score	0.0136 (0.0572)	0.0136 (0.0595)	0.0235 (0.0295)	0.0235 (0.0290)	-0.00894 (0.0150)	0.00159 (0.0189)	0.00160 (0.0193)
Depression score	-0.224 (0.155)	-0.224 (0.153)	-0.115 (0.0827)	-0.114 (0.0867)	-0.0463 (0.0360)	-0.0706 (0.0491)	-0.0706 (0.0479)
Neuropsychia tric symptoms score	0.0647 (0.0434)	0.0648 (0.0419)	0.0391* (0.0230)	0.0391* (0.0229)	0.0202 (0.0178)	0.00326 (0.0158)	0.00325 (0.0156)
ADL score	0.0870 (0.0539)	0.0869* (0.0528)	0.0332 (0.0229)	0.0331 (0.0241)	0.0413 (0.0319)	0.0153 (0.0125)	0.0154 (0.0114)
Care distress score	-0.00472 (0.119)	-0.00491 (0.122)	-0.0352 (0.0568)	-0.0354 (0.0550)	-0.0146 (0.0508)	0.0545 (0.0378)	0.0545 (0.0380)
Age	-0.0776 (0.0912)	-0.0778 (0.0920)	-0.0357 (0.0479)	-0.0359 (0.0484)	-0.0755* (0.0388)	0.0415 (0.0316)	0.0415 (0.0322)
Sex	-0.656 (1.037)	-0.656 (0.979)	-0.207 (0.626)	-0.208 (0.639)	-0.658 (0.402)	0.220 (0.330)	0.220 (0.344)
Married	0.157 (0.724)	0.157 (0.721)	0.271 (0.332)	0.270 (0.330)	0.377 (0.270)	-0.487 (0.490)	-0.487 (0.496)
Month 4- Interaction		-2.036* (1.096)		-1.107** (0.547)			-0.361 (0.417)
Month 9- Interaction		-3.403* (2.022)		-2.068** (0.930)			-0.490 (0.798)
After: Month 4		0.794 (0.706)		0.527 (0.339)			0.241 (0.181)
After: Month 9		2.147 (1.865)		1.300 (0.869)			0.208 (0.689)
Constant	6.778 (6.262)	6.804 (6.305)	1.800 (3.151)	1.833 (3.180)	6.276** (2.905)	-2.521 (1.648)	-2.524 (1.671)
Observations	411	411	411	411	411	411	411

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

Cluster bootstrapped standard errors in parentheses. Note: This table presents the ATET for informal care and categories of informal care. Columns (1), (3), (5), and (6) include an interaction term where we look at the post-treatment period as one period, while in columns (2) and (4) we look at each post-treatment period separately. Control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, change of cognition score over time, age and sex. Standard errors are bootstrapped and clustered on NH level.

ATET of COSMOS on use of psychotropic medication

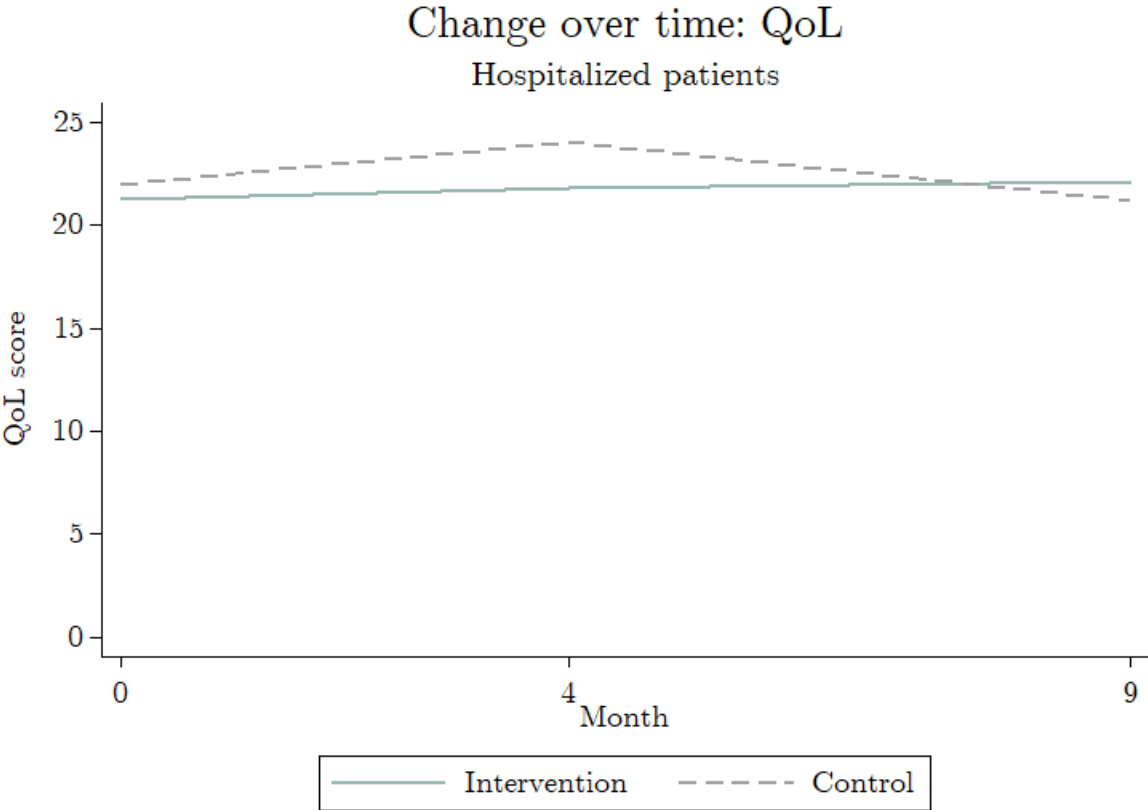
	(1)	(2)
After	-0.0902 (0.0639)	
Interaction	-0.0956 (0.0894)	
Pain score	0.0205 (0.0414)	0.0205 (0.0428)
Agitation score	-0.0126 (0.0114)	-0.0126 (0.0114)
Depression score	0.0325* (0.0187)	0.0324* (0.0182)
Neuropsychiatric symptoms score	0.0130 (0.0112)	0.0130 (0.0106)
ADL score	0.0128 (0.0141)	0.0129 (0.0141)
Care distress score	-0.0129 (0.0283)	-0.0129 (0.0286)
Age	-0.0253 (0.0193)	-0.0252 (0.0191)
Sex	-0.289 (0.264)	-0.289 (0.261)
Married	-0.207* (0.119)	-0.207* (0.119)
Month 4-Interaction		-0.131 (0.0876)
Month 9-Interaction		-0.0601 (0.112)
After: Month 4		-0.0482 (0.0547)
After: Month 9		-0.132 (0.0918)
Constant	4.032** (1.663)	4.027** (1.631)
Observations	411	411

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

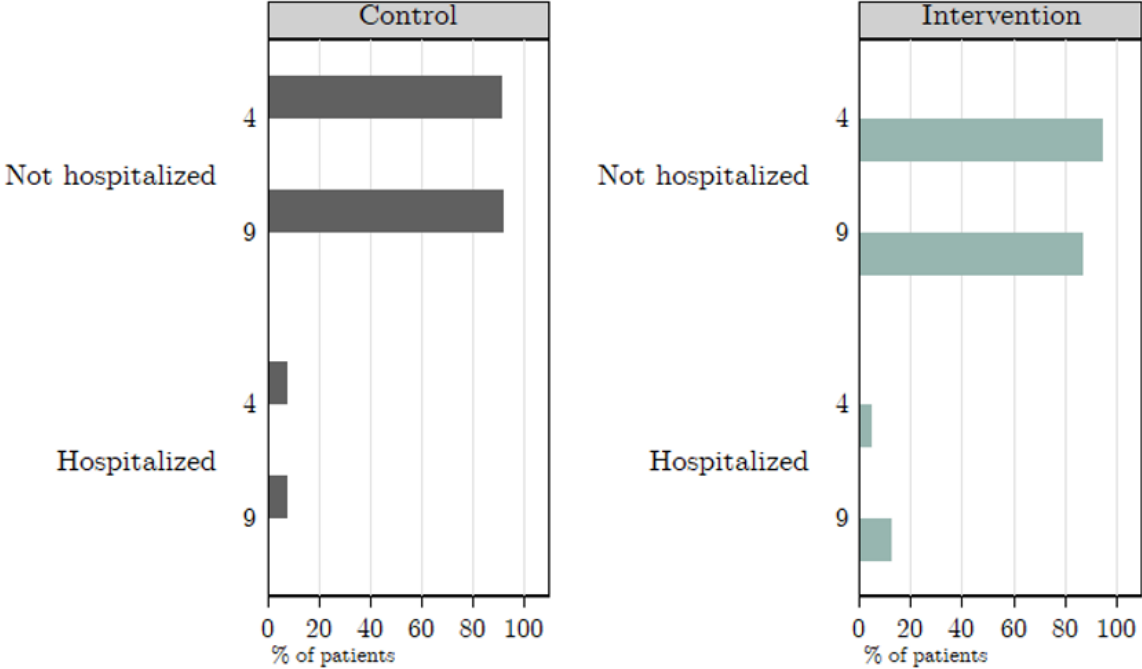
Cluster bootstrapped standard errors in parentheses.

Note: This table presents the ATET for use of psychotropic drugs. Column (1) include an interaction term where we look at the post-treatment period as one period, while in column (2) we look at each post-treatment period separately. Control variables include baseline measurement of pain, depression, agitation, ADL and neuropsychiatric symptoms score, change of cognition score over time, age, marital status and sex. Standard errors are bootstrapped and clustered on NH level.

Appendix 8 Change in QoL for hospitalized patients



Appendix 9 Change in hospital admissions



percent

This table shows the amount of patients (%) that have changed been admitted to a hospital for the months '4' and '9'.