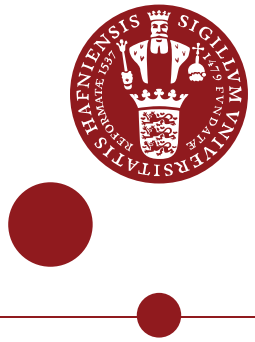
UNIVERSITY OF COPENHAGEN
DEPARTMENT OF ECONOMICS



# Essays on health care

## New insights into the economic burden of population ageing and the implementation of health policies aimed at promoting evidence-based clinical practice and consistency in the delivery of hospital care

**PhD Thesis**

**Anna Kollerup Iversen**

Supervisors: Mette Ejrnæs and Jakob Kjellberg

Submitted: 17 December 2021

# Contents

# Acknowledgements

I would like to thank my supervisors, Mette Ejrnæs and Jakob Kjellberg, for their outstanding guidance throughout my PhD studies. I thank my project group at VIVE (The Danish Center for Social Science Research), especially Mickael Bech and Jane Greve, for their assistance and establishing contacts to relevant researchers and collaborators. I would also like to thank my coauthors, Jakob Kjellberg, Sarah Wadmann, Rikke Ibsen and Toke Bek for great team work and useful feedback in writing the papers.

I would like to thank Meredith B. Rosenthal and Leemore Dafny for having me as a visiting PhD student at the Harvard T.H. Chan School of Public Health, and Søren Rud Kristensen for establishing contact to Meredith. Unfortunately, the research stay was shortened because of the COVID-19 pandemic, but the months I spent in Boston exceeded all my expectations. In this respect, I would also like to thank Torben og Alice Frimodts Fond, Knud Højgaards Fond, Augustinus Fonden and William Demant Fonden for pledging financial support for my research stay.

I thank colleagues and seminar participants at the University of Copenhagen and VIVE for helpful comments. Of course, a special thanks is given to Christine Halling, Ina Tapager and Lise Desireé Hansen with whom I have spent many hours both in and outside our office at VIVE. Thanks for all the fun, coffees and motivational talks that have kept me going. Finally, I would like to thank my friends and family, especially my husband Christian and my son Erik. Christian has supported me all the way and Erik has been the best distraction from research.

This PhD was made possible by a PhD position at VIVE funded by the Novo Nordisk Foundation. I am very grateful for all the opportunities, challenges and perspectives I have been given throughout this PhD, and I am looking forward to continuing my work within the field of health economics for many years to come.

*This thesis is part of an interdisciplinary research project titled "Socioeconomic impact of policy instruments for health research dissemination". The research project is funded by the Novo Nordisk Foundation and is divided into three work packages. The first is concerned with exploring the preconditions for and the influence of establishing research environments in hospitals on the uptake of new clinical guidelines. The second explores the impact of centralizing specialized hospital care and the third examines the impact of clinical practice guidelines on clinical practice and socioeconomic outcomes. With this PhD thesis, I contribute to the second and third work packages.*

# English introduction

The future delivery of universal health care coverage is increasingly challenged by an ageing population and a continuous introduction of expensive medical advances [1, 2, 5]. Without large budget expansions, national health care authorities and health care managers face complex organizational decisions to be on the frontier in providing high-quality health care. Correspondingly, tight health care budgets may challenge a fast and consistent adoption of new knowledge and medical advances.

This thesis consists of three self-contained chapters. Each chapter casts light on different aspects of the challenges that national health authorities face in providing universal health care coverage under resource constraints and under the continual introduction of new knowledge and new expensive medical advances. To inform the debate on the future economic burden of population ageing and possible threats to the sustainability of universal health care coverage, the first chapter examines static and dynamic properties of the past decade's health care expenditure patterns across age, disease groups and time in Denmark. The second and third chapters examine the implementation of two evidence-based health policies, namely the implementation of national clinical practice guidelines and a regulation of the supply of hospital care by centralization to fewer and larger clinical facilities. The policies had the common goal of promoting evidence-based clinical practice and procedures in hospital care by mitigating slow adoption of new knowledge and reducing variation in the adoption of new medical advances, routines and procedures. The second chapter analyzes the effects of centralizing breast cancer care to fewer high-volume clinics on both cost-saving metrics and the quality of care that closure-affected patients receive. The third chapter explores whether consistency in the delivery of health care is achieved when introducing a new treatment with both national clinical guidelines and care centralization.

From the writing of this thesis, I have learned that if the past decades' health care expenditure trends continue in the coming decades, it will become increasingly difficult to provide universal health care coverage without budget expansions. Using Danish registry data, my coauthors and I show that the individual-level age-distributed health care expenditure curve has steepened over a 12-year period from 2006 to 2018, and that expenditures are correlated with age and not only closeness to the time of death. We also identify diagnosis heterogeneity in the extent and age-distribution of the steepening. Hence, future implications of the steepening of the individual-level age-distributed health care expenditure curve will depend on the health status and disease pattern of the elderly, but also on the prioritization and allocation of health care to different patient groups.

I have also learned that there is no easy path to mitigating slow and non-uniform adoption of new knowledge and medical advances, and there is no one-size-fits-all policy solution across all clinical areas. Together with my coauthors, I show that the value of clinical practice guidelines is undermined when structural barriers limit the ability of clinicians and clinical managers to scale up necessary treatment. When studying patients diagnosed with wet age-related macular degeneration (wAMD), we find clinical practice variation attributable to regional differences in the preparedness for scaling up treatment activity to cope with substantial patient accumulation. We also find that the regions that managed to delegate and outsource treatment activity – rather than centralize – obtained higher treatment effects, presumably because this provided for higher treatment intensity. Conversely, when looking into the centralization of breast cancer surgery, I found that centralization may be a good solution to reduce unwarranted variation in surgical procedures and hospitalization days, as routines and professional abilities vary across clinical facilities, even though we have national clinical guidelines with clear definitions of state-of-the-art procedures.

National clinical guidelines and care centralization have been dominating health policies in Denmark in the past decades. Although these policies might increase knowledge uptake and decrease unwarranted variation when not subject to structural barriers, the policies do not include

exact solutions to the future challenges of rising health care expenditure levels. Therefore, future health policies should find a compromise between implementing the newest evidence-based knowledge while stimulating cost-effective medical advances, facilities and routines. I hope this thesis enlightens and informs the debate on the effects of the past decades' health policies and on the challenges that national health authorities face in providing high-quality universal health care for an ageing population.

**Chapter 1**
**Ageing and health care expenditures: the importance of age per se, steepening of the individual-level expenditure curve, and the role of morbidity**

*with Jakob Kjellberg and Rikke Ibsen*[1]

In chapter one, we examine static and dynamic health expenditure patterns across a 12-year period to shed light on future health care needs and threats to the sustainability of universal health care coverage. The demographic change towards a larger proportion of older individuals challenges universal health care systems in sustaining high-quality care and universal coverage without budget expansions. To build valuable predictions of the economic burden from population ageing, it is crucial to understand the determinants of individual-level health care expenditures. Often, studies have focused on the relative importance of an individual's age and time to death, while only a few newer studies highlight that individual-level health care expenditures seem to be increasing faster for the elderly – i.e. creating a steepening of the individual-level health care expenditure curve over time. If this steepening hypothesis is valid and the steepening tendencies continue in the coming decades, it will reinforce the economic challenges associated with population ageing. By applying individual-level administrative data from the entire Danish population, our study is the first to use a single data set to examine whether age, time to death and a steepening of the individual-level health care expenditure curve all contributed to individual-level health care expenditures over a 12-year observation pe-

[1] Chapter 1 is accepted for publication in *European Journal of Health Economics*, and it builds on initial work presented in Kjellberg and Ibsen [3] and Kollerup and Kjellberg [4].

riod (2006–2018). We find that individual-level expenditures are associated with an individual's age, an individual's time to death and a steepening of the expenditure curve, with the steepening driven by individuals above age 75. We observe heterogeneity in the extent and age-distribution of the steepening across disease groups. The threefold combination of an ageing population, the correlation between expenditures and age per se, and a steepening of the expenditure curve make establishing financially sustainable universal health care systems increasingly difficult. To mitigate budgetary pressure, we suggest that policy-makers encourage cost-effective medical advances and health care utilization in the treatment of elderly people. Moreover, we suggest that future health care expenditure forecasts include scenarios with a steepening of the expenditure curve.

**Chapter 2**
**Worth the trip? The effect of hospital clinic closures for patients undergoing scheduled surgery**

In chapter two, I examine the effects of hospital clinic closures on patients living in municipalities where their nearest breast cancer clinic closes. Recent decades have seen a large number of hospital closures and consolidations, which have been carried out to stimulate returns to volume and specialization in hospital care. In the non-acute setting of scheduled breast cancer surgery, I examine how hospital clinic closures affect cost-saving metrics and the quality of care that closure-affected patients receive. The effects are identified using closures of breast cancer clinics in Denmark from 2000 to 2011, during which time the number of clinics was more than halved. Using event study designs, I examine changes in surgical outcomes for patients living in municipalities where the nearest clinic had been closed. The results show that breast cancer clinic closures have been welfare-improving, as they have reduced the number of costly hospitalization days and shifted surgical procedures to state-of-the-art breast-conserving techniques without generating adverse health effects and without causing crowding in non-closing clinics. An examination of the mechanisms suggests that added volume returns at non-closing clinics were of less importance than simply reallocating patients to higher-quality clinics.

**Chapter 3**

**National clinical guidelines and treatment centralization do not guarantee consistency in healthcare delivery. A mixed-methods study of wet age-related macular degeneration treatment in Denmark**

*with Sarah Wadmann, Jakob Kjellberg and Toke Bek*[2]

As clinical practice variation has been problematized as a symptom of suboptimal care and inefficient resource spending, consistency in the delivery of healthcare is a recurring policy goal. We examine a case where the introduction of a new treatment is most likely to provide consistency in health care delivery because it was introduced with a national clinical practice guideline representing consensus about best clinical practice among leading clinicians, and because care delivery was highly centralized to few high-volume treatment units. Despite the consensus on best clinical practice and care centralization, this study shows pronounced regional variation in patient outcomes and treatment costs. Using a mixed-methods design, we find that the lack of consistency in care was largely unrelated to patient-specific characteristics, but seemed to reflect structural differences in the regional organization and financing of healthcare delivery. We conclude that the value of clinical practice guidelines is undermined when structural barriers limit the ability of clinicians and clinical managers to scale up treatment, and that some degree of decentralization may be a tool to maintain treatment intensity when the treatment effect is dependent on a high treatment intensity.

[2] A shorter version of chapter 3 has been submitted to the journal *Health Policy* October 2021.

# Dansk introduktion

Den aldrende befolkning og løbende introduktion af nye medicinske fremskridt, behandlinger og teknologier lægger økonomisk pres på muligheden for at tilbyde universel sundhedsdækning af høj kvalitet i fremtiden [1, 2, 5]. Uden tilføjelse af ekstra ressourcer til sundhedsvæsenet medvirker dette til, at der skal træffes stadig flere komplekse nationale beslutninger for at kunne tilbyde ydelser og behandlinger af høj kvalitet. Ligeledes vil stramme sundhedsbudgetter udfordre en hurtig og konsistent introduktion af ny forskning, nye behandlinger og teknologier.

Denne ph.d.-afhandling består af tre selvstændige kapitler. Hvert kapitel belyser forskellige aspekter af de udfordringer, som nationale sundhedsmyndigheder står over for ved at tilbyde og opretholde universel sundhedsdækning under begrænsede ressourcer og under en kontinuerlig introduktion af ny evidens og nye behandlingsmuligheder. Som bidrag til debatten om den fremtidige økonomiske byrde af den aldrende befolkning og trusler mod bæredygtigheden af universel sundhedsdækning undersøger mine medforfattere og jeg i første kapitel de statiske og dynamiske mønstre i de danske sundhedsudgifter i det seneste årti på tværs af aldersgrupper og sygdomsgrupper. I andet og tredje kapitel undersøges implementeringen af to sundhedspolitikker: implementering af nationale kliniske retningsliner og regulering af udbuddet af specialiseret hospitalsbehandling til færre og større afdelinger. De to sundhedspolitikker blev implementeret med henblik på at fremme evidensbaseret klinisk praksis og procedurer i hospitalsvæsenet ved at reducere omfanget af langsommelig implementering af ny evidens og ved at reducere variationen i implementeringen af denne. I det andet kapitel analyseres effekterne af at centralisere brystkræftbehandling til færre afdelinger. I det tredje kapitel undersøge det, om lighed i behandling kan opnås ved at introducere ny behandling samtidig med nationale kliniske retningslinjer og behandlingscentralisering.

Under udarbejdelsen af denne afhandling har jeg lært, at hvis tidligere årtiers trends i sundhedsudgifterne forsætter i de kommende årtier, så vil det blive endnu sværere at yde universel sundhedsdækning af høj kvalitet uden tilførsel af ekstra ressourcer til sundhedsvæsenet. Ved brug af danske registerdata viser mine medforfattere og jeg, at sundhedsudgifterne for det enkelte ældre individ er steget hurtigere end for det enkelte yngre individ over en 12-årig periode fra 2006 til 2018, dvs. at den individspecifikke aldersfordelte sundhedsudgiftskurve er blevet stejlere over tid (steepening). Samtidig finder vi, at sundhedsudgifterne er korreleret med et individs alder og ikke kun individets antal år til død. Vi finder også heterogenitet i de nævnte udgiftsmønstre på tværs af sygdomsgrupper, hvor den individspecifikke aldersfordelte sundhedsudgiftskurve er blevet stejlere for nogle sygdomsgrupper sammenlignet med andre. Implikationerne af disse fund vil afhænge af de fremtidige sygdomsmønstre og den generelle sundhed blandt ældre, men vil også afhænge af, hvordan sundhedsvæsenets ressourcer prioriteres og allokeres til forskellige patientgrupper.

Jeg har også lært, at der ikke er nogen let vej til at reducere langsommelig og varierende implementering af ny evidens i sundhedsvæsenet, og der er ikke én sundhedspolitik, som kan løse udfordringerne på tværs af alle kliniske områder. Sammen med mine medforfattere viser jeg, at værdien af nationale kliniske retningslinjer undermineres, når strukturelle barrierer begrænser sygehusafdelingers evne til at opskalere nødvendig behandling. Vi undersøger implementeringen af en ny behandling til patienter med våd aldersrelateret macula degeneration (vAMD) og finder klinisk praksisvariation, der kan kobles til regionale forskelle i paratheden til at opskalere behandlingsaktiviteten, når antallet af patienter, der skal modtage behandling, stiger kraftigt. Vi finder også, at de regioner, som formåede at decentralisere behandlingsaktiviteten – frem for at centralisere – opnåede højere behandlingseffekt, fordi de var i stand til at yde en højere behandlingsintensitet. I modsætning hertil finder jeg, når jeg undersøger centralisering af brystkræftkirurgi, at centralisering kan være en god løsning til at reducere uønsket variation i kirurgiske procedurer og indlæggelsesdage, da rutiner og professionelle evner varierer på tværs af sygehusafdelinger, selvom der eksisterer kliniske retningslinjer.

Nationale kliniske retningslinjer og behandlingscentralisering har været dominerende sundhedspolitikker i Danmark de seneste årtier. På trods af, at disse sundhedspolitikker kan øge vidensoptaget og mindske uønsket variation i fraværet af strukturelle barrierer, så har de ikke fokus på de fremtidige udfordringer med stigende sundhedsudgifter. Derfor bør der i fremtidige sundhedspolitiker findes et kompromis mellem at fordre en implementering af nyeste evidensbaserede viden, samtidig med at der tilskyndes til omkostningseffektiv brug af medicin, behandlinger og teknologier i behandlingen af især ældre individer. Jeg håber, at denne afhandling vil bidrage til debatten om effekterne af tidligere årtiers sundhedspolitikker og nationale sundhedsmyndigheders udfordringer med at opretholde en universel sundhedsdækning af høj kvalitet i en tid med en aldrende befolkning.

**Kapitel 1**
**Ageing and health care expenditures: the importance of age per se, steepening of the individual-level expenditure curve, and the role of morbidity**

*med Jakob Kjellberg og Rikke Ibsen*[3]

I det første kapitel undersøger vi statiske og dynamiske mønstre i sundhedsudgifterne over en 12-årig periode (2006-2018). Vi gør dette for at belyse de fremtidige behov for sundhedsydelser og trusler mod bæredygtigheden af universel sundhedsdækning. Den demografiske udvikling mod en større andel af ældre individer udfordrer den universelle sundhedsdækning i fremtiden, hvis ikke der tilføjes ekstra ressourcer til sundhedsvæsenet. For at kunne generere pålidelige fremskrivninger af den økonomiske byrde af den aldrende befolkning er det vigtigt at komme nærmere, hvad der driver de individspecifikke sundhedsudgifter. Ofte har fokus i fremskrivningen af sundhedsudgifterne været den relative vægtning af et individs alder og dets antal leveår til død, mens kun få studier beskriver, at sundhedsudgifterne kan være steget hurtigere for det enkelte ældre end det yngre individ, dvs. den aldersfordelte udgiftskurve er blevet stejlere over

[3]Kapitel 1 er accepteret til publikation i *European Journal of Health Economics*, og bygger videre på tidligere arbejde præsenteret i Kjellberg og Ibsen [3] og Kollerup og Kjellberg [4].

tid (steepening). Hvis sundhedsudgifterne fremadrettet også stiger hurtigere for det ældre end det yngre individ, vil det økonomiske pres stige på grund af den stigende andel af ældre i befolkningen. Ved at benytte administrativ data for hele den danske population på individniveau er dette studie det første til at benytte det samme datasæt til at undersøge, om både alder, antal leveår til død og en stejlere aldersfordelt udgiftskurve har bidraget til de individspecifikke sundhedsudgifter i perioden 2006-2018. Vores resultater viser, at de individspecifikke sundhedsudgifter både er korrelerede med et individs alder, et individs antal leveår til død og en stejlere aldersfordelt udgiftskurve. Vi observerer også heterogenitet i udviklingen i den aldersfordelte udgiftskurve over tid på tværs af sygdomsgrupper. Kombinationen af en aldrende befolkning, korrelationen mellem sundhedsudgifter og alder i sig selv samt en stejlere aldersfordelt udgiftskurve over tid lægger et øget pres på det fremtidige sundhedsvæsen. For at afbøde det store økonomiske pres foreslår vi, at beslutningstagere tilskynder til omkostningseffektiv brug af medicin, behandlinger og teknologier i behandlingen af især ældre individer. Derudover foreslår vi at tilføje scenarier for en stigende aldersfordelt udgiftskurve til fremskrivninger af sundhedsudgifterne.

**Kapitel 2**
**Worth the trip? The effect of hospital clinic closures for patients undergoing scheduled surgery**

Under hypotesen, at de hospitaler, der udfører flest behandlinger, har den største faglige ekspertise, har de seneste årtier været præget af et stort antal hospitalslukninger og konsolideringer. I kapitel 2 analyserer jeg effekterne af at lukke afdelinger, der foretager planlagte brystkræftoperationer. Effekterne er identificeret ved at benytte lukninger af danske afdelinger, der udfører brystkræftoperationer, fra 2000 til 2011, hvor mere end halvdelen af brystkræftafdelingerne i Danmark blev lukket. Ved at benytte et event-studie design undersøger jeg ændringerne i omkostningsbesparende mål og kvaliteten af behandling for patienter, der bor i kommuner, hvor den nærmeste brystkræftafdeling lukker. Resultaterne viser, at lukning af brystkræftafdelinger har været velfærdsforbedrende, da det for patienter i lukningsramte kommuner har reduceret antallet af hospitalsdage og medført et skift mod brystbevarende operationer uden at generere negative

sundhedseffekter og uden at medvirke til trængsel på ikke-lukningsramte brystkræftafdelinger. En undersøgelse af mekanismerne viser, at effekterne er drevet af, at patienter fra lukningsramte afdelinger efter lukningen bliver behandlet på afdelinger med højere kvalitet. Desuden finder jeg ikke evidens for, at der er forekommet en stor patientakkumulation på ikke-lukningsramte afdelinger, som kunne have givet anledning til yderligere stordriftsfordele.

**Kapitel 3**

**National clinical guidelines and treatment centralization do not guarantee consistency in healthcare delivery. A mixed-methods study of wet age-related macular degeneration treatment in Denmark**

*Med Sarah Wadmann, Jakob Kjellberg og Toke Bek*[4]

Variation i klinisk praksis bliver ofte problematiseret som et resultat af suboptimal behandling og ineffektiv brug af ressourcer. Derfor er lighed i behandling på tværs af hospitaler og geografiske regioner en tilbagevendende målsætning. I det tredje kapitel undersøger vi en case, hvor introduktionen af en ny behandling med stor sandsynlighed skulle have resulteret i ensartet behandling på tværs af hospitalsafdelinger, i og med at behandlingen blev introduceret med både nationale kliniske retningslinjer, og behandlingen var centraliseret til få, større afdelinger. På trods af behandlingscentralisering og konsensus om den bedste kliniske praksis viser analyserne i kapitlet, at der var udtalt variation i effekten af og omkostningen ved behandlingen. Vi benytter i kapitlet et mixed-methods-design og viser, at manglen på ensartethed i behandlingen ikke var drevet af forskelle i patientsammensætning, men reflekterer strukturelle forskelle i den regionale organisering og finansiering af behandlingen. Vi konkluderer, at værdien af kliniske retningslinjer undermineres, hvis der er strukturelle barrierer, som begrænser klinikere og lederes muligheder for at skalere behandlingen til stor patientakkumulation og høj behandlingsintensitet. Vi finder ligeledes, at nogen grad af decentralisering kan være et redskab til at opretholde behandlingsintensiteten, når behandlingseffekten afhænger af høj behandlingsintensitet.

[4]En kortere version af kapitel 3 er indsent til tidsskriftet *Health Policy* oktober 2021.

# Bibliography

[1] Jonathan Cylus, Josep Figueras, and Charles Normand. *Will poulation ageing spell the end of the welfare state? A review of evidence and policy options*. World Health Organization, 2019.

[2] Kamil Dybczak, Bartosz Przywara, et al. *The role of technology in health care expenditure in the EU*. Directorate General Economic and Financial Affairs (DG ECFIN), European Commission, 2010.

[3] Jakob Kjellberg and Rikke Ibsen. *Fremtidens sundhedsudgifter*. VIVE - The Danish Center for Social Science Research, 2020.

[4] Anna Kollerup and Jakob Kjellberg. *Flere ældre og nye behandlinger*. VIVE - The Danish Center for Social Science Research, 2018.

[5] Dorothée Rouzet, Aida Caldera Sánchez, Theodore Renault, and Oliver Roehn. *Fiscal challenges and inclusive growth in ageing societies*. OECD Economic Policy Papers. No 27, 2019.

# Chapter 1

# Ageing and health care expenditures: The importance of age per se, steepening of the individual-level expenditure curve, and the role of morbidity

*With Jakob Kjellberg and Rikke Ibsen*

The European Journal of Health Economics
https://doi.org/10.1007/s10198-021-01413-x

ORIGINAL PAPER


# Ageing and health care expenditures: the importance of age per se, steepening of the individual-level expenditure curve, and the role of morbidity

Anna Kollerup[1] · Jakob Kjellberg[1] · Rikke Ibsen[2]

Received: 31 January 2021 / Accepted: 11 November 2021


**Abstract**
The demographic change towards a larger proportion of older individuals challenges universal health care systems in sustaining high-quality care and universal coverage without budget expansions. To build valuable predictions of the economic burden from population ageing, it is crucial to understand the determinants of individual-level health care expenditures. Often, the focus has been on the relative importance of an individual's age and time to death, while only a few newer studies highlight that individual-level health care expenditures are increasing faster for the elderly—i.e., creating a *steepening* of the individual-level health care expenditure curve over time. Applying individual-level administrative data for the entire Danish population, our study is the first to use a single data set to examine whether age, time to death, and a steepening of the individual-level health care expenditure curve all contributed to individual-level health care expenditures over a 12 year observation period (2006–2018). We find that individual-level expenditures are associated with an individual's age, an individual's time to death, and a steepening of the expenditure curve, with the steepening driven by individuals above age 75. We observe heterogeneity in the extent and age distribution of steepening across disease groups. The threefold combination of an ageing population, the correlation between expenditures and age per se, and a steepening of the expenditure curve make establishing financially sustainable universal health care systems increasingly difficult. To mitigate budgetary pressure, policy-makers should stimulate cost-effective medical advances and health care utilization in the treatment of elderly. Moreover, steepening scenarios should be added to future health care expenditure forecasts.

**Keywords** Health expenditures · Steepening · Red herring · Healthy ageing · Demographic change

## Introduction

The predicted rising proportion of older individuals in most G20 countries, combined with rising longevity, pose severe challenges to the establishment of a financially sustainable infrastructure for the future delivery of universal health care coverage [1, 2]. Given that the large and growing share of elderly individuals is without precedent, predicting the specific gap between future demand and the supply of health care is difficult in the absence of updated knowledge about past decades health care expenditure patterns.

The future implications of population ageing on individual-level health care expenditures depend on the health status and disease pattern of the elderly, but also on the prioritization and allocation of health care to different patient groups. The extensive amount of expenditure determinants makes health care expenditure predictions increasingly difficult. For example, existing studies find that rising longevity may postpone the individual need for health care services to higher ages [3], whereas other studies emphasize that rising longevity may imply more years in ill health [4]. Moreover, politicians and health care administrators may prioritize or health care professionals may treat individuals with a high expected longevity more intensely than individuals with a low expected longevity [5], but health care resources may

✉ Anna Kollerup
aniv@vive.dk

Jakob Kjellberg
jakj@vive.dk

Rikke Ibsen
rikke@i2minds.dk

[1] VIVE-The Danish Center for Social Science Research, Herluf Trolles Gade 11, 1052 Copenhagen, Denmark

[2] I2minds, Åboulevarden 39, 1. Th, 8000 Aarhus, Denmark

Published online: 17 January 2022

also be directed towards the treatment of elderly as a rising proportion of elderly individuals may generate a political focus on the provision of specific health care services to the elderly population [6]. Hence, we also face a reverse causal pathway from expenditure levels to health and longevity if patients are prioritized based on their age or longevity.

The best knowledge about future implications of population ageing and the future need of health care resources must be obtained by studying properties of previous expenditure patterns; patterns of the age-distributed individual-level expenditure curve and the time-to-death-distributed individual-level expenditure curve. For example, the age-distributed curve may be parallel-shifted to the right if expenditures are simply postponed to older ages as longevity increases, and the slope of the curve might change if specific age groups are prioritized more than others over time.

In this study, we examine health care expenditure patterns on population-wide individual-level data across a 12 year period to shed light on future health care needs and threats to the sustainability of universal health care coverage. In particular, we examine both static and dynamic properties of previous expenditure patterns, and we have a particular focus on introducing individual-level health status in two strands of literature; (1) the *red herring* literature debating whether individual-level expenditures are best described by time to death or age, and, implicitly, whether time to death or age is the best proxy for individual-level health; (2) the literature on *steepening* describing the dynamic development in the shape of the age-distributed individual-level expenditure curve over time. Previous studies have focused on either *the red herring* or the *steepening* hypothesis, although the concepts theoretically are independent and could coexist where individual-level expenditures might increase faster for older age groups even if individual-level expenditures are postponed to higher ages as longevity increase [7, 8].

The *red herring* literature dates back to 1999 where Zweifel, Felder, and Meiers [3] began the "healthy ageing" debate by suggesting that individual-level health status, and, in turn, expenditure levels, should be approximated by an individual's time to death rather than the individual's age. The authors highlighted a risk of creating a 'red herring' if policy-makers base their decisions on high calendar age as a cause of expenditure growth instead of elderly cohorts' time to death. Since then, their argument has been referred to as *the red herring hypothesis*. The validity of *the red herring hypothesis* remains mixed [4, 9–13] and recent studies show that the importance of time to death is reduced when adding individual-level health status, such as clinical risk groups or morbidity, to the model [14, 15]. These studies reveal that adding measures of individual-level health status to the models reduces the importance of time to death, but that time to death is still a significant predictor of health care expenditures. If time to death was a perfect predictor for the end-of-life expenditures contributable to individual-level health status, the effect of time to death should reach 0 when adding measures of health status. This highlights that more research in obtaining good measures of individual-level health status is needed, and that some measures of individual-level health may be stronger related to end-of-life expenditures than others. With the present study, we contribute to the red herring literature by estimating the models on population-wide individual-level data and by including 28 chronic diseases to account for individual-level health status. We find that both age and time to death have an impact on the use of health care services and on individual-health care expenditure levels. Adding chronic diseases as measures of individual-level health status to the model lower the magnitude of the age and time to death coefficients, but both age and time to death remain important predictors for individual-level health care expenditures. Besides the importance of both age and time to death, we find that death-related expenditures increase over time, and from these findings together, we conclude that the red herring hypothesis cannot be confirmed.

In the literature on steepening, researchers test the hypothesis that individual-level expenditures for older individuals increase significantly faster than for younger ones [7, 16], i.e., creating a *steepening* of the individual-level expenditure curve over time. With this study, we are the first to test the steepening hypothesis with individual-level data and the first to examine disease-specific cost-drivers of steepening. We examine whether two major causes of death (cancer and cardiovascular disease) and chronic diseases in general are particular contributors to steepening or whether steepening occurs across a broad range of diseases. Applying individual-level administrative data on physician care and inpatient and outpatient care for the entire Danish population over a 12 year observation period, our study also expands the number of included health care services and the degree of detail compared to previous studies examining steepening of health care expenditures. Our results show that the age-distributed health care expenditure curve has become steeper over the 12 year period 2006–2018. The overall steepening is driven by the cost of treating the oldest individuals above 75 years. However, we find large heterogeneity in the extent and distribution of steepening across disease groups. We observe more steepening across the age distribution of individuals with a chronic disease compared to individuals without a chronic disease, but less steepening among individuals with a cardiovascular disease compared to individuals without a cardiovascular disease. When comparing individuals with and without a cancer disease, we find evidence of steepening in both groups, but with very different changes in the distribution of expenditures over time.

As we find a correlation between individual-level expenditures and age per se, and a steepening of the expenditure

curve, population ageing makes establishing financially sustainable universal health care systems increasingly difficult. To prevent negative consequences, policy-makers should stimulate cost-effective medical advances and health care utilization in the treatment of elderly. Moreover, steepening scenarios should be added to future health care expenditure forecasts, and preferably scenarios that take different disease patterns into account.

## Background

Since the late 1990s, researchers and policy-makers have focused on the relation between health expenditures and ageing, debating how much of individual-level health expenditures are determined by an individual's age and closeness to the time to death. As declining health status results in rising health care expenditures and ultimately in death, researchers have highlighted time to death as being a good predictor of health care expenditures, and possibly a better predictor than age. However, the evidence on the relative importance of age and time to death remains mixed. Some studies support the red herring hypothesis [3], proposing that individual-health care expenditures after age 65 are primarily determined by closeness to the time of death, not by calendar age per se [4, 10–13]. However, newer studies find that although increased longevity may postpone late-life health care expenditures, late-life expenditures are determined by both age and time to death [4, 9] along with mortality and 5 year survival rates [5]. In relation to the mixed evidence, some studies find that increasing longevity imply an increase in the number of required treatment years [4], while others find that health care expenditures may be concentrated in fewer years over time as longevity increases [7, 17] or that physicians treat patients with high expected longevity more intensely than individuals with low expected longevity, because treatments then pay off over a longer time span, dubbed "Eubie Blake effect". [5]. In recent years, studies have highlighted the importance of including individual-health status or morbidity indicators in the analyses of health care expenditures, along with age and time to death. These studies find that health care expenditure levels are strongly correlated with clinical risk groups [14], morbidities at the time of admission [15], and utilization of health care services (e.g., hospitalizations and drug expenditures) [18]. Carreras et al. [14] and Howdon and Rice [15] show that adding clinical risk groups or morbidities to the models reduces the importance of time to death, and that time to death can act as a proxy for individual-level health status in explaining health care expenditures. However, both time to death and age are crude measures for the underlying processes of health status [19], and it is very relevant to examine different measures of individual-level health status to explain health care demand, and, in turn, expenditures. The recent studies highlight that time to death is not a perfect proxy for the included measures of individual-level health, or at least that there is still room for interpretation on what part of individual-health is captured by time to death and what part is not captured.

Other studies have focus on the dynamic development in the age-distributed individual-level health expenditure curve over time. In 2006, Buchner and Wasem [16] introduced the term 'steepening,' meaning that the age-related health care expenditure curve has become steeper as the individual-level health care expenditures for older age groups increase significantly faster than for younger ones. This was later rejected by Felder and Werblow [20], but confirmed by Gregersen [7]. Gregersen [7] highlights the importance of accounting for death-related expenditures over time and across age groups as part of the steepening effect is driven by increased death-related expenditures. Empirical evidence for the mechanisms underlying steepening remains largely absent and theoretical suggestions are sparse. Some of the theoretical suggestions for the mechanism behind steepening include increased demand among the elderly, bias in the technological frontier towards the elderly, or simply more years spent in ill health [6, 20]. For example, increased demand among elderly may increase the supply of health care services to elderly as an increasing proportion of elderly voters creates political pressure and pharmaceutical companies become more aware of emerging markets. Besides an increasing proportion of elderly, changes in disease patterns may also affect the shape and slope of the age-distributed individual-level expenditure curve over time. For example, the number of elderly patients with multiple chronic conditions is rising, and researchers have shown that individual-level health care expenditures are positively correlated with both the number of comorbidities [21] and the combination of single diseases [22] an individual have. If elderly individuals live longer with multiple chronic conditions, we might observe that the age-distributed individual-level expenditure curve becomes steeper, but that the highest peak of the curve also becomes wider as individuals exhibit high expenditures for more years.

## Hypotheses

### Steepening of the individual-level health care expenditure curve over time

To examine some of the theoretical suggestions behind steepening, we hypothesize that some disease groups may contribute more to a steepening of the curve than others. In this paper, we assess whether two major causes of death, cancer and cardiovascular diseases, have contributed significantly more to steepening of the individual-level health

care expenditure curve than other diseases. Moreover, we aggregate all chronic diseases, including cancer and cardiovascular diseases, and assess chronic diseases relative to all other diseases. With respect to cancer, the expenditures related to treating cancer have increased relatively more than for other diseases in Denmark over the observation period [23]. If cancer expenditures increases are driven by treating the oldest cancer patients more intensively or for more years than younger patients, this will imply significantly more steepening among cancer diagnosed. For example, this may be the case if there has been a technological bias in the development of new medical advances towards the treatment of cancer diagnoses that are mostly prevalent among a large group of older individuals, e.g., pancreas and lung cancer patients. In addition, we hypothesize that the groups of individuals with a cardiovascular disease or a chronic disease in general may have contributed significantly more to steepening than other diagnosed as these proportionally large groups consists of many life-style-related diseases and chronically ill patients with several chronic diseases where increased longevity may imply more years spent in ill health and, in turn, cause steepening.

### The relative importance of age and time to death

As introduced in Sect. Background, the strand of literature considering the relative importance of age and time to death dates back to the 1990s and remains mixed. The relative importance in the present study cannot be predicted beforehand, but we hypothesize that the importance of time to death depends on the age groups, and therefore, we assess the importance of time to death for 5 year age groups. The most recent studies find that the effect of time to death is reduced when accounting for individual-level health status [14, 15]. We hypothesize that the inclusion of 28 chronic diagnoses as explanatory variables in our models yields similar interpretations as in Carreras et al. [14] and Howdon and Rice [15]. Our result might differ in magnitude from Howdon and Rice [15] as they look at pure decedent data, and as we examine another measure of individual-level health status that may be more or less correlated with time to death and age. We are interested in the overall problem of ageing, and therefore, we include both survivors and decedents in our analyses [24].

## Data and summary statistics

To calculate the annual individual-level health care expenditure, we combine expenditure data from the Danish National Patient Registry and the Danish National Health Insurance Registry. For each individual, we calculate the total annual health care expenditure as the sum of expenditures from inpatient and outpatient hospital care and physician care.[1] For inpatient and outpatient care, we apply the DRG-grouped National Patient register where all admitted patients are grouped according to the diagnosis related group (DRG) system and the Danish outpatient group system (DAGS).[2]

For inpatient care, we take the sum of the expenditures of all DRG-contacts for each individual in each year. The expenditure for each DRG-contact is calculated as the sum of a DRG-tariff and a fixed tariff for extended hospital stay if the hospital stay exceeded the maximum number of days covered by the DRG-tariff. For outpatient care, we take the sum of the expenditures of all DAGS-contacts for each individual in each year where each DAGS-contact has a DAGS-tariff. The DRG- and DAGS-tariffs cover procedures for all inpatient and outpatient contacts in somatic hospitals, including mental care contacts in the somatic hospital system, and the tariffs are provided directly in the National Patient register. To calculate annual health care expenditure per patient for physician care, we apply the National Health Insurance register that includes data on the expenditures on all individual services undertaken by physicians, and we take the sum of all expenditures for each individual in each year.

All expenditures have been deflated with the Danish consumer price index, such that all expenditures are measured in 2018-prices. We have not been able to include expenditures on prescription medicines from the primary sector (4 pct. of total HCE in Denmark in 2017 [25]) or expenditure data on Long-Term Care (LTC) as validated Danish data on LTC are not available for this type of longitudinal analysis. The expenditure data are merged with administrative registers from Statistics Denmark by means of personal identification numbers. The administrative registers contain data on date of death and demographic covariates.

In some of the steepening analyses, we include diagnostic information on the individuals to create diagnosis groups of cancer diagnosed, cardiovascular disease diagnosed, and chronic disease diagnosed. The diagnostic information is given by ICD-10 codes in the Danish National Patient Registry (in- and outpatient contacts), and when assigning individuals to diagnostic groups, we only use the main diagnosis. An individual may be present in several diagnosis groups if the individual, for example, is diagnosed with both cancer and a cardiovascular disease in the same year. When estimating the annual individual-level expenditure, we

[1] Physician care cover services undertaken in general practice, specialist doctors, dentists, physiotherapists, psychologists, and other services supported by the Danish Health Insurance system.

[2] The Danish Health Authority estimates DRG tariffs based on previous expenditure data, and actual individual expenditures may differ from the expected. However, the DRG system is particularly useful for its level of granularity where we are able to observe the expenditures for a given diagnosis over time.

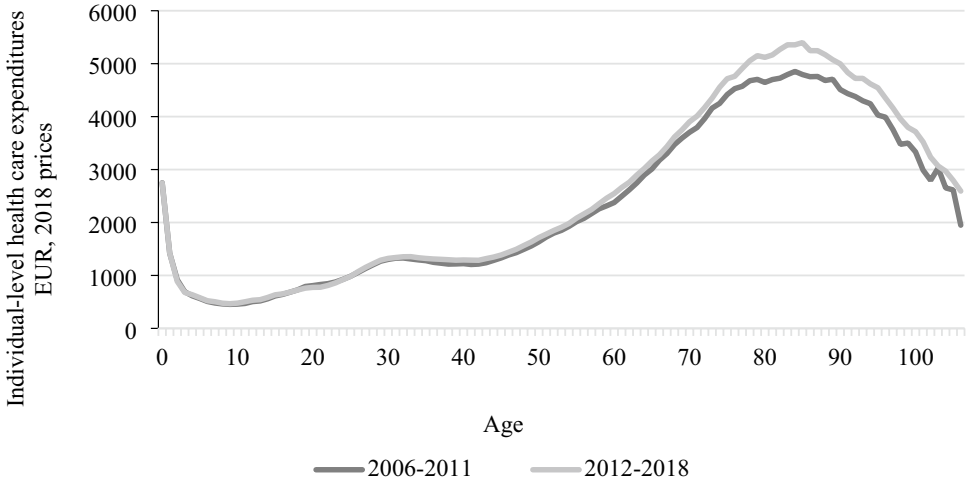



**Fig. 1** Average individual-level health care expenditure curve over a lifecycle

still sum across all expenditures in inpatient and outpatient care as well as physician care regardless of the diagnosis. In the analyses of the importance of age and time to death, we again use the ICD-10 codes from the Danish National Patient registry and include 28 indicators for whether an individual have been diagnosed with either of the 28 diagnoses. Again an individual may have more than one of the 28 diagnoses. In the steepening analyses, we include the entire Danish population over the observation period 2006–2018. This gives us a set of repeated cross sections where individuals are represented in the data set in a given year if they were alive and lived in Denmark in that year. Hence, some individuals will be included in the data set for all the years, while others will only be present in the data set for 1 year. In the analysis of the importance of age and time to death, we apply the same data, but narrow the period to 2006–2011 as we want to follow the individuals for at least 8 years prior to their death (until 2018). In the analyses, we include individuals who die within the 8 year window as well as those who do not. In all analyses, we include the entire Danish population and, hence, we avoid any sample selection issues.

Appendix Table 4 shows the average individual-health care expenditure per year as well as the number of individuals in the Danish population for each observation year from 2006 to 2018. The average individual-level expenditures vary from 1597 to 1956 EUR per year, and the number of observations varies from about 5.5 million to 5.9 million individuals per year from 2006 to 2018. Figure 1 shows that when examining the development of the average individual-level health care expenditure curve over a lifecycle in the periods 2006–2011 and 2012–2018, the curve has become steeper over time where the health care expenditures have increased relatively more for individuals above 70 years. In the following section, we present models to test whether the steepening has actually taken place, and whether this coexists with an importance of age and time to death.

## Methods

We present the models for the steepening analyses and the analyses of the importance of age and time to death, respectively.[3] In both analyses, we use identical two-part models and estimation techniques, but the right-hand side covariates and time periods differ across the models. Mechanically, the health care expenditure variable will be right skewed with zero inflation, as most individuals will have no or low health care expenditures, some will have low-to-moderate health care expenditures, and few will have high health care expenditures. Therefore, we have chosen models where we account for both the skewness, a large share of zeros as well as a non-constant variance. When examining health expenditure data, Generalized Linear Models (GLMs) have been suggested as a solution for similar data structures [27]. Therefore, we apply a two-part model where we first estimate a probit model for the probability that an individual receives some health care services in a given year, i.e., has positive health care expenditures. Second, we estimate a Poisson-distributed GLM model where only individuals who actually receive health care services in a given year are included. We use a modified Park test suggested by Manning and Mullahy [28] to choose the distribution for the positive health care expenditures. With the modified Park test, we conclude that we cannot reject that the variance is proportional to the mean, and we conclude that the Poisson distribution is the most appropriate distribution. The Poisson-distributed GLM model consists of a Poisson distribution for the expenditures and a log-link function where we assume that the logarithm of the expected value of the expenditures can be modelled by a linear combination of parameters on the right-hand side. The log-link function provides multiplicative effects and the

[3] The analyses extend initial work reported by Kjellberg and Ibsen [26] [source]

results can therefore be expressed as percentage increases in the average health care expenditure per unit increase in any of the covariates by taking the exponential of the covariates [29].

In the following, the dependent variable $y_{it}$ is thus equal to $Pr(Expenditures > 0)$ in the first part (the probit model) and $Ln(Expenditures)$ in the second part (the poisson-distributed GLM model). The right-hand side covariates are the same in both the probit specification and the GLM specification, but the right-hand side covariates vary over the steepening analyses and the analysis of the importance of age and time to death, respectively. The models are estimated using maximum-likelihood estimation (MLE), and we allow for clustering of repeated observations on each individual by correcting standard errors for intragroup correlation.

When interpreting the results of the two-part models, we calculate and present predicted expenditures as these are based on both model parts where we take both the probability of positive expenditures and the level among individuals with positive expenditures into account: $E(Y|X) = P(Y > 0|X) * E(Y|X, Y > 0)$.

## Empirical specification for steepening analyses

Buchner and Wasem [16] were the first to introduce the term steepening. Felder and Werblow [20] improved the methodological aspects of estimating steepening by defining it as a positive mixed derivative of per capita health care expenditures with respect to age and time. Gregersen [7] extended the model by Felder and Werblow [20] with interactions between mortality rates and age as well as interactions between mortality rates and time. The first element was introduced to capture that mortality-related expenditures are decreasing in age, and the second to capture potential compression of morbidity, affecting mortality-related expenditures over time. Gregersen [7] found that part of the steepening effect could be explained by increased mortality expenditures over time. Gregersen [7] estimated models based on averaged cell data. As we have detailed individual-level data for the entire Danish population, we have the exact date of death for each individual. Therefore, we are able to include an indicator for whether an individual dies in a given year and thus increase the precision in comparison to the use of mortality rates as in Gregersen [7].

In model (1)–(4), we follow the model by Gregersen [7] in terms of which explanatory variables to include. In model (5)–(7), we extend the model with three-way interaction terms to analyse disease-specific cost-drivers.

In the first model specification, we model the health care expenditures, $y_{it}$, of individual $i$ in calendar year $t$ by

$$y_{it} = \alpha + \beta \text{female}_i + \sum_{j=0, j\neq 5-9}^{90+} \gamma_j I(\text{age group} = j)_{it} + \theta \text{year}_{it} + \sum_{j=0, j\neq 5-9}^{90+} \eta_j I(\text{age group} = j)_{it} * \text{year}_{it} + \epsilon_{it}. \quad (1)$$

where $\alpha$ is a constant term and $\beta$ measures the average difference in health care expenditures among men and women over the observation period. $I(\text{age group} = j)$ is an indicator variable capturing whether individual $i$ in calendar year $t$ is in a given 5 year age group. To ensure identification of the parameters, the age group of 5–9 years is chosen as reference group, and therefore, $\gamma$ measures the impact of a given age group on individual-level health care expenditures relative to being 5–9 years old. $\theta$ captures the yearly linear growth rate in health care expenditures over the observation period. The coefficient of interest is $\eta$ as it measures the age group-specific growth rate as deviations from the linear trend $\theta$ and relative to being 5–9 years old. Hence, $\eta$ measures whether expenditures have increased significantly more for age group $j$ than for 5–9 years old from 2006 to 2018. We will in the following refer to model (1) as the basic model and sequentially add extensions to account for more covariates and analyse disease-specific cost-drivers.

In model (2), we extend model (1) by including an indicator capturing whether an individual died in a given calendar year. We do that to take into account that health care expenditures are higher in the year when an individual dies. In model (3), we extend model (2) by an interaction term between dying in a given calendar year and the age of the individual. We do this to take into account that death-related health care expenditures are a decreasing function of age, where, e.g., a 30-year-old dying individual is associated with higher expenditures than an 80-year-old dying individual in the year of death. This has also been suggested in the previous papers [9, 30, 31]. Model (4) is an extension of model (3) where we include an interaction between dying in a given year and the calendar year to capture that death-related expenditures may not have been constant over the observation period. For example, previous papers have suggested that health care expenditures may be concentrated to fewer years over time as longevity has increased [7, 17, 20]. The full model with all three extensions is given by

$$y_{it} = \alpha + \beta \text{female}_i + \sum_{j=0, j\neq 5-9}^{90+} \gamma_j I(\text{age group} = j)_{it} + \theta_j year_{it} + \sum_{j=0, j\neq 5-9}^{90+} \eta_j I(\text{age group} = j)_{it} * year_{it} + \kappa \text{died}_{it} + \tau \text{age} * \text{died}_{it} + \omega \text{year} * \text{died}_{it} + \epsilon_{it}. \quad (2–4)$$

where the added parameters are $\kappa, \tau$, and $\omega$, and where $\eta$ still captures the coefficients of relevance to the study of steepening.

Next, we propose models to exploit whether steepening can be explained by three disease-specific cost-drivers: having a cancer disease, having a cardiovascular disease, or having a chronic diseases. In the following, we will describe the model specification for cancer and refer to it as model (5). The models for cardiovascular diseases and chronic diseases are equivalent, but with indicators for these diseases instead of the indicator for cancer. We refer to the model for cardiovascular diseases as model (6) and for chronic diseases as (7). Model (5) is an extension of model (4) with an indicator for having or not having a cancer diagnosis in calendar year $t$, interaction terms between age groups and having a cancer diagnosis in year $t$ and a three-term interaction between age group, calendar year and having a cancer diagnosis in year $t$. The health care expenditures of individual $i$ in calendar year $t$ are thus given by

$$\begin{aligned} y_{it} = {} & \alpha + \beta \text{female}_i + \sum_{j=0, j\neq 0-49}^{90+} \gamma_j I(\text{age group} = j)_{it} + \theta \text{year}_{it} \\ & + \sum_{j=0, j\neq 0-49}^{90+} \eta_j I(\text{age group} = j)_{it} * \text{year}_{it} + \kappa \text{died}_{it} \\ & + \tau \text{ age} * \text{died}_{it} + \omega \text{year} * \text{died}_{it} + \pi \text{ cancer}_{it} \\ & + \sum_{j=0, j\neq 0-49}^{90+} \rho_j I(\text{age group} = j)_{it} * \text{cancer}_{it} \\ & + \sum_{j=0, j\neq 0-49}^{90+} \lambda I(\text{age group} = j)_{it} * year_{it} * \text{cancer}_{it} + \epsilon_{it}, \end{aligned} \tag{5–7}$$

where the coefficients of interest are given by $\eta$ and $\lambda$. The interaction terms between age group and year, captured by $\eta$, now measure the expenditure change over time for individuals who are not diagnosed with cancer relative to the age group of 0- to 49-years old, i.e., whether steepening has taken place among individuals without a cancer diagnosis. The three-way interaction terms between age group, year, and cancer, captured by $\lambda$, measure whether there is significantly more or less steepening among the group of individuals with a cancer diagnosis relative to the group with no cancer diagnosis. As very few young individuals are diagnosed with cancer, a cardiovascular disease, or a chronic disease, we need to expand the reference age group, and therefore, we use the age group of 0–49 years old as the reference group in the disease-specific steepening analyses.

The probability of positive health care expenditures within a diagnosis group (e.g., cancer) will naturally be very high (99 pct.), but it is still relevant to estimate the probit specification as some diagnosed may have a year with 0 expenditures due to the presence of DRG-tariffs of 0[4] and as some individuals in the reference groups of, e.g., non-cancer diagnosed not necessarily have positive health care expenditures.

## Empirical specification of the importance of age and time to death

The following model is inspired by the models presented by Seshamani and Gray [10, 11] and Gregersen and Godager [9]. Seshamani and Gray [11] criticised the original two-step Heckman selection model proposed by Zweifel et al. [3] for estimation of the relation between age, time to death, and expenditures as it suffered from multicollinearity in the second step. Instead, Seshamani and Gray [11] proposed a two-part model as presented in the beginning of Sect. Methods. From Gregersen and Godager [9], we are inspired by the use of 5 year age groups rather than including age as a continuous variable as done in the model by Seshamani and Gray [10, 11]. In a recent review of the red herring literature, 5 year age groups are also recommend above age and age squared [24]. As in the steepening analyses, this allows us to investigate patterns for specific age groups and not impose the assumption that health expenditures are linear or quadratic in age. Gregersen and Godager [9] apply averaged cell data, while we estimate the models on individual-level data. In the model, we include all individuals in Denmark in the period 2006–2011 where we are able to follow all individuals for at least 8 years (until 2018). We model the health care expenditures of individual $i$ in calendar year $t$ as a function of gender, age group and time to death

$$\begin{aligned} y_{it} = {} & \alpha + \beta \text{female}_i + \sum_{j=1, j\neq 1-19}^{90+} \gamma_j I(\text{age group} = j)_{it} \\ & + \sum_{m=0, m\neq 8+}^{8+} \eta_m I(\text{time to death} = m)_{it} \\ & + \sum_{j=1, j\neq 1-19}^{90+} \theta_j I(\text{age group} = j)_{it} * \text{female}_{it} \\ & + \sum \sum \lambda_{jm} I(\text{age group} = j)_{it} * I(\text{time to death} = m)_{it} \\ & + \sum_{t=2006, t\neq 2006}^{2011} \rho_t I(\text{year} = t)_{it} + \epsilon_{it}, \end{aligned} \tag{8}$$

[4] For example, A DRG-tariff equal to 0 enters if a patient is diagnosed in outpatient care December 31st 2008 but hospitalized January 1st 2009. In that case, the diagnosis may be given in 2008 with a DRG-tariff of 0, and in 2009, a DRG-tariff > 0 will be present in the data. In Denmark, DRG-groups without a cost are referred to as MG90-groups.

where $\alpha$, $\beta$, and $\gamma$ are the same as in the previous models, but where the reference age group is now the 1–19 years old. Moreover, we add a vector of indicator variables capturing time to death for individual $i$ in year $t$. The indicator variables for time to death run from 0 years to 8 + years to death (if the individual dies 8 or more years ahead of time $t$). Individuals who do not die during the observation period will thus be in the 8 + year group, and this group also functions as the reference group for the groups from 0 to 7 years of time to death. Positive time to death coefficients contained in the vector $\eta$ will capture whether each of the years 0–7 are correlated with individual-level health care expenditures in year $t$. If, for example, the coefficients in $\eta$ are significantly positive, but the coefficients in $\gamma$ are not, it suggests that time to death matters, but that age in itself does not matter for the individual-level health care expenditures. This would be in line with the red herring hypothesis. If both are significantly positive, then both age and time to death matter for the individual-level health care expenditures. Therefore, we also add interaction effects between age group and time to death captured in $\kappa$.

To follow recent trends in the literature on the relative importance of age and time to death [14, 15], we extend model (8) with indicator variables for 28 chronic diseases to account for individual-level health status:

$$\begin{aligned} y_{it} = {} & \alpha + \beta \text{female}_i + \sum_{j=1, j\neq 1-19}^{90+} \gamma_j I(\text{age group} = j)_{it} \\ & + \sum_{m=0, m\neq 8+}^{8+} \eta_m I(\text{time to death} = m)_{it} \\ & + \sum_{j=1, j\neq 1-19}^{90+} \theta_j I(\text{age group} = j)_{it} * \text{female}_{it} \\ & + \sum \sum \lambda_{jm} I(\text{age group} = j)_{it} * I(\text{time to death} = m)_{it} \\ & + \sum_{t=2006, t\neq 2006}^{2011} \rho_t I(\text{year} = t)_{it} \\ & + \sum_{k=0, k\neq 0}^{28} \omega_k I(\text{chronic disease} = k)_{it} + \epsilon_{it}, \end{aligned} \tag{9}$$

where the coefficients in the vector $\omega$ will capture whether each of the chronic diseases are associated with the probability of positive health care expenditures (in the probit model) and the expenditure level (in the Poisson-distributed GLM model) relative to having no chronic disease. The other coefficients have the same interpretation as in (8), but are now adjusted for the presence of a chronic disease. The chronic diseases are identified using ICD-10 codes as in Nexo et al. [32] and also presented in Appendix A.1.

## Results

### Steepening

In Table 1, the results of the steepening analyses are shown with the basic model (1) and a sequential inclusion of death-related expenditures from model (2) to model (4) for both the probit and the Poisson-distributed GLM specification. In the first column for each of the four models presented in Table 1, the probit regression results are shown, and in the second column, the GLM results are shown In Fig. 2, we show the predicted individual-level health care expenditure curves for men in 2006 and 2018 using the estimates from both the probit specification and the Poisson-distributed GLM specification from model (4). From the graph, it is very clear that the curve has become steeper over the 12 year period and it is in particular driven by individuals aged 75–89.

From the probit specification of model (1), we see that the probability of receiving health care services has increased over time for 0 years old and for all age groups from 50–90 +, while it has decreased for the 1–4 years old and the 10–49 years old, relative to 5–9 years old. The same is true when examining the probit specification for model (2)–(4), and the sizes of the coefficients are also similar across model (1)–(4). Hence, the inclusion of death-related expenditures and the relation between death and age and death and calendar year do not change the probability that an age group has become more or less likely to receive health care services from 2006 to 2018.

In the second column for each of the four models, we examine the change in health care expenditures for individuals who have positive health care expenditures over the observation period in the Poisson-distributed GLM specification. From model (1) without death-related expenditures, we show that the health care expenditures have increased relatively more for the age groups of 10–14, 35–44, and 80–90 + year olds relative to the reference age group of 5–9 years old, while no clear pattern is evident for the other age groups. Model (1) shows an indication of steepening after the age of 80, but in model (1), we do not take death-related expenditures into account, and therefore, model (1) must be referred to as a naïve model. In model (2), we take death-related expenditures into account, and here the pattern changes. First of all, we see that the indicator for death is significant and positive which suggests that individuals have higher health care expenditures in the year when they die. This was expected and it shows that it is very relevant to include death in the model. When examining model (2), it becomes clear that the health care expenditures have increased for all age groups from 30 years and up, where the expenditures have increased most and significantly for the 75–89 years

**Table 1** Regression results: steepening analyses

| | Model (1) | | | | | | Model (2) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp.) > 0[a] | | | Log (Exp.)[b] | | | *P* (Exp.) > 0 | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| Intercept | 0.958 | *** | 0.002 | 8.311 | *** | 0.008 | 0.958 | *** | 0.002 | 8.299 | *** | 0.008 |
| Year | −0.010 | *** | 0.000 | 0.008 | *** | 0.001 | −0.010 | *** | 0.000 | 0.008 | *** | 0.001 |
| Female | 0.395 | *** | 0.001 | 0.039 | *** | 0.001 | 0.396 | *** | 0.001 | 0.061 | *** | 0.001 |
| Death | | | | | | | 0.397 | *** | 0.005 | 1.607 | *** | 0.002 |
| Death*age | | | | | | | | | | | | |
| Death*year | | | | | | | | | | | | |
| Age group | | | | | | | | | | | | |
| 0 | 0.687 | *** | 0.005 | 1.626 | *** | 0.010 | 0.686 | *** | 0.005 | 1.613 | *** | 0.010 |
| 1–4 | 0.903 | *** | 0.004 | 0.515 | *** | 0.010 | 0.903 | *** | 0.004 | 0.515 | *** | 0.010 |
| 5–9 | Ref. | | | Ref. | | | Ref. | | | Ref. | | |
| 10–14 | −0.074 | *** | 0.002 | 0.026 | ** | 2.439 | −0.074 | *** | 0.002 | 0.025 | ** | 0.010 |
| 15–19 | 0.084 | *** | 0.003 | 0.320 | *** | 31.129 | 0.084 | *** | 0.003 | 0.319 | *** | 0.010 |
| 20–24 | 0.074 | *** | 0.003 | 0.550 | *** | 58.037 | 0.074 | *** | 0.003 | 0.548 | *** | 0.009 |
| 25–29 | 0.151 | *** | 0.003 | 0.803 | *** | 88.067 | 0.151 | *** | 0.003 | 0.801 | *** | 0.009 |
| 30–34 | 0.359 | *** | 0.003 | 0.908 | *** | 101.833 | 0.359 | *** | 0.003 | 0.905 | *** | 0.009 |
| 35–39 | 0.452 | *** | 0.003 | 0.831 | *** | 90.512 | 0.451 | *** | 0.003 | 0.827 | *** | 0.009 |
| 40–44 | 0.479 | *** | 0.003 | 0.826 | *** | 89.153 | 0.478 | *** | 0.003 | 0.819 | *** | 0.009 |
| 45–49 | 0.517 | *** | 0.003 | 0.979 | *** | 104.756 | 0.517 | *** | 0.003 | 0.968 | *** | 0.009 |
| 50–54 | 0.587 | *** | 0.003 | 1.189 | *** | 128.833 | 0.585 | *** | 0.003 | 1.170 | *** | 0.009 |
| 55–59 | 0.660 | *** | 0.003 | 1.372 | *** | 150.096 | 0.658 | *** | 0.003 | 1.344 | *** | 0.009 |
| 60–64 | 0.696 | *** | 0.003 | 1.558 | *** | 172.584 | 0.693 | *** | 0.003 | 1.517 | *** | 0.009 |
| 65–69 | 0.779 | *** | 0.004 | 1.778 | *** | 197.561 | 0.775 | *** | 0.004 | 1.716 | *** | 0.009 |
| 70–74 | 0.908 | *** | 0.005 | 1.960 | *** | 217.846 | 0.900 | *** | 0.005 | 1.859 | *** | 0.009 |
| 75–79 | 0.926 | *** | 0.006 | 2.098 | *** | 227.605 | 0.913 | *** | 0.006 | 1.932 | *** | 0.009 |
| 80–84 | 0.990 | *** | 0.007 | 2.116 | *** | 235.089 | 0.970 | *** | 0.007 | 1.861 | *** | 0.009 |
| 85–89 | 1.008 | *** | 0.009 | 2.109 | *** | 230.004 | 0.977 | *** | 0.009 | 1.730 | *** | 0.009 |
| 90+ | 1.020 | *** | 0.013 | 1.998 | *** | 208.925 | 0.963 | *** | 0.013 | 1.392 | *** | 0.010 |
| Year*age group | | | | | | | | | | | | |
| 0 | 0.024 | *** | 0.001 | −0.008 | *** | 0.002 | 0.024 | *** | 0.001 | −0.008 | *** | 0.002 |
| 1–4 | −0.004 | *** | 0.000 | −0.008 | *** | 0.002 | −0.004 | *** | 0.000 | −0.008 | *** | 0.002 |
| 5–9 | Ref | | | Ref | | | Ref | | | Ref | | |
| 10–14 | −0.002 | *** | 0.000 | 0.005 | ** | 0.002 | −0.002 | *** | 0.000 | 0.005 | ** | 0.002 |
| 15–19 | −0.005 | *** | 0.000 | −0.002 | | 0.002 | −0.005 | *** | 0.000 | −0.002 | | 0.002 |
| 20–24 | −0.003 | *** | 0.000 | −0.008 | *** | 0.002 | −0.003 | *** | 0.000 | −0.007 | *** | 0.002 |
| 25–29 | −0.007 | *** | 0.000 | −0.002 | | 0.001 | −0.007 | *** | 0.000 | −0.002 | | 0.001 |
| 30–34 | −0.015 | *** | 0.000 | 0.002 | | 0.001 | −0.015 | *** | 0.000 | 0.002 | | 0.001 |
| 35–39 | −0.012 | *** | 0.000 | 0.005 | *** | 0.001 | −0.012 | *** | 0.000 | 0.005 | *** | 0.001 |
| 40–44 | −0.006 | *** | 0.000 | 0.004 | ** | 0.001 | −0.006 | *** | 0.000 | 0.004 | *** | 0.001 |
| 45–49 | −0.003 | *** | 0.000 | 0.001 | | 0.001 | −0.003 | *** | 0.000 | 0.002 | | 0.001 |
| 50–54 | 0.001 | *** | 0.000 | −0.001 | | 0.001 | 0.001 | *** | 0.000 | 0.000 | | 0.001 |
| 55–59 | 0.003 | *** | 0.000 | 0.001 | | 0.001 | 0.003 | *** | 0.000 | 0.001 | | 0.001 |
| 60–64 | 0.012 | *** | 0.000 | 0.002 | | 0.001 | 0.012 | *** | 0.000 | 0.002 | | 0.001 |
| 65–69 | 0.018 | *** | 0.001 | 0.000 | | 0.001 | 0.018 | *** | 0.001 | 0.001 | | 0.001 |
| 70–74 | 0.020 | *** | 0.001 | 0.000 | | 0.001 | 0.020 | *** | 0.001 | 0.002 | | 0.001 |
| 75–79 | 0.032 | *** | 0.001 | 0.002 | | 0.001 | 0.032 | *** | 0.001 | 0.006 | *** | 0.001 |
| 80–84 | 0.033 | *** | 0.001 | 0.007 | *** | 0.001 | 0.033 | *** | 0.001 | 0.011 | *** | 0.001 |

**Table 1** (continued)

| | Model (1) | | | | | | Model (2) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp.) > 0[a] | | | Log (Exp.)[b] | | | *P* (Exp.) > 0 | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 85–89 | 0.033 | *** | 0.001 | 0.007 | *** | 0.001 | 0.033 | *** | 0.001 | 0.011 | *** | 0.001 |
| 90 + | 0.031 | *** | 0.002 | 0.006 | *** | 0.002 | 0.031 | *** | 0.002 | 0.008 | *** | 0.002 |
| Observations | 74,217,773 | | | 68,754,429 | | | 74,217,773 | | | 68,754,429 | | |
| | Model (3) | | | | | | Model (4) | | | | | |
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp.) > 0 | | | Log (Exp.) | | | *P* (Exp.) > 0 | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| Intercept | 0.958 | *** | 0.002 | 8.295 | *** | 0.008 | 0.958 | *** | 0.002 | 8.295 | *** | 0.008 |
| Year | −0.010 | *** | 0.000 | 0.008 | *** | 0.001 | −0.010 | *** | 0.000 | 0.008 | *** | 0.001 |
| Female | 0.396 | *** | 0.001 | 0.063 | *** | 0.001 | 0.396 | *** | 0.001 | 0.063 | *** | 0.001 |
| Death | 0.133 | *** | 0.019 | 3.989 | *** | 0.011 | 0.267 | *** | 0.021 | 3.937 | *** | 0.011 |
| Death*age | 0.004 | *** | 0.000 | −0.032 | *** | 0.000 | 0.004 | *** | 0.000 | −0.032 | *** | 0.000 |
| Death*year | | | | | | | −0.022 | *** | 0.001 | 0.010 | *** | 0.001 |
| Age group | | | | | | | | | | | | |
| 0 | 0.687 | *** | 0.005 | 1.465 | *** | 0.011 | 0.687 | *** | 0.005 | 1.473 | *** | 0.011 |
| 1–4 | 0.903 | *** | 0.004 | 0.507 | *** | 0.010 | 0.903 | *** | 0.004 | 0.507 | *** | 0.010 |
| 5–9 | Ref. | | | Ref. | | | Ref. | | | Ref. | | |
| 10–14 | −0.074 | *** | 0.002 | 0.025 | ** | 0.010 | −0.074 | *** | 0.002 | 0.025 | ** | 0.010 |
| 15–19 | 0.084 | *** | 0.003 | 0.314 | *** | 0.010 | 0.084 | *** | 0.003 | 0.315 | *** | 0.010 |
| 20–24 | 0.074 | *** | 0.003 | 0.541 | *** | 0.009 | 0.074 | *** | 0.003 | 0.542 | *** | 0.009 |
| 25–29 | 0.151 | *** | 0.003 | 0.796 | *** | 0.009 | 0.151 | *** | 0.003 | 0.796 | *** | 0.009 |
| 30–34 | 0.359 | *** | 0.003 | 0.900 | *** | 0.009 | 0.359 | *** | 0.003 | 0.900 | *** | 0.009 |
| 35–39 | 0.451 | *** | 0.003 | 0.820 | *** | 0.009 | 0.451 | *** | 0.003 | 0.821 | *** | 0.009 |
| 40–44 | 0.478 | *** | 0.003 | 0.809 | *** | 0.009 | 0.478 | *** | 0.003 | 0.810 | *** | 0.009 |
| 45–49 | 0.517 | *** | 0.003 | 0.954 | *** | 0.009 | 0.517 | *** | 0.003 | 0.955 | *** | 0.009 |
| 50–54 | 0.585 | *** | 0.003 | 1.151 | *** | 0.009 | 0.585 | *** | 0.003 | 1.153 | *** | 0.009 |
| 55–59 | 0.658 | *** | 0.003 | 1.323 | *** | 0.009 | 0.658 | *** | 0.003 | 1.326 | *** | 0.009 |
| 60–64 | 0.693 | *** | 0.003 | 1.498 | *** | 0.009 | 0.692 | *** | 0.003 | 1.502 | *** | 0.009 |
| 65–69 | 0.775 | *** | 0.004 | 1.702 | *** | 0.009 | 0.774 | *** | 0.004 | 1.707 | *** | 0.009 |
| 70–74 | 0.900 | *** | 0.005 | 1.857 | *** | 0.009 | 0.899 | *** | 0.005 | 1.863 | *** | 0.009 |
| 75–79 | 0.912 | *** | 0.006 | 1.956 | *** | 0.009 | 0.910 | *** | 0.006 | 1.965 | *** | 0.009 |
| 80–84 | 0.968 | *** | 0.007 | 1.934 | *** | 0.009 | 0.964 | *** | 0.007 | 1.946 | *** | 0.009 |
| 85–89 | 0.973 | *** | 0.009 | 1.881 | *** | 0.009 | 0.967 | *** | 0.009 | 1.898 | *** | 0.009 |
| 90 + | 0.954 | *** | 0.013 | 1.706 | *** | 0.010 | 0.944 | *** | 0.013 | 1.731 | *** | 0.010 |
| Year*age group | | | | | | | | | | | | |
| 0 | 0.024 | *** | 0.001 | −0.007 | *** | 0.002 | 0.024 | *** | 0.001 | −0.008 | *** | 0.002 |
| 1–4 | −0.004 | *** | 0.000 | −0.008 | *** | 0.002 | −0.004 | *** | 0.000 | −0.008 | *** | 0.002 |
| 5–9 | Ref | | | Ref | | | Ref | | | Ref | | |
| 10–14 | −0.002 | *** | 0.000 | 0.005 | ** | 0.002 | −0.002 | *** | 0.000 | 0.005 | ** | 0.002 |
| 15–19 | −0.005 | *** | 0.000 | −0.001 | | 0.002 | −0.005 | *** | 0.000 | −0.001 | | 0.002 |
| 20–24 | −0.003 | *** | 0.000 | −0.007 | *** | 0.002 | −0.003 | *** | 0.000 | −0.007 | *** | 0.002 |
| 25–29 | −0.007 | *** | 0.000 | −0.001 | | 0.001 | −0.007 | *** | 0.000 | −0.001 | | 0.001 |
| 30–34 | −0.015 | *** | 0.000 | 0.002 | | 0.001 | −0.015 | *** | 0.000 | 0.002 | | 0.001 |
| 35–39 | −0.012 | *** | 0.000 | 0.005 | *** | 0.001 | −0.012 | *** | 0.000 | 0.005 | *** | 0.001 |

**Table 1** (continued)

| | Model (3) | | | | | | Model (4) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp.) > 0 | | | Log (Exp.) | | | *P* (Exp.) > 0 | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 40–44 | −0.006 | *** | 0.000 | 0.004 | *** | 0.001 | −0.006 | *** | 0.000 | 0.004 | *** | 0.001 |
| 45–49 | −0.003 | *** | 0.000 | 0.002 | | 0.001 | −0.003 | *** | 0.000 | 0.002 | | 0.001 |
| 50–54 | 0.001 | *** | 0.000 | 0.001 | | 0.001 | 0.001 | *** | 0.000 | 0.000 | | 0.001 |
| 55–59 | 0.003 | *** | 0.000 | 0.002 | | 0.001 | 0.003 | *** | 0.000 | 0.001 | | 0.001 |
| 60–64 | 0.012 | *** | 0.000 | 0.003 | * | 0.001 | 0.012 | *** | 0.000 | 0.002 | | 0.001 |
| 65–69 | 0.018 | *** | 0.001 | 0.002 | | 0.001 | 0.018 | *** | 0.001 | 0.001 | | 0.001 |
| 70–74 | 0.020 | *** | 0.001 | 0.002 | | 0.001 | 0.020 | *** | 0.001 | 0.001 | | 0.001 |
| 75–79 | 0.032 | *** | 0.001 | 0.005 | *** | 0.001 | 0.033 | *** | 0.001 | 0.004 | *** | 0.001 |
| 80–84 | 0.034 | *** | 0.001 | 0.010 | *** | 0.001 | 0.034 | *** | 0.001 | 0.008 | *** | 0.001 |
| 85–89 | 0.033 | *** | 0.001 | 0.010 | *** | 0.001 | 0.034 | *** | 0.001 | 0.007 | *** | 0.001 |
| 90 + | 0.031 | *** | 0.002 | 0.007 | *** | 0.002 | 0.033 | *** | 0.002 | 0.003 | *** | 0.002 |
| Observations | 74,217,773 | | | 68,754,429 | | | 74,217,773 | | | 68,754,429 | | |

[a]The probability that an individual receives some health care services in a given year

[b]The log-link function provides multiplicative effects and the results can therefore be expressed as percentage increases in the average health care expenditure per unit increase in any of the covariates by taking the exponential of the covariates

*Statistically significant at the 0.10 level

**At the 0.05 level

***At the 0.01 level

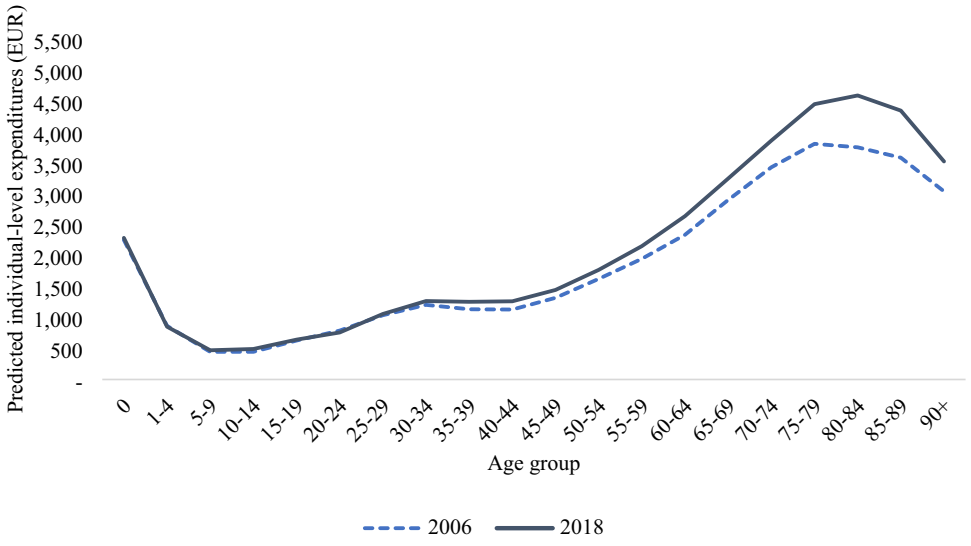



**Fig. 2** Predicted individual-level health care expenditure curves by calendar year for men. Note 1: estimates are predicted from model (4) using the coefficients estimated in the probit and the GLM-Poisson specification. The calculation follows $E(Y|X) = P(Y > 0|X) * E(Y|X, Y\rangle 0$ where $Y$ is the individual-level expenditures and $X$ represents covariates of interest from model (4). [E.g., for age group 65–69 in 2018, we calculate $P(Y > 0|X) = \Phi((\alpha + \gamma I(\text{age group} = 65 - 69) + (\theta \text{year} + \eta I(\text{age group} = 65 - 69) * \text{year})) * 12)$ using the estimated probit coefficients, and $E(Y|X, Y\rangle 0) = e^{\alpha+\gamma I(\text{age group}=65-69)+(\theta \text{year}+\eta I(\text{age group}=65-69)*\text{year})*12}$ using the estimated GLM-Poisson coefficients]

old relative to 5–9 years old. Exactly, the same pattern is evident in models (3) and (4). In model (3), we control for the fact that death-related expenditures are a decreasing function in age, which is confirmed by the significant estimate for the interaction between death and age, i.e., the older an individual becomes, the lower are the health care expenditures in the year of death. In model (4), we include the interaction between death and calendar year which shows that death-related expenditures have increased over time. Compared to model (2) and (3), the steepening estimates are slightly lower in model (4) from the age of 50 and in particular from the age of 80. This indicates that part of the expenditure growth for the older age groups is explained by increased death-related expenditures over the period, and, in turn, this has a higher impact on the estimates for the older age groups as the mortality rate is higher in these groups.

### Disease-specific cost-drivers

Table 2 shows the steepening results across disease groups: cancer, cardiovascular, and chronic diseases relative to no cancer, no cardiovascular, and no chronic disease. Overall, the analyses of disease-specific cost-drivers of steepening indicate that the individual-level health care expenditure curve have steepened more among individuals with a chronic disease than individuals without chronic diseases, and the treatment of cardiovascular diseases is associated with less steepening than other disease groups.

When examining the estimates for individuals with a cancer disease compared to those without a cancer disease in Table 2, we see that individuals with cancer in general have a higher probability of receiving health care services and are associated with significantly higher expenditures when examining individuals with positive expenditures in the GLM model. In the following, we describe the extent of steepening by examining Fig. 3. Figure 3 shows the predicted individual-level health care expenditure curves by calendar year and age group for individuals with cancer diseases and other diseases. From the figure, we see that steepening is evident among both individuals with and without cancer. The shape of the age-distributed expenditure curves for individuals with cancer differs from the curves of non-cancer diagnosed, and the shape changes over time. For individuals with cancer, we observe that individual-level expenditures among younger cancer diagnosed (50–74 years old) decrease, while the individual-level expenditure among elder cancer diagnosed (75–89 years old) increase from 2006 to 2018. Also, expenditures seem to be shifted to older age groups over time where the age-distributed curve decreased remarkably from age 69 in 2006, while the age-distributed curve in 2018 decrease from age 79. For non-cancer diagnosed, we see the same pattern as for the full population, where primarily older age groups drive the steepening of the curve (Fig. 4).

For individuals with a cardiovascular diagnosis, we observe that having a cardiovascular disease on average increases the probability that the individual will receive health care services as well as the level of expenditures among individuals with positive health care expenditures relative to those not having a cardiovascular disease. When examining the predicted individual-level expenditure curves in Fig. 4, it is clear that the curve has become steeper among individuals with no cardiovascular disease but not among individuals diagnosed with a cardiovascular disease from 2006 to 2018. Therefore, we conclude that less steepening has taken place in the group of individuals with a cardiovascular disease.

The last disease-specific group is individuals diagnosed with a chronic disease relative to individuals not diagnosed with a chronic disease. As for cancer and cardiovascular disease diagnosed, individuals with a chronic disease have a higher probability of receiving health care services and among individuals with positive health care expenditures, individuals with a chronic disease exhibit higher health care expenditures than individuals without a chronic disease. Figure 5 shows the predicted individual-level health care expenditure curves by calendar year and age group for individuals with chronic diseases and individuals without chronic diseases. From the figure, we see that the age-distributed curve steepens over time among chronic disease diagnosed, but not among individuals without a chronic disease. We also observe that the steepening primarily is driven by 50–74-year-old chronic disease diagnosed where we see a substantial expenditure reduction (about 500–700 EUR per individual) across the 12-year period while there is no large expenditure increase within the group of 75–89 years old.

**Table 2** Regression results: steepening analyses across diseases

| | 5 Cancer | | | | | | | 6 Cardiovascular | | | | | | | 7 Chronic | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | | Probit | | | GLM-Poisson | | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | | *P* (Exp. > 0) | | | Log (Exp.) | | | | *P* (Exp. > 0) | | | Log(Exp.) | | |
| | Coef | | SE | Coef | | SE | | Coef | | SE | Coef | | SE | | Coef | | SE | Coef | | SE |
| Intercept | 1.214 | *** | 0.001 | 8.932 | *** | 0.002 | Intercept | 1.209 | *** | 0.001 | 8.888 | *** | 0.002 | Intercept | 1.175 | *** | 0.001 | 8.575 | *** | 0.001 |
| Year | −0.016 | *** | 0.000 | 0.006 | *** | 0.000 | Year | −0.016 | *** | 0.000 | 0.008 | *** | 0.000 | Year | −0.017 | *** | 0.000 | 0.003 | *** | 0.000 |
| Female | 0.384 | *** | 0.001 | 0.046 | *** | 0.001 | Female | 0.389 | *** | 0.001 | 0.115 | *** | 0.001 | Female | 0.386 | *** | 0.001 | 0.052 | *** | 0.001 |
| Death | 0.283 | *** | 0.019 | 2.199 | ** | 0.014 | Death | 0.351 | *** | 0.020 | 3.590 | *** | 0.012 | Death | −0.032 | ** | 0.019 | 2.737 | *** | 0.010 |
| Death*age | 0.002 | *** | 0.000 | −0.017 | *** | 0.000 | Death*age | 0.002 | *** | 0.000 | −0.032 | *** | 0.000 | Death*age | 0.005 | *** | 0.000 | −0.024 | *** | 0.000 |
| Death*year | −0.023 | *** | 0.001 | 0.016 | *** | 0.001 | Death*year | −0.022 | *** | 0.001 | 0.014 | *** | 0.001 | Death*year | −0.026 | *** | 0.002 | 0.021 | *** | 0.000 |
| Cancer | 2.289 | *** | 0.111 | 2.722 | ** | 0.011 | Cardiovascular | 2.403 | *** | 0.063 | 1.820 | *** | 0.007 | Chronic | 2.229 | ** | 0.019 | 2.000 | *** | 0.003 |
| Cancer*year | 0.042 | *** | 0.015 | −0.001 | | 0.002 | Cardiovascular*year | −0.010 | | 0.008 | −0.001 | | 0.001 | Chronic*year | 0.009 | *** | 0.003 | −0.007 | *** | 0.000 |
| Age group | | | | | | | Age group | | | | | | | Age group | | | | | | |
| 0–49 | Ref. | | | Ref. | | | 0–49 | Ref. | | | Ref. | | | 0–49 | Ref. | | | Ref. | | |
| 50–54 | 0.326 | *** | 0.003 | 0.369 | *** | 0.004 | 50–54 | 0.318 | *** | 0.003 | 0.376 | *** | 0.004 | 50–54 | 0.290 | *** | 0.003 | −0.005 | | 0.003 |
| 55–59 | 0.396 | *** | 0.003 | 0.504 | *** | 0.004 | 55–59 | 0.386 | *** | 0.003 | 0.522 | *** | 0.004 | 55–59 | 0.348 | *** | 0.003 | 0.050 | *** | 0.003 |
| 60–64 | 0.425 | *** | 0.003 | 0.643 | *** | 0.004 | 60–64 | 0.413 | *** | 0.003 | 0.673 | *** | 0.004 | 60–64 | 0.367 | *** | 0.003 | 0.131 | *** | 0.004 |
| 65–69 | 0.501 | *** | 0.004 | 0.844 | *** | 0.004 | 65–69 | 0.486 | *** | 0.004 | 0.858 | *** | 0.004 | 65–69 | 0.422 | *** | 0.004 | 0.258 | *** | 0.004 |
| 70–74 | 0.622 | *** | 0.005 | 1.044 | *** | 0.004 | 70–74 | 0.603 | *** | 0.005 | 0.995 | *** | 0.004 | 70–74 | 0.521 | *** | 0.005 | 0.394 | *** | 0.005 |
| 75–79 | 0.631 | *** | 0.006 | 1.218 | *** | 0.005 | 75–79 | 0.606 | *** | 0.006 | 1.091 | *** | 0.005 | 75–79 | 0.502 | *** | 0.006 | 0.574 | *** | 0.005 |
| 80–84 | 0.688 | *** | 0.007 | 1.266 | *** | 0.004 | 80–84 | 0.660 | *** | 0.007 | 1.094 | *** | 0.004 | 80–84 | 0.556 | *** | 0.008 | 0.772 | *** | 0.005 |
| 85–89 | 0.699 | *** | 0.009 | 1.271 | *** | 0.004 | 85–89 | 0.666 | *** | 0.009 | 1.069 | *** | 0.005 | 85–89 | 0.575 | *** | 0.010 | 0.926 | *** | 0.006 |
| 90 + | 0.687 | *** | 0.013 | 1.150 | *** | 0.005 | 90+ | 0.657 | *** | 0.013 | 0.942 | *** | 0.006 | 90 + | 0.602 | *** | 0.013 | 0.989 | *** | 0.007 |
| Age group * year | | | | | | | Age group * year | | | | | | | Age group * year | | | | | | |
| 0–49 | Ref. | | | Ref. | | | 0–49 | Ref. | | | Ref. | | | 0–49 | Ref. | | | Ref. | | |
| 50–54 | 0.007 | *** | 0.000 | 0.001 | ** | 0.001 | 50–54 | 0.007 | *** | 0.000 | 0.001 | ** | 0.001 | 50–54 | 0.006 | *** | 0.000 | 0.004 | *** | 0.000 |
| 55–59 | 0.009 | *** | 0.000 | 0.003 | *** | 0.001 | 55–59 | 0.009 | *** | 0.000 | 0.002 | *** | 0.001 | 55–59 | 0.007 | *** | 0.000 | 0.004 | *** | 0.000 |
| 60–64 | 0.018 | *** | 0.000 | 0.003 | *** | 0.001 | 60–64 | 0.018 | *** | 0.000 | 0.003 | *** | 0.001 | 60–64 | 0.016 | *** | 0.000 | 0.003 | *** | 0.001 |
| 65–69 | 0.023 | *** | 0.001 | 0.000 | | 0.001 | 65–69 | 0.024 | *** | 0.001 | 0.001 | ** | 0.001 | 65–69 | 0.022 | *** | 0.001 | 0.002 | *** | 0.001 |
| 70–74 | 0.025 | *** | 0.001 | −0.003 | *** | 0.001 | 70–74 | 0.026 | *** | 0.001 | 0.002 | *** | 0.001 | 70–74 | 0.024 | *** | 0.001 | 0.001 | | 0.001 |
| 75–79 | 0.038 | *** | 0.001 | −0.002 | *** | 0.001 | 75–79 | 0.038 | *** | 0.001 | 0.004 | *** | 0.001 | 75–79 | 0.036 | *** | 0.001 | −0.002 | *** | 0.001 |
| 80–84 | 0.039 | *** | 0.001 | 0.002 | *** | 0.001 | 80–84 | 0.039 | *** | 0.001 | 0.007 | *** | 0.001 | 80–84 | 0.036 | *** | 0.001 | −0.005 | *** | 0.001 |
| 85–89 | 0.039 | *** | 0.001 | 0.003 | *** | 0.001 | 85–89 | 0.039 | *** | 0.001 | 0.007 | *** | 0.001 | 85–89 | 0.036 | *** | 0.001 | −0.003 | *** | 0.001 |
| 90 + | 0.038 | *** | 0.002 | 0.001 | | 0.001 | 90+ | 0.038 | *** | 0.002 | 0.004 | *** | 0.001 | 90 + | 0.037 | *** | 0.002 | −0.001 | | 0.001 |
| Age group * cancer | | | | | | | Age group * cardiovascular | | | | | | | Age group * chronic | | | | | | |
| 0–49 | Ref | | | Ref. | | | 0–49 | Ref. | | | Ref. | | | 0–49 | Ref. | | | Ref. | | |
| 50–54 | −0.088 | | 0.268 | −0.416 | *** | 0.016 | 50–54 | 0.128 | | 0.186 | −0.142 | *** | 0.012 | 50–54 | 0.100 | | 0.069 | 0.237 | *** | 0.007 |
| 55–59 | 0.799 | *** | 0.280 | −0.545 | *** | 0.014 | 55–59 | −0.376 | *** | 0.104 | −0.194 | *** | 0.011 | 55–59 | −0.060 | | 0.056 | 0.266 | *** | 0.006 |
| 60–64 | −0.046 | | 0.263 | −0.707 | *** | 0.013 | 60–64 | −0.204 | * | 0.122 | −0.285 | *** | 0.010 | 60–64 | −0.001 | | 0.065 | 0.275 | *** | 0.007 |
| 65–69 | 0.399 | | 0.515 | −0.914 | *** | 0.013 | 65–69 | −0.382 | *** | 0.114 | −0.392 | *** | 0.010 | 65–69 | 0.033 | | 0.071 | 0.211 | *** | 0.006 |
| 70–74 | −0.255 | | 0.254 | −1.179 | *** | 0.013 | 70–74 | −0.272 | ** | 0.131 | −0.488 | *** | 0.010 | 70–74 | −0.028 | | 0.080 | 0.096 | *** | 0.006 |

**Table 2** (continued)

| | 5 Cancer | | | | | | | 6 Cardiovascular | | | | | | | 7 Chronic | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | | Probit | | | GLM-Poisson | | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | | *P* (Exp. > 0) | | | Log (Exp.) | | | | *P* (Exp. > 0) | | | Log(Exp.) | | |
| | Coef | | SE | Coef | | SE | | Coef | | SE | Coef | | SE | | Coef | | SE | Coef | | SE |
| 75–79 | −0.024 | | 0.303 | −1.471 | *** | 0.014 | 75–79 | −0.266 | ** | 0.122 | −0.580 | *** | 0.010 | 75–79 | 0.000 | | 0.090 | −0.100 | *** | 0.007 |
| 80–84 | | | | −1.661 | *** | 0.014 | 80–84 | −0.125 | | 0.215 | −0.667 | *** | 0.010 | 80–84 | 0.101 | | 0.122 | −0.364 | *** | 0.007 |
| 85–89 | | | | −1.761 | *** | 0.015 | 85–89 | −0.265 | * | 0.147 | −0.715 | *** | 0.010 | 85–89 | −0.122 | | 0.126 | −0.564 | *** | 0.008 |
| 90+ | −0.624 | ** | 0.306 | −1.782 | *** | 0.018 | 90+ | −0.147 | | 0.232 | −0.668 | *** | 0.011 | 90+ | −0.404 | *** | 0.133 | −0.699 | *** | 0.009 |
| Age group* year * cancer | | | | | | | Age group* year * cardiovascular | | | | | | | Age group* year * chronic | | | | | | |
| 0–49 | Ref. | | | Ref. | | | 0–49 | Ref. | | | Ref | | | 0–49 | Ref. | | | Ref. | | |
| 50–54 | 0.037 | ** | 0.015 | −0.016 | *** | 0.002 | 50–54 | −0.028 | | 0.021 | −0.001 | | 0.002 | 50–54 | −0.017 | ** | 0.008 | −0.008 | *** | 0.001 |
| 55–59 | −0.135 | *** | 0.028 | −0.018 | *** | 0.002 | 55–59 | −0.002 | | 0.012 | −0.004 | ** | 0.002 | 55–59 | −0.011 | | 0.007 | −0.007 | *** | 0.001 |
| 60–64 | 0.055 | *** | 0.015 | −0.015 | *** | 0.002 | 60–64 | −0.020 | | 0.015 | −0.004 | *** | 0.001 | 60–64 | −0.016 | * | 0.008 | −0.007 | *** | 0.001 |
| 65–69 | −0.100 | * | 0.057 | −0.011 | *** | 0.002 | 65–69 | −0.008 | | 0.014 | −0.005 | *** | 0.001 | 65–69 | −0.026 | *** | 0.009 | −0.004 | *** | 0.001 |
| 70–74 | −0.015 | | 0.039 | −0.005 | *** | 0.002 | 70–74 | −0.024 | * | 0.015 | −0.007 | *** | 0.001 | 70–74 | −0.016 | * | 0.010 | −0.002 | * | 0.001 |
| 75–79 | −0.073 | ** | 0.036 | 0.001 | | 0.002 | 75–79 | −0.033 | ** | 0.013 | −0.009 | *** | 0.001 | 75–79 | −0.033 | *** | 0.011 | 0.004 | *** | 0.001 |
| 80–84 | | | | −0.002 | | 0.002 | 80–84 | −0.045 | * | 0.024 | −0.010 | *** | 0.001 | 80–84 | −0.045 | *** | 0.014 | 0.010 | *** | 0.001 |
| 85–89 | | | | −0.008 | *** | 0.002 | 85–89 | −0.035 | *** | 0.012 | −0.011 | *** | 0.001 | 85–89 | −0.031 | * | 0.016 | 0.007 | *** | 0.001 |
| 90+ | −0.051 | | 0.038 | −0.013 | *** | 0.002 | 90+ | −0.035 | *** | 0.008 | −0.012 | *** | 0.002 | 90+ | −0.017 | | 0.017 | 0.002 | | 0.001 |
| Observations | 73,981,710 | | | 68,754,429 | | | | 74,217,773 | | | 68,754,429 | | | | 74,217,773 | | | 68,754,429 | | |

[a]The probability that an individual receives some health care services in a given year

[b]The log-link function provides multiplicative effects and the results can therefore be expressed as percentage increases in the average health care expenditure per unit increase in any of the covariates by taking the exponential of the covariates

[c]All individuals with cancer aged 80–84 and 85–89 have positive health care expenditures, i.e., *P* (exp > 0) = 1, and therefore, the effective coefficient on the omitted covariates is infinity. We drop these covariates as dropping these has no effect on the likelihood or estimates of the remaining coefficients

*Statistically significant at the 0.10 level

**At the 0.05 level

***At the 0.01 level

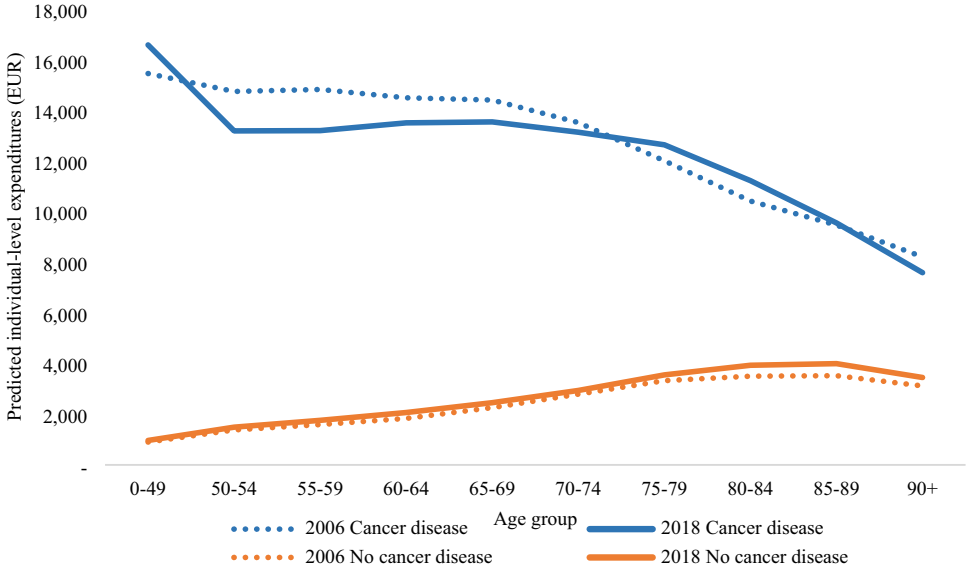



**Fig. 3** Predicted individual-level health care expenditure curves by cancer diagnosis and calendar year. Based on model specification 5 [Note 1: Estimates are predicted from model specification (5), shown in Table 2, using the coefficients estimated in the probit and the GLM-Poisson specification. The calculation follows $E(Y|X) = P(Y > 0|X) * E(Y|X, Y > 0)$ where $Y$ is the individual-level expenditures and $X$ represents covariates included in model (5). Note 2: The predicted expenditures are shown for survivors. It does not change the interpretation of the steepening results to include estimated expenditures from decedents. Likewise, model (6) and (7) are used when plotting the curves in Figs. 4 and 5]

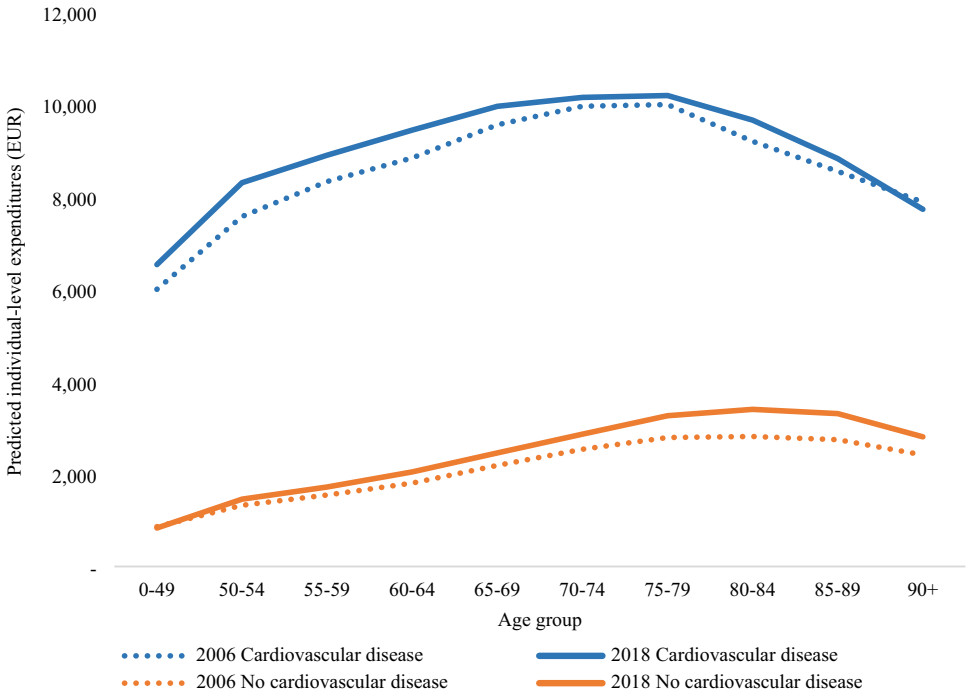



**Fig. 4** Predicted individual-level health care expenditure curves by cardiovascular diagnosis and calendar year. Based on model specification 6

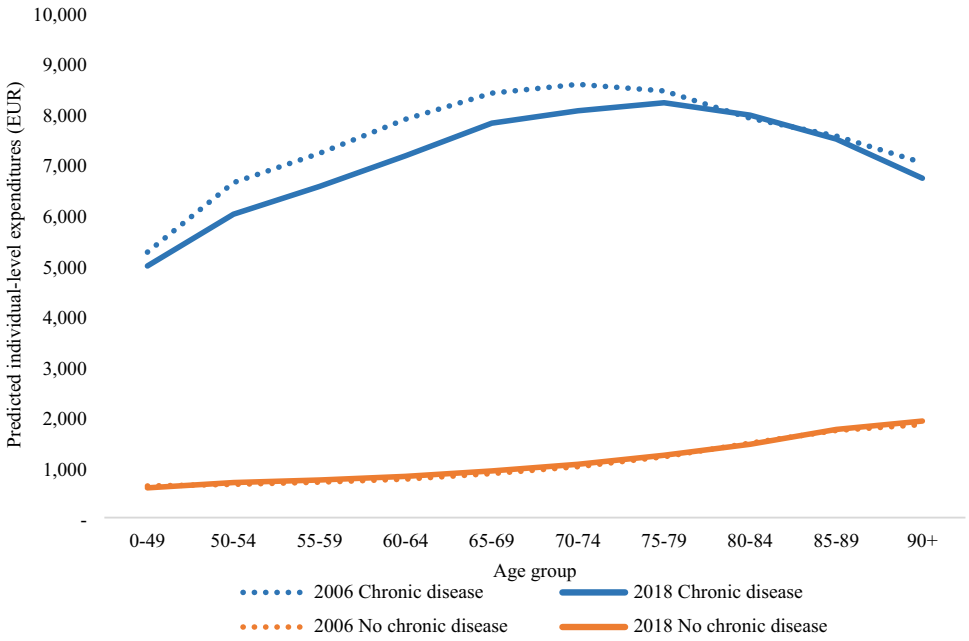



**Fig. 5** Predicted individual-level health care expenditure curves by chronic diagnosis and calendar year. Based on model specification 7

## Age and time to death

In Table 3, the results of estimating model (8) and (9) are shown for selected covariates. The full regression table, including interactions between age group indicators and time to death indicators, is presented in appendix Table 6. Overall, the results show that both age and time to death have an impact on the use of health care services and on the individual-health care expenditure levels. Therefore, we find that the red herring hypothesis cannot be confirmed. The conclusion is true for both model (8) and model (9) where we in model (9) include 28 chronic diseases as explanatory variables. Adding morbidity, in terms of the presence of 28 chronic diseases, lowers the magnitude of the age and time to death coefficients, but both age and time to death remain important predictors for the use and level of individual-level health care expenditures.

When examining the results of the probit model, we see that both age and time to death matters for the probability of receiving health care services. Not surprisingly, the estimates suggest that the probability of receiving health care services is increasing in age and decreasing in time to death. Hence, individuals in older age groups have a higher probability of receiving health care services than younger age groups (relative to the 1–19 years old), and individuals with only a few years to death have a higher probability of receiving health care services than individuals with more years to death (relative to 8 + years). When examining the Poisson-distributed GLM model, the same patterns are true. The GLM model shows that the health care expenditures increase with age, but from 75 years and up, the increase relative to the 1–19 years old is declining. When examining the effect of time to death on health care expenditures, we see that the health care expenditures decrease in time to death, although the effect of time to death on expenditures is significantly positive for all 7 years up to death relative to 8 + years. In Figs. 6 and 7, we illustrate predicted expenditures levels based on model specification (8) and (9) by age group and time to death for four age groups: 60–64, 70–74, 80–84, 90 + years old.

## Discussion and conclusion

In this paper, we find that death-related expenditures increase over time, and we demonstrate that both age and time to death are strongly associated with the individual-level health care expenditures. Hence, the red herring hypothesis cannot be confirmed. We also demonstrate that individual-level health care expenditures have increased relatively more for older age groups than younger age groups, and consequently, the individual-level age-distributed expenditure curve has steepened over the period 2006–2018. The steepening is driven by those aged 75–90 +, and we conclude that from 2006 to 2018, a proportionally higher share of the health care budget has been allocated to treat each of the oldest individuals. As the share of the elderly population continues to increase in the coming decades, spending more on the individual elderly increases the budgetary pressure of an ageing population even further.

Changes in the composition of individuals across age groups and disease groups over time may be important drivers of steepening as steepening may be caused by increased demand from large patient groups. An increasing share

**Table 3** Regression results: age and time to death

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| Intercept | 1.141 | *** | 0.001 | 8.469 | *** | 0.004 | 1.118 | *** | 0.001 | 8.451 | *** | 0.004 |
| 2006 (reference year) | | | | | | | | | | | | |
| 2007 | −0.009 | *** | 0.001 | 0.024 | *** | 0.002 | −0.010 | *** | 0.001 | 0.006 | *** | 0.002 |
| 2008 | −0.007 | *** | 0.001 | 0.028 | *** | 0.002 | −0.009 | *** | 0.001 | 0.000 | | 0.002 |
| 2009 | 0.015 | *** | 0.001 | 0.142 | *** | 0.002 | 0.010 | *** | 0.001 | 0.079 | *** | 0.002 |
| 2010 | −0.023 | *** | 0.001 | 0.196 | *** | 0.002 | −0.032 | *** | 0.001 | 0.108 | *** | 0.002 |
| 2011 | −0.003 | ** | 0.001 | 0.204 | ** | 0.002 | −0.013 | ** | 0.001 | 0.097 | ** | 0.002 |
| Female | 0.192 | *** | 0.002 | −0.038 | *** | 0.006 | 0.195 | *** | 0.002 | −0.038 | *** | 0.005 |
| Age group | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | −0.163 | *** | 0.002 | 0.054 | *** | 0.008 | −0.167 | *** | 0.002 | 0.047 | *** | 0.008 |
| 25–29 | −0.151 | *** | 0.002 | 0.057 | *** | 0.008 | −0.161 | *** | 0.002 | 0.030 | *** | 0.007 |
| 30–34 | 0.022 | *** | 0.002 | 0.140 | *** | 0.007 | 0.007 | *** | 0.002 | 0.092 | *** | 0.007 |
| 35–39 | 0.135 | *** | 0.002 | 0.238 | *** | 0.007 | 0.113 | *** | 0.002 | 0.171 | *** | 0.006 |
| 40–44 | 0.204 | *** | 0.002 | 0.359 | *** | 0.007 | 0.173 | *** | 0.002 | 0.259 | *** | 0.006 |
| 45–49 | 0.266 | *** | 0.003 | 0.504 | *** | 0.006 | 0.224 | *** | 0.003 | 0.362 | *** | 0.006 |
| 50–54 | 0.346 | *** | 0.003 | 0.674 | *** | 0.006 | 0.292 | *** | 0.003 | 0.480 | *** | 0.006 |
| 55–49 | 0.448 | *** | 0.003 | 0.854 | *** | 0.006 | 0.379 | *** | 0.003 | 0.594 | *** | 0.006 |
| 60–64 | 0.550 | *** | 0.003 | 1.024 | *** | 0.006 | 0.468 | *** | 0.003 | 0.694 | *** | 0.005 |
| 65–69 | 0.681 | *** | 0.004 | 1.210 | *** | 0.006 | 0.579 | *** | 0.004 | 0.781 | *** | 0.005 |
| 70–74 | 0.895 | *** | 0.006 | 1.355 | *** | 0.006 | 0.777 | *** | 0.006 | 0.834 | *** | 0.006 |
| 75–79 | 0.948 | *** | 0.007 | 1.450 | *** | 0.006 | 0.807 | *** | 0.008 | 0.845 | *** | 0.007 |
| 80–84 | 1.057 | *** | 0.011 | 1.470 | *** | 0.007 | 0.907 | *** | 0.011 | 0.846 | *** | 0.007 |
| 85–89 | 1.029 | *** | 0.016 | 1.476 | *** | 0.009 | 0.872 | *** | 0.017 | 0.867 | *** | 0.009 |
| 90 + | 0.860 | *** | 0.031 | 1.437 | *** | 0.016 | 0.711 | *** | 0.033 | 0.922 | *** | 0.015 |
| Time to death | | | | | | | | | | | | |
| Time to death 0 years | 0.690 | *** | 0.071 | 3.756 | *** | 0.051 | 0.434 | *** | 0.079 | 2.740 | *** | 0.055 |
| Time to death 1 year | 0.468 | *** | 0.063 | 3.615 | *** | 0.071 | 0.251 | *** | 0.069 | 2.664 | *** | 0.068 |
| Time to death 2 years | 0.176 | *** | 0.054 | 3.091 | *** | 0.085 | 0.005 | | 0.058 | 2.285 | *** | 0.083 |
| Time to death 3 years | 0.139 | *** | 0.053 | 2.631 | *** | 0.094 | −0.012 | | 0.057 | 1.814 | *** | 0.090 |
| Time to death 4 years | 0.222 | *** | 0.055 | 2.392 | *** | 0.111 | 0.084 | | 0.058 | 1.716 | *** | 0.103 |
| Time to death 5 years | 0.166 | *** | 0.052 | 2.470 | *** | 0.104 | 0.043 | | 0.055 | 1.716 | *** | 0.128 |
| Time to death 6 years | 0.138 | *** | 0.049 | 2.183 | *** | 0.120 | 0.026 | | 0.052 | 1.549 | *** | 0.120 |

**Table 3** (continued)

| | Model 8 | | | | | | Model 9 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | SE | Coef | | SE |
| Time to death 7 years | 0.119 | ** | 0.048 | 2.056 | *** | 0.138 | 0.012 | 0.050 | 1.548 | *** | 0.131 |
| Time to death 8 years + (reference group) | | | | | | | | | | | |
| 120 age group*time to death interactions | Yes | | | Yes | | | Yes | | Yes | | |
| 28 chronic diseases | | | | | | | Yes | | Yes | | |
| Observations | 33,260,167 | | | 30,986,554 | | | 33,260,167 | | 30,968,122 | | |

Full regression table can be seen in appendix table (Table 6)

[a]The probability that an individual receives some health care services in a given year

[b]The log-link function provides multiplicative effects and the results can therefore be expressed as percentage increases in the average health care expenditure per unit increase in any of the covariates by taking the exponential of the covariates

*Statistically significant at the 0.10 level

**at the 0.05 level

***at the 0.01 level

of elderly will increase the probability of steepening, but changes in the age distribution within and across disease groups may also imply different steepening patterns across diseases which, in turn, may impact the overall steepening; for example, increases in the number of elderly with a cancer diagnosis relative to younger cancer diagnosed and relative to other disease groups. To get more knowledge on potential disease-specific cost-drivers of steepening, we examine steepening patterns across disease groups. When examining steepening across individuals with cancer and individuals without cancer, we find evidence of steepening in both groups but with very different distributions over time. For cancer diagnosed, the steepening is driven by lower expenditures among 50–74 years old and higher expenditures among 75–89 years old over time. For non-cancer diagnosed, we see the same pattern as for the full population, where primarily older age groups drive the steepening of the curve. We find evidence of less steepening in the group of individuals with a cardiovascular disease compared to individuals without a cardiovascular disease. This is somewhat surprising if we relate this finding to earlier theoretically proposed mechanisms to steepening, namely that steepening may be caused by increased demand from large patient groups and increased supply through a technological bias towards the development of medical advances to treat these proportionally large groups. The group with a diagnosis of a cardiovascular disease includes many patients with lifestyle-related diseases, e.g., ischemic heart disease which is the most common cause of death in Europe among individuals over 65 years of age [33]. This group will presumably spend more years in ill health as longevity increases. Yet, we do observe less steepening in the group of cardiovascular diseases.

When examining individuals with a chronic disease and individuals without a chronic disease, we observe more steepening in the group of individuals with a chronic disease relative to not having a chronic disease. In fact, we do not find any evidence of steepening among individuals without a chronic disease. Our findings also reveal that within the group of individuals diagnosed with a chronic disease, the predicted individual-level expenditure of treating 75–89 years old have not increased much over the observation period, but the cost of treating 0–74 years old with a chronic disease have decreased. This may be because the underlying disease pattern within the group of individuals with a chronic disease has changed over time, and therefore, we suggest more research in the underlying cost distribution within the group of chronic disease diagnosed over time. Also, future studies should examine steepening mechanisms which have affected chronic diseases among the elderly. This might be supply mechanisms affecting treatment choice for chronic patients rather than increased patient demand in specific large diagnosis groups. Despite the disease complexity underlying the overall steepening effect, we do observe significantly more

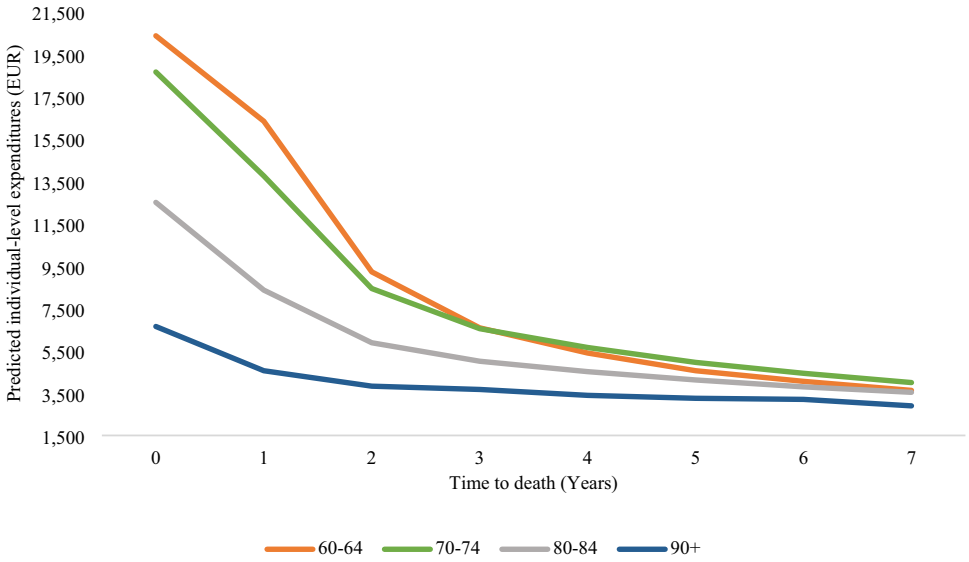



**Fig. 6** Predicted individual-level health care expenditure curves by time to death and age groups. Based on model specification 8

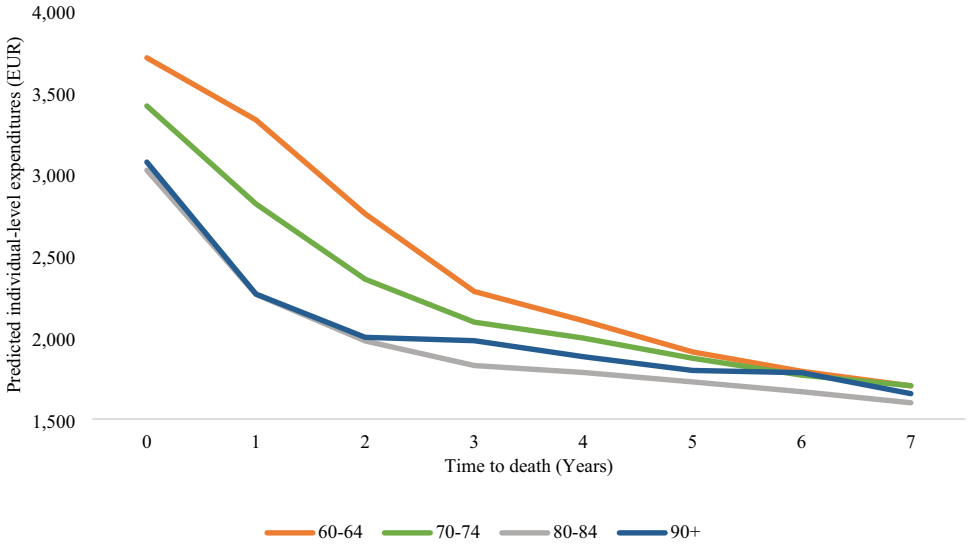



**Fig. 7** Predicted individual-level health care expenditure curves by time to death and age groups. Based on model specification 9

steepening over the 12 year period from 2006 to 2018, and if these trends are to continue, it will become increasingly difficult to establish a financially sustainable infrastructure for the future delivery of universal health care coverage.

In the study, we include inpatient and outpatient hospital care and general practitioner visits. In comparison to earlier studies of steepening where only inpatient care was included [7], we therefore include most of the universal coverage in the Danish health care system. We do not include prescription medicines which accounts for 4 pct. of total health care expenditures in Denmark [25]. Prescription medicines in Denmark are not covered by the universal coverage, although with some exceptions where partial costs are reimbursed. In the analyses of disease-specific cost-drivers, one might argue that prescription medicines should have been included as individuals diagnosed with a chronic disease have high prescription rates, and as the use of prescription medicines among older individuals with comorbidities may have increased, and, in turn, have contributed further to steepening over the period. However, it has not been possible to access data on prescription medicines in this study. Additionally, we do not include expenditure data on Long-Term Care (LTC). Therefore, we might have underestimated the extent of steepening as some hospital care may have been shifted to become LTC over the 12-year study period, and as it is primarily elderly individuals who utilize LTC. The implications of not including LTC in the analyses of the relative importance of age and time to death are not clear, but population ageing may have a different impact on LTC than non-LTC. From the perspective of the compression of morbidity hypothesis, LTC expenditures should primarily depend on time to death and age-specific LTC shares should fall as life-expectancy

increase [2]. If that is the case, we may underestimate the effect of time to death in the analyses of the relative importance of age and time to death. However, to our knowledge, there is no research on the relation between compression of morbidity and the development of LTC expenditures.

Related to the development in the utilization of LTC care, one limitation is that we only observe diagnostic information on individuals who have received health care services in the hospital sector. However, some individuals may have a diagnosis without utilizing health care services in the hospital sector, for example, because of uncovered needs or because they only utilize other health care services such as LTC or GP care. The lack of diagnostic information from other data sources may imply that some individuals are interpreted as having zero costs, although they are associated with, for example, high LTC cost. This may also reinforce the previously mentioned implications of not including LTC data in the present study.

To our knowledge, this is the first paper examining both steepening and the association between age, time to death, and individual-level health care expenditures in the same paper. In addition, we have added to the previous literature on steepening by applying individual-level data sets for the entire Danish population for an observation period of 12 years, and by examining disease-specific cost-drivers of steepening. Moreover, we provide a population-wide analysis of the importance of age and time to death where we account for individual-level health status by including 28 chronic conditions.

We conclude that the threefold combination of (1) an ageing population, (2) the positive correlation between expenditures and age per se, and (3) a steepening of the expenditure curve make the financial burden of population ageing substantial. As population ageing will be a fact in the coming decades and as previous steepening trends may continue, it may be relevant to include steepening scenarios in expenditure forecasts and we must work to identify the drivers of steepening. To mitigate budgetary pressure from an ageing population and steepening of the individual-level expenditure curve in universal health care systems, policymakers may stimulate cost-effective medical advances and health care utilization in the treatment of elderly patients.

## Appendix

## Appendix A: Definition of disease groups

### The disease groups are identified using ICD-10 codes and follow the grouping from Nexø et al. (2018)

#### Cancer diseases

All cancer diagnoses (C-codes).

**Table 4** Number of individuals and average individual-level health care expenditures by year (EUR, 2018-prices)

| | Mean | SD | Freq |
|---|---|---|---|
| 2006 | 1597 | 5886 | 5,538,154 |
| 2007 | 1632 | 5718 | 5,565,071 |
| 2008 | 1634 | 5819 | 5,599,681 |
| 2009 | 1830 | 6511 | 5,624,590 |
| 2010 | 1936 | 7002 | 5,648,337 |
| 2011 | 1942 | 6961 | 5,668,429 |
| 2012 | 1888 | 6870 | 5,690,953 |
| 2013 | 1919 | 7128 | 5,716,178 |
| 2014 | 2015 | 7582 | 5,749,618 |
| 2015 | 1901 | 6879 | 5,796,517 |
| 2016 | 1859 | 6939 | 5,840,397 |
| 2017 | 1855 | 6735 | 5,874,648 |
| 2018 | 1956 | 7555 | 5,905,200 |
| Total | 1845 | 6772 | 74,217,773 |

#### Cardiovascular diseases

Ischemic heart disease and heart failure, cardiac arrhythmia and valve disease, stroke, and vascular disease (I-codes).

#### Chronic diseases

Chronic infection, Cancer, Benign hematologic diseases including anemia, Diseases of the thyroid gland, Diabetes, Other endocrine diseases and malnutrition, Obesity, Neurological diseases, Paraplegia and hemiplegia, Eye diseases, Ear diseases, Hypertension, Cardiovascular (Ischemic heart disease and heart failure, Cardiac arrhythmia and valve disease, Stroke, Vascular disease), Chronic pulmonary disease including asthma, Inflammatory bowel disease, Diseases of the liver, Diseases of the skin, Inflammatory rheumatic disease, Degenerative rheumatic diseases and osteoarthritis, Osteoporosis, Diseases of the kidney, Gynecological diseases, Dementia, Substance abuse, Schizophrenia, Depression and anxiety, Eating disorders, and Personality disorders.

## Appendix B: Summary statistics

Table 5.

**Table 5** Number of individuals and average individual-level health care expenditures by year and disease group (EUR, 2018-prices)

| Cancer disease | | | | Cardiovascular disease | | | | Chronic disease | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | | | 0 | 1 | | | 0 | 1 |
| 2006 | Mean | 1348 | 17,001 | 2006 | Mean | 1303 | 10,258 | 2006 | Mean | 720 | 7286 |
| | SD | 4658 | 24,341 | | SD | 5040 | 15,154 | | SD | 2244 | 13,760 |
| | Freq | 5,449,843 | 88,311 | | Freq | 5,356,425 | 181,729 | | Freq | 4,798,311 | 739,843 |
| 2007 | Mean | 1383 | 16,217 | 2007 | Mean | 1340 | 10,141 | 2007 | Mean | 734 | 7314 |
| | SD | 4697 | 20,917 | | SD | 4877 | 14,721 | | SD | 2307 | 12,981 |
| | Freq | 5,471,452 | 93,619 | | Freq | 5,380,068 | 185,003 | | Freq | 4,805,775 | 759,296 |
| 2008 | Mean | 1374 | 16,081 | 2008 | Mean | 1344 | 10,243 | 2008 | Mean | 731 | 7326 |
| | SD | 4786 | 20,754 | | SD | 4934 | 15,471 | | SD | 2305 | 13,276 |
| | Freq | 5,500,765 | 98,916 | | Freq | 5,416,779 | 182,902 | | Freq | 4,832,788 | 766,893 |
| 2009 | Mean | 1535 | 16,840 | 2009 | Mean | 1500 | 11,109 | 2009 | Mean | 789 | 7908 |
| | SD | 5379 | 22,309 | | SD | 5432 | 17,776 | | SD | 2566 | 14,434 |
| | Freq | 5,516,252 | 108,338 | | Freq | 5,431,704 | 192,886 | | Freq | 4,802,696 | 821,894 |
| 2010 | Mean | 1619 | 17,655 | 2010 | Mean | 1588 | 11,454 | 2010 | Mean | 815 | 8200 |
| | SD | 5755 | 24,200 | | SD | 5904 | 18,517 | | SD | 2429 | 15,614 |
| | Freq | 5,536,698 | 111,639 | | Freq | 5,449,249 | 199,088 | | Freq | 4,791,103 | 857,234 |
| 2011 | Mean | 1606 | 17,832 | 2011 | Mean | 1587 | 11,356 | 2011 | Mean | 802 | 8117 |
| | SD | 5626 | 24,201 | | SD | 5862 | 18,173 | | SD | 2608 | 15,125 |
| | Freq | 5,551,093 | 117,336 | | Freq | 5,462,566 | 205,863 | | Freq | 4,784,626 | 883,803 |
| 2012 | Mean | 1559 | 16,923 | 2012 | Mean | 1537 | 11,117 | 2012 | Mean | 767 | 7789 |
| | SD | 5512 | 24,173 | | SD | 5750 | 18,161 | | SD | 2369 | 14,987 |
| | Freq | 5,569,050 | 121,903 | | Freq | 5,482,145 | 208,808 | | Freq | 4,782,299 | 908,654 |
| 2013 | Mean | 1571 | 17,196 | 2013 | Mean | 1559 | 11,438 | 2013 | Mean | 755 | 7835 |
| | SD | 5687 | 24,921 | | SD | 5918 | 19,336 | | SD | 2374 | 15,441 |
| | Freq | 5,588,633 | 127,545 | | Freq | 5,507,832 | 208,346 | | Freq | 4,776,235 | 939,943 |
| 2014 | Mean | 1639 | 17,638 | 2014 | Mean | 1625 | 12,104 | 2014 | Mean | 777 | 8080 |
| | SD | 6048 | 25,989 | | SD | 6241 | 20,757 | | SD | 2451 | 16,290 |
| | Freq | 5,614,386 | 135,232 | | Freq | 5,535,475 | 214,143 | | Freq | 4,774,921 | 974,697 |
| 2015 | Mean | 1555 | 15,800 | 2015 | Mean | 1540 | 11,055 | 2015 | Mean | 765 | 7355 |
| | SD | 5484 | 23,253 | | SD | 5668 | 18,578 | | SD | 2300 | 14,600 |
| | Freq | 5,655,417 | 141,100 | | Freq | 5,576,384 | 220,133 | | Freq | 4,797,261 | 999,256 |
| 2016 | Mean | 1521 | 15,049 | 2016 | Mean | 1504 | 10,833 | 2016 | Mean | 749 | 7141 |
| | SD | 5674 | 22,190 | | SD | 5776 | 18,397 | | SD | 2367 | 14,728 |
| | Freq | 5,694,492 | 145,905 | | Freq | 5,618,293 | 222,104 | | Freq | 4,826,374 | 1,014,023 |
| 2017 | Mean | 1515 | 15,023 | 2017 | Mean | 1500 | 10,611 | 2017 | Mean | 753 | 7042 |
| | SD | 5490 | 21,397 | | SD | 5551 | 18,018 | | SD | 2364 | 14,142 |
| | Freq | 5,726,898 | 147,750 | | Freq | 5,645,777 | 228,871 | | Freq | 4,845,421 | 1,029,227 |
| 2018 | Mean | 1588 | 16,039 | 2018 | Mean | 1588 | 10,952 | 2018 | Mean | 769 | 7465 |
| | SD | 5630 | 28,706 | | SD | 6396 | 19,170 | | SD | 2478 | 16,015 |
| | Freq | 5,754,655 | 150,545 | | Freq | 5,672,706 | 232,494 | | Freq | 4,857,784 | 1,047,416 |
| Total | Mean | 1525 | 16,497 | Total | Mean | 1502 | 10,992 | Total | Mean | 764 | 7600 |
| | SD | 5438 | 23,897 | | SD | 5667 | 18,045 | | SD | 2399 | 14,828 |
| | Freq | 72,629,634 | 1,588,139 | | Freq | 71,535,403 | 2,682,370 | | Freq | 62,475,594 | 11,742,179 |

Number of individuals by disease group and year

## Appendix C: extended regression results

Table 6.

**Table 6** Extended regression results for the importance of age and time to death—model (8) and (9)

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| Intercept | 1.141 | *** | 0.001 | 8.469 | *** | 0.004 | 1.118 | *** | 0.001 | 8.451 | *** | 0.004 |
| 2006 (reference year) | | | | | | | | | | | | |
| 2007 | − 0.009 | *** | 0.001 | 0.024 | *** | 0.002 | − 0.010 | *** | 0.001 | 0.006 | *** | 0.002 |
| 2008 | − 0.007 | *** | 0.001 | 0.028 | *** | 0.002 | − 0.009 | *** | 0.001 | 0.000 | | 0.002 |
| 2009 | 0.015 | *** | 0.001 | 0.142 | *** | 0.002 | 0.010 | *** | 0.001 | 0.079 | *** | 0.002 |
| 2010 | − 0.023 | *** | 0.001 | 0.196 | *** | 0.002 | − 0.032 | *** | 0.001 | 0.108 | *** | 0.002 |
| 2011 | − 0.003 | ** | 0.001 | 0.204 | ** | 0.002 | − 0.013 | ** | 0.001 | 0.097 | ** | 0.002 |
| Female | 0.192 | *** | 0.002 | − 0.038 | *** | 0.006 | 0.195 | *** | 0.002 | − 0.038 | *** | 0.005 |
| Age group | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | − 0.163 | *** | 0.002 | 0.054 | *** | 0.008 | − 0.167 | *** | 0.002 | 0.047 | *** | 0.008 |
| 25–29 | − 0.151 | *** | 0.002 | 0.057 | *** | 0.008 | − 0.161 | *** | 0.002 | 0.030 | *** | 0.007 |
| 30–34 | 0.022 | *** | 0.002 | 0.140 | *** | 0.007 | 0.007 | *** | 0.002 | 0.092 | *** | 0.007 |
| 35–39 | 0.135 | *** | 0.002 | 0.238 | *** | 0.007 | 0.113 | *** | 0.002 | 0.171 | *** | 0.006 |
| 40–44 | 0.204 | *** | 0.002 | 0.359 | *** | 0.007 | 0.173 | *** | 0.002 | 0.259 | *** | 0.006 |
| 45–49 | 0.266 | *** | 0.003 | 0.504 | *** | 0.006 | 0.224 | *** | 0.003 | 0.362 | *** | 0.006 |
| 50–54 | 0.346 | *** | 0.003 | 0.674 | *** | 0.006 | 0.292 | *** | 0.003 | 0.480 | *** | 0.006 |
| 55–49 | 0.448 | *** | 0.003 | 0.854 | *** | 0.006 | 0.379 | *** | 0.003 | 0.594 | *** | 0.006 |
| 60–64 | 0.550 | *** | 0.003 | 1.024 | *** | 0.006 | 0.468 | *** | 0.003 | 0.694 | *** | 0.005 |
| 65–69 | 0.681 | *** | 0.004 | 1.210 | *** | 0.006 | 0.579 | *** | 0.004 | 0.781 | *** | 0.005 |
| 70–74 | 0.895 | *** | 0.006 | 1.355 | *** | 0.006 | 0.777 | *** | 0.006 | 0.834 | *** | 0.006 |
| 75–79 | 0.948 | *** | 0.007 | 1.450 | *** | 0.006 | 0.807 | *** | 0.008 | 0.845 | *** | 0.007 |
| 80–84 | 1.057 | *** | 0.011 | 1.470 | *** | 0.007 | 0.907 | *** | 0.011 | 0.846 | *** | 0.007 |
| 85–89 | 1.029 | *** | 0.016 | 1.476 | *** | 0.009 | 0.872 | *** | 0.017 | 0.867 | *** | 0.009 |
| 90 + | 0.860 | *** | 0.031 | 1.437 | *** | 0.016 | 0.711 | *** | 0.033 | 0.922 | *** | 0.015 |
| Age group*female | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.256 | *** | 0.003 | 0.462 | *** | 0.010 | 0.253 | *** | 0.003 | 0.443 | *** | 0.009 |
| 25–29 | 0.412 | *** | 0.004 | 0.849 | *** | 0.009 | 0.408 | *** | 0.004 | 0.820 | *** | 0.009 |
| 30–34 | 0.507 | *** | 0.004 | 0.892 | *** | 0.009 | 0.502 | *** | 0.004 | 0.846 | *** | 0.008 |
| 35–39 | 0.455 | *** | 0.004 | 0.619 | *** | 0.009 | 0.449 | *** | 0.004 | 0.557 | *** | 0.008 |
| 40–44 | 0.375 | *** | 0.004 | 0.359 | *** | 0.009 | 0.371 | *** | 0.004 | 0.291 | *** | 0.008 |
| 45–49 | 0.321 | *** | 0.004 | 0.263 | *** | 0.009 | 0.318 | *** | 0.004 | 0.196 | *** | 0.008 |

**Table 6** (continued)

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 50–54 | 0.337 | *** | 0.005 | 0.245 | *** | 0.008 | 0.335 | *** | 0.005 | 0.172 | *** | 0.008 |
| 55–49 | 0.273 | *** | 0.005 | 0.156 | *** | 0.008 | 0.275 | *** | 0.005 | 0.103 | *** | 0.008 |
| 60–64 | 0.168 | *** | 0.005 | 0.074 | *** | 0.008 | 0.173 | *** | 0.005 | 0.051 | *** | 0.008 |
| 65–69 | 0.080 | *** | 0.006 | 0.023 | *** | 0.008 | 0.086 | *** | 0.006 | 0.036 | *** | 0.008 |
| 70–74 | −0.102 | *** | 0.008 | −0.024 | *** | 0.008 | −0.102 | *** | 0.008 | 0.023 | *** | 0.008 |
| 75–79 | −0.143 | *** | 0.009 | −0.037 | *** | 0.008 | −0.138 | *** | 0.010 | 0.067 | *** | 0.009 |
| 80–84 | −0.222 | *** | 0.011 | −0.052 | *** | 0.008 | −0.210 | *** | 0.012 | 0.104 | *** | 0.008 |
| 85–89 | −0.195 | *** | 0.015 | −0.059 | *** | 0.008 | −0.170 | *** | 0.016 | 0.128 | *** | 0.009 |
| 90 + | −0.146 | *** | 0.022 | −0.104 | *** | 0.010 | −0.114 | *** | 0.023 | 0.084 | *** | 0.011 |
| Time to death | | | | | | | | | | | | |
| Time to death 0 year | 0.690 | *** | 0.071 | 3.756 | *** | 0.051 | 0.434 | *** | 0.079 | 2.740 | *** | 0.055 |
| Time to death 1 year | 0.468 | *** | 0.063 | 3.615 | *** | 0.071 | 0.251 | *** | 0.069 | 2.664 | *** | 0.068 |
| Time to death 2 years | 0.176 | *** | 0.054 | 3.091 | *** | 0.085 | 0.005 | | 0.058 | 2.285 | *** | 0.083 |
| Time to death 3 years | 0.139 | *** | 0.053 | 2.631 | *** | 0.094 | −0.012 | | 0.057 | 1.814 | *** | 0.090 |
| Time to death 4 years | 0.222 | *** | 0.055 | 2.392 | *** | 0.111 | 0.084 | | 0.058 | 1.716 | *** | 0.103 |
| Time to death 5 years | 0.166 | *** | 0.052 | 2.470 | *** | 0.104 | 0.043 | | 0.055 | 1.716 | *** | 0.128 |
| Time to death 6 years | 0.138 | *** | 0.049 | 2.183 | *** | 0.120 | 0.026 | | 0.052 | 1.549 | *** | 0.120 |
| Time to death 7 years | 0.119 | ** | 0.048 | 2.056 | *** | 0.138 | 0.012 | | 0.050 | 1.548 | *** | 0.131 |
| Time to death 8 years + (reference group) | | | | | | | | | | | | |
| Age group*time to death 0 year | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | −0.269 | *** | 0.099 | −1.065 | *** | 0.093 | −0.167 | | 0.108 | −0.887 | *** | 0.116 |
| 25–29 | −0.103 | | 0.106 | −0.951 | *** | 0.086 | −0.031 | | 0.116 | −0.973 | *** | 0.095 |
| 30–34 | −0.201 | ** | 0.100 | −0.938 | *** | 0.078 | −0.141 | | 0.110 | −1.014 | *** | 0.087 |
| 35–39 | −0.298 | *** | 0.091 | −0.942 | *** | 0.064 | −0.296 | *** | 0.101 | −1.150 | *** | 0.072 |
| 40–44 | −0.171 | ** | 0.086 | −0.863 | *** | 0.057 | −0.215 | ** | 0.096 | −1.122 | *** | 0.064 |
| 45–49 | −0.166 | ** | 0.082 | −0.919 | *** | 0.055 | −0.260 | *** | 0.091 | −1.170 | *** | 0.059 |
| 50–54 | −0.141 | * | 0.079 | −1.066 | *** | 0.053 | −0.270 | *** | 0.088 | −1.327 | *** | 0.059 |
| 55–49 | −0.210 | *** | 0.077 | −1.176 | *** | 0.052 | −0.383 | *** | 0.086 | −1.461 | *** | 0.057 |
| 60–64 | −0.153 | ** | 0.076 | −1.307 | *** | 0.052 | −0.341 | *** | 0.085 | −1.611 | *** | 0.056 |
| 65–69 | −0.187 | ** | 0.076 | −1.518 | *** | 0.052 | −0.362 | *** | 0.084 | −1.761 | *** | 0.056 |
| 70–74 | −0.200 | *** | 0.076 | −1.733 | *** | 0.052 | −0.313 | *** | 0.085 | −1.860 | *** | 0.056 |

**Table 6** (continued)

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 75–79 | −0.173 | ** | 0.076 | −1.980 | *** | 0.052 | −0.189 | ** | 0.084 | −1.958 | *** | 0.056 |
| 80–84 | −0.115 | | 0.076 | −2.251 | *** | 0.052 | −0.015 | | 0.084 | −2.009 | *** | 0.056 |
| 85–89 | −0.023 | | 0.078 | −2.516 | *** | 0.052 | 0.168 | * | 0.085 | −2.032 | *** | 0.056 |
| 90 + | 0.158 | * | 0.081 | −2.851 | *** | 0.053 | 0.431 | *** | 0.089 | −2.073 | *** | 0.057 |
| Age group*time to death 1 year | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | −0.061 | | 0.094 | −0.908 | *** | 0.113 | 0.012 | | 0.102 | −1.099 | *** | 0.255 |
| 25–29 | 0.007 | | 0.097 | −0.928 | *** | 0.114 | 0.081 | | 0.105 | −0.818 | *** | 0.107 |
| 30–34 | −0.041 | | 0.091 | −1.037 | *** | 0.095 | 0.006 | | 0.100 | −1.089 | *** | 0.091 |
| 35–39 | −0.187 | ** | 0.081 | −0.895 | *** | 0.084 | −0.185 | ** | 0.089 | −1.054 | *** | 0.082 |
| 40–44 | −0.207 | *** | 0.074 | −0.897 | *** | 0.077 | −0.248 | *** | 0.081 | −1.038 | *** | 0.077 |
| 45–49 | −0.301 | *** | 0.070 | −0.967 | *** | 0.074 | −0.373 | *** | 0.077 | −1.192 | *** | 0.078 |
| 50–54 | −0.367 | *** | 0.067 | −1.103 | *** | 0.073 | −0.456 | *** | 0.074 | −1.319 | *** | 0.071 |
| 55–49 | −0.386 | *** | 0.066 | −1.225 | *** | 0.072 | −0.497 | *** | 0.073 | −1.453 | *** | 0.070 |
| 60–64 | −0.439 | *** | 0.065 | −1.356 | *** | 0.072 | −0.565 | *** | 0.072 | −1.584 | *** | 0.069 |
| 65–69 | −0.452 | *** | 0.065 | −1.596 | *** | 0.072 | −0.573 | *** | 0.072 | −1.776 | *** | 0.069 |
| 70–74 | −0.417 | *** | 0.066 | −1.884 | *** | 0.071 | −0.483 | *** | 0.072 | −1.949 | *** | 0.069 |
| 75–79 | −0.347 | *** | 0.066 | −2.215 | *** | 0.071 | −0.339 | *** | 0.072 | −2.129 | *** | 0.069 |
| 80–84 | −0.355 | *** | 0.066 | −2.505 | *** | 0.071 | −0.266 | *** | 0.072 | −2.207 | *** | 0.069 |
| 85–89 | −0.292 | *** | 0.067 | −2.786 | *** | 0.072 | −0.134 | * | 0.073 | −2.263 | *** | 0.069 |
| 90 + | −0.081 | | 0.071 | −3.082 | *** | 0.073 | 0.145 | * | 0.077 | −2.292 | *** | 0.069 |
| Age group*time to death 2 years | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.314 | *** | 0.092 | −0.942 | *** | 0.170 | 0.381 | *** | 0.097 | −0.784 | *** | 0.132 |
| 25–29 | 0.383 | *** | 0.094 | −0.884 | *** | 0.123 | 0.440 | *** | 0.100 | −0.846 | *** | 0.121 |
| 30–34 | 0.210 | ** | 0.083 | −1.031 | *** | 0.109 | 0.226 | ** | 0.090 | −1.051 | *** | 0.105 |
| 35–39 | 0.056 | | 0.072 | −0.941 | *** | 0.102 | 0.069 | | 0.078 | −1.058 | *** | 0.117 |
| 40–44 | −0.007 | | 0.064 | −0.862 | *** | 0.092 | −0.012 | | 0.069 | −0.887 | *** | 0.089 |
| 45–49 | −0.125 | ** | 0.060 | −0.973 | *** | 0.089 | −0.155 | ** | 0.065 | −1.049 | *** | 0.089 |
| 50–54 | −0.168 | *** | 0.058 | −1.128 | *** | 0.088 | −0.203 | *** | 0.062 | −1.146 | *** | 0.085 |
| 55–49 | −0.212 | *** | 0.057 | −1.255 | *** | 0.087 | −0.250 | *** | 0.061 | −1.284 | *** | 0.084 |
| 60–64 | −0.256 | *** | 0.056 | −1.393 | *** | 0.086 | −0.301 | *** | 0.060 | −1.398 | *** | 0.084 |

**Table 6** (continued)

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 65–69 | −0.279 | *** | 0.056 | −1.594 | *** | 0.086 | −0.323 | *** | 0.060 | −1.588 | *** | 0.084 |
| 70–74 | −0.183 | *** | 0.056 | −1.846 | *** | 0.086 | −0.192 | *** | 0.060 | −1.754 | *** | 0.084 |
| 75–79 | −0.155 | *** | 0.056 | −2.115 | *** | 0.086 | −0.126 | ** | 0.060 | −1.889 | *** | 0.084 |
| 80–84 | −0.143 | ** | 0.056 | −2.332 | *** | 0.086 | −0.057 | | 0.060 | −1.961 | *** | 0.084 |
| 85–89 | −0.071 | | 0.057 | −2.549 | *** | 0.086 | 0.063 | | 0.062 | −2.010 | *** | 0.083 |
| 90 + | 0.100 | | 0.062 | −2.730 | *** | 0.087 | 0.279 | *** | 0.066 | −2.033 | *** | 0.084 |
| Age group*time to death 3 years | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.290 | *** | 0.091 | −0.675 | *** | 0.149 | 0.343 | *** | 0.096 | −0.383 | *** | 0.133 |
| 25–29 | 0.357 | *** | 0.093 | −0.685 | *** | 0.134 | 0.397 | *** | 0.099 | −0.438 | *** | 0.121 |
| 30–34 | 0.227 | *** | 0.081 | −0.871 | *** | 0.124 | 0.258 | *** | 0.086 | −0.708 | *** | 0.126 |
| 35–39 | 0.002 | | 0.069 | −0.774 | *** | 0.110 | 0.020 | | 0.074 | −0.642 | *** | 0.105 |
| 40–44 | −0.060 | | 0.062 | −0.782 | *** | 0.102 | −0.053 | | 0.066 | −0.676 | *** | 0.113 |
| 45–49 | −0.119 | ** | 0.059 | −0.813 | *** | 0.098 | −0.130 | ** | 0.063 | −0.689 | *** | 0.094 |
| 50–54 | −0.216 | *** | 0.056 | −0.997 | *** | 0.096 | −0.226 | *** | 0.060 | −0.854 | *** | 0.093 |
| 55–49 | −0.236 | *** | 0.056 | −1.128 | *** | 0.096 | −0.246 | *** | 0.060 | −0.988 | *** | 0.093 |
| 60–64 | −0.267 | *** | 0.055 | −1.264 | *** | 0.095 | −0.278 | *** | 0.059 | −1.118 | *** | 0.092 |
| 65–69 | −0.242 | *** | 0.055 | −1.425 | *** | 0.095 | −0.253 | *** | 0.059 | −1.243 | *** | 0.091 |
| 70–74 | −0.198 | *** | 0.055 | −1.638 | *** | 0.094 | −0.194 | *** | 0.059 | −1.399 | *** | 0.091 |
| 75–79 | −0.165 | *** | 0.055 | −1.843 | *** | 0.094 | −0.133 | ** | 0.059 | −1.500 | *** | 0.091 |
| 80–84 | −0.129 | ** | 0.055 | −2.032 | *** | 0.094 | −0.049 | | 0.059 | −1.569 | *** | 0.091 |
| 85–89 | −0.069 | | 0.057 | −2.185 | *** | 0.094 | 0.052 | | 0.060 | −1.575 | *** | 0.091 |
| 90 + | 0.103 | * | 0.062 | −2.310 | *** | 0.095 | 0.262 | *** | 0.066 | −1.571 | *** | 0.092 |
| Age group*time to death 4 years | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.215 | ** | 0.095 | −0.753 | *** | 0.178 | 0.261 | *** | 0.100 | −0.556 | *** | 0.172 |
| 25–29 | 0.296 | *** | 0.094 | −0.847 | *** | 0.140 | 0.336 | *** | 0.099 | −0.652 | *** | 0.128 |
| 30–34 | 0.119 | | 0.081 | −0.982 | *** | 0.135 | 0.147 | ** | 0.086 | −0.799 | *** | 0.128 |
| 35–39 | −0.136 | ** | 0.069 | −0.793 | *** | 0.123 | −0.112 | | 0.073 | −0.661 | *** | 0.114 |
| 40–44 | −0.170 | *** | 0.063 | −0.805 | *** | 0.117 | −0.164 | ** | 0.067 | −0.682 | *** | 0.109 |
| 45–49 | −0.243 | *** | 0.060 | −0.794 | *** | 0.115 | −0.240 | *** | 0.064 | −0.716 | *** | 0.107 |
| 50–54 | −0.321 | *** | 0.058 | −0.953 | *** | 0.113 | −0.320 | *** | 0.062 | −0.848 | *** | 0.106 |

**Table 6** (continued)

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 55–49 | −0.349 | *** | 0.057 | −1.102 | *** | 0.113 | −0.349 | *** | 0.061 | −0.987 | *** | 0.105 |
| 60–64 | −0.367 | *** | 0.056 | −1.222 | *** | 0.112 | −0.369 | *** | 0.060 | −1.101 | *** | 0.105 |
| 65–69 | −0.324 | *** | 0.056 | −1.376 | *** | 0.111 | −0.322 | *** | 0.060 | −1.236 | *** | 0.104 |
| 70–74 | −0.295 | *** | 0.057 | −1.543 | *** | 0.111 | −0.283 | *** | 0.060 | −1.349 | *** | 0.104 |
| 75–79 | −0.243 | *** | 0.057 | −1.726 | *** | 0.111 | −0.204 | *** | 0.060 | −1.441 | *** | 0.104 |
| 80–84 | −0.185 | *** | 0.057 | −1.897 | *** | 0.111 | −0.104 | * | 0.061 | −1.496 | *** | 0.104 |
| 85–89 | −0.129 | ** | 0.059 | −2.025 | *** | 0.111 | −0.016 | | 0.062 | −1.509 | *** | 0.104 |
| 90 + | −0.066 | | 0.063 | −2.146 | *** | 0.112 | 0.077 | | 0.067 | −1.519 | *** | 0.105 |
| Age group*time to death 5 years | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.128 | | 0.088 | −0.899 | *** | 0.155 | 0.153 | * | 0.093 | −0.514 | *** | 0.169 |
| 25–29 | 0.260 | *** | 0.087 | −1.091 | *** | 0.141 | 0.289 | *** | 0.092 | −0.768 | *** | 0.159 |
| 30–34 | 0.022 | | 0.075 | −1.091 | *** | 0.130 | 0.051 | | 0.078 | −0.860 | *** | 0.187 |
| 35–39 | −0.057 | | 0.066 | −0.908 | *** | 0.120 | −0.031 | | 0.070 | −0.651 | *** | 0.138 |
| 40–44 | −0.143 | ** | 0.059 | −1.016 | *** | 0.112 | −0.120 | * | 0.063 | −0.747 | *** | 0.134 |
| 45–49 | −0.214 | *** | 0.057 | −1.032 | *** | 0.108 | −0.218 | *** | 0.060 | −0.812 | *** | 0.131 |
| 50–54 | −0.273 | *** | 0.055 | −1.251 | *** | 0.107 | −0.266 | *** | 0.058 | −1.002 | *** | 0.130 |
| 55–49 | −0.302 | *** | 0.054 | −1.376 | *** | 0.106 | −0.296 | *** | 0.057 | −1.114 | *** | 0.129 |
| 60–64 | −0.325 | *** | 0.054 | −1.467 | *** | 0.105 | −0.324 | *** | 0.057 | −1.198 | *** | 0.129 |
| 65–69 | −0.283 | *** | 0.054 | −1.583 | *** | 0.105 | −0.281 | *** | 0.057 | −1.293 | *** | 0.128 |
| 70–74 | −0.271 | *** | 0.054 | −1.753 | *** | 0.104 | −0.257 | *** | 0.057 | −1.412 | *** | 0.128 |
| 75–79 | −0.149 | *** | 0.054 | −1.914 | *** | 0.105 | −0.104 | * | 0.057 | −1.498 | *** | 0.128 |
| 80–84 | −0.148 | *** | 0.054 | −2.065 | *** | 0.104 | −0.072 | | 0.058 | −1.529 | *** | 0.128 |
| 85–89 | −0.135 | ** | 0.056 | −2.168 | *** | 0.104 | −0.033 | | 0.059 | −1.538 | *** | 0.128 |
| 90 + | −0.037 | | 0.061 | −2.265 | *** | 0.106 | 0.089 | | 0.065 | −1.563 | *** | 0.129 |
| Age group*time to death 6 years | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.154 | * | 0.087 | −0.898 | *** | 0.169 | 0.190 | ** | 0.090 | −0.688 | *** | 0.158 |
| 25–29 | 0.256 | *** | 0.084 | −0.967 | *** | 0.148 | 0.283 | *** | 0.088 | −0.715 | *** | 0.147 |
| 30–34 | 0.067 | | 0.072 | −0.843 | *** | 0.148 | 0.098 | | 0.076 | −0.664 | *** | 0.144 |
| 35–39 | −0.085 | | 0.063 | −0.791 | *** | 0.135 | −0.074 | | 0.066 | −0.633 | *** | 0.132 |
| 40–44 | −0.132 | ** | 0.057 | −0.881 | *** | 0.129 | −0.111 | * | 0.060 | −0.691 | *** | 0.129 |

**Table 6** (continued)

| | Model 8 | | | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | | GLM-Poisson | | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | | Log (Exp.)[b] | | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | | SE | Coef | | SE | Coef | | SE | Coef | | SE |
| 45–49 | −0.234 | *** | 0.054 | −0.868 | *** | 0.126 | −0.239 | *** | 0.057 | −0.703 | *** | 0.125 |
| 50–54 | −0.205 | *** | 0.053 | −1.087 | *** | 0.123 | −0.191 | *** | 0.056 | −0.933 | *** | 0.130 |
| 55–49 | −0.287 | *** | 0.052 | −1.177 | *** | 0.122 | −0.281 | *** | 0.055 | −0.990 | *** | 0.122 |
| 60–64 | −0.331 | *** | 0.051 | −1.292 | *** | 0.121 | −0.331 | *** | 0.054 | −1.091 | *** | 0.121 |
| 65–69 | −0.280 | *** | 0.051 | −1.433 | *** | 0.121 | −0.275 | *** | 0.054 | −1.201 | *** | 0.121 |
| 70–74 | −0.230 | *** | 0.051 | −1.577 | *** | 0.121 | −0.211 | *** | 0.054 | −1.305 | *** | 0.121 |
| 75–79 | −0.158 | *** | 0.051 | −1.722 | *** | 0.120 | −0.115 | ** | 0.054 | −1.363 | *** | 0.121 |
| 80–84 | −0.158 | *** | 0.052 | −1.860 | *** | 0.120 | −0.085 | | 0.055 | −1.396 | *** | 0.121 |
| 85–89 | −0.126 | ** | 0.053 | −1.955 | *** | 0.121 | −0.034 | | 0.056 | −1.420 | *** | 0.121 |
| 90+ | −0.022 | | 0.060 | −1.994 | *** | 0.122 | 0.091 | | 0.063 | −1.403 | *** | 0.122 |
| Age group*time to death 7 years | | | | | | | | | | | | |
| 1–19 (reference group) | | | | | | | | | | | | |
| 20–24 | 0.249 | *** | 0.087 | −0.688 | *** | 0.181 | 0.269 | *** | 0.092 | −0.473 | *** | 0.166 |
| 25–29 | 0.232 | *** | 0.079 | −0.722 | *** | 0.164 | 0.266 | *** | 0.083 | −0.555 | *** | 0.151 |
| 30–34 | 0.039 | | 0.069 | −0.773 | *** | 0.161 | 0.059 | | 0.073 | −0.656 | *** | 0.154 |
| 35–39 | −0.098 | | 0.060 | −0.765 | *** | 0.151 | −0.074 | | 0.063 | −0.649 | *** | 0.142 |
| 40–44 | −0.165 | *** | 0.055 | −0.808 | *** | 0.145 | −0.148 | ** | 0.057 | −0.712 | *** | 0.139 |
| 45–49 | −0.181 | *** | 0.052 | −0.834 | *** | 0.142 | −0.172 | *** | 0.055 | −0.728 | *** | 0.135 |
| 50–54 | −0.245 | *** | 0.051 | −1.097 | *** | 0.140 | −0.228 | **** | 0.054 | −0.971 | *** | 0.133 |
| 55–49 | −0.257 | *** | 0.050 | −1.191 | *** | 0.139 | −0.243 | *** | 0.053 | −1.077 | *** | 0.133 |
| 60–64 | −0.300 | *** | 0.049 | −1.278 | *** | 0.139 | −0.292 | *** | 0.052 | −1.145 | *** | 0.132 |
| 65–69 | −0.233 | *** | 0.049 | −1.409 | *** | 0.139 | −0.216 | *** | 0.052 | −1.260 | *** | 0.132 |
| 70–74 | −0.201 | *** | 0.050 | −1.554 | *** | 0.138 | −0.174 | *** | 0.052 | −1.342 | *** | 0.132 |
| 75–79 | −0.118 | ** | 0.050 | −1.668 | *** | 0.138 | −0.064 | | 0.053 | −1.389 | *** | 0.132 |
| 80–84 | −0.119 | ** | 0.050 | −1.803 | *** | 0.138 | −0.043 | | 0.053 | −1.437 | *** | 0.131 |
| 85–89 | −0.076 | | 0.052 | −1.887 | *** | 0.139 | 0.021 | | 0.055 | −1.448 | *** | 0.132 |
| 90+ | −0.045 | | 0.058 | −1.965 | *** | 0.140 | 0.062 | | 0.061 | −1.473 | *** | 0.133 |
| Chronic diseases | | | | | | | | | | | | |
| Chronic infection | | | | | | | 2.241 | *** | 0.110 | 0.831 | *** | 0.031 |
| Cancer | | | | | | | 2.362 | *** | 0.069 | 1.459 | *** | 0.004 |
| Benign hematologic diseases including anemia | | | | | | | 1.418 | *** | 0.047 | 0.621 | *** | 0.016 |
| Diseases of the thyroid gland | | | | | | | 1.942 | *** | 0.074 | 0.477 | *** | 0.009 |

**Table 6** (continued)

| | Model 8 | | | | Model 9 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Probit | | GLM-Poisson | | Probit | | | GLM-Poisson | | |
| | *P* (Exp. > 0)[a] | | Log (Exp.)[b] | | *P* (Exp. > 0) | | | Log (Exp.) | | |
| | Coef | SE | Coef | SE | Coef | | SE | Coef | | SE |
| Diabetes | | | | | 2.279 | *** | 0.059 | 0.487 | *** | 0.007 |
| Other endocrine diseases and malnutrition | | | | | 2.218 | *** | 0.096 | 0.701 | *** | 0.017 |
| Obesity | | | | | 2.077 | *** | 0.088 | 1.086 | *** | 0.011 |
| Neurological diseases | | | | | 2.340 | *** | 0.053 | 0.883 | *** | 0.006 |
| Paraplegia and hemiplegia | | | | | 2.149 | *** | 0.196 | 1.386 | *** | 0.029 |
| Eye diseases | | | | | 2.382 | *** | 0.101 | 0.427 | *** | 0.006 |
| Ear diseases | | | | | 2.159 | *** | 0.063 | 0.335 | *** | 0.007 |
| Hyptertension | | | | | 0.027 | | 0.134 | −0.235 | *** | 0.007 |
| Cardiovascular | | | | | 2.282 | *** | 0.044 | 1.083 | *** | 0.003 |
| Chronic pulmonary disease including asthma | | | | | 2.338 | *** | 0.054 | 0.835 | *** | 0.006 |
| Inflammatory bowel disease | | | | | 2.137 | *** | 0.056 | 0.828 | *** | 0.008 |
| Diseases of the liver | | | | | 1.975 | *** | 0.106 | 0.741 | *** | 0.015 |
| Diseases of the skin | | | | | 2.321 | *** | 0.151 | 0.672 | *** | 0.015 |
| Inflammatory rheumatic disease | | | | | 2.313 | *** | 0.072 | 0.880 | *** | 0.006 |
| Degenerative rheumatic diseases and osteoarthritis | | | | | 2.291 | *** | 0.030 | 0.881 | *** | 0.003 |
| Osteoporosis | | | | | 2.194 | *** | 0.144 | 0.420 | *** | 0.008 |
| Diseases of the kidney | | | | | 2.223 | *** | 0.109 | 1.052 | *** | 0.010 |
| Gynaecological diseases | | | | | | | | 0.932 | *** | 0.016 |
| Mental: dementia | | | | | 2.081 | *** | 0.258 | 0.315 | *** | 0.023 |
| Mental: substance abuse | | | | | 1.988 | *** | 0.095 | 0.881 | *** | 0.016 |
| Mental: schizophrenia | | | | | 1.797 | *** | 0.309 | 0.797 | *** | 0.068 |
| Mental: depression and anxiety | | | | | 1.690 | *** | 0.106 | 0.762 | *** | 0.026 |
| Mental: eating disorders | | | | | 1.811 | *** | 0.209 | 1.455 | *** | 0.095 |
| Mental: personality disorders | | | | | 1.289 | *** | 0.201 | 0.957 | *** | 0.133 |

[a]The probability that an individual receives some health care services in a given year

[b]The log-link function provides multiplicative effects and the results can therefore be expressed as percentage increases in the average health care expenditure per unit increase in any of the covariates by taking the exponential of the covariates

*Statistically significant at the 0.10 level

**At the 0.05 level

***At the 0.01 level

**Author contributions** All authors contributed to the study conception and design. Material preparation and data collection were performed by RI and JK. Analyses were performed by AK and RI. The first draft of the manuscript and the final manuscript were written by AK. All authors approved the final manuscript.

**Funding** The work was funded by a grant from the Novo Nordisk Foundation. Grant number NNF17OC0026466.

**Availability of data and material/code availability** We use proprietary data sets from Statistics Denmark. Access to the data and code can only be obtained with approval from Statistics Denmark—Division of Research Services.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial conflicts of interests to disclose.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

## References

1. Rouzet, D., Sánchez, A.C., Renault, T., Roehn, O.: Fiscal challenges and inclusive growth in ageing societies. In: *OECD Economic Policy Papers*, No. 27, OECD Publishing (2019). https://doi.org/10.1787/c553d8d2-en
2. World Health Organization, Global health and ageing, 2021 (2016)
3. Zweifel, P., Felder, S., Meiers, M.: Ageing of population and health care expenditure: a red herring? Health Econ. **8**, 485–496 (1999)
4. Arnberg, S., Bjørner, T.B.: Sundhedsudgifter og levetid. Natl Tidsskr **148**, 43–66 (2010)
5. Breyer, F., Lorenz, N., Niebel, T.: Health care expenditures and longevity: is there a Eubie Blake effect? Eur. J. Health Econ. **16**, 95–112 (2015)
6. Barer, M.L., Evans, R.G., Hertzman, C., Lomas, J.: Aging and health care utilization: new evidence on old fallacies. Soc. Sci. Med. **24**, 851–862 (1987)
7. Gregersen, F.A.: The impact of ageing on health care expenditures: a study of steepening. Eur. J. Health Econ. **15**, 979–989 (2014)
8. Kollerup, A., Kjellberg, J.: Flere ældre og nye behandlinger. VIVE - The Danish Center for Social Science Resarch (2018)
9. Gregersen, F.A., Godager, G.: The association between age and mortality related hospital expenditures: evidence from a complete national registry, Nordic J. Health Econ. **2**(1), (2014). https://doi.org/10.5617/njhe.656
10. Seshamani, M., Gray, A.M.: A longitudinal study of the effects of age and time to death on hospital costs. J. Health Econ. **23**, 217–235 (2004)
11. Seshamani, M., Gray, A.: Ageing and health-care expenditure: the red herring argument revisited. Health Econ. **13**, 303–314 (2004)
12. Stearns, S.C., Norton, E.C.: Time to include time to death? The future of health care expenditure predictions. Health Econ. **13**, 315–327 (2004)
13. Werblow, A., Felder, S., Zweifel, P.: Population ageing and health care expenditure: a school of 'red herrings'? Health Econ. **16**, 1109–1126 (2007)
14. Carreras, M., Ibern, P., Inoriza, J.M.: Ageing and healthcare expenditures: exploring the role of individual health status. Health Econ. **27**, 865–876 (2018)
15. Howdon, D., Rice, N.: Health care expenditures, age, proximity to death and morbidity: Implications for an ageing population. J. Health Econ. **57**, 60–74 (2018)
16. Buchner, F., Wasem, J.: "Steeping" of health expenditure profiles. Geneva Pap. Risk Insur.-Issues Pract. **31**, 581–599 (2006)
17. von Wyl, V.: Proximity to death and health care expenditure increase revisited: a 15-year panel analysis of elderly persons. Heal. Econ. Rev. **9**, 1–16 (2019)
18. Gray, A.: Population ageing and health care expenditure. Annu. Rev. Public Health. **2**, 15–20 (2004)
19. Fries, J.F.: The compression of morbidity, the Milbank memorial fund quarterly. Health Soc. **61**, 397–419 (1983)
20. Felder, S., Werblow, A.: Does the age profile of health care expenditure really steepen over time? New evidence from Swiss Cantons. Geneva Pap. Risk Insur.-Issues Pract. **33**, 710–727 (2008)
21. Lehnert, T., Heider, D., Leicht, H., Heinrich, S., Corrieri, S., Luppa, M., et al.: Health care utilization and costs of elderly persons with multiple chronic conditions. Med. Care Res. Rev. **68**, 387–420 (2011)
22. Eckardt, M., Brettschneider, C., van den Bussche, H., König, H.: MultiCare Study Group. Analysis of health care costs in elderly patients with multiple chronic conditions using a finite mixture of generalized linear models. Health Econ. **26**, 582–599 (2017)
23. The Danish Health Authority, Faktaanalyse—kræftområdet 2007–2014, (2016)
24. Breyer, F., Lorenz, N.: The "red herring" after 20 years: ageing and health care expenditures. Eur. J. Health Econ. **22**, 661–667 (2020)
25. OECD, Health at a Glance 2019. OECD Indicators, (2019)
26. Kjellberg, J., Ibsen, R.: Fremtidens sundhedsudgifter. VIVE - The Danish Center for Social Science Researh, (2020)
27. Blough, D.K., Ramsey, S.D.: Using generalized linear models to assess medical care costs. Health Serv. Outcomes Res. **1**, 185–202 (2000)
28. Manning, W.G., Mullahy, J.: Estimating log models: to transform or not to transform? J. Health Econ. **20**, 461–494 (2001)
29. Barber, J., Thompson, S.: Multiple regression of cost data: use of generalised linear models. J. Health Serv. Res. Policy. **9**, 197–204 (2004)
30. Melberg, H.O., Godager, G., Gregersen, F.A.: Hospital expenses towards the end of life. Tidsskr. Den Nor Legeforen **133**, 841 (2013)
31. Levinsky, N.G., Yu, W., Ash, A., Moskowitz, M., Gazelle, G., Saynina, O., et al.: Influence of age on Medicare expenditures and medical care in the last year of life. JAMA **286**, 1349–1355 (2001)
32. Nexo, M.A., Carlsen, K., Pedersen, J., Hetland, M.L., Watt, T., Hansen, S.M., et al.: Long-term sickness absence of 32 chronic conditions: a Danish register-based longitudinal study with up to 17 years of follow-up. BMJ Open **8**, e020874 (2018)
33. Eurostat, Major causes of death, EU-28, 2020 (2016)

# Chapter 2

# Worth the trip? The effect of hospital clinic closures for patients undergoing scheduled surgery

# Worth the trip? The effect of hospital clinic closures for patients undergoing scheduled surgery

Anna Kollerup Iversen

December 2021

**Abstract**

Recent decades have seen a large number of hospital closures and consolidations, which have been carried out to stimulate returns to volume and specialization in hospital care. In the non-acute setting of scheduled breast cancer surgery, I examine how hospital clinic closures affect cost-saving metrics and the quality of care that closure-affected patients receive. The effects are identified using closures of breast cancer clinics in Denmark from 2000 to 2011, during which time the number of clinics was more than halved. Using event study designs, I examine changes in surgical outcomes for patients living in municipalities where the nearest clinic had been closed. The results show that breast cancer clinic closures have been welfare-improving, as they have reduced the number of costly hospitalization days and shifted surgical procedures to state-of-the-art breast-conserving techniques without generating adverse health effects and without causing crowding in non-closing clinics. An examination of the mechanisms suggests that added volume returns at non-closing clinics were of less importance than simply reallocating patients to higher-quality clinics.

# 1 Introduction

The supply of hospital services has become increasingly consolidated to fewer and larger hospitals as a result of clinic closures and consolidations in many OECD countries. While policy recommendations on closures and consolidations have been rationalized with reference to arguments of efficiency and quality returns to scale and scope, the resulting decrease in reduced access to nearby hospital care meet public outcry from patients living in areas affected by such closures (Christiansen and Vrangbæk 2018). This is especially the case if the accompanying increases in travel time are not offset by a higher quality of care.

Previous research on consolidation and access to acute care has established an undeniable trade-off between increases in patients' travel time to acute care and an ensuing lower probability of survival (Buchmueller et al. 2006; Avdic 2016; Bertoli and Grembi 2017; Carroll 2019). However, in the case of consolidation of hospital clinics that perform less complex and non-acute procedures, we lack knowledge about the trade-off between reduced access to nearby care and the quality of care that patients receive. One study from 2003 conjectures that there may be low returns to scale when consolidating low-complexity procedures (Shahian and Normand 2003) while a newer study of maternity ward closures find positive quality effects on infant health but negative effects on maternal health because of ward overcrowding (Avdic, Lundborg, et al. 2020). Despite limited evidence, clinics performing low-complexity and non-acute procedures have been closed and consolidated, and national health authorities propose that non-acute and relatively low-complexity care to be consolidated further, for example in France (French National Health Insurance 2018; Huguet 2020).

In a non-acute setting with scheduled breast cancer surgery, I analyze the effects of hospital clinic closures on the quality of care that closure-affected patients receive. I also analyze the effects on cost-saving metrics that stakeholders might use when deciding where hospital services should be delivered, and I explore what is driving the results. First, I investigate whether non-closing clinics[1] are affected by the patient inflow from nearby closures, and, in turn, whether non-closing clinics experience an increased potential of benefiting from scale effects or whether they might be exposed to crowding. Second, I explore whether the closing and non-closing clinics differ significantly in the years up to closure, and, in turn, whether a simple reallocation of patients to higher-quality clinics drives the effects.

Over the observation period 1996 to 2014, I follow municipalities where the nearest breast cancer clinic[2] closes between 2000 and 2011. From each closure-affected municipality, I extract detailed individual level data from Danish administrative registers on patients who have under-

[1] I define 'non-closing' clinics as the clinics that have been kept in service following a consolidation programme and closure of other clinics.

[2] I define 'nearest breast cancer clinic' as the clinic to which a municipality on average has the shortest travel time to.

gone breast cancer surgery from 1996 to 2014. The individual level data include information on the date and type of surgical procedure and other treatments, date of discharge, admission type, patient demographics and comorbidities as well as drug prescriptions. Based on the individual level data, I explore outcomes that capture both the quality of care and cost-saving metrics: hospitalization days, outpatient treatment days, mortality rates, opioid prescriptions, surgical procedure and the extent of re-surgery.

To identify the effects, I follow closure-affected municipalities and exploit the variation in the timing of the clinic closures. In an event-study design, I estimate within-municipality changes in event-time relative to a linear pre-trend in the outcomes. Conditional on closing a clinic in the observation period and on the included control variables, the identifying assumption is that the timing of a closure is uncorrelated with deviations in the outcomes from a linear trend in event time. Also, there should be no unobserved shock that occurs both in the same calendar year as the clinic closure and is also correlated with the outcomes.

The results indicate that breast cancer clinic closures have been welfare-improving, as they have reduced the number of costly hospitalization days and shifted surgical procedures to state-of-the-art breast-conserving techniques without generating adverse health effects on mortality rates or pain levels (use of opioids) and without causing crowding in non-closing clinics. The results show that the number of hospitalization days for patients living in closure affected municipalities on average decreases from five to four days, while the number of outpatient treatment days is unaffected. In the full observation period, breast-conserving surgery was the state-of-the-art surgery as opposed to mastectomies where the whole breast is removed. The results indicate a 7-10 pct. points decrease in the share of mastectomies (down from 62-65 pct.). In turn, more patients in closure-affected municipalities undergo a breast-conserving surgery following closure of their nearest clinic. This finding may reassure patients in closure-affected areas that closures can help to ensure that patients will receive state-of-the-art care in higher-quality hospital clinics, and thus counterbalance patients' concerns that closures of their local clinics might only be carried out for cost-saving reasons. Mechanically, for patients living in a closure-affected municipality, closures imply that their travel time on average increases by 22 minutes.

I provide evidence for two mechanisms that can explain the results: 1) Breast cancer incidence increases over time and the closing clinics are very small compared to non-closing clinics. In turn, the patient flow from closing to non-closing clinics is not large enough to generate a significant increase in patient volume in non-closing clinics. 2) Low-quality clinics are the ones to close. Together, the mechanisms suggest that the results are driven by reallocating patients to higher-quality clinics and that there is no potential for added economics of scale and scope in non-closing clinics, nor a risk of congestion or crowding in non-closing clinics.

The findings in this paper contribute to the existing literature by examining a non-acute

setting of hospital clinic closures, and by investigating utilization of surgical procedure and opioid prescriptions as outcomes in addition to more traditional outcomes, such as hospitalization days and mortality rates. The paper adds to the strand of literature that examines the effects of hospital clinic closures among patients living in closure-affected areas. Most of the existing literature examines high-risk, highly complex and acute procedures where travel time is essential for survival (e.g., road accidents and strokes) (Buchmueller et al. 2006; Avdic 2016; Bertoli and Grembi 2017; Carroll 2019). While, in the context of acute care, these papers find negative health effects among patients directly affected by hospital clinic closures, I, in the context of non-acute care, find no negative health effects among patients affected by clinic closures. Furthermore, I find positive effects on health care utilization (decrease in hospitalization days) and a shift towards patients receiving state-of-the-art surgical procedures.

The findings of this paper also relate to the literature on the returns to hospital size. In this literature, it is broadly established that larger hospitals typically perform better than smaller hospitals (Chowdhury et al. 2007; Halm et al. 2002; Birkmeyer et al. 2002), and evidence suggest that closures and consolidations may generate added potential for economies of scale and scope at non-closing clinics, for example, through learning-by-doing from physicians' own experience or through knowledge spillovers across physician specialities (Luft et al. 1979; Hamilton and Ho 1998; Gaynor, Seider, et al. 2004; Huesch and Sakakibara 2009; Avdic, Lundborg, et al. 2019). Previous papers have studied the effects of patient inflow from closing clinics to non-closing clinics to explore scale and scope returns to volume and adverse effects in terms of crowding (Gaynor, Seider, et al. 2004; Gaynor, Seider, et al. 2005; Hentschker and Mennicken 2018; Avdic, Lundborg, et al. 2019; Avdic, Lundborg, et al. 2020). While recent papers find both a potential for returns to volume (Avdic, Lundborg, et al. 2019) and a risk of crowding (Avdic, Lundborg, et al. 2020) in non-closing clinics, this paper find no potential for added returns to volume in non-closing clinics, and there is no risk of congestion or crowding. The reasoning behind these conclusions is that closing clinics are very small compared to non-closing clinics and the number of breast cancer surgeries varies a lot over time due to changes in screening patterns and a general increase in breast cancer incidence.

# 2 Background: Breast cancer clinic closures

The Danish hospital sector primarily consists of tax-financed publicly owned hospitals; only 3 pct. of the Danish hospital activity are delivered by private providers (The Commonwealth Fund 2020). Therefore, only public hospitals are included in the analyses in this paper. The Danish hospital sector is generally a decentralised system in which five regions (and before 2007, 16 counties) are responsible for financing, priority setting and planning of the hospital sector. However, for specialised care, for example, most of the cancer care and acute care, deciding on the location of the hospital care is planned by the the Danish National Board of Health. The regional governments are asked to submit hospital service plans on where services ought to be delivered, while the Danish National Board of Health is responsible for determining the organizational structure and which hospitals can deliver which services (OECD 2013; Christiansen and Vrangbæk 2018). Thus, the planning is carried out in a collaboration between the national and regional governments, which also draws on input from expert committees, including hospital administrators and health care professionals (OECD 2013; Christiansen and Vrangbæk 2018). As a result of this planning, Denmark has seen a substantial number of clinic closures in the hospital sector, and only 25 pct. of the Danish breast cancer clinics remain open over the period examined in this study (1996 to 2014). From 1996 to 2014, 48 clinics have provided breast cancer surgery at some time during the study period. Among these clinics, 32 clinics have closed, 4 clinics have opened and closed again, and 12 clinics have remained open during the whole period from 1996 to 2014. In the present study, I investigate the effects of breast cancer clinic closures where it has been possible to follow a clinic four years before a closure, and where the catchment area (municipality) for a closing clinic has not been directly affected by other closures or openings within the event window of four years before and after a closure. Table 1 shows that 17 clinics met these criteria, and these 17 clinic closures resulted in that 29 municipalities lost their nearest clinic. Empirically, I use the 17 clinic closures as the exogenous variation in the clinics that patients nested within municipalities are admitted to, and follow patients living in the 29 closure-affected municipalities over time. The timing of the 17 closures is presented in figure 1 with the surgery volume in the year before the closure year shown on the y-axis. Figure 2 shows the closure affected municipalities and the breast cancer clinics defined as closing and non-closing, as well as the clinics with opened or had a short event-window. A full overview of how the breast cancer clinics are categorised into closing and non-closing, and how each municipality is categorised as closure-affected, closure unaffected and uncertain treatment exposure is presented in Appendix A.

Table 1: Distribution of municipalities and clinics

| Treatment exposure | Closure-affected (nearest clinic closes) | Unaffected (nearest clinic does not close) | Uncertain (Short event-window or openings) |
|---|---|---|---|
| Municipalities | 29 | 18 | 51 |
| Clinics | 17 | 12 | 19 |

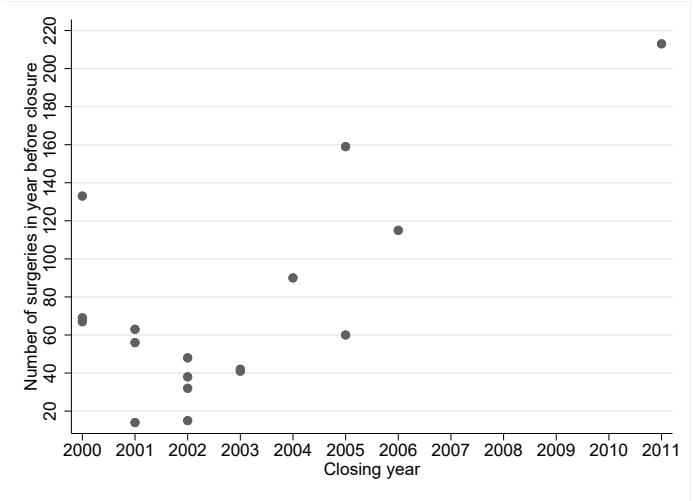



Figure 1: Timing and volume of each clinic closure

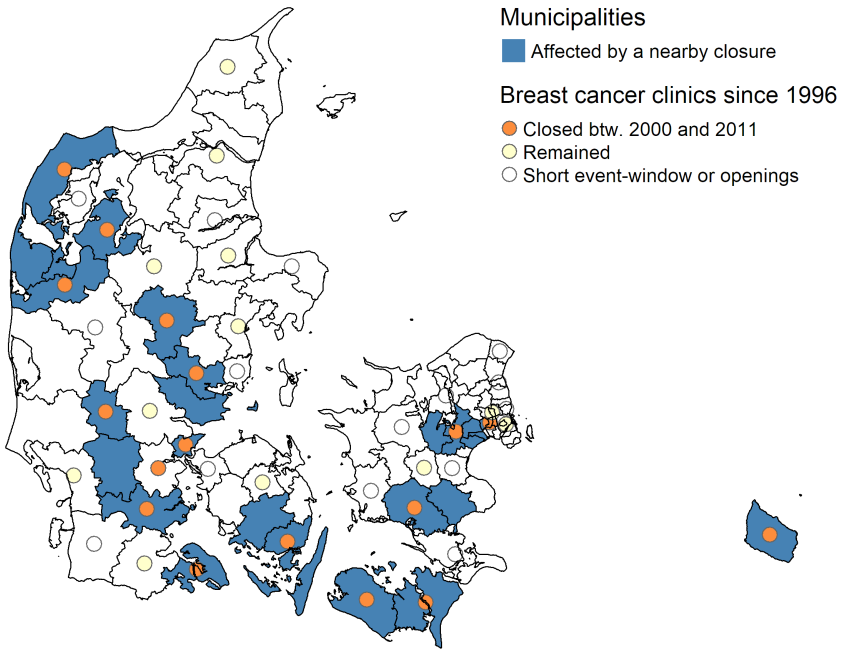



Figure 2: Municipalities affected by clinic closures

# 3 Data and Summary Statistics

For each of the 29 closure-affected municipalities, the full Danish population of women undergoing breast cancer surgery in the period from 1996 to 2014 is included in the analyses. I extract detailed individual level data from nation-wide administrative registers provided by Statistics Denmark, the Danish Health Data Authority (The Danish national patient registry) and the Danish Medicines Agency (The Danish national prescription registry). The individual level data include information on the date and type of surgical procedure and other treatments, date of discharge, admission type, patient demographics and comorbidities as well as drug prescriptions.

The location and the number of clinics performing breast cancer surgery are gathered from a combination of yearly hospital department classification books and the Danish Address Register[3]. From these sources it has been possible to geographically pin the clinics to parish level. The clinic location data are merged to the individual level data by clinic identification numbers. To identify the municipalities that are affected by clinic closures, I estimate the average travel time from each municipality to all breast cancer clinics, and rank the travel time from lowest to highest. For each municipality, the clinic-municipality match with the lowest travel time is kept, and if the clinic closes over the observation period, the municipality is categorized as closure-affected.[4] Travel time is calculated from parish centroid to parish centroid[5], and parishes are nested within municipalities. To estimate the average travel time between a given municipality and each of the clinics in Denmark, I take the average over the distance between the parishes within a given municipality and each of the clinics.

The sample consists of women diagnosed with breast cancer who undergo breast cancer surgery and who live in a municipality where their nearest breast cancer clinic closes during the examination period. Sample characteristics are shown in table 2[6]. The sample is not restricted by previous diagnoses or first-time diagnoses, but I control for variation in patient health 1-5 years prior to the surgery year by including the Charlson Comorbidity Index as control variables, which takes into account the number and the seriousness of patients' comorbid diseases (Charlson et al. 1987)[7]. I also account for patient heterogeneity by including covariates

[3]Sources: The Danish Health Authority (1999-2004), The Danish Health Authority (2004-2013), and Datafordeler (2018). I thank Kasper Wennervaldt, Line Flytkjær Virgilsen and Søren Neermark for valuable guidance in my work on identifying closed clinics. All errors are my own.

[4]A few other measures have been taken to ensure that a municipality that is assigned as closure-affected is not affected by several nearby closures or openings of clinics. These measures are stated in appendix A.3.

[5]The data set on travel time between Danish parishes was prepared by Hviid and Kristensen (2021), where distances and travel times between parishes were calculated using Google API. The Google API returns information based on the recommended driving route as calculated by the Google Maps API.

[6]A patient can be represented in the sample more than once, which is controlled for in the regressions by adding a dummy for whether the patient has gone through breast cancer surgery in the 365 days before a given surgery. As seen in table 2, this is the case for only 5 pct. of the sample.

[7]The Charlson Comorbidity Index is a weighted count of the presence and seriousness of 17 diseases. The

that capture demography (age, foreign born, marital status) and educational levels. The average age in the sample is 61.4 years old and the majority of the sample are married (60 pct.), and have a lower or upper secondary education (80 pct.). 3 pct. of the sample is foreign born. In comparison to patients living in municipalities that are unaffected by clinic closures, the sample of patients living in closure-affected municipalities are married to a higher extent (3 pct. points difference) and have a slightly lower educational level.

In the main analyses, I include six outcome variables. First of all, I include the number of hospitalization days following surgery[8] as Denmark has had a general goal of lowering the number of inpatient days and increasing the number of outpatient visits by accelerating discharge after surgical treatment. This goal was implemented to secure an effective delivery of hospital care and because patient recovery can benefit from patients being discharged to their own homes as soon as possible after surgery (Kambouris 1996). Hence, fewer hospitalization days should have welfare-improving effects due to cost reductions gained by shifting to more outpatient care and by ensuring increased (or at least unchanged) health benefits by allowing patients to recover in their own homes (McManus et al. 1994; Margolese and Lasry 2000; Marla and Stallard 2009). To explore whether there are some opposite effects due to more frequent outpatient visits, I include the number of days that the patient receives outpatient care in the first 30 days after surgery. The average number of hospitalization days is 5.98 days in the year before a closure and the number of outpatient treatment days is 0.95.

I also include the type of surgical procedure as a measure of surgical performance. Breast cancer surgery involves the physical removal of the tumour and some of the surrounding tissue. Standard surgerical procedures include mastectomy and lumpectomy; the whole breast is removed in a mastectomy, and a smaller part of the breast is removed in a lumpectomy (breast-conserving). As a breast-conserving lumpectomy is a less invasive procedure and as there is no difference in overall survival between the breast-conserving lumpectomy and a mastectomy (Fisher et al. 2002), guidelines recommend that breast-conserving lumpectomies should be per-

diseases are weighted from 0 to 6, with 6 being the most serious. For example, diabetes without chronic complications are weighted 0, while diabetes with organ damage is weighted 1, AIDS/HIV is weighted 4 and metastatic solid tumour is weighted 6 (Quan, Li, et al. 2011). For each individual, the weights are summed and the index is generated from 0 to 2. All individuals with a weight above 1 is assigned an index value of 2. To construct the index, I use the ICD-10 coding algorithm proposed by Quan, Sundararajan, et al. (2005) and corresponding STATA command written by Stagg (2006), where I update the weighting in accordance with Quan, Li, et al. (2011).

[8]Hospitalization days are counted as the number of continual inpatient days from the day the patient is admitted to the hospital for surgery to the day that the patient leaves the hospital again. If a hospitalization starts before the surgery takes place, these days are included in the count. Only inpatient stays for breast cancer diagnoses are included, such that inpatient stays for other diseases and treatments do not influence the count. If the patient is transferred to another department or hospital with no more than four hours since the last hospitalization, the days are added together as a continual inpatient stay. For patients undergoing surgery as an outpatient visit, the number of days is set to 0, unless the outpatient visit is followed by an inpatient stay within two days after surgery. If the latter is the case, the length of the inpatient stay is used as the number of hospitalization days.

formed whenever possible across the whole observation period (Wolters et al. 2012; Keelan et al. 2021). However, performing a breast-conserving surgery compared to a mastectomy requires surgical training which may explain why some surgeons do not perform breast-conserving surgery whenever possible (Wyld et al. 2019). In the year before a closure, on average 70 pct. of the performed surgeries in closure-affected municipalities were mastectomies, while 30 pct. were breast-conserving lumpectomies.

A shift in surgical procedure or a shift to another breast clinic clinic may also have an impact on the number of patients undergoing re-surgery, and therefore I also include re-surgery as an outcome. Re-surgery is defined as an occurrence of breast surgery after primary breast surgery due to complications within 30 days after primary breast surgery (Larsen and Schiøler 2005). The average share undergoing re-surgery in the year before closure is 4 pct. To examine the well-being of the patients, I include prescription of opioids to capture the level of pain that patients experience after surgery. This share is 7 pct. in the first 30 days after surgery when measured in the year before a closure. Finally, the two-year mortality rate is included to measure the direct health effect. The two-year mortality rate in the year before closure is 8 pct. in this study.

Breast cancer clinic closures may also imply changes in the use of neoadjuvant and adjuvant treatments, for example, radiation therapy and chemotherapy. Neoadjuvant treatment takes place before surgery and might help in downstaging the cancer, and thereby reduce the extent of surgery such that breast-conserving surgery, rather than a mastectomy, becomes possible (Thompson and Moulder-Thompson 2012). Adjuvant treatment takes place after surgery to improve overall survival by lowering the risk of relapse (Chew 2001). In the present study, adjuvant and neoadjuvant treatment is aggregated as very few patients receive neoadjuvant treatment in the years leading up to closure. Hence, I examine whether the patients receive radiation therapy or chemotherapy in a period of six months before their primary breast cancer surgery to 12 months after their primary breast cancer surgery. Unfortunately, it has only been possible to follow radiation therapy and chemotherapy consistently over time from 2001 (and not 1996 as with the other outcomes), and therefore I only include the results and summary statistics on these outcomes in appendix C.

Table 2: Summary Statistics

| Patient characteristics | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Nearest clinic closes (treatment) | | No closures | | | | |
| | Mean | SD | Mean | SD | Difference[a] | Min | Max |
| Previous surgery[b] | 0.05 | 0.23 | 0.06 | 0.24 | 0.01* | 0 | 1 |
| Surgery year | 2004 | 3.94 | 2005 | 5.39 | 1.62*** | 1996 | 2014 |
| *Charlson comorbidity index*[c] | | | | | | | |
| 0 | 0.83 | 0.38 | 0.83 | 0.37 | 0.01 | 0 | 1 |
| 1 | 0.04 | 0.20 | 0.04 | 0.20 | 0.00 | 0 | 1 |
| 2 | 0.13 | 0.34 | 0.13 | 0.33 | 0.00 | 0 | 1 |
| Age | 61.38 | 12.45 | 61.41 | 12.63 | 0.03 | 21 | 100 |
| Foreign born | 0.03 | 0.18 | 0.04 | 0.19 | 0.00 | 0 | 1 |
| Married | 0.60 | 0.49 | 0.57 | 0.50 | -0.03*** | 0 | 1 |
| Single | 0.34 | 0.47 | 0.37 | 0.48 | 0.03*** | 0 | 1 |
| Cohabiting | 0.06 | 0.24 | 0.06 | 0.24 | 0.00 | 0 | 1 |
| *Education* | | | | | | | |
| Lower secondary | 0.41 | 0.49 | 0.39 | 0.49 | -0.03*** | 0 | 1 |
| Upper secondary and vocational | 0.39 | 0.49 | 0.38 | 0.49 | -0.01* | 0 | 1 |
| Short and middle further | 0.17 | 0.38 | 0.19 | 0.40 | 0.02*** | 0 | 1 |
| Long further | 0.03 | 0.16 | 0.04 | 0.20 | 0.01*** | 0 | 1 |
| Number of patients | 9790 | | 19194 | | | | |
| Number of municipalities | 29 | | 18 | | | | |

| Outcomes in the year before a closure, $t = -1$ | | | | |
|---|---|---|---|---|
| | Nearest clinic closes | | | |
| | Mean | SD | Min | Max |
| Hospitalization day[d] | 5.98 | 3.09 | 0 | 21 |
| Outpatient treatment days[e] | 0.95 | 1.61 | 0 | 15 |
| Had a mastectomy | 0.70 | 0.46 | 0 | 1 |
| Had a re-surgery within 30 days | 0.04 | 0.19 | 0 | 1 |
| Prescribed opioids within 30 days | 0.07 | 0.26 | 0 | 1 |
| 2-year mortality rate | 0.08 | 0.27 | 0 | 1 |
| Travel time in minutes[f] | 21.39 | 28.46 | 0 | 262 |
| Number of observations | 846 | | | |
| Number of municipalities | 29 | | | |

[a] Significance determined with t-test for two independent samples
[b] Patient underwent another breast cancer surgery in the prior 365 days
[c] Measured in the 5 years prior to surgery year
[d] Continual hospitalization days post surgery
[e] Outpatient treatment days within 30 days after surgery
[f] From hospital parish to residence parish

# 4 Empirical Framework: Event study

I exploit the differential timing of the clinic closures over the period 2000-2011, where I for each municipality index all years relative to the event of experiencing that the nearest clinic closes. This enables me to estimate intention-to-treat (ITT) effects in relative time to the event of a clinic closure, where I assume that the majority of the patients living in a closure-affected municipality are exposed to the effects of the closure. In section 7.1, I show a sharp timing of the closure event where patient travel time increases in the year following a clinic closure. The compliance to treatment in the nearest breast cancer clinic is discussed in detail in section 6.

I estimate both non-parametric and parametric event study specifications where the choice of functional form in the parametric specification is guided by the results of the non-parametric specification. The non-parametric event studies reveal that most of the outcomes trend linearly over time, for example, due to progress in surgical competence as surgeons learn to perform less-invasive surgical procedures and because of the ongoing goals of lowering the number of inpatient days through accelerated discharge after surgical treatment. Hence, based on the visual inspection in the non-parametric specifications, it seems reasonable to introduce a linear pre-trend in the parametric specifications and estimate the effects in the post period as deviations from a linear pre-trend.

When examining the effects of clinic closures, one needs to deal with patients sorting into high-quality clinics, and potentially bad control groups of unaffected patients or non-closed clinics. First, I tackle patient sorting into high-quality clinics by following the same geographical units over time, i.e. patients living in closure-affected municipalities. Second, attempts to make comparisons with a control group will often suffer from selection problems as the patients, municipalities or clinics differ in unobserved ways that affect outcomes of interest (Dafny 2009), for example, if the clinics that are to be closed exhibit decreasing performance in the years up to closure, or if one compares patients from closure-affected rural areas with unaffected patients in urban areas. By estimating within-municipality changes in event time relative to a linear pre-trend, I overcome the empirical challenge of relying on an unaffected control group to take out linear trends within the unit of observation (e.g., municipalities or hospital clinics) (Dobkin et al. 2018; Borusyak and Jaravel 2017).

## 4.1 Non-parametric event study

The non-parametric event study specification consists of indicator variables that capture time $k$ since the year before a closure in municipality $m$ at time $t$ for all closure-affected municipalities.

The outcome, $y_{imt}$, that patient $i$ living in municipality $m$ faces at time $t$ is thus given by

$$y_{imt} = \sum_{k=-5,\ k\neq -1}^{5} \beta_k 1(\textit{Time since closure}{=}k)_{mt} + X_{it}\delta + \lambda_m + \lambda_t + \epsilon_{imt} \quad (1)$$

where $\beta_k$ represents the key coefficients of interest and measures the change in the outcome for patients living in closure-affected municipalities relative to being treated one year before their nearest breast cancer clinic closes in the same municipality ($k = -1$). The vector $X_{it}$ represents individual-level health and sociodemographic covariates for patient $i$ treated in calendar year $t$. Calendar year fixed effects are denoted by $\lambda_t$ that captures calendar year changes in the outcome common to all municipalities. Municipality fixed effects are denoted by $\lambda_m$ that captures time-invariant differences across municipalities and therefore the key coefficients, $\beta_k$, are identified using only within-municipality changes in event time. Due to colinearity between municipality fixed effects, calendar year and event time, an additional zero normalization is required and therefore two calendar year fixed effects are omitted in the estimation of model (1).

### 4.1.1 Identification and interpretation

Event time $k = -1$ is the year before a closure is registered, and is chosen as the reference event time over $k = 0$, as a closure can happen any time in event time 0. Hence, the effects of a closure may enter in event time 0, if the closure takes place in the beginning of event time 0. If a closure takes place late in event time 0, then the first effects are more likely to be observed in event time 1. Therefore, immediate effects are potentially observed in both year 0 and year 1.

To interpret the coefficients in $\beta_k$ for $k > -1$ as the causal effects of breast cancer clinic closures, the identifying assumption is that, conditional on closing a clinic in the observation period and the included control variables, the timing of closure is uncorrelated with the surgical outcomes for the patients. Hence, there should be no unobserved shock that both occurs in the same calendar year as the clinic closure and which is also correlated with the outcomes. There could be reasons a priori to be concerned about changes in the outcomes in the years prior to a closure, and therefore it seems plausible that the timing of clinic closures is not completely unanticipated. For example, clinics that have shown low performance may be the clinics to be closed. Hence, a closure that is preceded by changes in outcomes in the years leading up to a closure would potentially violate the identifying assumption. To examine the risk of endogenous timing, I plot differences in outcomes between closing and non-closing clinics in the four years leading up to closure in section 7.3. The results show that clinics to be closed perform worse in the four years leading up to a closure, but the differences between closing and non-closing clinics are consistent in the four years leading up to closure, and there is no large increase or

decrease in the difference when approaching the closure year. Therefore, there is no strong evidence of endogenous timing.

## 4.2 Parametric event study

Another threat to identification is that some outcomes are linearly trending over time. In the non-parametric specification, I include both calendar year fixed effects and municipality fixed effects, and therefore I will not be able to disentangle an underlying *linear* calendar time trend from dynamic treatment effects within a municipality (Borusyak and Jaravel 2017). To ease the interpretation of the dynamic effects in case of linear trends in the outcomes, I follow Dobkin et al. (2018) and restrict the pre-trend to be linear and estimate deviations from the linear pre-trend in the post period. In the parametric specification, the outcome, $y_{imt}$, that patient $i$ living in municipality $m$ faces at time $t$ becomes

$$y_{imt} = \gamma k + \sum_{k=0}^{5} \beta_k 1(\textit{Time since closure}{=}k)_{mt} + X_{it}\delta + \lambda_m + \lambda_t + \epsilon_{imt} \tag{2}$$

where $\gamma k$ is a linear pre-trend in event time. The key coefficients of interest, $\beta_k$, now measure the change in the outcome following a breast cancer clinic closure relative to any preexisting linear trend $\gamma k$, while $X_{it}$, $\lambda_m$ and $\lambda_t$ are the same as in the non-parametric specification. With this specification, I assume that if the municipalities were not hit by a clinic closure, the outcomes would have evolved as the linear pre-trend in the post period. Clearly, it is a simplification to restrict the pre-trend to be linear, but in the presence of linearly trending outcomes in the pre-period the parametric specification allows for more conservative estimates.

### 4.2.1 Identification and interpretation

Compared to the non-parametric specification, the identifying assumption in the parametric specification is weaker and requires that, conditional on closing a clinic in the observation period and the included control variables, the timing of a closure is uncorrelated with deviations in the outcome from a linear trend in event time. Also, there should be no unobserved shock that occurs both in the same calendar year as the clinic closure and is also correlated with the outcome.

The choice of specifying a parametric model is guided by: 1) a visual inspection of the results in the non-parametric specification, and 2) a choice of not excluding municipality fixed effects as a solution to identify the linear trend, as this most certainly would result in omitted variable bias. Besides specifying a parametric event study, Borusyak and Jaravel (2017) suggest excluding unit fixed effects or adding a control group to pin down the year effects in a non-parametric event study. The first solution requires a stronger identifying assumption, that is

that there are no time-invariant municipality level covariates that are correlated with closing a nearby breast cancer clinic and the outcomes measured within a given municipality. The latter solution requires that it is plausible to assume that the calendar year effects are the same in municipalities that experience a nearby closure and in municipalities that do not experience a nearby closure. In the robustness section, I include a model that omits the municipality fixed effects and a model that includes unaffected municipalities as a control group to examine how the results change when imposing other identifying assumptions.

# 5 Results

## 5.1 Main results

To visually assess the parametric assumption of a linear pre-trend in the outcomes, figure 3 presents the event time coefficients estimated in the non-parametric event studies and the linear pre-trends estimated in the parametric event studies. In the figure, the linear pre-trends are normalized to the level of the non-parametric coefficients from period $k = -5$ to $k = -1$, and the vertical distance between the linear pre-trend and the non-parametric coefficients for $k > -1$ constitutes the magnitude of the coefficients estimated in the parametric specification. Table 3 summarizes the magnitude and significance of these deviations from the linear pre-trend.

When examining the non-parametric results in figure 3, it is clear that an unidentified linear trend is present in most outcomes, and in table 3 we can see that the coefficients for the pre-trends are significant for three out of six outcomes. This justifies estimating the effects as the deviations from a linear pre-trend in the outcomes in a parametric event study specification. For the three outcomes with an insignificant pre-trend, we still summarize the effects relative to a linear pre-trend in table 3 as the non-parametric and parametric specification yield very similar results.

The closures take place during year 0, and therefore the effects are interpreted from year 1 and onwards. The coefficients from the parametric specification in table 3 show a decrease in hospitalization days at 1.18 days in the year following a closure and 1.36 days in the second year, suggesting that the patients are hospitalized on average one day less (down from around five days). There is no opposite effect on the number of outpatient treatment days, suggesting there is no important shift in resources or treatment need from inpatient to outpatient care. Moreover, there is a strong indication that the share of mastectomies decreases and the share of breast-conserving surgery increases following closures. In the first two years after a closure, the share of mastectomies decreases by 7-10 pct. points on average (down from 62-65 pct.). There is no impact on the share undergoing re-surgery, the two-year mortality rate or the use of opioids. This suggests that there are no adverse health effects of being reallocated to another hospital. The results show that breast cancer clinic closures had welfare-improving effects in terms of reducing the number of costly hospitalization days and shifting surgical procedures to state-of-the-art breast-conserving techniques without generating adverse health effects on mortality rates, the extent of re-surgery or pain levels (use of opioids).

Table 3: Results from parametric event study specification by outcome and event time. *Interpretation: Deviation from linear pre-trend in the outcome*

| | Hospitalisation days | Outpatient treatment days | Mastectomy share | Re-surgery share | Two-year mortality rate | Prescribed opioids share |
|---|---|---|---|---|---|---|
| Pre-trend | -0.34*** | 0.20*** | -0.03*** | -0.00 | -0.01 | 0.00 |
| | [0.10] | [0.05] | [0.01] | [0.00] | [0.00] | [0.00] |
| Year 0 | -0.58*** | 0.01 | -0.04 | 0.01 | -0.01 | -0.00 |
| | [0.20] | [0.11] | [0.03] | [0.01] | [0.02] | [0.01] |
| Year 1 | -1.18*** | -0.06 | -0.07 | 0.00 | -0.00 | 0.00 |
| | [0.42] | [0.17] | [0.04] | [0.02] | [0.02] | [0.01] |
| Year 2 | -1.36** | 0.11 | -0.10** | -0.00 | -0.00 | -0.02 |
| | [0.50] | [0.29] | [0.05] | [0.02] | [0.02] | [0.02] |
| Year 3 | -1.37** | 0.29 | -0.09* | 0.02 | -0.01 | -0.01 |
| | [0.55] | [0.39] | [0.05] | [0.02] | [0.02] | [0.02] |
| Year 4 | -1.62** | 0.58 | -0.13* | 0.01 | -0.00 | 0.00 |
| | [0.66] | [0.48] | [0.07] | [0.02] | [0.02] | [0.02] |
| Year 5 | -1.76** | 0.42 | -0.15* | -0.01 | -0.00 | 0.02 |
| | [0.80] | [0.55] | [0.08] | [0.03] | [0.03] | [0.03] |
| Pre-closure mean $t-1$ | 5.98 | 0.95 | 0.70 | 0.04 | 0.08 | 0.07 |
| Mean in $t+1^{+}$ | 5.30 | 1.35 | 0.65 | 0.03 | 0.07 | 0.07 |
| Mean in $t+2^{+}$ | 4.97 | 1.54 | 0.62 | 0.03 | 0.06 | 0.07 |

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

Number of observations: 9790. Number of municipalities: 29.

All models include municipality fixed effects, calendar year fixed effects and patient covariates. Standard errors are clustered on the municipality level and presented in brackets. Extended regression table with patient covariates is presented in appendix B.

$^{+}$Means in $t+1$ and $t+2$ are the expected means where I assume that if the municipalities were not hit by a clinic closure, the outcomes would have evolved as the linear pre-trend in the post period. Hence, the mean in $t+1$ for the number of hospitalzation days is given by $\textit{pre-closure mean}_{t-1} + 2 * \textit{pretrend} = 5.98 + 2 * (-0.34) = 5.30$. The expected mean levels in $t+1$ and $t+2$ are the relevant levels to interpret the estimated coefficients relative to in $t+1$ and $t+2$.

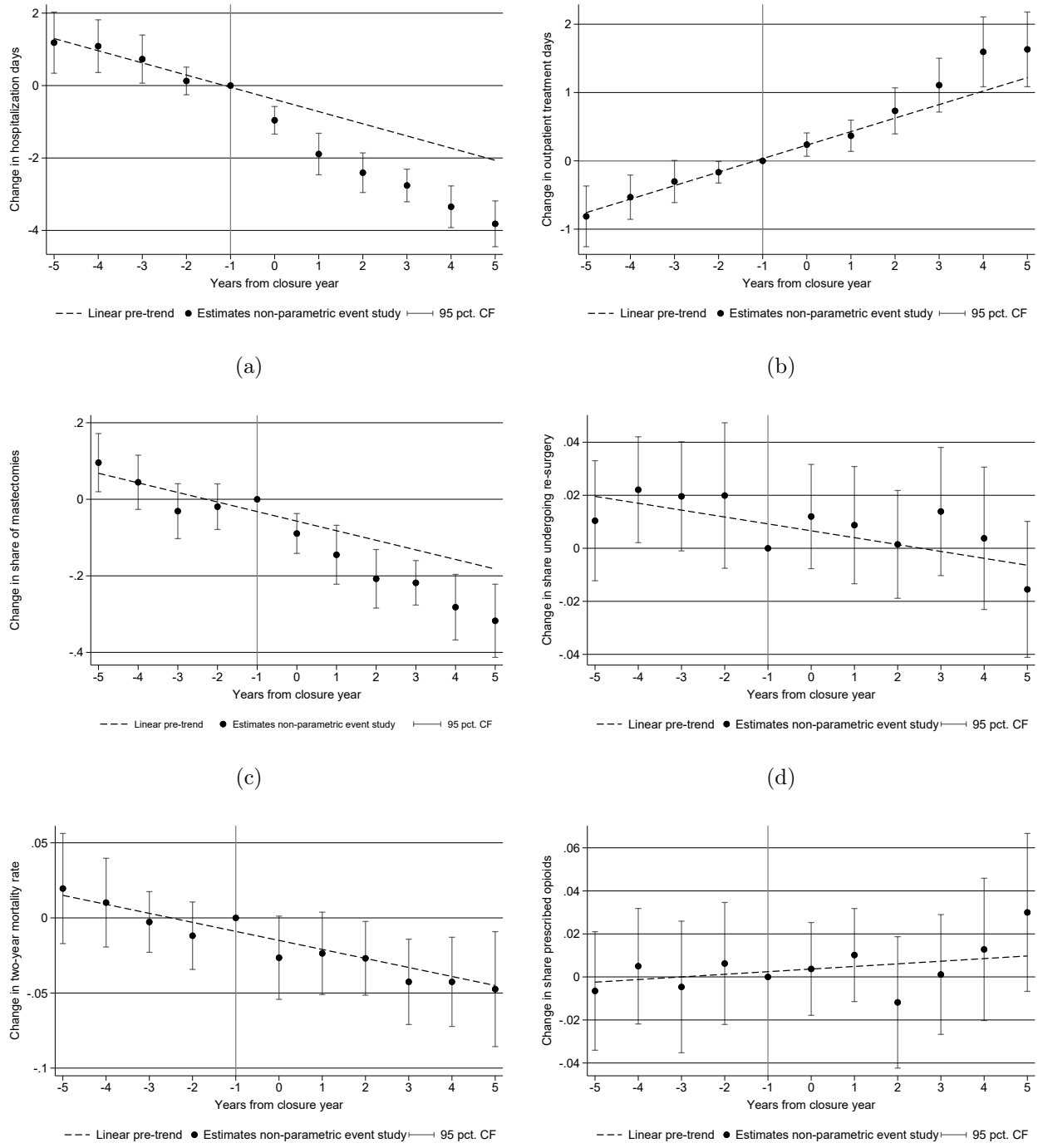



Figure 3: The plotted coefficients are $\beta_k$ from the non-parametric event study. The trend is the linear pre-trend estimated in the parametric event study and carried forward in the post period. The trend is normalized to the level of the non-parametric coefficients in the pre-period. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids within 30 days.

## 5.2 Other outcomes: Neoadjuvant and adjuvant treatment

In appendix C, I also examine changes in the use of radiation therapy and chemotherapy. The results show strong tendencies of an increase in the share receiving radiation therapy where the share increases with 11-16 pct. points (up from 51 pct.) in the first two years after closure. The share receiving chemotherapy remains unchanged. For both outcomes, the confidence intervals are fairly wide because the sample size is reduced from 9,790 observations to 5,418 observations, and the number of municipalities is reduced from 29 to 17 municipalities[9].

## 5.3 Cost changes from one year before closure year to one year after closure year

The shift towards breast-conserving surgery and fewer hospitalization days leads to a cost reduction per patient of approximately EUR 800. Figure 4 shows that the average cost per surgery is reduced by EUR 262 and the average cost per hospitalization period is reduced by EUR 528 from one year before closure year to one year after closure year. As appendix C also shows evidence of a 11-16 pct. points increase in the share receiving radiation therapy, the total cost reduction may be lower than 800 EUR. Although the changes in the share receiving radiation therapy is estimated on a smaller sample and fewer closures, it is relevant to examine the costs associated with changes in the use of radiation therapy. The average patient cost related to the use of radiation therapy increases with EUR 256 from one year before closure year to one year after closure year. Therefore, when taking radiation therapy into account, the average total cost per patient is reduced with approximately EUR 530 per patient from one year before closure year to one year after closure year.

The cost calculation is an approximation which may differ from actual costs[10], and the calculation does not take changes in other procedures, routines, locales or staff resources into account than the three outcomes included in figure 4. For example, it could have been very relevant to examine changes in the number and types of breast reconstruction surgery, but this has not been possible due to changes in reporting practices over the observation period.

[9]Unfortunately, it has only been possible to follow radiation therapy and chemotherapy consistently over time from 2001 (and not 1996 as with the other outcomes). In the main sample, I follow municipalities at least four years before the first clinic closure enters. In this subsample, I follow municipalities at least two years before the first clinic closure enters, and therefore the first closure enters in year 2003 (compared to 2000 in the main sample).

[10]The average change in costs per patient is calculated by using DRG tariffs for the respective types of surgery (mastectomy and lumpectomy), hospitalization days and type of radiation therapy. The calculation is based on the mean age in the sample and takes the sample distribution of different procedure codes for respectively surgeries and radiation therapy into account, but the actual cost may differ from the one calculated with the DRG tariffs. The calculations are based on DRG tariffs extracted from an interactive platform provided by the Danish Health Data Authority: https://interaktivdrglpr2.sundhedsdata.dk/.

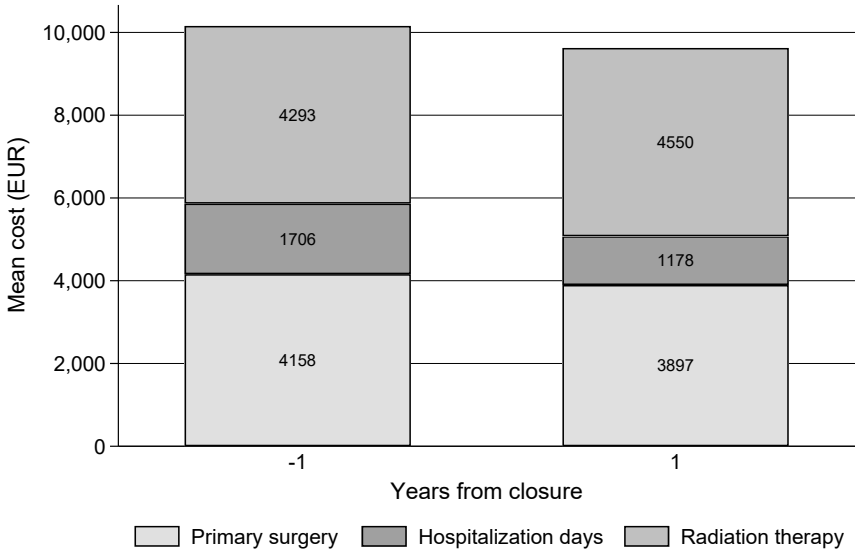



Figure 4: Mean cost per patient one year before closure and one year after closure

# 6 Robustness of results

In the following, I examine the robustness of the results across alternative econometric specifications and sample extractions. All robustness analyses yield results similar to the main results presented in section 5. However, the robustness analyses also show that the main econometric specification captures the outcome pattern in the pre-period better than several of the robustness specifications, especially for the share of mastectomies.

The robustness analyses are described in the following, and the results are presented in figure 5 and 6. In both figures, the baseline model represents the main results presented in section 5. Together with the baseline model, figure 5 presents four other econometric specifications and figure 6 presents four sample extractions. In figure 5 and 6, I show the coefficients in year 0, 1 and 2 across the different robustness specifications to illustrate whether the effect sizes and confidence intervals differ from the baseline model. Two models are estimated with the non-parametric specification while the other models are estimated with the parametric specification. If nothing else is stated, the model is estimated with the parametric specification, and the coefficients in figure 5 and 6 are to be interpreted as deviations from a linear pre-trend in the outcome. The parametric assumption of a linear pre-trend in the outcomes is visually assessed in appendix D.3 to D.8. In appendix D.1 to D.2, the full event window for the two non-parametric specifications are shown.

In figure 5, I show that the main results are robust to a non-parametric specification with a control group of unaffected municipalities and a non-parametric specification without municipality fixed effects. Moreover, I show that the main results are robust to parametric specifications with fewer lags and leads. The only important difference across the model specifications is for the share of mastectomies, where the coefficients are slightly higher in magnitude when estimating the model with three lags and leads compared to five lags and leads in the baseline model. Appendix D.4 shows that this is because the linear pre-trend is upward sloping when only estimating the pre-trend with three lags, but from a visual inspection of the full event window with five lags in figure 3, a downward sloping pre-trend is a more correct representation of the share of mastectomies in the pre-period. Also, it is counter-intuitive that an upward sloping trend would continue in the post-period when breast-conserving surgery is preferred over mastectomies as the state-of-the-art surgery technique in the full observation period.

The estimated effects are intention-to-treat (ITT) effects. All patients living in a closure-affected municipality lose their nearest breast cancer clinic during the observation period. Nevertheless, all patients are included according to the municipality they live in, regardless of the clinic at which they undergo surgery, and therefore some patients may have been treated in clinics further away than their nearest clinic before the closure. Although some patients may have bypassed the nearest clinic even before the closure, an abrupt travel time increase in the year after closure (see fig. 7) shows that most patients in closure-affected municipalities travel

longer when their municipality was hit by a clinic closure and thus the ITT effects must capture the implications that the majority in the closure-affected municipalities face. In figure 6, I show that the results are robust to other sample extractions where I only focus on the municipalities where the treatment exposure is most intense, i.e. where the patients are most likely to be affected by a nearby closure and where it is most unlikely that the patients bypass their nearest clinic.

In the main analyses, I include municipalities based on whether the municipality experiences a closure of the nearest clinic, but I also correct for the fact that some municipalities may have two or three clinics which are all placed near to the municipality, for example, if living in the capital municipality, Copenhagen, where several clinics are located. In the main analyses, I exclude an otherwise closure-affected municipality if there is a second- and/or third-nearest future-closing clinic less than ten minutes further away than the nearest clinic within the event window (see restriction 3 and 4 in appendix A.3). In the first robustness specification in figure 6, I increase this travel time to the second- and third-nearest clinic from 10 to 30 minutes. Thereby, I lower the number of included municipalities, but at the same time, the ITT effect should be more precise as I leave out municipalities where patients potentially could be treated in another clinic just as often as the nearest clinic. In the second robustness specification in figure 6, I narrow the number of included municipalities by excluding municipalities where the second- or third-nearest clinic is a non-closing clinic. This is again to see what happens to the ITT effect when only focusing on the municipalities where the probability of being affected by a closure is highest. As seen from figure 6, the magnitude of the effects is not sensitive to changing these assumptions, but the confidence intervals are slightly wider as the number of included municipalities and patients are lower in the two robustness specifications than in the baseline model.

Next, I examine the extent of compliance to treatment in the nearest breast cancer clinic, that is, whether patients living in a closure-affected municipality are in fact treated in the clinic to be closed in the year before the clinic closes. In the main analyses, patients living in a municipality in which the nearest clinic closes are assigned as closure-affected. However, some patients living in a closure-affected municipality may have bypassed the nearest clinic even before the closure of the nearest clinic. This may be the case if a patient is referred to another breast cancer clinic by their general practitioner, or because of the free choice of hospital in Denmark that allows patients to choose among all public hospitals in Denmark and some private non-profit-making hospitals with the same level of specialization (Olejaz et al. 2012). When calculating the compliance rate, I define a compliant patient as a patient that lives in a closure-affected municipality and is treated in the clinic nearest to that municipality before the closure. If a patient lives in a closure-affected municipality but is not treated in the clinic nearest to that municipality before the closure, then the patient does not comply. In

this study, I find a compliance rate of 74 pct., and therefore a high share of the patients that live in closure-affected municipalities are treated in the clinic nearest to the municipality that they live in one year before the closure of their nearest clinic. A total of 19 of the 29 closure-affected municipalities have a compliance rate above 75 pct., while only six municipalities have a compliance below 50 pct. in the year before closure. The last two robustness analyses in figure 6 show the results where I first exclude the six municipalities with a compliance rate below 50 pct. in the year before closure ($k = -1$), and second, I exclude nine municipalities where the compliance rate on average is below 50 pct. in the five years leading up to closure ($k < 0$). Again, the results are very similar to the results presented in the baseline model.

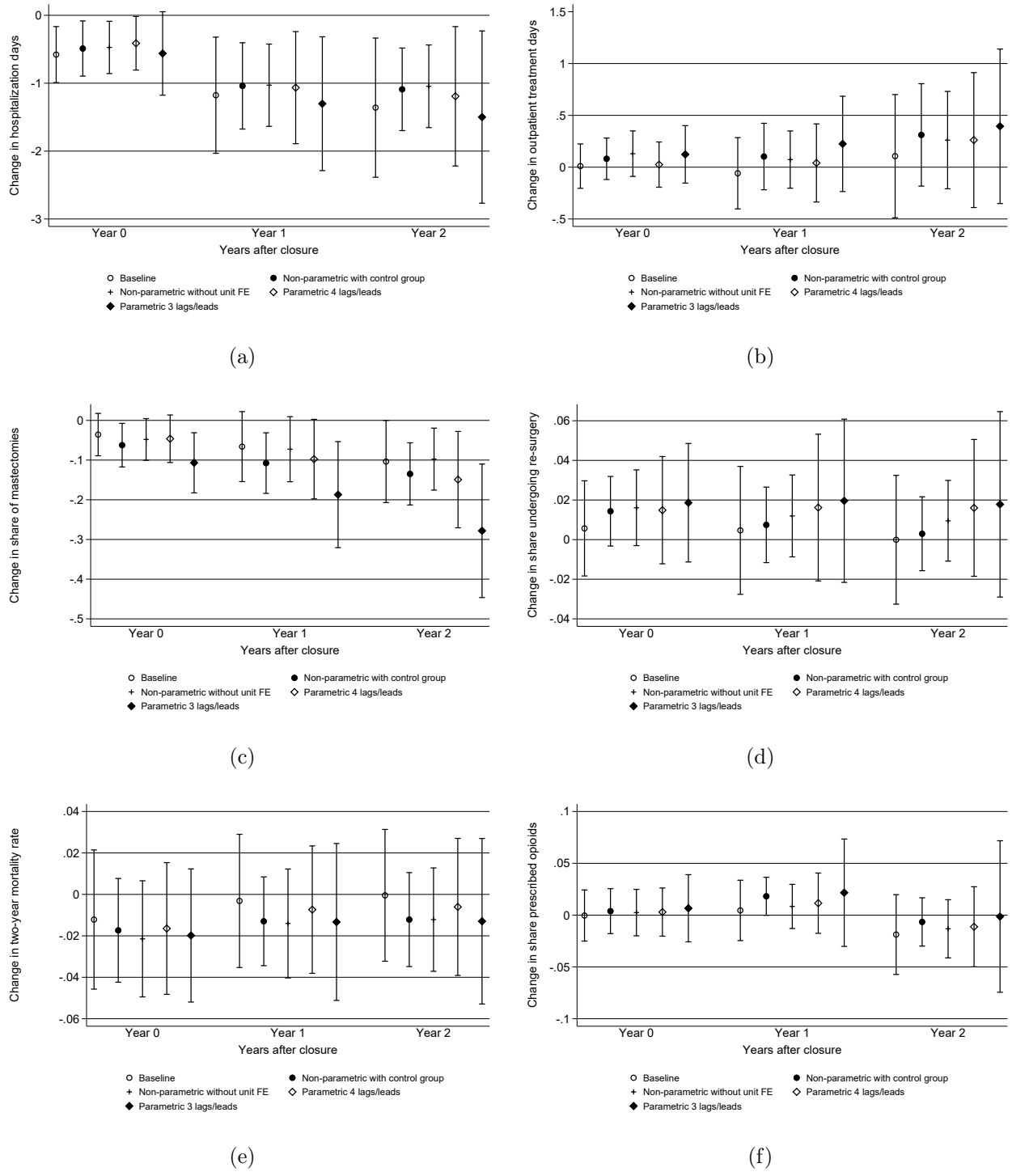
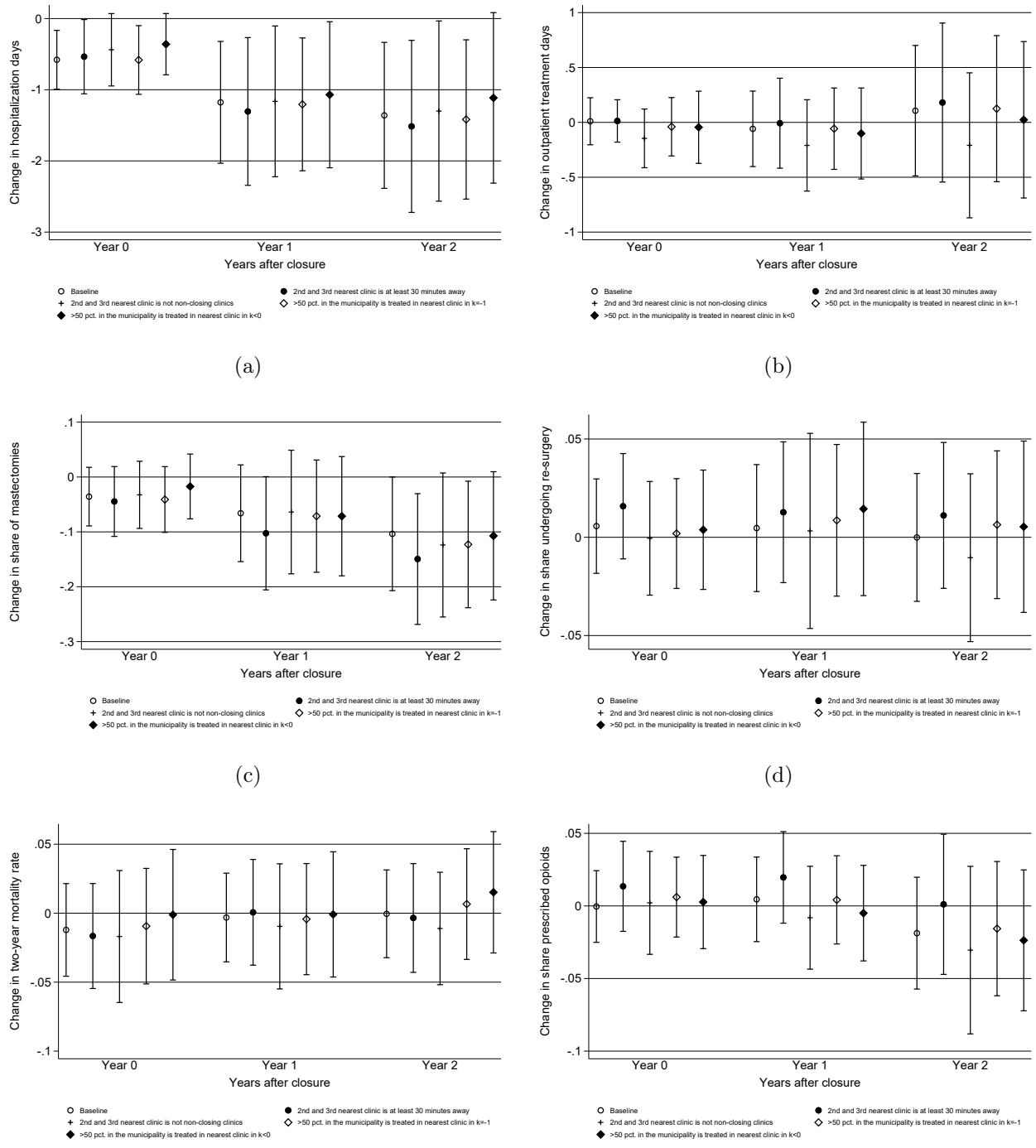



Figure 5: Robustness analyses across different econometric specifications with 95 pct. confidence intervals. The coefficients are interpreted relative to a linear pre-trend except for the two non-parametric specifications where the coefficients are interpreted relative to $k = -1$. Baseline is the main results from table 3. *(a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids within 30 days. The full event window for each outcome and each robustness analysis is shown in appendix D.*




(a) (b)

(c) (d)

(e) (f)

Figure 6: Robustness analyses across different sample extractions with 95 pct. confidence intervals. The coefficients are interpreted relative to a linear pre-trend. Baseline is the main results from table 3. *(a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids within 30 days. The full event window for each outcome and each robustness analysis is shown in appendix D.*

# 7 Mechanisms

In the following, I explore what is driving the results. First, I show that patient travel time increases when a municipality is exposed to a breast cancer clinic closure, and, in turn, that patients are in fact reallocated to other clinics. Second, I provide evidence for two channels, that is that the closures do not generate significant patient volume shocks to non-closing clinics, and that closing clinics to a lower extent provide state-of-the-art surgery and that they hospitalize patients longer in the years leading up to closure. Together, the channels suggest that added volume returns at non-closing clinics were of less importance for the results than simply reallocating patients to hospital clinics that provide higher-quality surgeries and perform better on cost-saving metrics such as the length of hospitalizations.

## 7.1 Travel time increases

Figure 7 shows that travel time increases by about 26 minutes in the year following closure of a clinic compared to the year before a closure. When examining the travel time relative to a linear pre-trend, travel time increases about 22 minutes. The results are hence almost similar in the nonparametric and parametric specification, and patients are in fact reallocated to other clinics following a nearby closure. The travel time results also show that patients on average experience a rather small increase in travel time.

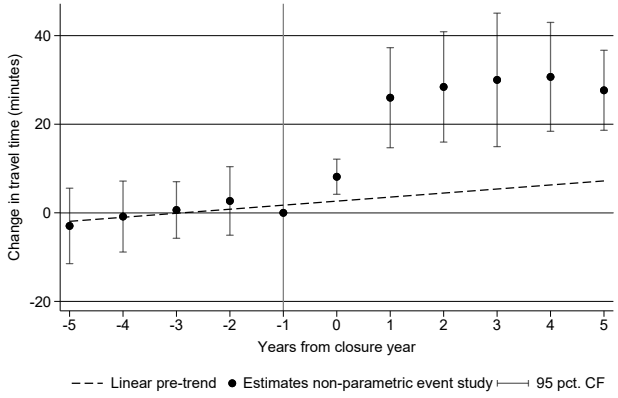



Figure 7: The plotted coefficients are $\beta_k$ from the non-parametric event study. The trend is the linear pre-trend estimated in the parametric event study and carried forward in the post period. The trend is normalized to the level of the non-parametric coefficients in the pre-period.

## 7.2 No volume shocks at non-closing clinics

Breast cancer clinic closures might affect the patient volume at non-closing clinics as patients from closure-affected municipalities are reallocated and referred to the non-closing clinics. If the patient volume at non-closing clinics increase significantly, the added patient volume may generate a potential for economies of scale and scope in the non-closing clinics, but also a risk of crowding if capacity is constrained in the non-closing clinics. Therefore, I examine whether closures generated patient volume shocks at non-closing breast cancer clinics.

To estimate the volume shocks to non-closing clinics, I am inspired by the literature on the returns to hospital volume where authors have applied instrument variable models to address the endogeneity concern between volume and quality (Gaynor, Seider, et al. 2004; Gaynor, Seider, et al. 2005; Hentschker and Mennicken 2018; Avdic, Lundborg, et al. 2019). All these papers apply an instrument for volume in a first stage, and in a second stage estimate the returns to volume on hospital outcomes. As an instrument for patient volume, Gaynor, Seider, et al. (2004), Gaynor, Seider, et al. (2005), and Hentschker and Mennicken (2018) use the number of potential patients and the number of other hospitals in the area surrounding each hospital where they exploit the fact that patients choose hospitals that are closer to their residence. Hentschker and Mennicken (2018) also use the ratio between the number of potential patients and the number of hospitals in the area to take into account that the case volume of a hospital should increase if patients have few versus many nearby hospitals. As an instrument for volume in non-closing clinics, Avdic, Lundborg, et al. (2019) apply a vector of dummy variables that captures whether a closure took place within a given regional catchment area.

In contrast to existing studies, I use the patient outflow from closing clinics to predict expected patient inflow at non-closing clinics, where I distribute the patient volume from closing clinics across the nearest non-closing clinics. This allows me to take into account that non-closing clinics might be affected by several subsequent closures with different expected patient inflows each time, and sometimes a non-closing clinic is affected by two closing clinics in the same year. I apply travel time instead of regional catchment areas as the regional structure in Denmark changes over time.

The annual surgery volume faced by patient $i$ at non-closing breast cancer clinic $c$ in calendar year $t$ is given by

$$ln(Volume)_{ict} = \alpha_1 Exp.Inflow_{ct} + \beta X_{ict} + \lambda_t + \lambda_c + \epsilon_{ict} \tag{3}$$

where $Exp.Inflow_{ct}$ is the expected patient inflow from nearby closing clinics to non-closing clinic $c$ in calendar year $t$. $X_{ict}$ is a vector of individual level health and sociodemographic covariates. $\lambda_t$ is calendar year fixed effects that capture calendar year changes in the volume common to all non-closing clinics, and $\lambda_c$ is clinic fixed effects that capture time-invariant

differences across clinics. Hence, the key coefficient, $\alpha_1$, is identified using only within-clinic changes. Standard errors are clustered at the clinic level.

The expected patient inflow, $Exp.Inflow_{ct}$, is given by the probability that non-closing clinic $c$ will experience a patient inflow in year $t$ multiplied by the patient volume from the closing clinic(s) in year $t-1$

$$\begin{aligned}\text{Exp. Inflow}_{ct} &= Pr(Inflow_{ct}) \times \text{Volume at closing clinic}_{t-1} \\ &= \frac{\frac{1}{\text{Travel time to closed clinic}_c}}{\sum_{c=1}^{4} \frac{1}{\text{Travel time to closed clinic}_c}} \times \text{Volume at closing clinic}_{t-1} \end{aligned} \tag{4}$$

where the denominater in the first term is the sum of the inverse travel time to the closed clinic among the four nearest clinics to the closing clinic. Hence, the nearest clinic is assigned the highest probability of patient inflow and the second nearest clinic is assigned the second highest probability of patient inflow etc. Following this, the expected inflow will be highest for the nearest clinic and lowest for the fourth nearest clinic. For all other clinics than the four nearest clinics, the probability of inflow is set to 0. In the years where no closures take place, the probability of inflow in the following year is set to 0 for all clinics. The number of non-closing clinics that the patient volume from the closed clinic is assumed to be distributed across is varied from one to four clinics when presenting the results, but the calculation of expected patient inflow is only shown for four clinics in equation 4.

Table 4 shows that clinic closures do not generate a sufficiently large patient inflow to the non-closing clinics to affect the volume in non-closing clinics. This is also true when varying the number of non-closing clinics that the patient volume from the closing clinics is distributed across. The results show that there is no potential for added economies of scale and scope at non-closing clinics, and there is no risk of congestion or crowding at non-closing clinics. This is most likely because the closing clinics have a very low volume of patients compared with the non-closing clinics and because the breast cancer incidence and the number of surgeries increase and vary over time, for example, due to changes in screening patterns. Figure 8 shows the average number of surgeries in non-closing clinics and the mean volume shock to the nearest non-closing clinics over time.

Table 4: Results from estimating equation (3): Volume shocks in non-closing clinics by the number of clinics that the patient volume from closing clinics is distributed across. *Dependent variable is ln(volume) in non-closing clinic.*

| | Nearest clinic | 2 nearest clinics | 3 nearest clinics | 4 nearest clinics |
|---|---|---|---|---|
| Exp. Inflow | 0.002 | 0.003 | 0.003 | 0.003 |
| | [0.001] | [0.003] | [0.003] | [0.008] |
| ln(Exp. Inflow)[a] | 0.040 | 0.029 | 0.022 | 0.023 |
| | [0.003] | [0.022] | [0.027] | [0.026] |
| Number of observations | 27,191 | 41,068 | 49,426 | 51,449 |
| Number of clinic shocks | 9 | 15 | 20 | 22 |
| Mean volume $t = -1$ [b] | 205.9 | 213.8 | 220.2 | 217.1 |
| Mean shock $t = 0$ [c] | 53.9 | 28.9 | 19.3 | 15.3 |

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

Based on model 3 and includes a vector of patient covariates, clinic FE and calendar year FE. Standard errors are clustered at the clinic level.

[a] For zero expected inflow, the expected inflow is set to 1 patient to be able to take log.

[b] Mean volume in $t = -1$ is the mean patient volume in the non-closing clinics in the year before they are affected by patient inflow from a closure.

[c] Mean shock in $t = 0$ is the mean expected patient inflow in the non-closing clinics in year $t$.

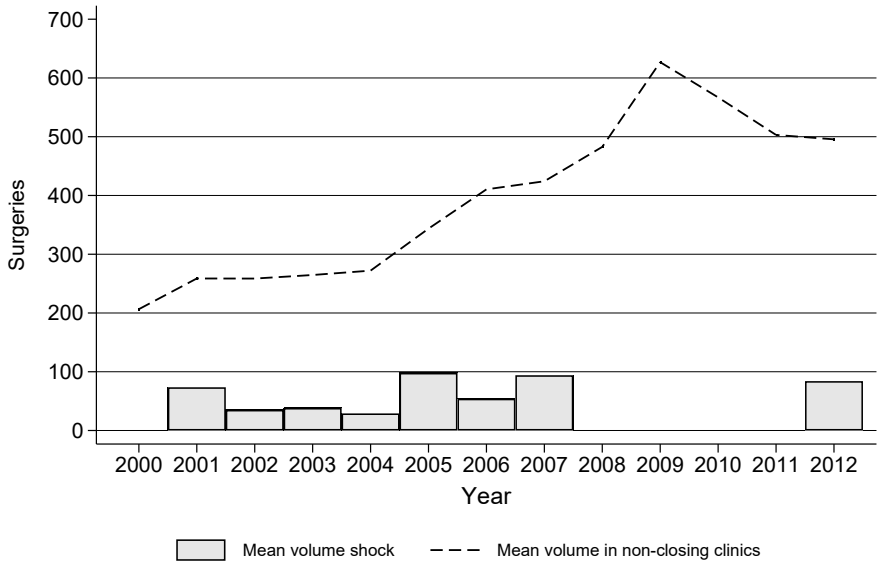



Figure 8: Expected patient inflow (volume shock) to nearest non-closing clinic

## 7.3 Which clinics close?

To examine differences between closing and non-closing clinics, I compare the study outcomes in the four years leading up to a closure across the closing and non-closing clinics. To compare the outcomes of closing and non-closing clinics, I estimate the difference in each of the outcomes for each event time $k = -4$ to $k = -1$. For event time $k = -1$, I examine the outcome, $y_{ict,k=-1}$, that patient $i$ undergoing surgery in clinic $c$ faces in event time $k = -1$

$$y_{ict,k=-1} = \alpha 1(\textit{Closing clinic=1})_{c,k=0} + \delta X_{i,k=-1} + \lambda_t + \epsilon_{ictk} \tag{5}$$

where $\alpha$ represents the coefficient of interest and measures the difference in the outcome in $k = -1$ between patients treated in a clinic that closes in $k = 0$ relative to patients being treated in a clinic that remains open. The vector $X_{it}$ represents individual level health and sociodemographic covariates for patient $i$ treated in event time $k = -1$.The calendar year fixed effects $\lambda_t$ capture calendar year changes in the outcomes common to all clinics. Likewise, I estimate equation (5) for $k = -4$, $k = -3$ and $k = -2$.

The charts presented in figure 9 show the results of equation (5), which are the differences between closing clinics and non-closing clinics for each outcome and for each event time leading up to a closure year. Chart (a) of figure 9 shows that the number of hospitalization days is significantly higher (approx. 0.8-1.1 days) in closing clinics than in non-closing clinics in the four years leading up to a closure. Chart (b) shows that the number of outpatient treatment days is slightly lower (0.1-0.3 days) in the closing clinics, and not significantly lower in $k = -3$. Chart (c) of figure 9 shows that the share of patients undergoing a mastectomy, compared to a breast-conserving surgery, on average is between 3 and 11 pct. points higher in the closing clinics compared to the non-closing clinics. Although the difference is lower and not significantly different in $k = -3$ at a 5 pct. significance level, the higher share of mastectomies in the closing clinics in the years leading up to a closure highlights that fewer patients receive state-of-the-art surgery in the closing clinics. Chart (d) shows no important differences in the share undergoing re-surgery in the four years leading up to a closure. Chart (e) and (f) show no significant differences in the two-year mortality rate or the use of opioids between closing and non-closing clinics.

Despite minor fluctuations, chart (a) and (c) in figure 9 show that the differences in the number of hospitalization days and the share of patients undergoing a mastectomy are persistent in the four years leading up to closure year. Based on the dynamic patterns of the outcomes in the years leading up to a closure, I find no strong evidence of endogenous timing of closures. However, the dynamics of the outcomes in the pre-event years highlight that the closing clinics are clinics that perform worse over a period of at least four years preceding a closure. The magnitude of the differences in the share of mastectomies and the number of hospitalization

days between closing and non-closing clinics in the years leading up to a closure is very similar to the effects of closing clinics shown in section 5. Hence, this indicates that the effects of closing clinics are driven by differences between closing- and non-closing clinics in the years leading up to a closure.

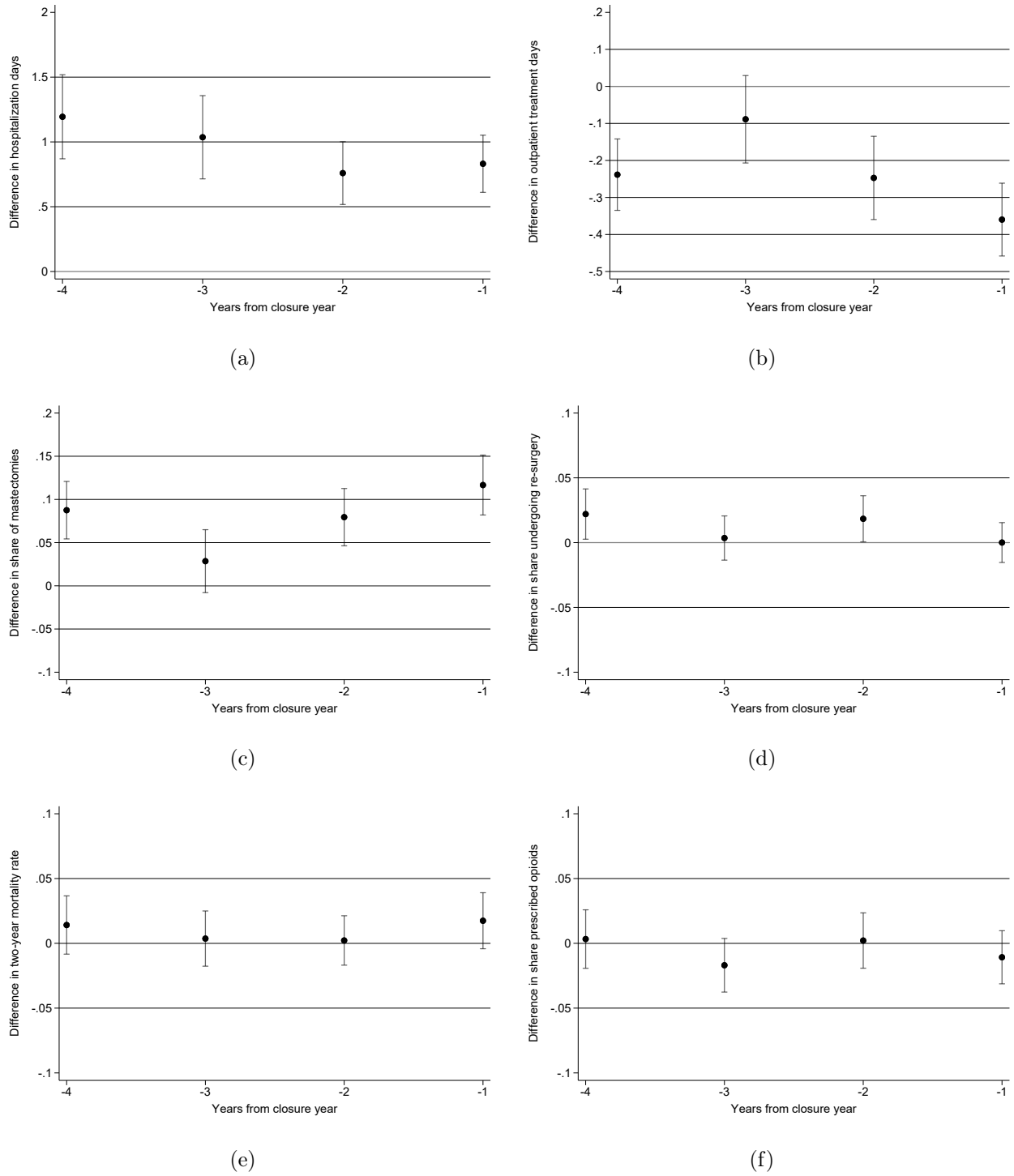



Figure 9: Difference in outcomes between closing and non-closing clinics in the years leading up to a closure with 95 pct. confidence intervals. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share undergoing a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids within 30 days.

# 8 Discussion

The results show that the number of hospitalization days for patients living in closure-affected municipalities on average decreases from five to four days, while the number of outpatient treatment days is unaffected. More patients in closure-affected municipalities receive breast-conserving surgery following closure of their nearest clinic. In the first two years after closure, the share of mastectomies decreases by 7-10 pct. points on average (down from 62-65 pct.). In turn, patients from closure-affected areas experience an increase in their access to state-of-the-art care. The closure-affected patients do not experience adverse health effects on re-surgery share, mortality rate or pain levels (use of opioids).

The papers that are most comparable to the present study are those that analyze disease-specific clinic closures. Most disease-specific papers study high-risk, complex and acute procedures and find negative health effects among patients directly affected by hospital closures (Buchmueller et al. 2006; Avdic 2016; Bertoli and Grembi 2017; Carroll 2019). A very sparse literature studies less acute conditions. Avdic, Lundborg, et al. (2020) study closures of maternity wards in Sweden and find positive health effects for infants as the births are shifted to larger wards of higher quality. However, Avdic, Lundborg, et al. (2020) find negative effects on maternal obsteric trauma because of ward overcrowding. Buchmueller et al. (2006) study closure and distance effects on mortality for subgroups of both acute and non-acute diseases, and find no impact on mortality for non-acute diseases. A broader literature examines closures of entire hospitals where consolidated hospitals are compared to non-consolidated hospitals. By using a matched control group of non-consolidated hospitals in the US, Joynt et al. (2015) and Beaulieu et al. (2020) find no overall effect of hospital consolidations on mortality rates, however Joynt et al. (2015) do find a small drop in readmission rates. Beaulieu et al. (2020) find that the quality of the patient experience decreases. Similarly, Gaynor, Laudicella, et al. (2012) do not find any indications of quality increases when examining mergers of short-term general hospitals in England, but they do find adverse effects such as rising waiting times and increased financial deficits.

When studying the mechanism underlying the results in this paper, I find that the results are driven by allocating patients to clinics that hospitalize patients for fewer days and that more often perform breast-conserving surgery. I also find that the patient flow from closing to non-closing clinics is not large enough to generate a significant increase in patient volume in non-closing clinics because the closed clinics are very small (about 1/4 of the size of a non-closing clinic) and because the breast cancer incidence and number of surgeries increase and vary over time in non-closing clinics. Previous studies also show that closed hospital clinics tend to be smaller than non-closing clinics (Lillie-Blanton et al. 1992; Lindrooth et al. 2003; Ciliberto and Lindrooth 2007). When the volume in non-closing clinics does not increase significantly, there is no potential for added economies of scale or scope (volume-outcome effects) in non-closing

clinics, nor does a high risk of crowding exist. However, even in the absence of volume-outcome effects at the non-closing clinics, breast cancer clinic closures may have been welfare-improving, as they have reduced the number of costly hospitalization days and shifted surgical procedures to state-of-the-art breast-conserving techniques because the lower-performing clinics are the ones to close. The closure of lower-performing clinics is also supported in previous studies, for example, Harrison (2007) finds that capacity utilization and length of stay are important determinants for closure. The present study's finding of positive effects of closures and being reallocated to another breast cancer clinic also adds to the literature that provides evidence that referring patients to higher-quality physician teams and hospital clinics leads to positive health effects and less expensive hospital stays (Doyle et al. 2019; Doyle Jr et al. 2010).

This paper has some limitations. First, some municipalities were excluded from the analyses. In Denmark, we have 98 municipalities. A total of 51 municipalities were excluded from the analyses, 29 were assigned status as closure-affected and the remaining 18 municipalities were unaffected. The excluded municipalities cover municipalities where it was not possible to identify whether the municipality was affected by a closure or not, or if subsequent closures (or openings) affected the municipality within the event window of five years before and after closure. In cases where the municipalities were excluded because they could not be assigned as closure or non-closure affected, the difficulty in assigning them was due to them having almost the same travel distance to a closing and non-closing clinic.

A second limitation is the lack of comparable data over time for other treatment and procedures than the main surgery (breast conserving vs. mastectomy). It would, for example, have been interesting to examine changes in the number and types of breast reconstruction surgeries, as changes in the main surgery type (breast conserving vs. mastectomy) may have an effect on reconstruction surgery. With the present data, it has not been possible to follow surgery procedure codes over time for breast reconstruction surgery due to changes in procedure codes and reporting practices over the observation period. Additionally, it is a limitation that we are only able to observe neoadjuvant and adjuvant treatment across some of the observation period.

For future research, it would be interesting to examine characteristics of the surgeons across closing and non-closing clinics, and whether surgeons in closing clinics move to the non-closing breast cancer clinics together with the patients or leave the labour market early because of being close to retirement age.

# 9 Conclusions

Closures, consolidations and the trade-off between access to nearby care and health care quality have been topics of health care policy debates for more than two decades. On the one side, patients in closure-affected areas will be worse off if increases in travel time are not offset by higher health care quality. On the other side, stakeholders have reasons to value cost-saving and efficiency higher than patient travel time in order to meet health care budgets. This paper contributes to the debate on whether to close hospital clinics, and if they are to be closed, which ones should be closed. This paper does so by studying the impact of breast cancer clinic closures on the quality of care that closure-affected patients receive and on cost-saving metrics. The case of breast cancer clinic closures serves as a non-acute scheduled setting in contrast to previous studies of acute care. The paper also contributes to the existing literature by investigating utilization of surgical procedure and opioid prescriptions in addition to more traditional outcomes, such as hospitalization days and mortality rates.

The results show that closing breast cancer clinics reduces hospitalization days and shifts surgical procedures to state-of-the-art breast-conserving techniques for patients living in municipalities affected by a closure. There are no adverse health effects on the share undergoing re-surgery, mortality rates or the use of opioids, and the patient flow from closed hospital clinics is too small to cause crowding in non-closed clinics. The findings in this paper may reassure patients living in closure-affected areas that closing hospital clinics may be a pathway to receiving state-of-the-art care in higher-quality clinics. For policy makers and hospital administrators, the results reveal that closures of clinics performing scheduled surgery may be an effective policy instrument if the goal is to reduce variation in the delivery of hospital care, but if the clinics to be closed are small compared to non-closing clinics then there is no potential for added economies of scale or scope (volume-outcome effects) in non-closing clinics.

# References

Avdic, Daniel (2016). "Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden". In: *Journal of Health Economics* 48, pp. 44–60.

Avdic, Daniel, Petter Lundborg, and Johan Vikström (2019). "Estimating returns to hospital volume: Evidence from advanced cancer surgery". In: *Journal of Health Economics* 63, pp. 81–99.

— (2020). "How do Closures of Maternity Wards Affect Birth Outcomes?" Unpublished manuscript from October 15, 2020.

Beaulieu, Nancy D., Leemore S. Dafny, Bruce E. Landon, Jesse B. Dalton, Ifedayo Kuye, and J. Michael McWilliams (2020). "Changes in Quality of Care after Hospital Mergers and Acquisitions". In: *New England Journal of Medicine* 382.1, pp. 51–59.

Bertoli, Paola and Veronica Grembi (2017). "The life-saving effect of hospital proximity". In: *Health economics* 26, pp. 78–91.

Birkmeyer, John D, Andrea E Siewers, Emily VA Finlayson, Therese A Stukel, F Lee Lucas, Ida Batista, H Gilbert Welch, and David E Wennberg (2002). "Hospital volume and surgical mortality in the United States". In: *New England Journal of Medicine* 346.15, pp. 1128–1137.

Borusyak, Kirill and Xavier Jaravel (2017). "Revisiting event study designs". Available at SSRN: https://ssrn.com/abstract=2826228.

Buchmueller, Thomas C, Mireille Jacobson, and Cheryl Wold (2006). "How far to the hospital?: The effect of hospital closures on access to care". In: *Journal of health economics* 25.4, pp. 740–761.

Carroll, Caitlin (2019). "Impeding Access or Promoting Efficiency? Effects of Rural Hospital Closure on the Cost and Quality of Care". Unpublished manuscript from March 19, 2019.

Charlson, Mary E., Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie (1987). "A new method of classifying prognostic comorbidity in longitudinal studies". In: *Journal of Chronic Diseases* 40, pp. 373–383.

Chew, Helen K (2001). "Adjuvant therapy for breast cancer: who should get what?" In: *Western Journal of medicine* 174.4, p. 284.

Chowdhury, MM, H Dagash, and A Pierro (2007). "A systematic review of the impact of volume of surgery and specialization on patient outcome". In: *British Journal of Surgery: Incorporating European Journal of Surgery and Swiss Surgery* 94.2, pp. 145–161.

Christiansen, Terkel and Karsten Vrangbæk (2018). "Hospital centralization and performance in Denmark—Ten years on". In: *Health Policy* 122.4, pp. 321–328.

Ciliberto, Federico and Richard C Lindrooth (2007). "Exit from the hospital industry". In: *Economic Inquiry* 45.1, pp. 71–81.

Dafny, Leemore (2009). "Estimation and identification of merger effects: An application to hospital mergers". In: *The Journal of Law and Economics* 52.3, pp. 523–550.

Datafordeler (2018). *Dansk Adresseregister*. URL: `https://datafordeler.dk/vejledning/brugervilkaar/danmarks-adresseregister-dar/` (visited on 08/30/2018).

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J Notowidigdo (2018). "The economic consequences of hospital admissions". In: *American Economic Review* 108.2, pp. 308–52.

Doyle Jr, Joseph J, Steven M Ewer, and Todd H Wagner (2010). "Returns to physician human capital: Evidence from patients randomized to physician teams". In: *Journal of health economics* 29.6, pp. 866–882.

Doyle, Joseph, John Graves, and Jonathan Gruber (2019). "Evaluating measures of hospital quality: Evidence from ambulance referral patterns". In: *Review of Economics and Statistics* 101.5, pp. 841–852.

Fisher, Bernard, Stewart Anderson, John Bryant, Richard G Margolese, Melvin Deutsch, Edwin R Fisher, Jong-Hyeon Jeong, and Norman Wolmark (2002). "Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer". In: *New England Journal of Medicine* 347.16, pp. 1233–1241.

French National Health Insurance (2018). *Améliorer la qualité du système de santé et maîtriser les dépenses - propositions de l'Assurance Maladie pour 2019. Rapport Charges et Produits.* `https://assurance-maladie.ameli.fr/sites/default/files/rapport-charges-et-produits-2019-web.pdf`. Accessed: 2021-04-19.

Gaynor, Martin, Mauro Laudicella, and Carol Propper (May 2012). "Can governments do it better? Merger mania and hospital outcomes in the English NHS". In: *Journal of Health Economics* 31.3, pp. 528–543.

Gaynor, Martin, Harald Seider, and William B. Vogt (2004). "Volume-outcome and antitrust in US health care markets". In: *Unpublished manuscript, Carnegie Mellon University.*

— (2005). "The volume-outcome effect, scale economies, and learning-by-doing". In: *American Economic Review* 95.2, pp. 243–247.

Halm, Ethan A, Clara Lee, and Mark R Chassin (2002). "Is volume related to outcome in health care? A systematic review and methodologic critique of the literature". In: *Annals of internal medicine* 137.6, pp. 511–520.

Hamilton, Barton H and Vivian Ho (1998). "Does practice make perfect?: Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec". In: *Medical care* 36.6, pp. 892–903.

Harrison, Teresa D. (2007). "Consolidations and closures: an empirical analysis of exits from the hospital industry". In: *Health economics* 16.5, pp. 457–474.

Hentschker, Corinna and Roman Mennicken (2018). "The volume–outcome relationship revisited: practice indeed makes perfect". In: *Health services research* 53.1, pp. 15–34.

Huesch, Marco D and Mariko Sakakibara (2009). "Forgetting the learning curve for a moment: how much performance is unrelated to own experience?" In: *Health economics* 18.7, pp. 855–862.

Huguet, Marius (2020). "Centralization of care in high volume hospitals and inequalities in access to care". In: *Social Science & Medicine* 260, p. 113177.

Hviid, Simon J. and Nicolai Kristensen (2021). "Rural Decline and Residential Sorting". Unpublished Manuscript, VIVE.

Joynt, Karen E, Paula Chatterjee, E John Orav, and Ashish K Jha (2015). "Hospital closures had no measurable impact on local hospitalization rates or mortality rates, 2003–11". In: *Health affairs* 34.5, pp. 765–772.

Kambouris, Angelos (1996). "Physical, psychological, and economic advantages of accelerated discharge after surgical treatment for breast cancer." In: *The American Surgeon* 62.2, pp. 123–127.

Keelan, Stephen, Michael Flanagan, and Arnold DK Hill (2021). "Evolving trends in surgical management of breast cancer: An analysis of 30 years of practice changing papers". In: *Frontiers in Oncology*, p. 1608.

Larsen, Ole B. and Gunnar Schiøler (2005). *NCSP Nomesko Classification of Surgical Procedures*. The Danish Health Authority. Munksgaard Danmark.

Lillie-Blanton, Marsha, Suzanne Felt, Patrick Redmon, Steven Renn, Steve Machlin, and Elizabeth Wennar (1992). "Rural and urban hospital closures, 1985-1988: operating and environmental characteristics that affect risk". In: *Inquiry*, pp. 332–344.

Lindrooth, Richard C, Anthony T Lo Sasso, and Gloria J Bazzoli (2003). "The effect of urban hospital closure on markets". In: *Journal of health economics* 22.5, pp. 691–712.

Luft, Harold S, John P Bunker, and Alain C Enthoven (1979). "Should operations be regionalized? The empirical relation between surgical volume and mortality". In: *New England Journal of Medicine* 301.25, pp. 1364–1369.

Margolese, Richard G and Jean-Claude M Lasry (2000). "Ambulatory surgery for breast cancer patients". In: *Annals of Surgical Oncology* 7.3, pp. 181–187.

Marla, S and S Stallard (2009). "Systematic review of day surgery for breast cancer". In: *International Journal of Surgery* 7.4, pp. 318–323.

McManus, Susan A, Denise A Topp, and Chris Hopkins (1994). "Advantages of outpatient breast surgery." In: *The American Surgeon* 60.12, pp. 967–970.

OECD (2013). *OECD Reviews of Health Care Quality: Denmark 2013; Raising Standards*.

Olejaz, Maria, Annegrete Juul Nielsen, Andreas Rudkjøbing, Hans Okkels Birk, Allan Krasnik, and Cristina Hernández-Quevedo (2012). "Denmark health system review". In: *Health systems in transition* 14.2.

Quan, Hude, Bing Li, Chantal M Couris, Kiyohide Fushimi, Patrick Graham, Phil Hider, Jean-Marie Januel, and Vijaya Sundararajan (2011). "Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries". In: *American journal of epidemiology* 173.6, pp. 676–682.

Quan, Hude, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L. Duncan Saunders, Cynthia A. Beck, Thomas E. Feasby, and William A. Ghali (2005). "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data". In: *Medical care*, pp. 1130–1139.

Shahian, David M and Sharon-Lise T Normand (2003). "The volume-outcome relationship: from Luft to Leapfrog". In: *The Annals of thoracic surgery* 75.3, pp. 1048–1058.

Stagg, Vicki (2006). "CHARLSON: Stata module to calculate Charlson index of comorbidity". Boston College Department of Economics. Revised: September 13, 2017.

The Commonwealth Fund (2020). *International Health Care System Profiles Denmark*. `https://www.commonwealthfund.org/international-health-policy-center/countries/denmark`. Accessed: 2021-03-24.

The Danish Health Authority (2004-2013). *Afdelingsklassifikationer*.

— (1999-2004). *Sygehusklassifikation og kommunekoder*. Munksgaard.

Thompson, AM and SL Moulder-Thompson (2012). "Neoadjuvant treatment of breast cancer". In: *Annals of Oncology* 23.

Wolters, R, AC Regierer, L Schwentner, V Geyer, K Possinger, R Kreienberg, MB Wischnewsky, and A Wöckel (2012). "A comparison of international breast cancer guidelines–do the national guidelines differ in treatment recommendations?" In: *European journal of cancer* 48.1, pp. 1–11.

Wyld, Lynda, Isabel T Rubio, and Tibor Kovacs (2019). "Education and Training in Breast Cancer Surgery in Europe". In: *Breast Care* 14.6, pp. 366–372.

# A Appendix A: Data selection

## A.1 Population

The population consists of all women diagnosed with breast cancer who underwent breast cancer surgery from 1996-2014 and and who live in a municipality where the nearest breast cancer clinic closes between 2000 and 2011. A patient can be represented in the sample more than once, which is controlled for in the regressions by adding a dummy for whether the patient has undergone breast cancer surgery 365 days prior to their current surgery. As seen in table 2, this is the case for only 5 pct. of the sample. The sample is not restricted to first-time diagnoses, but I control for variation in patient health 1-5 years prior to the surgery year by including the Charlson Comorbidity Index.

## A.2 Breast cancer clinics

In the following, I first describe the general criteria for including and excluding clinics in a given year. Second, I describe how the different breast cancer clinics are categorized as *closing*, *non-closing* and *others*.

### A.2.1 Left out/not authorized breast cancer clinics

1. Clinics performing surgery on less than 10 patients annually

   (a) Clinics performing surgery on less than 10 patients in all observation years are left out of the analyses as these are not categorized as authorized breast cancer clinics. The patient boundary at 10 is chosen as this is the boundary that the Danish Health Authority uses. If a clinic performs less than 10 surgeries, then it is assumed to be a surgery taking place together with another surgery or a specific case where it for instance has not been responsible to move the patient to an authorized clinic due to transport limitations, time constraints or other characteristics of the patient.

   (b) Clinics performing surgery on less than 10 patients annually in the tails of its presence are deleted in tail years.

      i. If a clinic has performed less than 10 surgeries in a given year (e.g., 2004) and has experienced a decrease of more than 10 patients since last year (e.g., 2003), then the clinic is set to close in the year before the patient volume decreased below 10 (e.g., 2003). If a clinic in one year has for example 100 patients and the next year only 9 patients, then the clinic is set to close in the year of the 100 patients.

ii. If a clinic opens in a given year (e.g., 2000), but has below 10 patients, and in the next year have above 10 patients (e.g., 2001), then the next year is set to the opening year (e.g., 2001)

2. Private clinics are left out of the analyses. There is only one private clinic over the observation period and it has a very low patient volume.

### A.2.2 Closing, non-closing and other clinics

1. *Closing clinics* are clinics which close during the period 2000 to 2011, and which can be followed at least 4 years prior to the closure year. For example, a clinic that closes in 2000 and is present in the data since at least 1996.

2. *Non-closing clinics* are clinics which are present and perform breast cancer surgery in all observation years from 1996 to 2014.

3. *Other clinics* are clinics that closed before 2000 and therefore cannot be followed four years prior to their closure and clinics that opened after 1996 or that closed and later reopened during the observation period.

## A.3 Assignment of treatment exposure to municipalities

In the following, I describe how each municipality is categorized as treatment (closure-affected), potential control (closure unaffected) and uncategorized (uncertain treatment exposure).

1. I include all municipalities and all clinics with more than 10 patients annually when assigning travel time between breast cancer clinics and municipalities. In 2007, the number of Danish municipalities decreased from 283 municipalities to 98, where the former old municipalities were nested within new municipalities. I use the present municipality structure of 98 municipalities for the full period 1996-2014.

2. By using the identified breast cancer clinics (independent of their status as either closing, non-closing or other), I calculate the average travel time among the parishes in each municipality to each clinic. For each municipality, I rank the clinics with the nearest clinic first and keep the three nearest clinics for each municipality.

3. **If the nearest clinic is a closing clinic**, I exclude a municipality based on the characteristics of the second and third nearest clinics. If a municipality is excluded, the municipality will be left out of all analyses. Non-excluded municipalities will be classified

as closure-affected municipalities and the event time will be calculated around the year that the nearest clinic closes. Exclusion criteria:

(a) Exclude if second and/or third nearest clinic is less than 10 minutes further away than the nearest clinic and second and/or third nearest clinic closes within the same event window ($k < 5$ or $k > -5$ and $k \neq 0$) as the nearest clinic. If closing year in second or third nearest clinic is in the same year ($k = 0$) as the nearest clinic, the municipality is not excluded.

(b) Exclude if second and/or third nearest clinic is less than 10 minutes further away than the nearest clinic and second and/or third nearest clinic opens or reopens within the same event window ($k < 5$ or $k > -5$) as the nearest clinic.

4. **If the nearest clinic is a non-closing clinic**, the municipality could be a valid control municipality. I again exclude a potential control municipality based on the characteristics of the second and third nearest clinics. If the municipality is not excluded by the following characteristics, the municipality is categorized as a control municipality.

   (a) Exclude if second and/or third nearest clinic is less than 10 minutes further away than the nearest clinic and second and/or third nearest clinic closes/opens or reopens. This is to ensure that a municipality classified as an unaffected control municipality is in fact unaffected and not treated by nearby closures/openings/reopenings.

5. **If the nearest clinic opens or reopens after 1996**, I exclude the municipality.

6. **If the nearest clinic closes before 2000**, I exclude the municipality as I want a pre-event window of at least four years.

## A.4 Definitions of diagnosis and procedures

**Breast cancer diagnosis**

*Definition: Breast cancer as main diagnosis (in Denmark "aktionsdiagnose")*

*Diagnosis code (ICD-10): DC50*

**Primary breast cancer surgery**

*Definition:* Lumpectomies and mastectomies in breast

*Procedure codes:*

Lumpectomy: KHAB40

Mastectomy: KHAC10, KHAC15, KHAC20, KHAC25

**Re-surgery**

*Definition:* Re-surgery after primary breast surgery due to complication within 30 days after primary surgery.

*Procedure codes:* KHWA, KHWB, KHWC, KHWD, KHWE, KHWF, KHWW

*Source:* Sundhedsstyrelsen (2005): NCSP, Nomesko Classification of Surgical Procedures Munksgaard Danmark (red Ole B. Larsen and Gunnar Schiøler)

**Use of opioids**

*Definition:* Prescribed opiods within 30 days from surgery

*ATC-codes:* N02A

**Neoadjuvant and adjuvant treatment**

*Definition:* Treated with radiation therapy or chemotherapy in the six months before and/or in the 12 months after primary breast cancer surgery.

*Procedure codes:*

Radiation therapy: BWGC, BWGE, BWGG, BWGJ, BNGD, BWHD, BNGC1, BHGA, BWGA

Chemotherapy: BOHJ, BWHA, BWHB, BWHC, BOHE, BJCZ01, BJHE11, BJHE12, BHHK, BHHL

*Source:* Sundhedsdatastyrelsen (2016): Faktaanalyse kræft, page 38.

# B Appendix B: Extended regression table for table 3

Table 5: Results from parametric event study specification by outcome and event time. *Interpretation: Deviation from linear pre-trend in the outcome*

| | Hospitalisation days | Outpatient treatment days | Mastectomy share | Re-surgery share | Two-year mortality rate | Prescribed opioids share |
|---|---|---|---|---|---|---|
| Pre-trend | -0.34*** | 0.20*** | -0.03*** | -0.00 | -0.01 | 0.00 |
| | [0.10] | [0.05] | [0.01] | [0.00] | [0.00] | [0.00] |
| Year 0 | -0.58*** | 0.01 | -0.04 | 0.01 | -0.01 | -0.00 |
| | [0.20] | [0.11] | [0.03] | [0.01] | [0.02] | [0.01] |
| Year 1 | -1.18*** | -0.06 | -0.07 | 0.00 | -0.00 | 0.00 |
| | [0.42] | [0.17] | [0.04] | [0.02] | [0.02] | [0.01] |
| Year 2 | -1.36** | 0.11 | -0.10** | -0.00 | -0.00 | -0.02 |
| | [0.50] | [0.29] | [0.05] | [0.02] | [0.02] | [0.02] |
| Year 3 | -1.37** | 0.29 | -0.09* | 0.02 | -0.01 | -0.01 |
| | [0.55] | [0.39] | [0.05] | [0.02] | [0.02] | [0.02] |
| Year 4 | -1.62** | 0.58 | -0.13* | 0.01 | -0.00 | 0.00 |
| | [0.66] | [0.48] | [0.07] | [0.02] | [0.02] | [0.02] |
| Year 5 | -1.76** | 0.42 | -0.15* | -0.01 | -0.00 | 0.02 |
| | [0.80] | [0.55] | [0.08] | [0.03] | [0.03] | [0.03] |
| Age | 0.03*** | -0.01*** | 0.00*** | 0.00 | 0.00*** | 0.00** |
| | [0.00] | [0.00] | [0.00] | [0.00] | [0.00] | [0.00] |
| Foreign born | 0.24 | 0.05 | 0.04 | 0.02 | -0.02** | -0.01 |
| | [0.14] | [0.11] | [0.03] | [0.01] | [0.01] | [0.01] |
| Married | -0.42*** | 0.04 | -0.05*** | 0.00 | -0.03*** | -0.01** |
| | [0.08] | [0.04] | [0.01] | [0.01] | [0.01] | [0.01] |
| Cohabiting | -0.50*** | -0.11* | -0.04 | -0.01 | -0.01 | -0.02 |
| | [0.10] | [0.06] | [0.02] | [0.01] | [0.01] | [0.01] |
| Charlson index 1 | 0.14 | 0.01 | 0.03 | 0.02 | 0.03** | 0.06*** |
| | [0.11] | [0.08] | [0.02] | [0.02] | [0.01] | [0.01] |
| Charlson index 2 | 0.28*** | -0.06 | 0.11*** | 0.01 | 0.07*** | 0.01 |
| | [0.10] | [0.05] | [0.02] | [0.01] | [0.01] | [0.01] |
| Previous surgery | -0.13 | -0.04 | 0.09*** | 0.00 | -0.01 | 0.01 |
| | [0.17] | [0.08] | [0.03] | [0.01] | [0.01] | [0.01] |
| *Education* | | | | | | |
| Upper secondary and vocational | -0.23*** | 0.01 | -0.02* | 0.01** | -0.00 | -0.02*** |
| | [0.05] | [0.03] | [0.01] | [0.00] | [0.01] | [0.01] |
| Short and middle further | -0.22*** | 0.08 | -0.04** | 0.01* | -0.02*** | -0.02** |
| | [0.07] | [0.05] | [0.02] | [0.01] | [0.01] | [0.01] |
| Long further | -0.28** | 0.08 | -0.05* | 0.03** | -0.04*** | -0.03* |
| | [0.12] | [0.15] | [0.03] | [0.01] | [0.01] | [0.01] |
| Constant | 3.28*** | 2.73*** | 0.37*** | -0.00 | -0.05 | 0.09*** |
| | [0.54] | [0.35] | [0.06] | [0.03] | [0.03] | [0.03] |
| Pre-closure mean | 5.98 | 2.72 | 0.70 | 0.04 | 0.08 | 0.07 |

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

Number of observations: 9790. Number of municipalities: 29. All models include municipality fixed effects and calendar year fixed effects.

# C Appendix C: Other outcomes

## C.1 Neoadjuvant and adjuvant treatment: Radiation therapy and chemotherapy

I examine whether the patient receives radiation therapy or chemotherapy in a period of six months before their primary breast cancer surgery to 12 months after their primary breast cancer surgery. Unfortunately, it has only been possible to follow radiation therapy and chemotherapy consistently over time from 2001 (and not 1996 as with the other outcomes). In the main sample, I follow municipalities at least four years before the first clinic closure enters. In this subsample, I follow municipalities at least two years before the first clinic closure enters, and therefore the first closure enters in year 2003 (compared to 2000 in the main sample). This reduces the sample from 9,790 observations to 5,418 observations, and the number of municipalities is reduced from 29 to 17 municipalities.

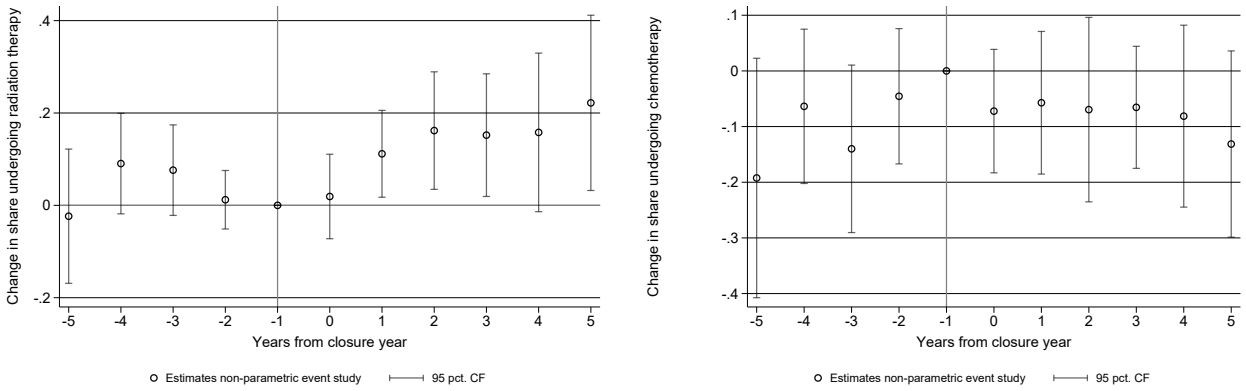
I examine the outcomes with non-parametric event study specifications as the non-parametric pre-event coefficients are not significantly different from zero. Compared to the main analyses, the coefficients are estimated on a smaller sample (fewer observations and fewer municipalities) and on a shorter pre event-window for some of the municipalities, and therefore I do not estimate the effects relative to a linear pre-trend.

Figure 10 and table 7 show strong tendencies towards an increase in the share of radiation therapy where the share increases with 11-16 pct. points in the first two years after closure (up from 51 pct.). The share of chemotherapy remains unchanged. For both outcomes, fairly wide confidence intervals are observed and the uncertainty is greater than in the main analyses.

Table 6: Outcomes in the year before a closure, $t = -1$)

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| Radiation therapy[a] | 0.51 | 0.50 | 0 | 1 |
| Chemotherapy[a] | 0.57 | 0.50 | 0 | 1 |

[a] Undergoes the treatment 6 months before and/or 12 months after primary breast cancer surgery
[b] Number of observations in $t = -1$: 496, Number of municipalities: 17




(a) (b)

Figure 10: The plotted coefficients are $\beta_k$ from the non-parametric event study. (a) Share receiving radiation therapy, (b) Share receiving chemotherapy. Observations: 5,418. Municipalities: 17.

Table 7: Results from non-parametric event study specification by outcome and event time.

| | Radiation therapy share | Chemotherapy share |
|---|---|---|
| Year 1 | 0.11*** | -0.06 |
| | [0.04] | [0.06] |
| Year 2 | 0.16*** | -0.07 |
| | [0.06] | [0.08] |
| Pre-closure mean $t-1$ | 0.51 | 0.57 |

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

Number of observations: 5,418. Number of municipalities: 17.

All models include municipality fixed effects, calendar year fixed effects and patient covariates. Standard errors are clustered on the municipality level and presented in brackets.

# D Appendix D: Robustness analyses

## D.1 Including a control group of unaffected municipalities

Some of the outcomes show a linear pre-trend. As the municipality fixed effects subsume linear trends in the event study setup, the effects of breast cancer clinic closures are in the main analysis estimated by comparing each post-period estimate to a linear pre-trend through a parametric event study specification. However, other possible solutions exists. To remove linear time trends in the outcomes, I now assume that the calendar time effects are identical in municipalities that experience a nearby closure and municipalities that do not. Hence, I now include municipalities where the nearest breast cancer clinic *does not close* as control municipalities and set all event time periods equal to 0 for the control municipalities. I then estimate a regression similar to the non-parametric event study specification in equation (1). The distribution of affected municipalities (treatment) and unaffected municipalities (control) is shown in figure 11. In figure 12, the results of estimating model (1) with inclusion of control municipalities are shown. In all plots in figure 12, the pre-trends are stable and do not indicate any anticipation effects or differences in time trends between treatment and control municipalities - except for the share of mastectomies. For the share of mastectomies, it does not seem to be a valid assumption that the time trend in the pre-period is the same in treatment and control municipalities.

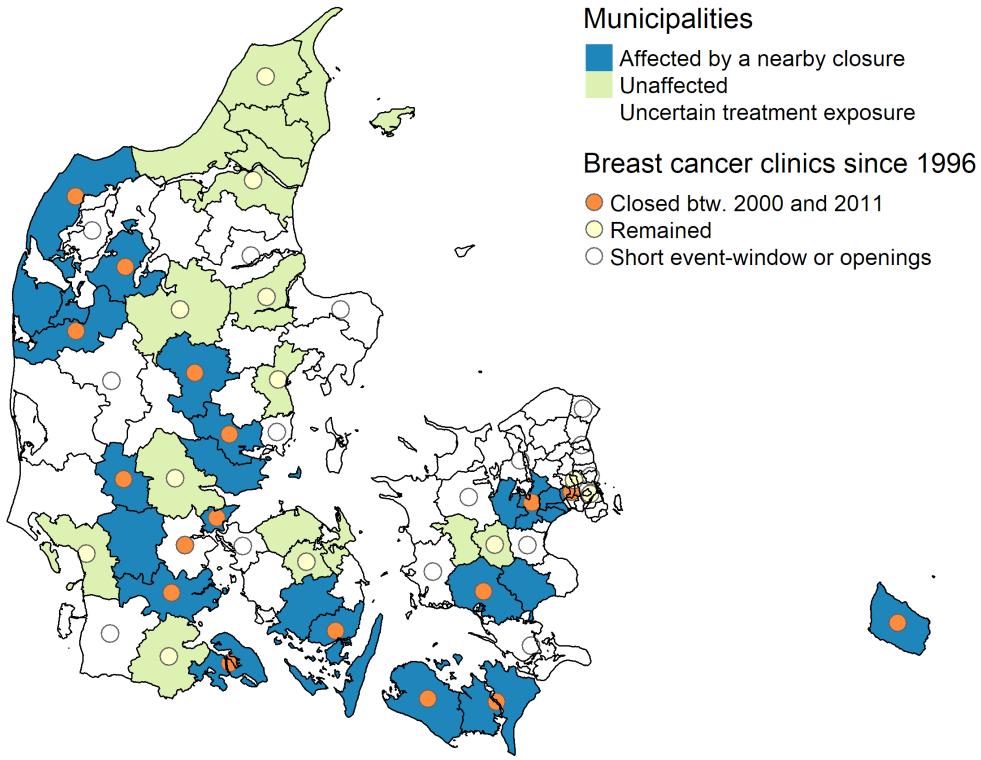



Figure 11: Treatment (closure-affected) and control (unaffected) municipalities

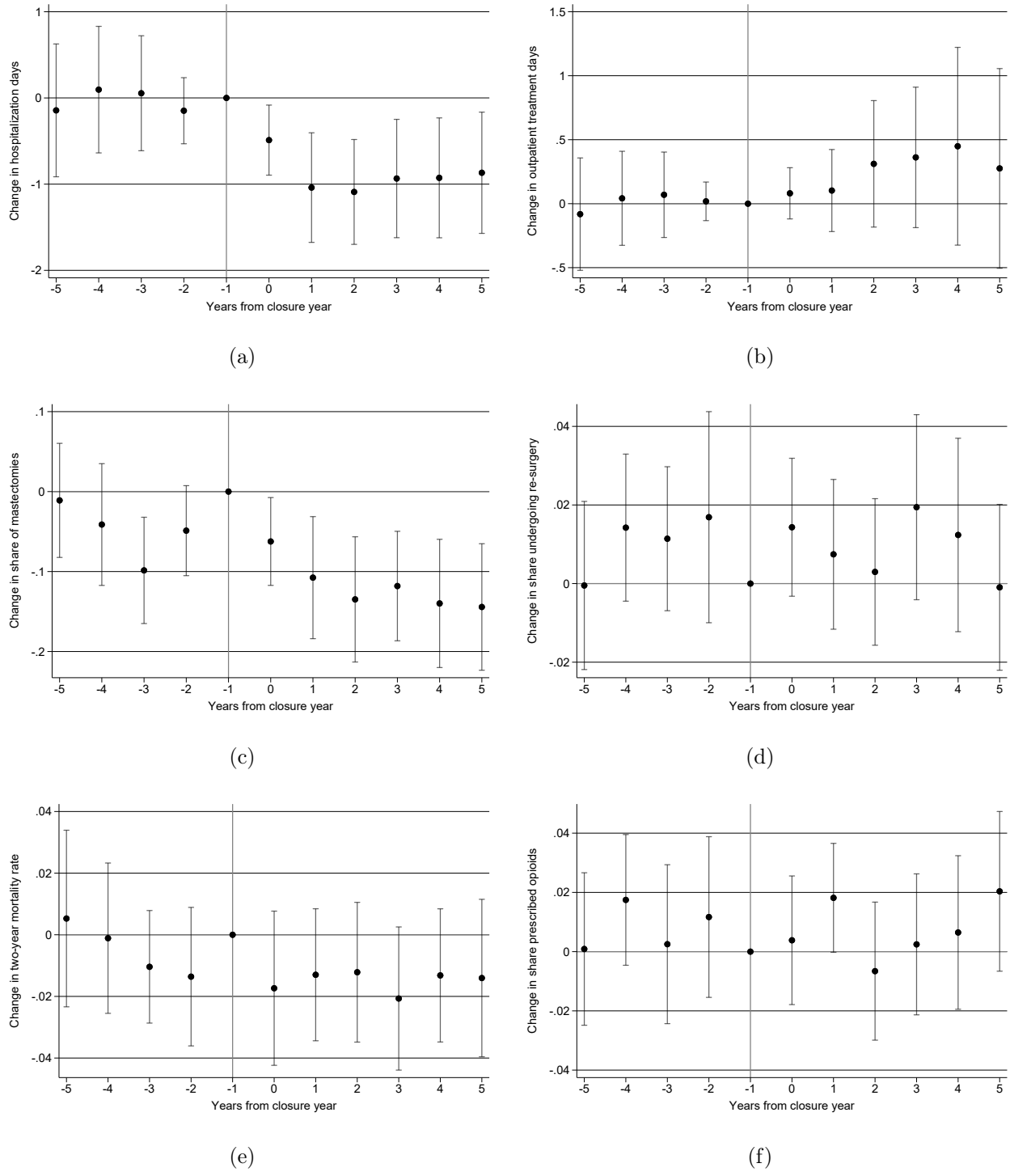



Figure 12: Non-parametric event study with a control group of closure *unaffected* municipalities for which $1(\textit{Time since closure}=k)_{mt} = 0$ for all $k$. The plotted coefficients are $\beta_k$ from the non-parametric specification. (a) Continual hospitalisation days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids within 30 days

## D.2 Excluding municipality fixed effects

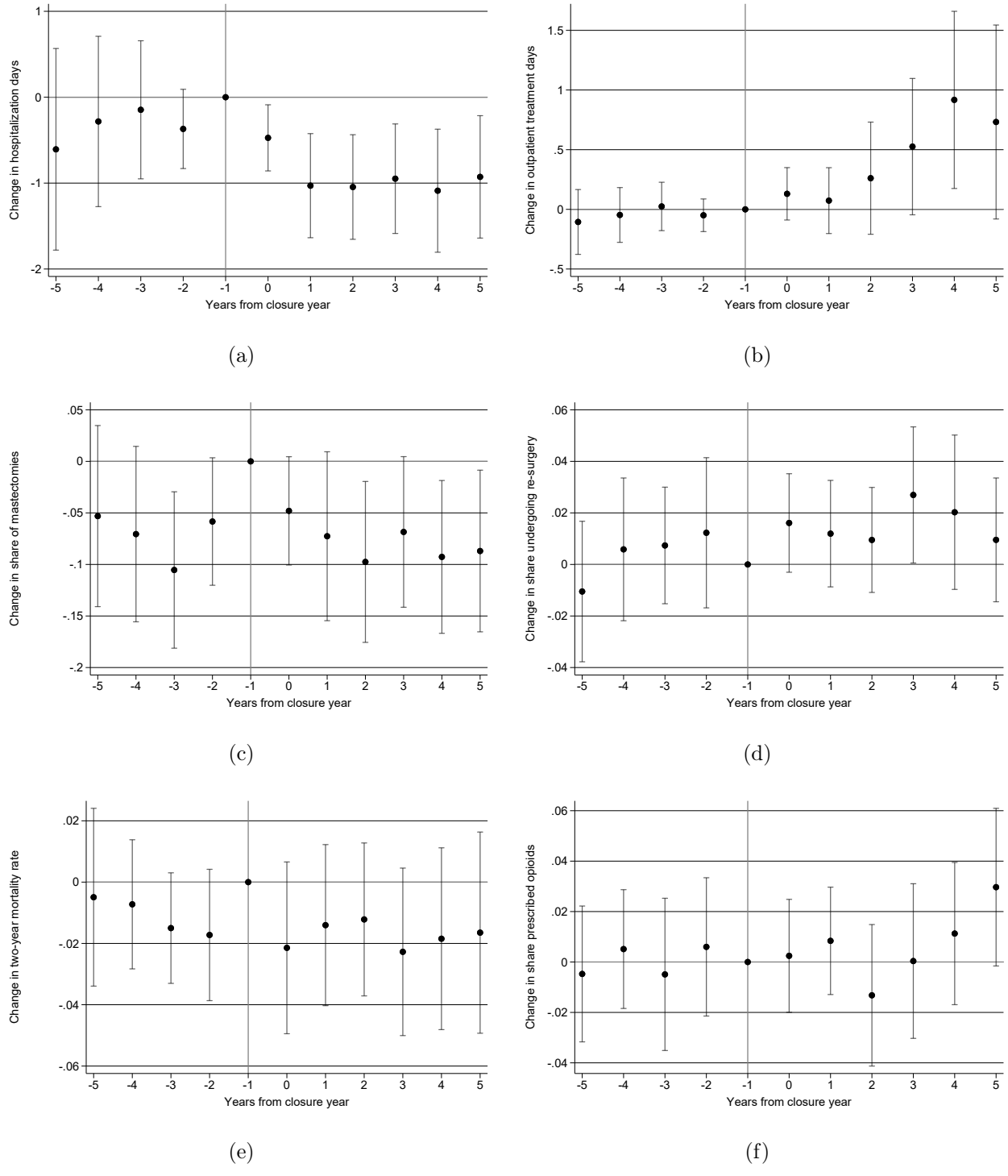



Figure 13: Non-parametric event study without municipality fixed effects. (a) Continual hospitalisation days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

## D.3 4 lags/leads

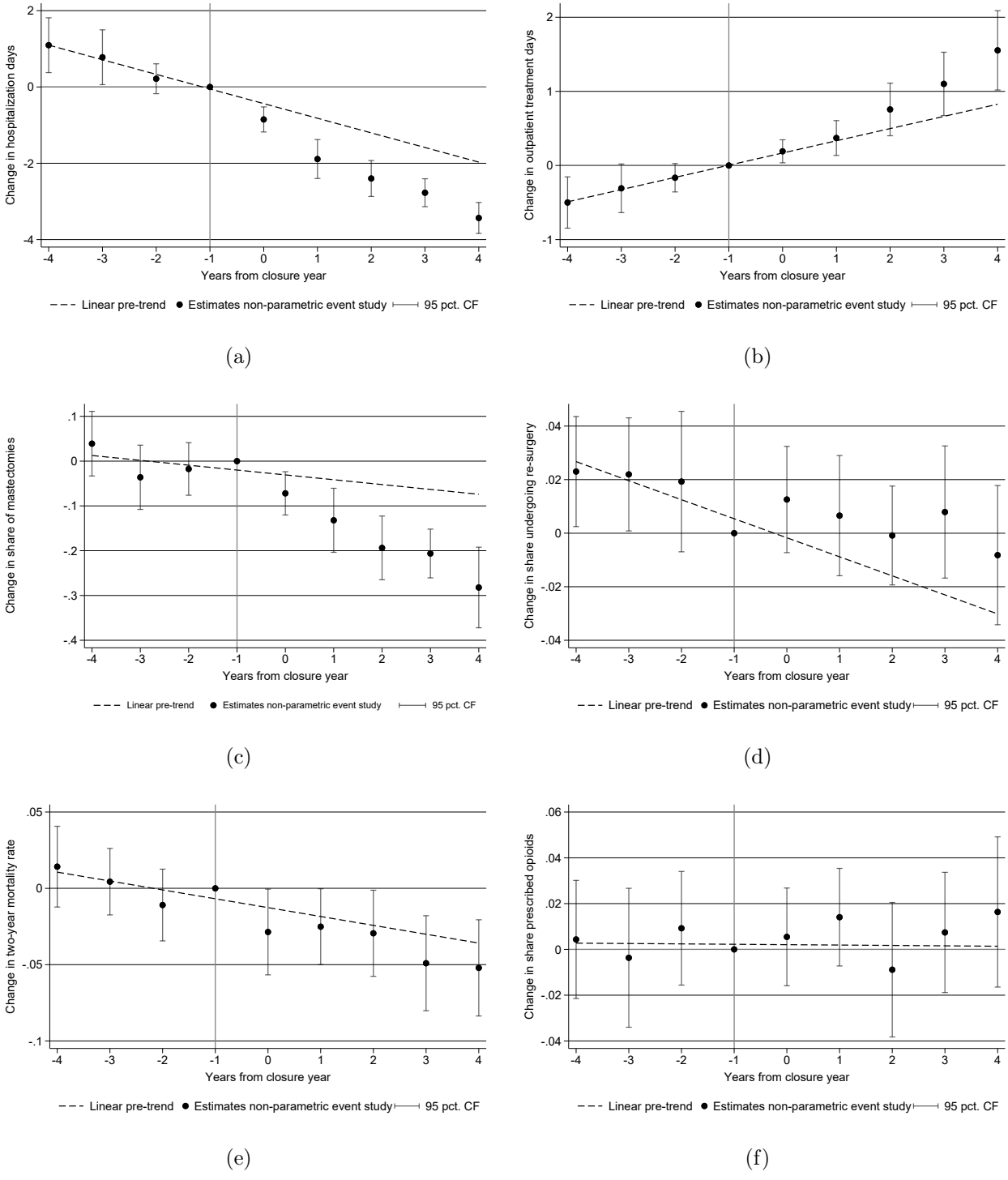



Figure 14: Fit of pre-trend to coefficients from non-parametric event study. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

## D.4 3 lags/leads

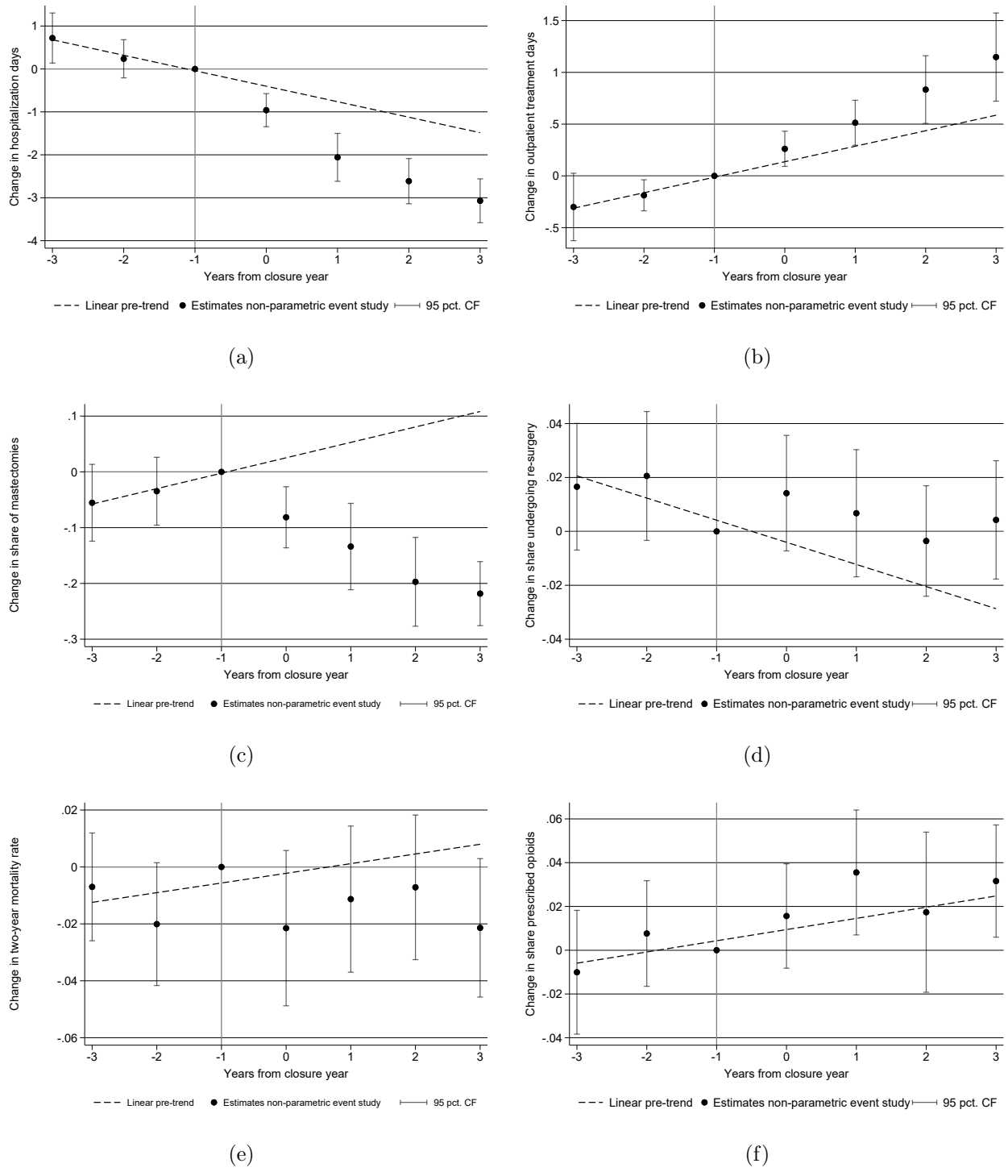



Figure 15: Fit of pre-trend to coefficients from non-parametric event study. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

## D.5 Distance to 2nd and 3rd nearest clinic is at least 30 minutes

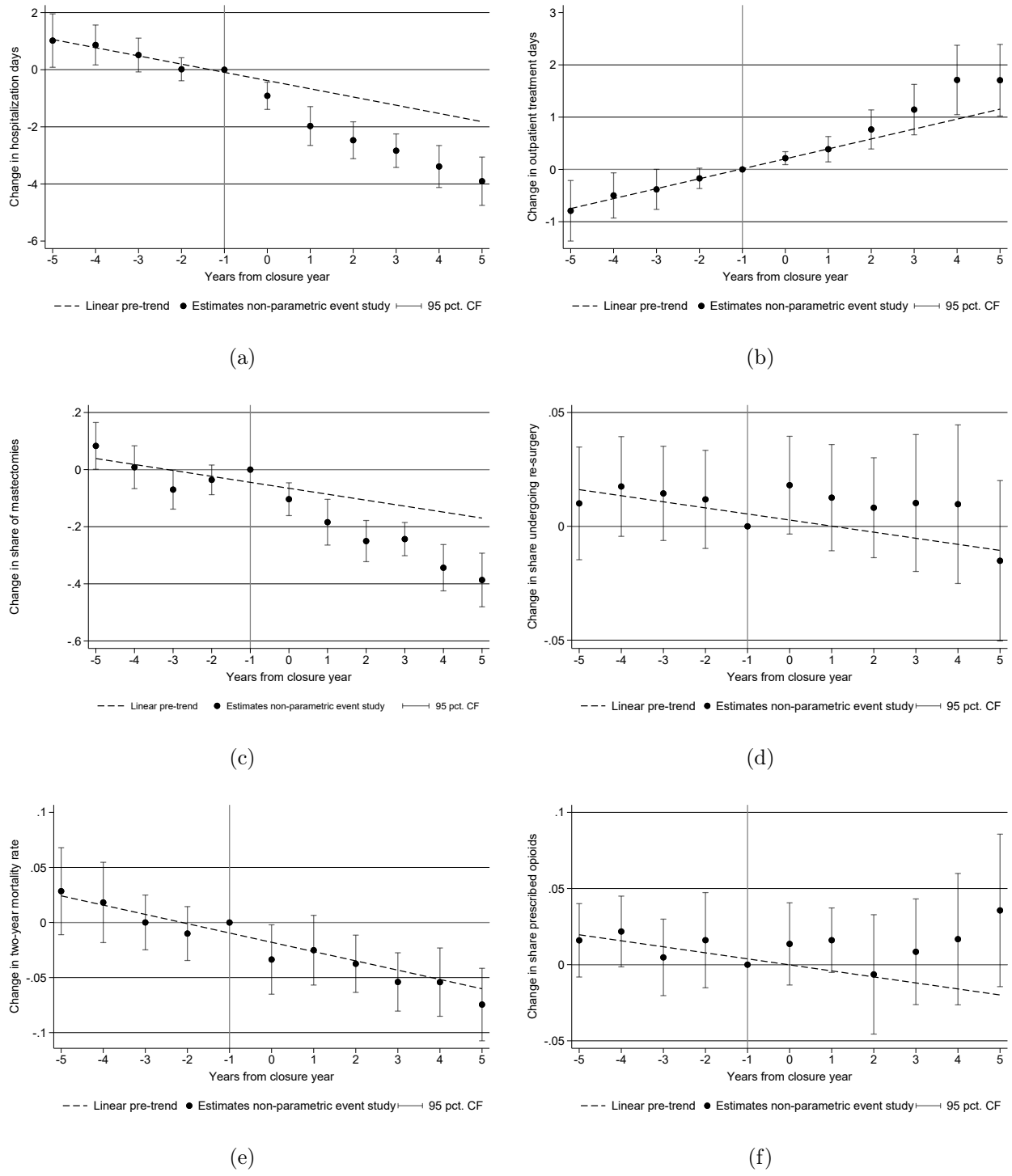



Figure 16: Fit of pre-trend to coefficients from non-parametric event study. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

## D.6 2nd and 3rd nearest clinic is not a non-closing clinic

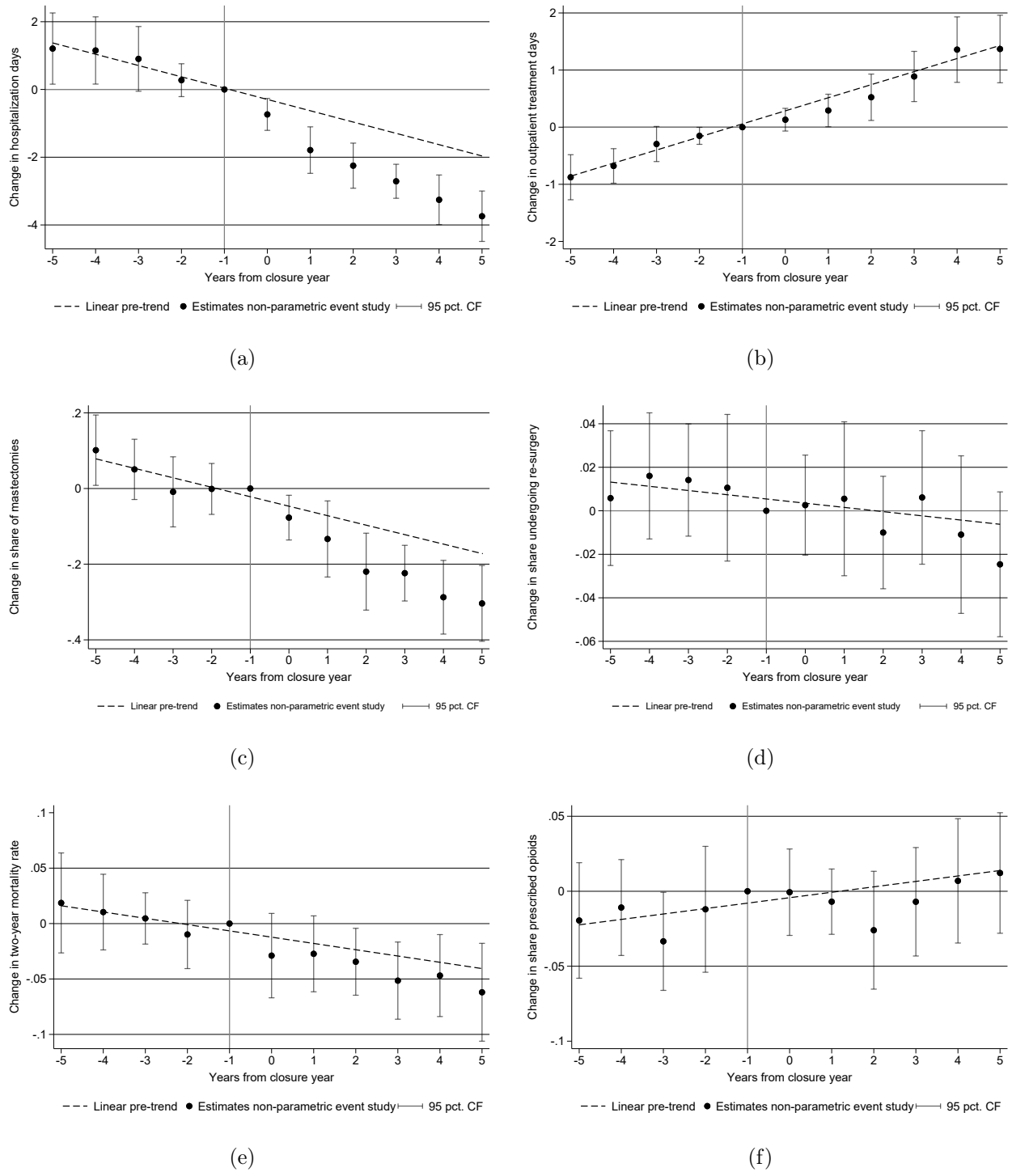



Figure 17: Fit of pre-trend to coefficients from non-parametric event study. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

## D.7 Exclude municipalities with a compliance rate <50 pct. in $k = -1$

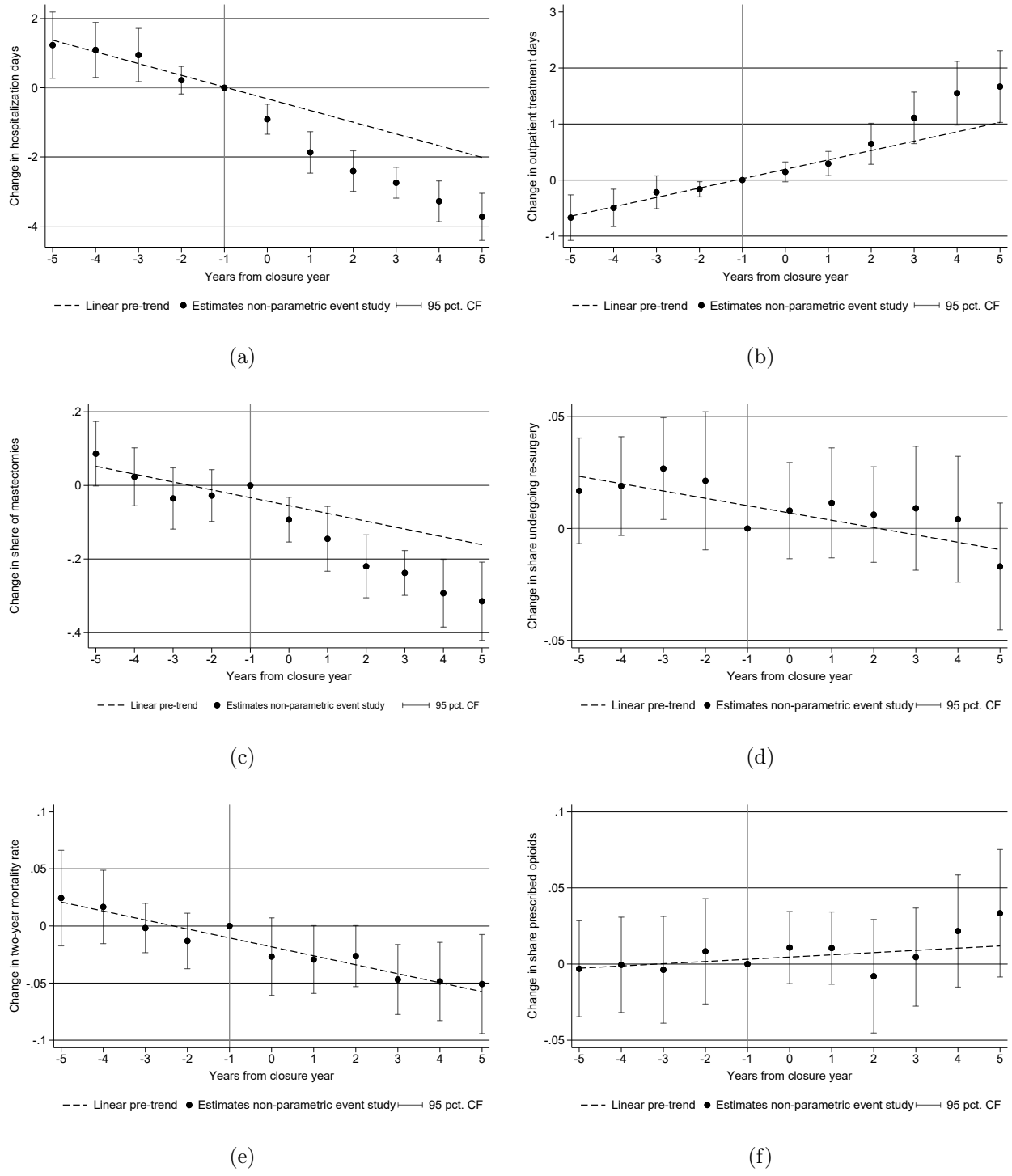



Figure 18: Fit of pre-trend to coefficients from non-parametric event study. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

## D.8 Exclude municipalities with a compliance rate <50 pct. in $k < 0$

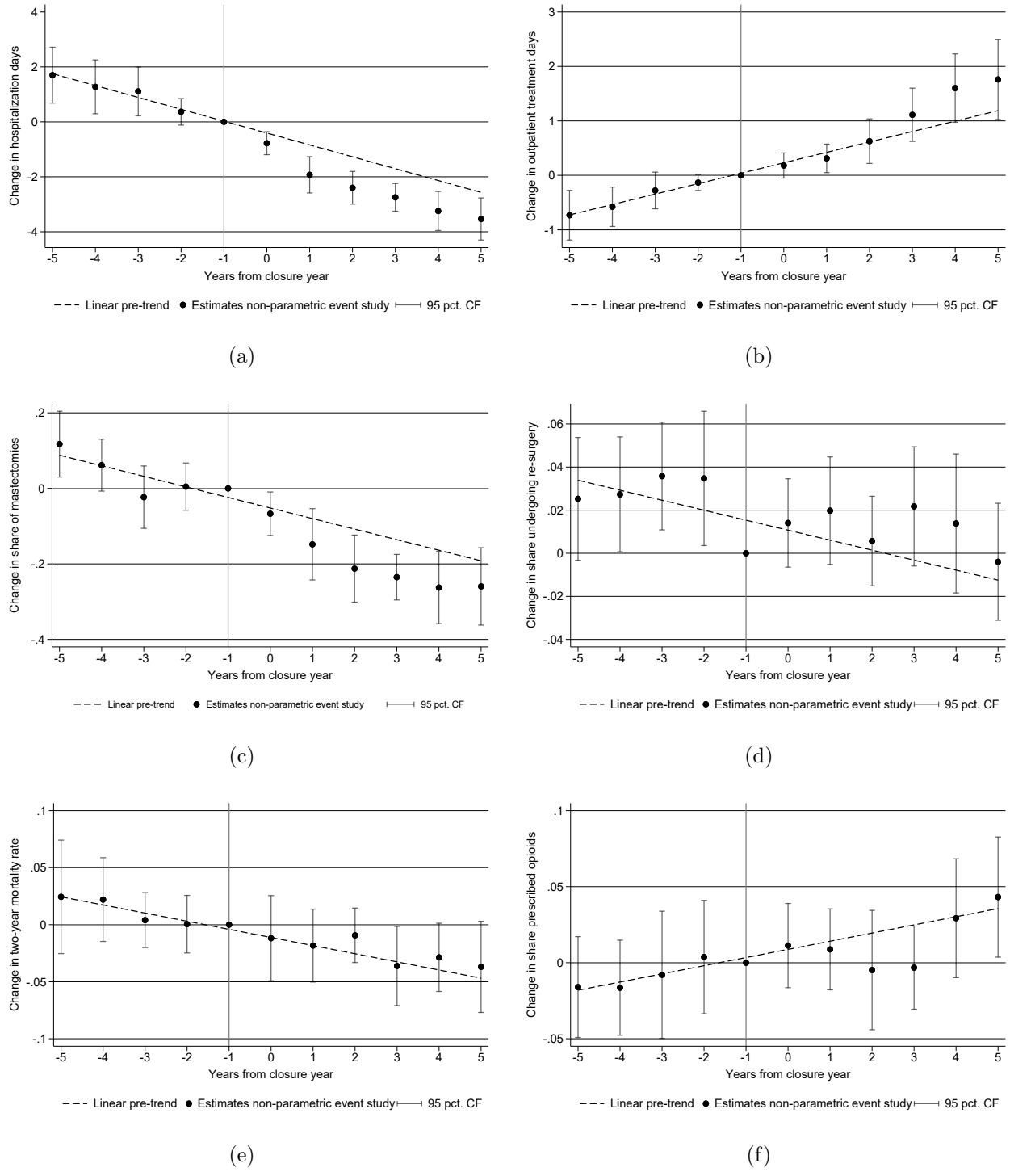



Figure 19: Fit of pre-trend to coefficients from non-parametric event study. (a) Continual hospitalization days post surgery, (b) Outpatient treatment days within 30 days, (c) Share receiving a mastectomy, (d) Share undergoing re-surgery, (e) Two-year mortality rate, (f) Share prescribed opioids.

# Chapter 3

# National clinical guidelines and treatment centralization do not guarantee consistency in health care delivery. A mixed-methods study of wet age-related macular degeneration treatment in Denmark

*With Sarah Wadmann, Toke Bek and Jakob Kjellberg*

# National clinical guidelines and treatment centralization do not guarantee consistency in healthcare delivery. A mixed-methods study of wet age-related macular degeneration treatment in Denmark

Anna Kollerup Iversen, Sarah Wadmann, Toke Bek and Jakob Kjellberg

December 2021

**Abstract**

As clinical practice variation has been problematized as a symptom of suboptimal care and inefficient resource spending, consistency in the delivery of healthcare is a recurring policy goal. We examine a case where the introduction of a new treatment is most likely to provide consistency in health care delivery because it was introduced with a national clinical practice guideline representing consensus about best clinical practice among leading clinicians, and because care delivery was highly centralized to few high-volume treatment units. Despite the consensus on best clinical practice and care centralization, this study shows pronounced regional variation in patient outcomes and treatment costs. Using a mixed-methods design, we find that the lack of consistency in care was largely unrelated to patient-specific characteristics, but seemed to reflect structural differences in the regional organization and financing of healthcare delivery. We conclude that the value of clinical practice guidelines is undermined when structural barriers limit the ability of clinicians and clinical managers to scale up treatment, and that some degree of decentralization may be a tool to maintain treatment intensity when the treatment effect is dependent on a high treatment intensity.

## 1 Introduction

Clinical practice variation has been of scientific interest the last forty years [1-4] and has been documented throughout OECD countries [5]. Variation has been problematized as a symptom of suboptimal care and inefficient resource spending [6-9] related to a risk of underuse, overuse or misuse of services [5]. Unwanted variation is a contested term and, given patients' varying preferences and the uncertainty involved in managing many clinical conditions, an acceptable level of variation is difficult to determine [7]. Policy goals for clinical practice are therefore often formulated as an ambition of obtaining 'consistency in the delivery of care', i.e. "patients with identical clinical conditions should receive identical care, irrespective of the healthcare professional, the healthcare institution, or the (socio-geographic) practice setting" [7].

Two common strategies to foster consistency in healthcare delivery are the development of clinical practice guidelines (CPGs) and the centralization of healthcare delivery in high-volume units. The CPGs strategy assumes that clinical practice variation stems from lacking or variable knowledge translation, such as clinicians' difficulties in keeping up-to-date with rapidly developing scientific knowledge [7,8]. The idea that centralization should foster quality and consistency is supported by a long list of theoretical mechanisms that can be broadly grouped into scale and scope effects [10]. Scale effects refer to the idea that larger units have more resources for establishing well-functioning infrastructures, for example more resources to improve care processes and quality control. Scope effects cover for example learning-by-doing and knowledge spillovers from teams working together [10]. Centralization may contribute to reaching consistency in care by reducing the number of departments delivering specialist services with low-skilled staff or in potentially unsafe circumstances [11,12].

This paper explores possible explanations for inconsistencies in healthcare delivery. Based on a longitudinal case study of anti-vascular endothelial growth factor (anti-VEGF) therapy in Danish ophthalmology specialist care, we investigate whether and how patient outcomes and medicine costs varied across geographical regions over time, and whether patient, professional or structural determinants can have contributed to such inconsistencies.

In 2007, the groundbreaking and costly anti-VEGF therapy was introduced for the treatment of wet age-related macular degeneration (wAMD) in Denmark, decreasing the incidence of legal blindness

attributable to wAMD substantially [13]. The introduction of anti-VEGF therapy can be considered a 'most likely' case [14,15] for the obtainment of consistency in healthcare delivery as it was introduced with two research-based strategies to ensure consistency. The therapy came with a national clinical guideline representing consensus about best practice among leading ophthalmologists, and care delivery was centralized to a few high-volume treatment units. Despite the consensus on best clinical practice and care centralization, this study shows a lack of consistency: regional variation in patient outcomes and medicine costs were pronounced, and the variation in patient outcomes increased over time. We find that the lack of consistency was largely unrelated to patient-specific characteristics, instead it seems to reflect regional differences in the preparedness and strategies for scaling up treatment activity to cope with substantial patient accumulation. We also find that the regions that managed to delegate and outsource treatment activity – rather than centralize – obtained higher treatment effects; presumably because this provided for higher treatment intensity.

We make empirical contributions to two strands of literature: 1) the literature on the drivers of clinical practice variation, and 2) the literature on the impact of research-based strategies to ensure consistency in care. To the first strand of literature, our main contribution is to employ a mixed-methods design, as previous studies examine drivers of clinical practice variation with either quantitative or qualitative methods. Previous quantitative methods range from purely descriptive to more advanced regression analyses [16,17], while qualitative studies include interview-based analyses, scoping reviews and theoretical discussions [7-9,18-20]. Quantitative studies often result in general decompositions with broad categories of determinants (e.g. patient and structural), while qualitative studies may point to case-specific explanations without being able to clearly isolate the contribution of one determinant from another. In this study, our mixed-methods design enables us to quantitatively assess to what extent patient-level data can explain regional variation in patient outcomes and medicine costs, while we qualitatively explore specific structural and professional-level determinants after having adjusted for differences in patient case-mix.

The second strand of literature on the impact of strategies to ensure consistency in care is sparse. Although studies suggest that centralization may contribute to reaching consistency in care [11,12], the empirical relationship between the number of units and the degree of consistency in care across units is largely undocumented. A few studies indicate that centralization of care temporarily implies more inconsistency in the delivery of care as patients living in rural areas may be exposed to increased travel distance to acute care necessary for survival [21,22]. For non-acute care, the empirical relation

between centralization and consistency in the delivery of care remains an open question. Only a few studies empirically evaluate the impact of and adherence to CPGs as a means of ensuring consistency in healthcare delivery [23-25]. To contribute to the literature on the impact of research-based strategies to provide consistency in care delivery, this study analyzes patient-level longitudinal data across eight years, and explores a most-likely case where we would expect to find consistency in care. The longitudinal data structure enables us to study the development between healthcare regions over time, and the patient-level data allows us to exploit the variation across patients within regions and avoid using averaged patient data that often mask patient heterogeneity and lead to incorrect conclusions about relative performance [26,27]. Moreover, previous studies are often based on survey data [25], while this study includes both clinical data, administrative data and qualitative data sources (documents and interviews).

## 2 Theoretical framework

### 2.1 Determinants of clinical practice variation

Inconsistency in healthcare delivery can be conceptualized as a result of demand- and supply-side determinants given by patient needs and preferences (demand), and structural factors and professional assessment by specialists (supply), see Figure 1.

On the demand side, a patient's need for a given treatment may stem from illness severity, the existence of comorbidity, or socioeconomic characteristics contributing to specific needs. For example, a patient's need for treatment depends on the patient's response to treatment (e.g. captured by illness severity, comorbidity and age) and the patient's investment in their own health (e.g. captured by educational level) [28,29]. Differences in patient demand may also stem from differences in patient preferences [30,31]. In the case of anti-VEGF treatment for wAMD, we expect differences in patient preferences to be limited, as the refusal of treatment may result in severe visual loss [32] and treatment improves prognosis markedly with a minimal risk of adverse effects [33]. Moreover, earlier studies have shown that differences in patient preferences do not explain a large part of the regional variation in specialist care [34,35].

On the supply side, studies suggest that important drivers of inconsistency in healthcare delivery at a professional level are competing evidence [20] and inadequate professional knowledge, but also the extent of research engagement [7,8], and skills, beliefs and attitudes among professionals [7,9]. Hence, inconsistency can be expected in "gray areas of medicine" where consensus about best

practice is not well-established, leading to clinicians adopting faith in different technologies or guidelines [35].

At the structural level, variation in clinical practice may reflect 'system failures' attributable to inadequately organized care practices [18,36] and failures among regulators to provide adequate behavioral change stimuli [7,9,18,36]. For example, professionals experience different resource and capacity constraints, regulatory structures and differences in healthcare teams and organizations [18]. Studies have shown that healthcare delivery is influenced by waiting lists, shortage of facilities, equipment and staff resources [8,9,19], where an inappropriate skills mix among staff [8] and a lack of operating rooms may cause professionals to introduce informal criteria to prioritize patients for surgery [19]. For ophthalmic patients, evidence shows that patients are exposed to risk from delays [37], and that treatment delay harms patients by permanently reducing vision [38].

Financial resources and payment systems may also affect clinical decisions and, in turn, cause inconsistency in healthcare delivery. Provider and patient incentives from payment systems are important for which and how many healthcare services patients with identical needs receive [39]. Such payment responses may reflect supplier-induced demand mechanisms, i.e. the provision of more services if fees are reduced or the number of procedures per patient in diagnostic-related group (DRG) rates are lowered [35].

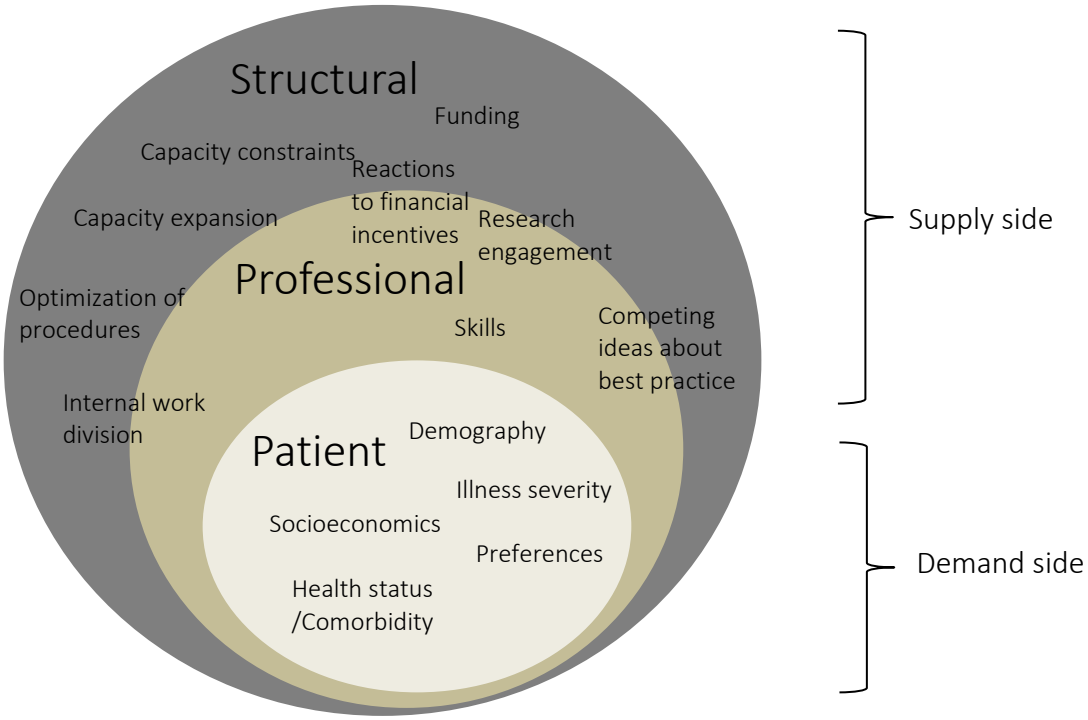



Figure 1: Determinants of clinical practice variation

### 2.2 Efficiency analysis to explore clinical practice variation

Efficiency analysis can be used to explore variation in performance across hospital units relative to a best practice, for example, lowest cost or best health outcome. Evaluating agents based on their performance relative to a best practice stems from the term 'yardstick competition' [40]. The modern definition of 'yardstick competition' was developed by Shleifer (1985) [40] to describe competitive incentives for local monopolies by regulating prices relative to the costs observed in other firms in the same industry. Although the analyses in this paper focus on the variation across non-competitive public hospitals units, Shleifer's thoughts on measuring efficiency relative to other units are generalizable to issues faced in public health care systems. For example, national health authorities call for reliable information about the relative performance of hospital units to identify scope for efficiency improvements and to inform debates about value for money [41,42].

As both demand- and supply-side factors are important for treatment costs and the health outcomes that a patient experiences, it is difficult for health authorities to distinguish between the contributions from different factors when comparing hospital units. The issues faced by health authorities are closely related to non-identical objectives and information asymmetries that originate in the principal-agent framework [43,44]. For example, imperfect information and different objectives arise when national health authorities (the principal) have difficulties in distinguishing between contributions made by hospital administrators and physicians (the agents) and other exogenous determinants outside the agents' control such as patient needs and preferences [44]. When concatenating the yardstick competition theory and the principal-agent framework, the principal's problem is to accurately assess agents' performance relative to best practice, for example, lowest costs or highest quality.

To accurately assess relative performance, one must limit information asymmetries and systematic heterogeneity, including factors over which organizations have little control [40,45]. Empirically, stochastic frontier analysis (SFA) has emerged as a regression method to control for exogenous heterogeneous influences on performance (e.g. medicine cost and outcome) that organizations (e.g. regional hospital units) face within an industry (e.g. the public hospital sector). In SFA, one controls for exogenous influences on performance and decomposes the residual into a random error and a component that is interpreted as a negative deviation from the efficiency frontier (inefficiency) [46]. The interpretation of this residual decomposition is that the random error captures unobservable factors outside the organization's control, for example, luck or unfavorable external events. The

component that is interpreted as a negative deviation from the efficiency frontier is a result of factors under the organization's control, for example structural or physician level factors [46]. Applications of SFA require a large number of observations when regressions are estimated at department or hospital level. However, in many health care systems, there will only be a small number of units to compare, for example, in countries with small populations like Denmark and in specialized care where treatment is concentrated to few units [45]. In the case of wAMD disease in Denmark, anti-VEGF treatment was centralized to few units, and to draw conclusions about relative efficiency, we follow applications of SFA where inference is improved by exploiting patient-level data [45]. We return to the empirical specification in section 3.2.

## 3 Data and methods

### 3.1 A mixed-methods framework

We integrate quantitative and qualitative methods for data collection and analysis in an explanatory, sequential mixed-methods design [47,48], see Figure 2.

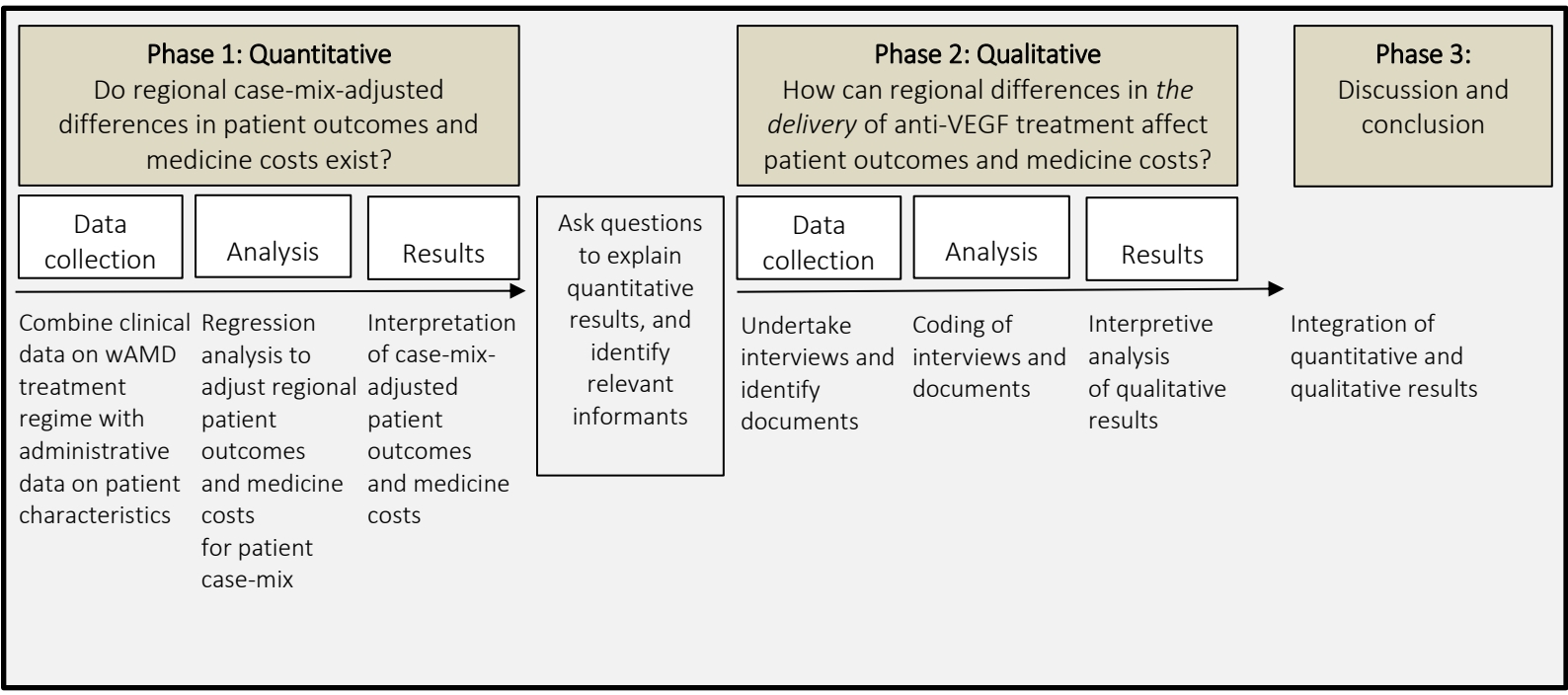



Figure 2: Sequential mixed-methods design

We first undertook regression analyses to examine whether variation could be attributed to regional differences in patient case-mix. To examine possible explanations for the observed variation that was not explained by patient case-mix, we did a qualitative analysis based on interviews and documents. The quantitative analyses allowed us to examine detailed patient-level data to describe patient case-

mix-adjusted regional variation (demand side), while the qualitative analyses enabled us to map regional variation in the organization and funding of anti-VEGF treatment to explore structural and professional-level determinants (supply-side).

## 3.2 Quantitative data and methods

### *3.2.1 Data and summary statistics*

The quantitative analyses combine a unique clinical data set on wAMD patients over the period 2008 to 2015 with Danish administrative data. The year 2008 was the first full treatment year, and 2015 was the last year in which all three participating regions provided clinical data on visual acuity. The clinical data include the number of anti-VEGF injections that each patient received during the first 12 months of treatment and their visual acuity at the beginning and the end of the 12-month period. The latter enables us to calculate the gain in visual acuity during the first 12 months of treatment and to adjust regional differences for baseline visual acuity. Adjusting for baseline visual acuity is very important as visual acuity rapidly decreases from the start of symptoms, and therefore the treatment effect depends on the time between the first symptoms and treatment onset. Visual acuity is measured in Early Treatment Diabetic Retinopathy Study (ETDRS) letters. In cases where the regions used other measures for visual acuity, the regions converted the visual acuity to ETDRS letters. The direct cost of the treatment over the 12-month period is given by the number of injections multiplied by the cost per injection. To be able to relate the medicine cost to the treatment intensity over a 12-month period, we only focus on the direct cost of the anti-VEGF treatment and not costs related to, for example, clinical facilities or productivity losses. From the administrative data, we extract patient covariates such as socioeconomic, demographic and health covariates (comorbidity).

The sample consists of 3,260 unique patients distributed across three regions. In Table *1*, we show the sample extraction criteria, and we show that the sample represents 20 pct. of the total Danish AMD population and 37 pct. of the population in the three included regions. Importantly, we asked the regions to submit data on the date and the visual acuity at treatment onset and at the first follow-up visit after 12 months. To ensure that we compare treatment courses of approximately the same length across the regions, we exclude patients for whom the 12-month follow up visit is much shorter or longer than 12 months (below 330 days or above 400 days). This sample restriction reduces the sample size a lot, so in the robustness analyses in section 4.2, we widen the 12-month interval to 265-465 days, and it does not change the interpretation of the results. In the discussion section 5.2, we discuss the implications of excluding patients without a 12-month follow up.

Table 1: Data extraction from population to sample

| Sample extraction criteria | Observations | Pct. of population | Pct. of population in the three included regions |
|---|---|---|---|
| 1. Population of AMD patients receiving minimum one injection in the period 2008-2015 in Denmark (distributed across five Danish regions) | 16,354 | | |
| 2. Population of AMD patients from three Danish regions 2008-2015 | 8,820 | 54 pct. | |
| 3. Patients are treated over approx. 12 months[a] and clinical data on start and end visual acuity is available | 3,301 | 20 pct. | 37 pct. |
| 4. First 12-month course of treatment per patient is included | 3,265 | 20 pct. | 37 pct. |
| 5. Administrative data available | 3,260 | 20 pct.[b] | 37 pct.[c] |
| **Final sample** | **Observations** | | |
| Region 1 | 1,334 | | |
| Region 2 | 1,425 | | |
| Region 3 | 501 | | |
| Total | 3,260 | | |

[a] 12 months is between 330 and 400 treatment days from measured start to end visual acuity
[b] Final sample represents 20 pct. of total population (3,260/16,354)
[c] Final sample represents 37 pct. of population in the three included regions (3,260/8,820)

In Table 2 and Table 3, we show summary statistics of the representativeness of the sample and differences across regions. Table 2 shows that the sample and population are similar with respect to gender, age and educational level while, in the sample, the share of foreign born is 1 pct. point lower and the share of married persons is 3 pct. points higher. For 9 out of 17 comorbidities, the five-year incidence prior to diagnosis is 1-2 pct. points lower in the sample. Table 3 shows regional variation in the outcome variables – the patient level costs over a 12-month period (EUR) and the patient-level change in visual acuity (Change in ETDRS letters). These regional differences in the outcome variables are also shown in Figure 3 where the regional differences are plotted for the time periods that we examine in the regression analyses, 2008-2011 and 2012-2015. The graphs highlight regional differences, and for changes in visual acuity over the 12-month observation period, the differences are amplified over time. In the second part of Table 3, we present variables extracted from the clinical data. We see that the regional differences in the outcome variables are driven by differences in visual acuity at treatment onset and after 12 months, and by the number of injections each patient is given. The medicine cost over 12 months is perfectly correlated with the number of injections because the cost is calculated as the number of injections multiplied by the price of an anti-VEGF injection in a given year. As we have the number of injections and the days between treatment onset and first visit after 12 months, we are also able to calculate the average treatment interval (the number of days between each injection). The regional variation in the average number of injections and the injection interval is illustrated in Figure 4, which shows that the injection interval is longer in region 3 compared to regions 1 and 2. This illustrates a lower treatment intensity and lower medicine costs in region 3 compared with regions 1 and 2. In the last part of Table 3, we also show the covariates that

we use to case-mix adjustment in the regression analyses. We find regional variation in the share of foreign born individuals, the share of married individuals and in educational levels. We also see that comorbidity five years prior to diagnosis is higher in region 1. This highlights the importance of adjusting for patient case-mix when comparing the regions. In the regression analyses, we also include the presence of each of the comorbid diseases presented in Table 2.

Table 2: Sample comparison to population mean

| | **Sample** | | **Population** | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Difference |
| Female | 0.64 | 0.48 | 0.64 | 0.48 | -0.00 |
| Age | 78.47 | 8.06 | 78.49 | 10.67 | 0.02 |
| Foreign born | 0.03 | 0.17 | 0.04 | 0.20 | 0.01** |
| Married | 0.45 | 0.50 | 0.42 | 0.49 | -0.03** |
| Single | 0.52 | 0.50 | 0.55 | 0.50 | 0.03** |
| Cohabiting | 0.03 | 0.18 | 0.03 | 0.17 | -0.00 |
| Lower secondary education | 0.44 | 0.50 | 0.44 | 0.50 | 0.00 |
| Upper secondary and vocational education | 0.39 | 0.49 | 0.40 | 0.49 | 0.00 |
| Short and middle further education | 0.13 | 0.33 | 0.12 | 0.32 | -0.01 |
| Long further education | 0.04 | 0.20 | 0.04 | 0.20 | -0.00 |
| *Comorbidity index in year of diagnosis* | | | | | |
| 0 | 0.91 | 0.29 | 0.84 | 0.37 | -0.07*** |
| 1 | 0.03 | 0.18 | 0.05 | 0.23 | 0.02*** |
| 2 | 0.06 | 0.23 | 0.10 | 0.31 | 0.05*** |
| *Comorbidity index 5 years prior to year of diagnosis* | | | | | |
| 0 | 0.77 | 0.42 | 0.72 | 0.45 | -0.04*** |
| 1 | 0.08 | 0.28 | 0.10 | 0.30 | 0.01* |
| 2 | 0.15 | 0.35 | 0.18 | 0.38 | 0.03*** |
| *5-year incidence prior to year of diagnosis* | | | | | |
| AMI (Acute Myocardial) | 0.03 | 0.17 | 0.04 | 0.20 | 0.01** |
| CHF (Congestive Heart) | 0.03 | 0.18 | 0.05 | 0.22 | 0.02*** |
| PVD (Peripheral Vascular) | 0.04 | 0.20 | 0.05 | 0.23 | 0.01** |
| CEVD (Cerebrovascular | 0.07 | 0.25 | 0.08 | 0.27 | 0.01* |
| Dementia | 0.01 | 0.08 | 0.02 | 0.13 | 0.01*** |
| COPD (Chronic Obstructive Pulmonary) | 0.07 | 0.26 | 0.09 | 0.29 | 0.02*** |
| Rheumatoid Disease | 0.02 | 0.15 | 0.02 | 0.16 | 0.00 |
| PUD (Peptic Ulcer) | 0.02 | 0.14 | 0.02 | 0.15 | 0.00 |
| Mild LD (Liver) | - | | - | | |
| Diabetes | 0.05 | 0.22 | 0.07 | 0.26 | 0.02*** |
| Diabetes + Complications | 0.01 | 0.12 | 0.03 | 0.16 | 0.01*** |
| HP/PAPL (Hemiplegia or Paraplegia) | - | | - | | |
| RD (Renal) | 0.01 | 0.10 | 0.02 | 0.14 | 0.01*** |
| Cancer | 0.10 | 0.30 | 0.11 | 0.31 | 0.01 |
| Moderate/Severe LD (Liver) | - | | - | | |
| Metastatic Cancer | 0.01 | 0.09 | 0.01 | 0.11 | 0.00* |
| AIDS | - | | - | | |
| Observations | 3,260 | | 16,354 | | |

Table 3: Sample characteristics across regions

| | **Region 1** | | **Region 2** | | **Region 3** | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Min | Max |
| ***Outcome variables from clinical data*** | | | | | | | | |
| Change in visual acuity over 12 months | 5.90 | 16.00 | 3.64 | 13.63 | 0.92 | 15.04 | -66 | 62 |
| Medicine costs per 12 months (EUR) | 7650.31 | 2080.42 | 9055.50 | 2706.36 | 4557.91 | 1666.68 | 841 | 21749 |
| | | | | | | | | |
| ***Variables from clinical data*** | | | | | | | | |
| Visual acuity at first visit | 55.69 | 15.37 | 59.94 | 13.97 | 59.63 | 14.94 | 5 | 90 |
| Visual acuity at first visit after 12 months | 61.59 | 17.04 | 63.58 | 16.16 | 60.54 | 17.37 | 3 | 95 |
| Number of anti-VEGF injections | 8.22 | 1.92 | 9.95 | 1.90 | 5.21 | 1.69 | 1 | 19 |
| Number of days between first inj. and first inj. after 12 months | 354.40 | 14.06 | 379.80 | 13.40 | 385.68 | 8.76 | 330 | 399 |
| First treatment in 2008 | 0.13 | 0.34 | 0.11 | 0.31 | 0.10 | 0.29 | 0 | 1 |
| First treatment in 2009 | 0.18 | 0.38 | 0.12 | 0.33 | 0.08 | 0.27 | 0 | 1 |
| First treatment in 2010 | 0.16 | 0.36 | 0.12 | 0.32 | 0.11 | 0.31 | 0 | 1 |
| First treatment in 2011 | 0.10 | 0.30 | 0.13 | 0.34 | 0.12 | 0.32 | 0 | 1 |
| First treatment in 2012 | 0.12 | 0.33 | 0.13 | 0.34 | 0.16 | 0.36 | 0 | 1 |

Continued table 3: Sample characteristics across regions

| | Region 1 | | Region 2 | | Region 3 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Min | Max |
| First treatment in 2013 | 0.11 | 0.32 | 0.12 | 0.33 | 0.13 | 0.34 | 0 | 1 |
| First treatment in 2014 | 0.10 | 0.30 | 0.15 | 0.35 | 0.14 | 0.34 | 0 | 1 |
| First treatment in 2015 | 0.09 | 0.29 | 0.13 | 0.33 | 0.17 | 0.38 | 0 | 1 |
| ***Covariates from administrative data*** | | | | | | | | |
| Female | 0.63 | 0.48 | 0.65 | 0.48 | 0.66 | 0.47 | 0 | 1 |
| Age | 78.42 | 8.31 | 78.65 | 7.91 | 78.09 | 7.81 | 28 | 102 |
| Foreign born | 0.05 | 0.21 | 0.02 | 0.13 | 0.02 | 0.14 | 0 | 1 |
| Married | 0.43 | 0.49 | 0.44 | 0.50 | 0.52 | 0.50 | 0 | 1 |
| Single | 0.54 | 0.50 | 0.53 | 0.50 | 0.45 | 0.50 | 0 | 1 |
| Cohabiting | 0.03 | 0.18 | 0.04 | 0.18 | 0.02 | 0.15 | 0 | 1 |
| Lower secondary education | 0.34 | 0.47 | 0.51 | 0.50 | 0.48 | 0.50 | 0 | 1 |
| Upper secondary and vocational education | 0.46 | 0.50 | 0.34 | 0.47 | 0.37 | 0.48 | 0 | 1 |
| Short and middle further education | 0.14 | 0.35 | 0.12 | 0.32 | 0.12 | 0.33 | 0 | 1 |
| Long further education | 0.06 | 0.24 | 0.03 | 0.17 | 0.03 | 0.17 | 0 | 1 |
| *Comorbidity index in year of diagnosis* | | | | | | | | |
| 0 | 0.90 | 0.30 | 0.92 | 0.27 | 0.90 | 0.31 | 0 | 1 |
| 1 | 0.04 | 0.19 | 0.03 | 0.16 | 0.04 | 0.20 | 0 | 1 |
| 2 | 0.06 | 0.24 | 0.05 | 0.23 | 0.06 | 0.24 | 0 | 1 |
| *Comorbidity index 5 years prior to year of diagnosis* | | | | | | | | |
| 0 | 0.75 | 0.43 | 0.77 | 0.42 | 0.79 | 0.41 | 0 | 1 |
| 1 | 0.08 | 0.27 | 0.09 | 0.28 | 0.09 | 0.28 | 0 | 1 |
| 2 | 0.16 | 0.37 | 0.14 | 0.35 | 0.12 | 0.33 | 0 | 1 |
| Observations | 1,334 | | 1,425 | | 501 | | | |

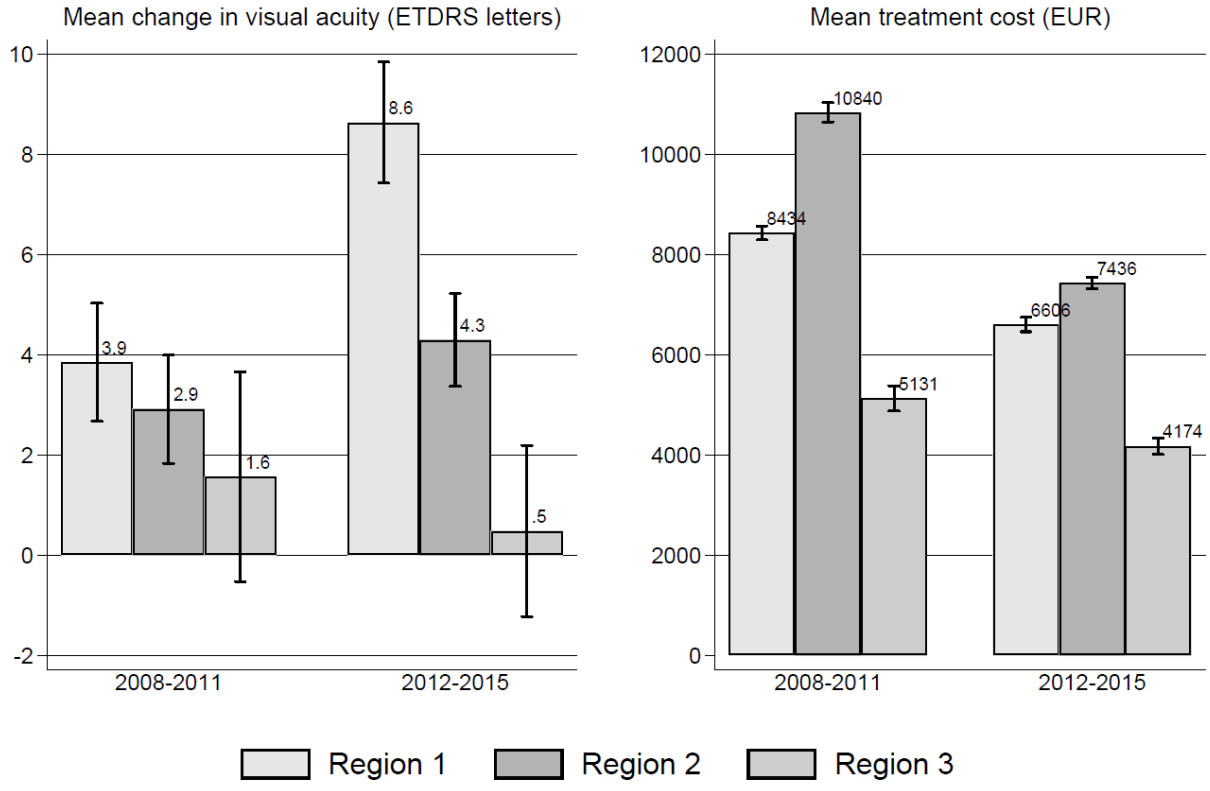



Figure 3: Outcome variables by time period and region - Mean change in visual acuity and mean medicine cost over 12 months

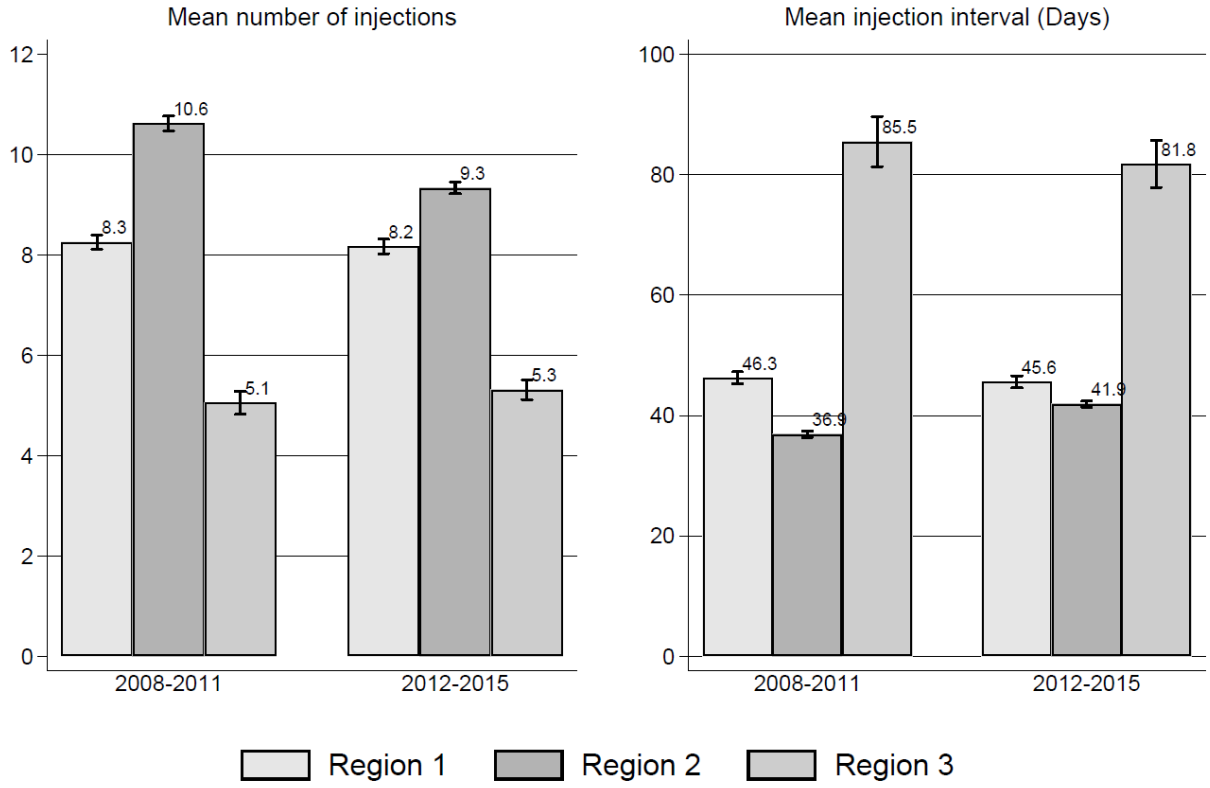



Figure 4: Mean number of injections and number of days between injections over 12 months by time period and region

### 3.2.2 *Quantitative method*

In this study, we use patient level data nested within regions where we follow previous studies specifying cost functions on patient-level data clustered within hospital lines, (see e.g. [16,45,49,50]). As we are able to exploit the variation across patients within regions, we compensate for the fact that we have a small number of comparators (3 regions) to assess statistical significance [45] and we avoid using averaged patient data, which often masks patient heterogeneity and leads to incorrect conclusions about relative performance [26,27].

With the patient-level data we are able to control for exogenous influences from patient heterogeneity (patient case-mix) on the outcomes, enabling us to statistically assess the regions relative to the region with lowest case-mix-adjusted cost or highest case-mix-adjusted quality gain. Inspired by Schmidt and Sickles (1984) [48] and Olsen and Street (2008) [45], we specify stochastic frontier production functions as fixed effects (FE) models with two different outcomes: 1) the patient level medicine costs over a 12-month period, 2) the patient-level gain in treatment outcome (change in visual acuity)

over a 12-month period. We will present the model for the first outcome in detail, and present the second more superficially, as the model specifications and assumptions are equivalent in the two models. For each outcome, the models are specified for two time periods, 2008-2011 and 2012-2015. We have chosen to aggregate data across the two four-year periods to reach a sufficient amount of observations for each region. Although estimating the models across two four-year time periods, we only include each patient once as the analyses are based on the patients' first 12-month course of treatment.

The model for the cost of anti-VEGF treatment is specified as a two-level multilevel model where patients are clustered within regions and where the residual is decomposed into a region effect $u_j$ and a residual patient effect $v_{ij}$. Hence, the medicine cost, $c_{ij,}$ of patient *i* treated in region *j* is given by

$$c_{ij} = \alpha + \beta x_{ij} + u_j + v_{ij}\,, \qquad i = 1, \dots, n, j = 1, \dots k \qquad (1)$$

where $\alpha$ is a constant term and $x_{ij}$ is a matrix of socioeconomic, demographic and health covariates of patient *i* receiving treatment in department *j.* The error term $v_{ij}$ captures the unexplained random variation in patient costs and is assumed *iid* with mean zero and constant variance, $\sigma_v^2$. The vector of interest is $u_j$ and consists of region effects. In the SFA terminology $u_j$ captures inefficiency, i.e. a negative deviation from the efficiency frontier as a result of factors that are under the regions' control [46]. We estimate equation $u_j$ with the least squares dummy variable (LSDV) estimator. We do this by keeping the constant and adding a dummy variable for *N-1* regions[1], and then estimate equation (1) with ordinary least squares (OLS). Then $u_j$ captures the difference in average patient-level cost in region *j* relative to the region left out of equation (1).

We have chosen to estimate equation (1) using FE specifications with an LSDV estimator as it provides us with explicit estimates of the region effects without making any distributional assumptions about the region effects. With this approach, there is no need to assume that the region effects are independent of the included patient covariates as the region effects will still be unbiased. Hence, the regions may act upon the risk of their patients, and it is also realistic that they do so. The

[1] Equivalently, one can apply a within transformation where regressand and regressors are expressed in deviations from their respective (group) means, or one can suppress the constant term and add a dummy variable for each of the N regions [51]. Some papers have chosen to set the weighted sum of the fixed effects equal to zero, and the fixed effects are then interpreted as deviations from an overall expected mean rather than from a baseline region, see e.g. [49,50].

mentioned reasons for using FE specifications emphasize the clear advantage compared to using a random effects (RE) specification. A RE specification requires both strong distributional assumptions about the region effects and requires that no correlation exists between the region effects and the patient covariates [51]. Again, especially the latter seems unreasonable, as the physicians within each region may have chosen their treatment strategy from the patient case-mix they face. Also, the FE specification has previously been validated as a reasonable choice compared to estimators used to estimate RE models (e.g. maximum likelihood and generalized least squares estimators) even when the number of organizations is low [45].

We argue that $u_j$ is an unbiased estimate of the true but unknown region effect if adjustment covariates are strictly exogenous $(v_{ij}|x_{1j}, \dots, x_{ij}) = 0$ [52]. Hence, to interpret the region effects as the true case-mix-adjusted systematic differences across regions, we assume that the rich set of adjustment covariates are exogenous and also sufficient to capture all patient variation - and thus leave only the remaining variation specific to regional treatment decisions. All adjustment covariates have been carefully chosen such that we only include covariates which are associated to pre-treatment characteristics of the patients. Therefore, we do not include any covariates which are within the control of the regions. For example, an endogenous variable would be number of visits to the ophthalmologist department.

From estimation of (1), we generate a set of region-specific intercepts derived as $\widehat{\alpha_j} = \hat{\alpha} + \widehat{u_j}$ , where the region that was left out in the regression is identified by the constant term $\alpha$. The region-specific intercept for region *j* can be interpreted as the average case-mix-adjusted patient-level cost in region *j.* In line with the SFA terminology, we define that the efficiency frontier is located by assuming that the region with the lowest region-specific intercept (lowest case-mix-adjusted costs), $\widehat{\alpha_j} = \hat{\alpha} + \widehat{u_j}$, is fully efficient and can serve as reference region. The new estimates of relative inefficiency are thus derived as deviations from the efficiency frontier

$$\widehat{u_j} = \widehat{\alpha_j} - \min_j(\widehat{\alpha_j}) \qquad (2)$$

Confidence intervals around the inefficiency estimates in equation (2) are derived by applying the multiple comparison with a control (MCC) approach [53-55]. The MCC approach takes into account the fact that each comparison region has the same control region in common. This is in contrast to previous studies that compare hospital-level effects with F-tests [50]. Specifically, the calculation of the confidence intervals follows Dunnett's two-sided MCC method [55], which provides

simultaneous confidence intervals to determine whether the costs in region *j* are significantly different from the costs in control region *k.* We set the control region *k* to be the region with the lowest costs, and the confidence intervals are then given by:

$$\widehat{u_j} - \widehat{u_k} \pm |d|\ \hat{\sigma}\sqrt{\frac{1}{n_j}+\frac{1}{n_k}} \qquad j = 1, \dots, k-1 \qquad (3)$$

where $\widehat{u_j}$ is the inefficiency estimate of region *j* and $\widehat{u_k}$ is the inefficiency estimate in the control region *k* and equals *0* as the control region *k* is assumed fully efficient. The critical value is $|d|$ and $\hat{\sigma}$ is the pooled standard deviation (RMSE). The critical value $|d|$ is the two-sided critical value reported by Dunnett (1964) where $|d|$ is adjusted for unequal sample sizes in region *j* and *k* [56]. In this paper, the critical values for a 95 pct. confidence interval lie between 2.13 and 2.23. Hence, the confidence intervals are wider than using the Student-t distribution and should detect more false-positives.

Similar to equation (1), we now specify a model where the quality given by the gain in visual acuity over a 12-month period is the dependent variable

$$q_{ij} = \gamma + \beta x_{ij} + \rho_j + \tau_{ij}\,, \qquad i = 1, \dots, n, j = 1, \dots, k \qquad (4)$$

All independent variables on the right hand side are equivalent to model (1). However, when calculating the estimates of relative inefficiency, we now compare the regions to the region which has the *highest* region-specific intercept, i.e. the region that exhibits the largest improvement in case-mix-adjusted visual acuity. Hence, the relative inefficiency estimates are calculated as

$$\widehat{\rho_j} = \widehat{\gamma_j} - \max_j(\widehat{\gamma_j}) \qquad (5)$$

and the confidence intervals are calculated as

$$\widehat{\rho_j} - \widehat{\rho_k} \pm |d|\ \hat{\sigma}\sqrt{\frac{1}{n_j}+\frac{1}{n_k}} \qquad j = 1, \dots, k-1 \qquad (6)$$

where k is the control region with the largest improvement in case-mix-adjusted visual acuity.

### 3.3 Qualitative materials and methods

Eight semi-structured telephone interviews were conducted with clinical and managerial leaders in the three regions during fall 2020 (see informant characteristics in appendix table A *1*). The interviews lasted 40-80 min. and were recorded with approval from the informants. Before the interviews, the informants were asked to prepare and share a timeline (2007-2015) identifying events they considered

important for patient outcomes and costs in their region (e.g. changes in clinical guidelines, procurement of new equipment, changes in funding mechanisms). To facilitate open dialogue, the informants were first invited to develop the points listed on their timeline. Subsequently, the interviews followed a semi-structured guide that included questions about capacity, care organization, and possible differences in clinical practice. During interviews, informants were encouraged to reflect on how the organization and funding of AMD diagnostics and treatment in their region might differ from other regions. When possible differences were identified, follow-up questions were posed to informants in the other regions, mainly through e-mail correspondence. To reduce the risk of recall bias, informants were encouraged to identify documents (e.g. e-mails or meeting minutes) to support their explanations. The documents that could be shared were included in the analysis along with interview transcripts as a means of triangulation [57]. We received 19 documents, including budget notes, letters from unit managers to hospital managers and regional managers, collaboration agreements, and meeting minutes from a national clinical network that developed national guidelines for AMD treatment. See document overview in Appendix table *A 3*.

The aim of the qualitative analysis was to identify regional differences in the funding, organization or delivery of anti-VEGF-treatment that may affect patient outcomes and medicine costs through a qualitative content analysis [58]. First, possible reasons for clinical practice variation were identified in existing literature and developed into a coding scheme (deductive approach) (Appendix table *A 2*). During the coding process, additional sub-codes were added to the scheme to account for new sub-themes identified by the informants (marked with bold in Appendix table *A 2*) (inductive approach). The analysis was structured with the resulting codes and sub-codes.

## 4 Results

### 4.1 Quantitative results

Table *4* shows the regional differences in medicine costs (in EUR) and treatment effect (change in visual acuity) over the 12-month period from 2008-2015, and separately for the periods 2008-2011 and 2012-2015. The regional differences are shown with and without case-mix adjustment, and the differences are assessed relative to the most efficient region (region with lowest costs and largest visual-acuity change). It appears that case-mix adjustment alters the magnitude and significance of some of the estimates, but the interpretation of the results remains the same: Differences in patient case-mix explain some of the regional variation, but not in a consistent manner, and even after case-

mix adjustment, the regional differences are noteworthy. The estimated regional differences in treatment outcomes are between one and eight ETDRS letters, and the mean visual acuity is 58 letters at treatment onset. Whether a change in visual acuity is of clinical importance depends on visual acuity at treatment onset. A change of, for example, fivet letters may be less important if the baseline visual acuity is very low, but of clinical importance if the improvement implies that a legal or practical threshold for better visual performance is crossed, e.g. keeping one's driving license. An improvement of five letters from the mean of 58 may ensure that the patient can maintain independence and does not need specialized visual aids such as low-vision magnifiers or electronics for visually impaired.

In the following, we only describe the case-mix-adjusted differences, as these differences reflect remaining structural or professional-level variation in the delivery of care. For a visual overview of the case-mix-adjusted differences, see Figure 5 and Figure 6. On average, **region 3** exhibits a significantly lower medicine cost per patient from 2008 to 2015, although the difference to **region 1** and **2** decreases over the period. On average, the mean medicine cost in region 3 was 4,558 EUR from 2008 to 2015 (Table 3), and in Table 4, we see that medicine costs are 2,409 to 5,753 EUR higher in **region 1 and 2** depending on the period. With respect to patient outcomes, **region 1** reaches a higher treatment effect over the period 2008 to 2015, which is driven by the last period 2012-2015. Hence, the regional differences in treatment outcome are amplified over time, with **region 1** being the region with highest treatment effect and **region 3** being the region with the lowest treatment effect. Combined with the cost estimates, we see that **region 1** on average has medium costs and highest treatment effect while **region 3** has lowest costs and lowest treatment effect.

Table 4: Regression results

| | Difference from region with lowest medicine costs over 12 months | | | | Difference from region with largest gain in ETDRS letters over 12 months (Absolute difference in letters) | | |
|---|---|---|---|---|---|---|---|
| | Region 1 | Region 2 | Region 3 | | Region 1 | Region 2 | Region 3 |
| **Raw** | | | | | | | |
| 2008-2015 (n=3,260) | 3092.4** (2829.4;3355.4) | 4497.6** (4236.9-4758.3) | Ref | | Ref | -2.3** (-3.5;-1.0) | -5.0** (-6.7;-3.3) |
| 2008-2011 ( n=1,641) | 3303.6** (2929.3.0;3677.8) | 5709.5** (5328.7-6090.3) | Ref | | Ref | -0.9 (-2.8;0.9) | -2.3 (-5.0;0.4) |
| 2012-2015 (n=1,619) | 2431.9** (2180.6;2683.2) | 3261.5** (3021.7;3501.4) | Ref | | Ref | -4.3** (-6.1;-2.6) | -8.2** (-10.4;-6.0) |
| **Case-mix-adjusted** | | | | | | | |
| 2008-2015 (n=3,260) | 3065.9** (2806.1;3325.8) | 4534.6** (4277.0; 4792.2) | Ref | | Ref | -0.7 (-1.9;0.5) | -3.6** (-5.2;-2.0) |
| 2008-2011 ( n=1,641) | 3316.5** (2946.1;3686.9) | 5753.1** (5376.1;6129.9) | Ref | | -0.4 (-2.1;1.4) | Ref | -2.0 (-4.6;0.7) |
| 2012-2015 (n=1,619) | 2408.5** (2158.5;2658.9) | 3277.9** (3039.0:3516.9) | Ref | | Ref | -2.5** (-4.1;-1.0) | -6.4** (-8.3;-4.4) |

***Significant difference from region with lowest cost at a 95% confidence level*

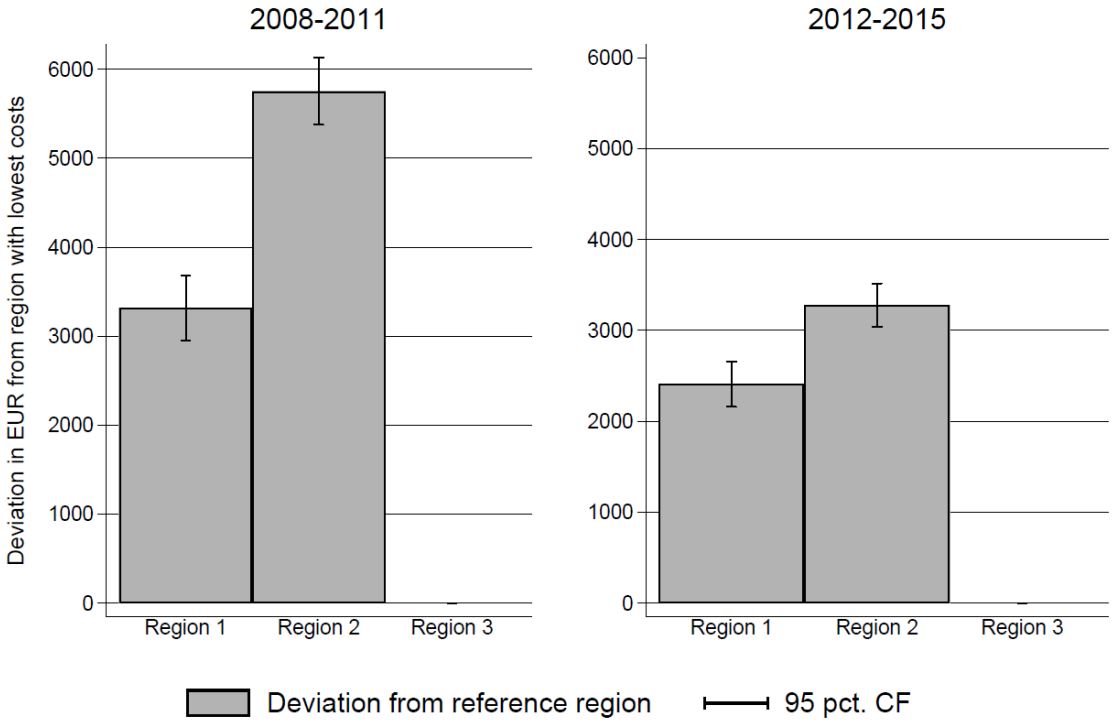



Figure 5: Case-mix-adjusted regression results. Deviation in costs from region with lowest individual-level cost over 12 months by time period and region

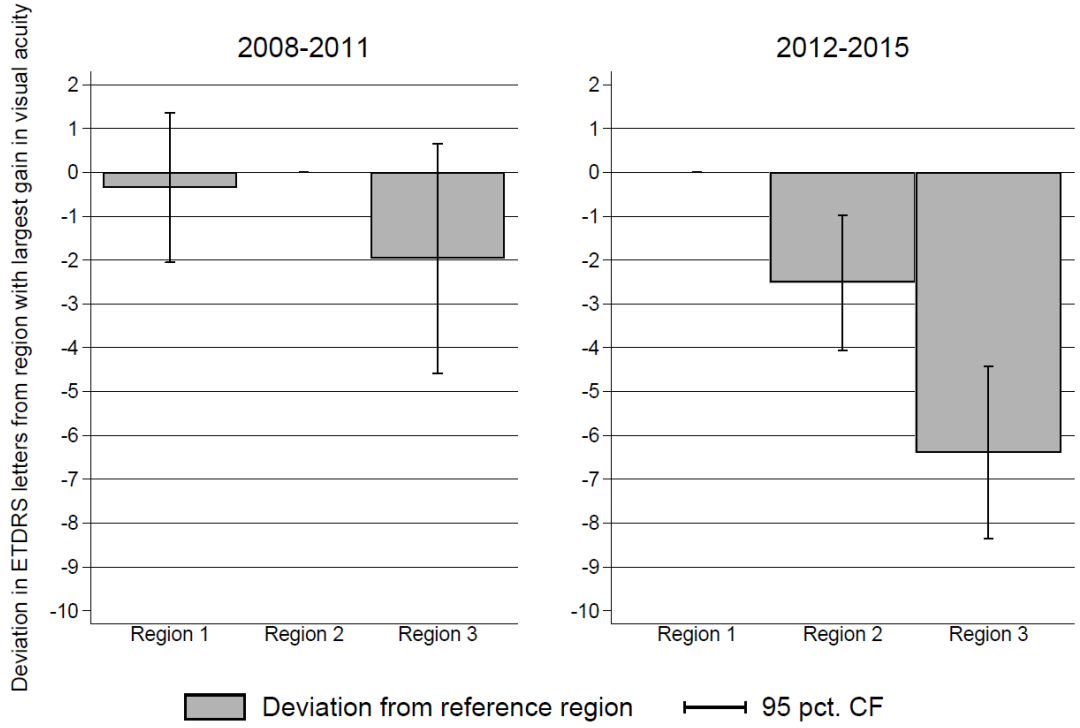



Figure 6: Case-mix-adjusted regression results. Deviation in ETDRS letters from region with largest gain in visual acuity over 12 months by time period and region

### 4.2 Robustness analyses

In the following, we examine the robustness of the results across three alternative specifications and sample restrictions.

*Sensitivity to chosen time periods*

In the main analyses, we estimate the regional differences for each of the periods 2008-2011 and 2012-2015. Therefore, we assume that within each four-year period, the regional variation is constant, but naturally, a region may perform better in 2008 than in 2011, although the regression is based on the full period 2008-2011. It might be more relevant to examine, for example, four time periods of 2-years, i.e. 2008-2009, 2010-2011, 2012-2013, 2014-2015. In the main analyses, we have chosen two periods of four years as we want a sufficient amount of observations in each region for each time period to be able to adjust for patient case-mix. Moreover, we are interested in whether there is a general tendency to more or less variation over time, which we are able to examine with the two four-year periods, and we are not interested in year-to-year comparisons per se. However, we still find it relevant to examine whether the interpretation of our main results is sensitive to the choice of time periods. Therefore, we have estimated the models with four time periods. Table 5 shows us that the interpretation of the results still holds with four periods as there is a tendency towards less cost-variation over time but more variation in the visual acuity gain.

*Sensitivity to length of treatment period*

To compare patients treated for approximately the same amount of time, we have restricted the sample to only include patients for whom the duration of the first course of treatment is approximately 12 months (330-400 days). As seen in Table 1 this excludes more than half of the patients from the three regions, and therefore we are interested in examining how sensitive the results are to this restriction. We do this by including patients with a first-time course of treatment between 265 and 465 days. When doing this, we see from Table 5 that the cost differences are slightly lower and the quality differences are slightly higher than in the main analyses. However, the interpretation of the results does not change. We discuss this restriction in more detail in the discussion section.

*Sensitivity to cost distribution*

As the effect of a given covariate is not necessarily linear in medicine costs, we also examine a log-transformed cost distribution. For example, individuals of high age may receive more injections than individuals of lower ages. Moreover, the log transformation removes the sensitivity of outliers arising

from patients with very high or low costs. From the results in Table 5, the interpretation of the results does not change when applying the log-transformed cost distribution.

Table 5: Robustness analyses. All results are adjusted for patient case-mix.

| | Difference from region with lowest medicine costs over 12 months | | | | Difference from region with largest gain in ETDRS letters over 12 months (Absolute difference in ETDRS letters) | | |
|---|---|---|---|---|---|---|---|
| | Region 1 | Region 2 | Region 3 | | Region 1 | Region 2 | Region 3 |
| **Main results** | | | | | | | |
| 2008-2015 (n=3,260) | 3065.9** (2806.1;3325.8) | 4534.6** (4277.0;4792.2) | Ref | | Ref | -0.7 (-1.9;0.5) | -3.6** (-5.2;-2.0) |
| 2008-2011 ( n=1,641) | 3316.5** (2946.1;3686.9) | 5753.1** (5376.1;6129.9) | Ref | | -0.4 (-2.1;1.4) | Ref | -2.0 (-4.6;0.7) |
| 2012-2015 (n=1,619) | 2408.5** (2158.5;2658.9) | 3277.9** (3039.0:3516.9) | Ref | | Ref | -2.5** (-4.1;-1.0) | -6.4** (-8.3;-4.4) |
| **Robustness analysis 1: Sensitivity to chosen time periods** | | | | | | | |
| 2008-2009 (n=830) | 2563.3** (2047.5;3079.1) | 6579.0** (6043.1;7114.9) | Ref | | -1.6 (-5.4;2.3) | -0.5 (-5.4;4.4) | Ref |
| 2010-2011 (n=811) | 3934.0** (3495.3;4372.7) | 4815.2** (4377.9;5252.5) | Ref | | Ref | -0.9 (-3.2;1.5) | -5.3** (-8.6;-2.0) |
| 2012-2013 (n=817) | 2799.6** (2471.7;3127.5) | 3614.0** (3294.2;3933.9) | Ref | | Ref | -2.5** (-4.6;-0.3) | -6.2** (-9.0;-3.4) |
| 2014-2015 (n=802) | 1793.1** (1438.6;2102.6) | 2914.5** (2627.8;3201.1) | | | Ref | -2.6** (-4.9;-0.4) | -6.3** (-9.0;-3.5) |
| **Robustness analysis 2: Sensitivity to length of treatment period Keep if treatment end date is between 265 and 465 days after first treatment (compared to 330-400 days in main analysis)** | | | | | | | |
| 2008-2015 (n=5,469) | 2047.6** (1872.2;2223.0) | 3721.0** (3546.9;3895.2) | Ref | | Ref | -1.4** (-2.4;-0.4) | -4.0** (-5.0;2-.9) |
| 2008-2011 (n=2,652) | 2211.1** (1963.6;2458.5) | 4841.8** (4586.3;5097.4) | Ref | | Ref | -0.1 (-1.6;1.3) | -1.3 (-2.9;0.4) |
| 2012-2015 (n=2,817) | 1601.3** (1424.5;1778.0) | 2738.1** (2569.3;2906.9) | Ref | | Ref | -3.2** (-4.5;-1.9) | -7.1** (-8.5;-5.7) |
| **Robustness analysis 3: Sensitivity to cost distribution Convert costs into logarithmic form (a log-linear model)** | | | | | | | |
| 2008-2015 (n=3,260) | 0.54** (0.50;0.57) | 0.71** (0.68;0.75) | Ref | | | | |
| 2008-2011 ( n=1,641) | 0.53** (0.48;0.57) | 0.78** (0.74;0.82) | Ref | | | | |
| 2012-2015 (n=1,619) | 0.48** (0.44;0.52) | 0.62** (0.58;0.66) | Ref | | | | |

***Significant difference from region with lowest cost or highest patient outcome at a 95% confidence level*

### 4.3 Qualitative results

*Small variations in treatment principles*

According to the interviewees, the national guidelines provided clinical consensus on diagnostic criteria and treatment strategies. This was also the case when the national guidelines were changed in 2013 when a new product, Aflibercept (Eylea), was recommended as first line treatment instead of Ranibizumab (Lucentis) due to greater cost-effectiveness. We identified only small variations in treatment principles across the three regions. Overall, the treatment regime appeared slightly more intense in **region 1** compared to the other two regions. In the loading phase, the practice in **region 1**

was to prescribe a treatment series of three injections followed by a new treatment series of two doses in cases of persistent disease activity (Interviews 1-3). Subsequently, patients would be followed at regular control visits if no disease activity was identified (Interview 1, Doc. 18). In **region 2** and **region 3**, one dose would be prescribed at a time if disease activity persisted after the loading phase (Interviews 4, 6, 7). For some patients, treatment may have been continued for a longer time in **region 1** compared to the other regions. In cases when vision was stabilized and sub-retinal fluid was present, patients were treated actively in **region 1** (Interview 1, 5), whereas treatment would be paused and disease activity monitored regularly for all patients in **region 2** (Interview 5) and for some patients in **region 3** (Interview 7). It has not been possible to trace the identified variations in clinical practice to local treatment guidelines, as the hospitals could not retrieve historical guidelines.

*Capacity constitutes a major concern across the regions*

The medical success of the anti-VEGF treatment quickly came to constitute a major organizational challenge across all regions. Patients who had previously not received any specialist treatment turned into chronic patients who needed continuous treatment and control. Accordingly, patient volumes accumulated, turning AMD clinics into 'factories' focused on the 'organization of large-scale production' (Interview 5). According to the minutes from the annual meetings of the ophthalmology specialists, the capacity of the clinics in several regions was already under strain in 2008 (Doc. 16). Waiting time ensued. The specialists' meeting minutes report that the increasing volume of patients led to a prolongation of control intervals beyond four weeks (Doc 16, 17). The goal of anti-VEGF-treatment was to postpone the loss of vision for as long time as possible by halting the progressive disease activity of wAMD. Treatment results therefore depended on the ability to detect disease activity and initiate treatment as early as possible to prevent further vision loss. Prolonged control intervals may increase disease activity and some patients may therefore lose vision before re-treatment. In 2011, patient volumes increased even more as anti-VEGF treatment was approved for new indications. From March and June 2011, patients with diabetic macular edema (DME) and patients with central and branch retinal vein occlusions (CVO and BRVO) were also entitled to anti-VEGF treatment (Doc. 10). The minutes from the ophthalmology specialists' meeting in November 2011 report that the number of diabetes patients treated with anti-VEGF treatment varied considerably among the regions (Doc. 19). Our informants mainly attributed challenges of waiting times to limited personnel. It has not been possible to quantify differences in capacity (e.g. patient accumulation and limited personnel) due to incomparable data across the regions over time.

*Variation in regional responses to capacity challenges*

The introduction of anti-VEGF treatment led to patient accumulation in all regions. A detailed study of patient accumulation has been published for one of the regions included in this study [59]. However, the regions adopted somewhat different strategies to handle the capacity challenges.

The expansion of capacity occurred differently in the regions. **Region 1** gradually scaled up rooms, screening equipment and personnel at the main hospital from 2009 to 2014. From 2011-2015, satellite clinics were established to further increase capacity – in some clinics by means of telemedicine solutions (Interview 1, 3, doc.1). **Region 2** established a satellite function in 2007 and kept it throughout the study period (Interview 6, doc. 4). Until 2011, rooms, screening equipment and personnel at the main hospital were scaled up (Interviews 4-6, doc. 2). After this, when dealing with capacity constraints, management and clinicians had to rely on "creative thinking" and optimizing procedures (Interview 6). **Region 3** gradually scaled up rooms, screening equipment and personnel from 2008-2013. A temporary satellite clinic was established from 2010 to 2013.

Variations in internal work divisions meant that personnel resources were used differently among the regions. **Region 1** was quick to delegate tasks from specialist physicians to specially trained nurses. Relying on organizational set-ups and training programs from diabetic eye diseases, nurses became responsible for anti-VEGF injections from 2009. From 2012, nurses also undertook scans and evaluated the need for re-treatment under supervision of specialist physicians (Interview 1, 3, doc. 1). By contrast, **region 2** and **region 3** delegated few tasks from specialist physicians to nurses. In **region 2**, nurses undertook some scans from 2008, but specialist physicians evaluated scan results and gave injections (Interviews 4-6). In **Region 3**, early-career physicians were recruited to do injections while nurses conducted scans and specialist physicians evaluated scanning results (Interviews 7-8).

Organizational measures to reduce waiting time from scans to treatment were introduced at a different pace in the regions. This may reflect different reactions to economic incentives. The regional remuneration systems worked against efforts to reduce waiting time because the hospitals would be remunerated for one visit only instead of two (Interviews 1, 6). Despite this incentive, **region 1** and **region 2** decided to combine scans and treatment into one patient visit instead of two in 2008-2009 (Interviews 1-3, 5-6). This markedly reduced waiting time from diagnosis to treatment initiation and from control scans to re-treatment, thereby preventing further vision loss. In **region 3**, the economic

incentive contributed to sustaining a practice where scans and injections were conceived of and remunerated as different activities delivered at different points in time. Towards the end of the study period, Region 3 also combined scans and injections into one visit (Interviews 7-8).

*Funding mechanisms*

In all regions, the regional authorities reimbursed hospitals and departments 100% for anti-VEGF-medicine costs. Even if the remuneration systems were comparable, some interviewees indicated that different political priorities in the regions may have made it easier for some clinical managers to obtain resources for capacity expansion and more burdensome for others. In **region 1**, a political agreement prior to the implementation of the treatment meant that funds were secured early on for the gradual capacity expansion, whereas in **region 3,** local managers needed to continuously apply for resources.

*Research engagement*

Informants from **region 1** suggested that participation in clinical drug testing prepared them to engage in early dialog with hospital managers and regional authorities about funding and capacity (Interviews 2-3). We were unable to verify whether research engagement differed among the regions or if it had implications for the organization and funding of AMD treatment.

## 5 Discussion

### 5.1 Interpretation of results

We identify regional differences in treatment outcomes and medicine costs, and the differences in treatment outcomes are amplified over time. The regional differences persist when adjusting for patient case-mix across the regions, and therefore the regional differences most likely reflect structural or professional-level differences. Our results indicate that difficulties in reducing waiting lists and keeping up with desired treatment intensity stem from structural capacity constraints and lack of organizational preparedness to handle patient accumulation, rather than stemming from competing ideas about best practice or differences in payment systems. *See joint display of findings in Figure 7.*

| FINDINGS | | | |
|---|---|---|---|
| *QUANTITATIVE RESULTS* | *QUALITATIVE RESULTS* | *INTERPRETATION* | *CONCLUSIONS* |
| ▪ Regional variation in medicine costs and outcomes before and after adjustment for patient case-mix<br>▪ Regional variation in outcomes increased over time<br>▪ The magnitude of the variation is of clinical importance for the patients | Regional variation arises from<br>▪ Capacity constraints<br>▪ Lack of organizational preparedness<br>Solutions<br>▪ Delegation of tasks to nurses and residents<br>▪ Establishment of permanent satellite clinics<br>▪ Early and continued expansion of clinical facilities, equipment and personnel | ▪ Regional variation reflects differences in capacity and organizational preparedness to handle major patient accumulation<br>▪ Regional variation does not reflect large differences in patient case-mix, competing ideas about best practice or differences in payment systems | ▪ The value of CPGs is undermined when structural conditions do not support the organizational changes necessary to ensure optimal treatment<br>▪ The ability to delegate treatment responsibilities may be important – in particular when the treatment outcome depends on high treatment intensity |

*INTEGRATION OF FINDINGS*

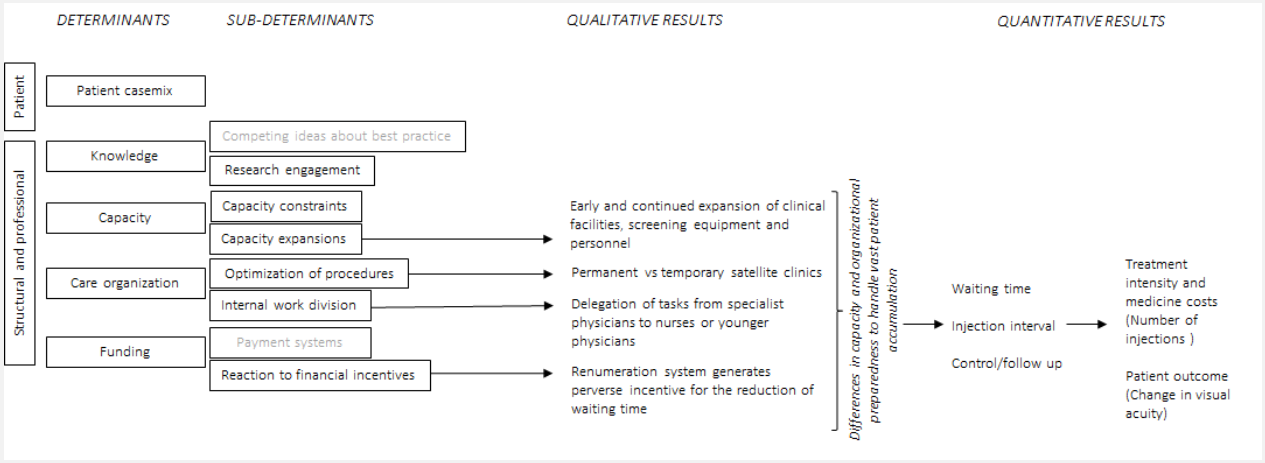



Figure 7: Joint display of findings

The results show that **region 1** might have been better equipped to expand capacity and organize work procedures. According to informants, **region 1** prepared the "organizational set-up" to handle patient accumulation through early dialog with hospital managers and regional authorities to secure funding and capacity (early and continued expansion of clinical facilities, equipment and personnel, establishment of permanent satellite clinics, delegation of tasks from specialist physicians to nurses). **Region 2** obtained treatment effects in between region 1 and region 3 from 2012 to 2015, but at the

highest cost. Although **region 2** took some of the same measures as region 1 (early and continued expansion of clinical facilities, equipment and personnel, and establishment of permanent satellite clinics), only few tasks were delegated from specialist physicians to nurses. In **region 3**, we observe the lowest medicine costs (i.e. fewer injections per patients) but also a lower treatment effect. Informants describe that they struggled with keeping waiting lists low, and, in turn, had difficulties ensuring early treatment initiation and short injection intervals due to capacity constraints. While a temporary satellite clinic was established, there was limited delegation of tasks from senior physicians. In contrast to **region 1** and **2**, **region 3** organized the diagnostic procedures and treatment initiation into two separate visits, which might have resulted in prolonged waiting lists and injection intervals. This organizational difference may reflect varying responses to economic incentives stemming from the DRG-based reimbursement system.

Varying effects of CPGs have been documented before, and the impact of and adherence to CPGs are dependent on how and where CPGs are implemented [7,24,60-62]. This study adds to the existing literature by pointing out that reaping benefits from centralization and CPGs is difficult when the organizational and financial basis contributing to economies of scale and scope are not ensured a priori. Moreover, this study indicates that regions that managed to decentralize and outsource treatment activity – rather than centralize – obtained higher treatment effects, presumably because this provided for higher treatment intensity

### 5.2 Limitations

This paper has some limitations. One limitation is that we only measure the outcome variables in the first 12-month course of treatment for wAMD patients. The purpose of this was to obtain a comparable sample across regions where the patients had been treated for approximately the same duration of time, and where no patients had been treated with anti-VEGF injections before the 12-month period. The drawback of this data restriction is that 63 pct. of the population in the three included regions did not have a 12-month follow-up, but had a first course of treatment that was much shorter or longer than 12 months. On the one hand, patients for whom the first course of treatment was much longer than 12 months may be patients who have been lost to follow-up or who have experienced very long waiting times. On the other hand, patients for whom the first course of treatment was much shorter than 12 months may be patients who have benefited much from receiving treatment very quickly. Patients with a shorter or longer course of treatment than approximately 12

months may differ from the included sample both in terms of the patient case-mix and in terms of the treatment outcome and cost. In the summary statistics in Table 2, we compared the sample and the population of wAMD patients. The comparison showed that for nine out of 17 comorbidities, the five-year incidence prior to diagnosis was 1-2 pct. points lower in the sample. Although this is not a large difference, the sample might be slightly healthier, in terms of comorbidity, than the population. In the robustness analyses, we include a wider time span around the 12-month window such that we include patients treated for a period between 265 and 465 days after first treatment, and only exclude 38 pct. of the population in the three included regions. This does not change the interpretation of the results, and it shows that the regional differences are pronounced also among patients who undergo treatment for a shorter or longer period than 12-months.

A second limitation is that we are not able to perform any quantitative analyses on the structural or professional-level determinants of importance for regional variation. Ideally, we would have been able to compare capacity loads (e.g. personnel and location resources, number of scanners per patient) across the regions. Moreover, we could have gained knowledge on how close the regions were to their capacity constraint by comparing the extent and timing of patient accumulation across the regions. We have tried to make comparable data analyses with both administrative data and by collecting data from the ophthalmology departments in the three regions, but this has not been possible. With the administrative data, we have not been able to make valid data comparisons across regions, as some departments have changed their unique department classification number in the Danish classification of hospital departments. Also, some ophthalmology departments have had several department classification numbers in some years, and we have not been able to validate whether the same procedures have been registered more than once in two different departments by mistake. We have also tried to gather new quantitative data from the three regions, where two out of the three regions have provided us with their own logs of patient accumulation. However, the methods behind the logs have not been comparable. Due to the low data quality and the lack of comparable data sources across the regions, we have not been able to inform or validate the findings from the qualitative analyses with quantitative data. While this validation of the findings from the qualitative part of the analyses would have been beneficial in interpreting and discussing the results, the lack of comparable quantitative data at a structural-level emphasizes the importance of the sequential mixed-methods framework in which we are able to utilize both quantitative and qualitative data sources.

A third limitation is that we only examine direct medicine costs, i.e. the cost of giving anti-VEGF injections. We do not examine costs related to hospital resources (e.g. scanners, clinical facilities and personnel), costs related to the patients' and relatives' resource use (patients' and relatives' time and transport expenses), resources in other sectors (e.g. productivity losses and social worker visits) or GP visits. We have chosen to focus on the direct medicine costs as we are mostly interested in examining variation in the treatment intensity (the number of injections) across regions, as an important drawback of large regional variation in wAMD treatment is the large variation in costs caused by the price per anti-VEGF injection. Naturally, other costs than the direct cost of anti-VEGF injections may generate regional variation, but we presume that other costs (e.g. use of clinical facilities and patients' time) are positively correlated with the number of injections the patients are given, and, in turn, that the regional differences will remain when adding more cost measures. However, we also acknowledge that some cost measures are negatively correlated with the number of injections given in the hospital, as the patients' vision may improve during their treatment period, and thereby reduce the number of GP visits and social worker visits.

## 6 Conclusion

We conclude that even when the patient case-mix is the same across regions, clinical practice variation across geographical regions may arise from structural regional differences in the organization and financing of healthcare delivery. Furthermore, we conclude that the value of CPGs is undermined when structural conditions do not support the organizational changes necessary to ensure optimal treatment, and that some degree of decentralization may be warranted when treatment effects depend on high treatment intensity.

## References

[1] J Wennberg, J Freeman, W Culp. Are hospital services rationed in New Haven or over-utilised in Boston? The Lancet. 329 (1987) 1185-1189.

[2] J Wennberg, A Gittelsohn. Variations in medical care among small areas, Sci.Am. 246 (1982) 120-135.

[3] JE Wennberg, L Blowers, R Parker, AM Gittelsohn. Changes in tonsillectomy rates associated with feedback and review, Pediatrics. 59 (1977) 821-826.

[4] J Wennberg, Gittelsohn. Small area variations in health care delivery, Science. 182 (1973) 1102-1108.

[5] AN Corallo, R Croxford, DC Goodman, EL Bryan, D Srivastava, TA Stukel. A systematic review of medical practice variation in OECD countries, Health Policy. 114 (2014) 5-14.

[6] SH Woolf, R Grol, A Hutchinson, Eccles M FAGrimshaw J., J Grimshaw. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines, BMJ. 318 (1999) 527-530.

[7] R Fijn, J Brouwers, J Timmer. Rational pharmacotherapy and clinical practice guidelines Theories and perspectives on implementing pharmacotherapeutic treatment guidelines, Pharmacy World and Science. 22 (2000) 152-158.

[8] ID Graham, J Logan, MB Harrison, SE Straus, J Tetroe, W Caswell, et al. Lost in knowledge translation: time for a map? J.Contin.Educ.Health Prof. 26 (2006) 13-24.

[9] JM Grimshaw, MP Eccles, JN Lavis, SJ Hill, JE Squires. Knowledge translation of research findings, Implementation science. 7 (2012) 1-17.

[10] MD Huesch, M Sakakibara. Forgetting the learning curve for a moment: how much performance is unrelated to own experience? Health Econ. 18 (2009) 855-862.

[11] DM Shahian, ST Normand. The volume-outcome relationship: from Luft to Leapfrog, Ann.Thorac.Surg. 75 (2003) 1048-1058.

[12] OECD, *OECD reviews of health care quality: Denmark 2013: raising standards*, (2013).

[13] SB Bloch, M Larsen, IC Munch. Incidence of legal blindness from age-related macular degeneration in Denmark: year 2000 to 2010, Am.J.Ophthalmol. 153 (2012) 209-213. e2.

[14] R Rogowski, How Inference in the Social (But Not the Physical) Sciences Neglects Theoretical Anomaly, in: Brady HE, Collier D (Eds.), Rethinking social inquiry: Diverse tools, shared standards, Rowman & Littlefield Publishers, 2004, pp. 75-84.

[15] H Eckstein, Case Study and Theory in Political Science, in: Fred I. Greenstein and Nelson W. Polsby (Ed.), Handbook of Political Science, Volume 7: Strategies of Inquiry. 1975, pp. 79-138.

[16] B Dormont, C Milcent. The sources of hospital cost variability, Health Econ. 13 (2004) 927-939.

[17] P Eibich, NR Ziebarth. Analyzing regional variation in health care utilization using (rich) household microdata, Health Policy. 114 (2014) 41-53.

[18] R Grol, J Grimshaw. From best evidence to best practice: effective implementation of change in patients' care, The lancet. 362 (2003) 1225-1230.

[19] PL Hudak, P Grassau, RH Glazier, G Hawker, H Kreder, P Coyte, et al. ``Not Everyone Who Needs One Is Going to Get One'': The Influence of Medical Brokering on Patient Candidacy for Total Joint Arthroplasty, Medical Decision Making. 28 (2008) 773-780.

[20] A Grove, R Johnson, A Clarke, G Currie. Evidence and the drivers of variation in orthopaedic surgical work: a mixed method systematic review, Health Syst Policy Res. 3 (2016).

[21] TC Buchmueller, M Jacobson, C Wold. How far to the hospital?: The effect of hospital closures on access to care, J.Health Econ. 25 (2006) 740-761.

[22] D Avdic. Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden, J.Health Econ. 48 (2016) 44-60.

[23] JM Grimshaw, IT Russell. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations, The Lancet. 342 (1993) 1317-1322.

[24] AG Ellrodt, L Conner, M Riedinger, S Weingarten. Measuring and improving physician compliance with clinical practice guidelines: a controlled interventional trial, Ann.Intern.Med. 122 (1995) 277-282.

[25] CL Egholm, HK Rossau, P Nilsen, G Bunkenborg, MH Rod, P Doherty, et al. Implementation of a politically initiated national clinical guideline for cardiac rehabilitation in hospitals and municipalities in Denmark, Health Policy. 122 (2018) 1043-1051.

[26] JP Newhouse. Frontier estimation: how useful a tool for health economics? J.Health Econ. 13 (1994) 317-322.

[27] N Rice, A Jones. Multilevel models and health economics, Health Econ. 6 (1997) 561-575.

[28] A Lleras-Muney. The relationship between education and adult mortality in the United States, The Review of Economic Studies. 72 (2005) 189-221.

[29] DM Cutler, A Lleras-Muney. Understanding differences in health behaviors by education, J.Health Econ. 29 (2010) 1-28.

[30] MJ Barry, FJ Fowler, AG Mulley, JV Henderson, JE Wennberg. Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplagia. Med.Care. (1995).

[31] K Sepucha, AG Mulley. A perspective on the patient's role in treatment decisions, Medical Care Research and Review. 66 (2009) 53S-74S.

[32] J Bonastre, C Le Pen, P Anderson, A Ganz, P Berto, G Berdeaux. The epidemiology, economics and quality of life burden of age-related macular degeneration in France, Germany, Italy and the United Kingdom, The European Journal of Health Economics. 3 (2002) 94-102.

[33] AA Plyukhova, MV Budzinskaya, KM Starostin, R Rejdak, C Bucolo, M Reibaldi, et al. Comparative safety of bevacizumab, ranibizumab, and aflibercept for treatment of neovascular age-related macular degeneration (AMD): a systematic review and network meta-analysis of direct comparative studies, Journal of clinical medicine. 9 (2020) 1522.

[34] DL Anthony, MB Herndon, PM Gallagher, AE Barnato, JP Bynum, DJ Gottlieb, et al. How much do patients' preferences contribute to resource use? Health Aff. 28 (2009) 864-873.

[35] A Chandra, D Cutler, Z Song. Who ordered that? The economics of treatment choices in medical care, Handbook of health economics. 2 (2011) 397-432.

[36] R Grol. Personal paper. Beliefs and evidence in changing clinical practice, BMJ. 315 (1997) 418-421.

[37] A Davis, A Baldwin, M Hingorani, A Dwyer, D Flanagan. A review of 145 234 ophthalmic patient episodes lost to follow-up, Eye. 31 (2017) 422-429.

[38] B Foot, C MacEwen. Surveillance of sight loss due to delay in ophthalmic treatment or review: frequency, cause and outcome, Eye. 31 (2017) 771-775.

[39] M McClellan. Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality? Journal of Economic Perspectives. 25 (2011) 69-92.

[40] A Shleifer. A theory of yardstick competition, Rand J.Econ. (1985) 319-327.

[41] AB Atkinson, The Atkinson review: final report. Measurement of government output and productivity for the national accounts, Palgrave Macmillan 2005.

[42] R Jacobs, PC Smith, A Street, Measuring efficiency in health care: analytic techniques and health policy, Cambridge University Press 2006.

[43] C Propper. Agency and incentives in the NHS internal market, Soc.Sci.Med. 40 (1995) 1683-1690.

[44] PC Smith, A Street, Concepts and challenges in measuring the performance of health care organizations, The Elgar Companion to Health Economics, Second Edition, Edward Elgar Publishing, 2012,.

[45] KR Olsen, A Street. The analysis of efficiency among a small number of organisations: how inferences can be improved by exploiting patient-level data, Health Econ. 17 (2008) 671-681.

[46] D Aigner, CK Lovell, P Schmidt. Formulation and estimation of stochastic frontier production function models, J.Econ. 6 (1977) 21-37.

[47] A Bryman. Integrating quantitative and qualitative research: how is it done? Qualitative research. 6 (2006) 97-113.

[48] JW Creswell, VLP Clark, Core mixed methods design , in: Creswell JW, Clark VLP (Eds.), Designing and conducting mixed methods research, Sage publications, 2017, pp. 51-99.

[49] A Hvenegaard, A Street, TH Sørensen, D Gyrd-Hansen. Comparing hospital costs: What is gained by accounting for more than a case-mix index? Soc.Sci.Med. 69 (2009) 640-647.

[50] A Hvenegaard, JN Arendt, A Street, D Gyrd-Hansen. Exploring the relationship between costs and quality: Does the joint evaluation of costs and quality alter the ranking of Danish hospital departments? The European Journal of Health Economics. 12 (2011) 541-551.

[51] P Schmidt, RC Sickles. Production frontiers and panel data, Journal of Business & Economic Statistics. 2 (1984) 367-374.

[52] JM Wooldridge, Econometric analysis of cross section and panel data, MIT press 2010.

[53] WC Horrace, P Schmidt. Confidence statements for efficiency estimates from stochastic frontier models, Journal of Productivity Analysis. 7 (1996) 257-282.

[54] WC Horrace, P Schmidt. Multiple comparisons with the best, with economic applications, J.Appl.Econometrics. 15 (2000) 1-26.

[55] J Hsu, Multiple comparisons: theory and methods, CRC Press 1996.

[56] CW Dunnett. New tables for multiple comparisons with a control, Biometrics. 20 (1964) 482-491.

[57] TD Jick. Mixing qualitative and quantitative methods: Triangulation in action, Adm.Sci.Q. 24 (1979) 602-611.

[58] F Kohlbacher. The use of qualitative content analysis in case study research, Forum Qualitative Sozialforschung/Forum: Qualitative Social Research. 7 (2006) 1-30.

[59] I Potapenko, M la Cour. Modelling and prognostication of growth in the number of patients treated for neovascular age-related macular degeneration, Acta Ophthalmol. (2021).

[60] AL Francke, MC Smit, AJ de Veer, P Mistiaen. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review, BMC medical informatics and decision making. 8 (2008) 1-11.

[61] M Lugtenberg, JS Burgers, GP Westert. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review, Qual.Saf.Health.Care. 18 (2009) 385-392.

[62] F Fischer, K Lange, K Klose, W Greiner, A Kraemer. Barriers and strategies in guideline implementation - a scoping review, Healthcare. 4 (2016) 36.

# Appendix A: Qualitative materials and methods

A 1 Informant characteristics

| Region | Informant characteristics | No. |
|---|---|---|
| Region 1 | Consultant, clinical responsibility for AMD treatment | 1 |
| | Consultant, professor AMD | 2 |
| | Department head, professor, managerial responsibility for AMD treatment | 3 |
| Region 2 | Senior registrar, clinical responsibility for AMD treatment | 4 |
| | Senior registrar, clinical responsibility for AMD treatment | 5 |
| | Department head, professor, managerial responsibility for AMD treatment | 6 |
| Region 3 | Consultant, professor AMD | 7 |
| | Department head, managerial responsibility for AMD treatment | 8 |

A 2 Coding scheme

| Code | Sub-codes | Code description |
|---|---|---|
| Knowledge | Competing ideas about best practice | Uncertainties in the evidence base, for example because of inconsistent research results, may lead clinicians to emphasize different research results or develop various convictions about best practice. |
| | Research engagement | Difficulties of keeping up-to-date with research results may lead clinicians in less research-oriented environments to adapt new treatment practices later than those in research-intensive environments. Participation in drug testing may influence product preferences because of early experience and practical experience with particular treatments. |
| Capacity | Capacity constraints | Lack of facilities, staff or medical equipment may prolong waiting time in diagnostics or treatment processes. Inadequate skill mix may also prolong waiting time or cause suboptimal treatment. |
| | Capacity expansion | To cope with capacity constraints, various measures may be taken to expand existing capacity, for example by expansion of facilities, hiring of extra staff or establishing sub-units |
| Care organization | **Optimization of procedures** | **Various measures may be taken to use the existing capacity most efficiently to reduce waiting time** |
| | **Internal work division** | **Various groups of professionals may be involved in diagnostic and treatment practices depending on internal work divisions. Task delegation may reduce waiting time and/or affect treatment outcomes.** |
| Funding | Payment systems | Differences in payment systems or mechanisms may incentivize clinicians or provider organizations to offer treatments of different types or intensity |
| | **Reactions to financial incentives** | **Clinician managers may react differently to the same financial incentives, leading to differences in care delivery** |

## A 3 Overview of received documents from the three regions

| Region | Title | No. |
|---|---|---|
| Region 1 | Summary note developed by IP1 based on meeting minutes, e-mail correspondence etc. | 1 |
| Region 2 | Note, Intravitreal anti-VEGF treatment of wAMD, 1st of January 2007 | 2 |
| | Working paper, Estimation of budget implications for treatment of wAMD in 2007 | 3 |
| | Collaboration agreement about treatment of wAMD with Lucentis, 2nd of April 2007 | 4 |
| Region 3 | Letter to hospital management about anti-VEGF treatment of wAMD, 29th of June, 2007 | 5 |
| | Letter to regional authorities about anti-VEGF treatment of wAMD, 7th of November, 2007 | 6 |
| | Note, Consolidation and expansion of capacity for treatment of AMD, 16th of July 2009 | 7 |
| | Note, Payment model for wAMD 2017, 3rd of February 2010 | 8 |
| | Note, Payment model for wAMD 2017, 15th of February 2011 | 9 |
| | Budget note, Expansion of the ophthamology department, 4th of August 2011 | 10 |
| | Budget note, Funding of wAMD since 2015, 17th of June 2015 | 11 |
| | Budget note, Funding of wAMD 2014, basis honorarium, 28th of May 2014 | 12 |
| | Budget note, Funding of wAMD final settlement 2014, 16th of February 2015 | 13 |
| National | Letter from the National Board of Health, 29th of November 2006. Intravitreal angiostatic treatment of wAMD (anti-VEGF-treatment) | 14 |
| | Meeting minutes, 5th of February 2007. Protocol for the treatment of wAMD with Lucentis | 15 |
| | Meeting minutes, 10th of March 2008. 6th national macula meeting | 16 |
| | Meeting minutes, 29th of September 2008, 7th national AMD meeting | 17 |
| | Meeting minutes, 15th of March 2010, 9th national AMD/macula meeting | 18 |
| | Meeting minutes, 23rd of November 2011, Meeting in the national AMD group | 19 |