

Det Nationale Indikatorprojekt



LUNGECANCER

Supplerende analyse af mortalitet og udredningsomstændigheder

Udarbejdet af:

Anders Green, klinisk epidemiolog, professor, overlæge, dr.med.

Maria Iachina, cand.scient. (statistik), PhD

Peter Gustav, cand.scient. (datalogi)

Kompetencecenter Syd for Landsdækkende Kliniske Databaser

Odense Universitetshospital

Udgivelsesdato: 05. maj 2011

Det Nationale Indikatorprojekt til måling og forbedring af de sundhedsfaglige kerneydelser er et samarbejdsprojekt mellem regionerne i Danmark, Danske Regioner, Dansk Medicinsk Selskab, De faglige sammenslutninger på sygeplejeområdet, Danske Fysioterapeuter, Ergoterapeutforeningen samt Sundhedsstyrelsen.

Postadresse:

NIP-sekretariatet

Det nationale Indikatorprojekt

c/o Regionshuset Århus, Olof Palmes Allé 15, DK-8200 Århus N

Telefon: (+45) 8728 4981

E-mail: Fagligkvalitet@rm.dk

Web-adresse: www.nip.dk

INDHOLD

1. INDIKATORGRUPPENS KOMMENTARER.....	3
2. SAMMENFATNING	4
3. BAGGRUND.....	5
3. PROBLEMFOMULERING.....	6
4. DATAGRUNDLAG	6
Oversigt over data	6
Justering for komorbiditet.....	11
Dataanalyse.....	12
5. RESEKTIONSFORHOLD	13
Data og analysestrategi	13
Resultater.....	14
Konklusion.....	16
6. MISKLASSIFIKATION	17
Data og analysestrategi	17
Resultater.....	17
Konklusion.....	21
7. MORTALITET I DET TOTALE MATERIALE.....	22
Data og analysestrategi	22
Resultater.....	23
Konklusion.....	24
8. MORTALITET UDEN RESEKTION.....	26
Data og analysestrategi	26
Resultater.....	27
Konklusion:	28
9. MORTALITET EFTER RESEKTION.....	29
Data og analysestrategi	29
Resultater.....	30
Konklusion.....	32
APPENDIX: BEREGNING AF CHARLSON INDEX	33
Klargøring af data	33
Metode	34
Bilag 1. Gruppering af diagnosekoder (ICD8 og ICD10) i Charlson Index.....	38
Bilag 2. SAS-program til beregning af Charlson Index.....	39
Bilag 3. SQL algoritme til beregning af Charlson Index, trin C1 – C4	44
Bilag 4. SQL algoritme til beregning af Charlson Index, trin C5 – C7	45
Bilag 5. SQL algoritme til beregning af Charlson Index, trin C8.....	46
Bilag 6. SQL algoritme til beregning af Charlson Index, trin C9 – C10.....	47
Bilag 7. SQL algoritme til beregning af Charlson Index, trin C11	48

1. INDIKATORGRUPPENS KOMMENTARER

Dette er den 3. udgave af "Den supplerende mortalitetsanalyse", der oprindeligt i 2008 blev bestilt hos Kompetencecenter Syd af Indikatorgruppen for NIP Lungecancer, idet de primære indikatorer vedrørende overlevelse generelt, samt med og uden operation umiddelbart tydede på betydende regionale forskelle. Den første udgave af mortalitetsrapporten udført på et DLCC / NIP patientmateriale fra 2003 – 2007 konkluderede, at der faktisk var signifikante forskelle i overlevelsen mellem regionerne. I analysen af disse forskelle blev komorbiditet foreslået som en prognostisk faktor der kunne forklare en del af denne forskel, hvilket gav anledning til 2. udgave af rapporten, der dog afkræftede en sådan forklaringsmodel. Denne 2. udgave rejste dog hypotese om, at de konstaterede regionale forskelle i misklassifikation mellem udredningsstadiet (pTNM) og operationsstadiet (pTNM) var relateret til regionale forhold omfattende volumen af udredende afdelinger, ligesom datas validitet blev diskuteret på baggrund af stærk varierende datakomplethed og kvalitet i den anvendte periode. Dette gav så igen anledning til, at Indikatorgruppen bad Kompetencecenter Syd gentage undersøgelsen, men nu med de mere komplette og valide data fra perioden 2007 – 2010 og med fokus på afdelingernes volumens eventuelle betydning i relation til forskelle i overlevelse.

Indledningsvis er det ikke lykkedes rapporten yderligere at afdække betydningen af volumen af de udredende afdelinger, idet analyserne demonstrerer, at da de store afdelinger findes i 2 regioner og de mindre i de øvrige 3 regioner, er det ikke muligt i en statistisk model at skelne mellem effekt af regionale forhold af andre årsager og effekt af udredningsvolumens størrelse. Rapporten har således ikke - uafhængigt af andre regionale karakteristika - kunnet påvise en eventuel effekt af udredende afdelings volumen på sandsynligheden for resektion. Det konkluderes, at den tidligere påviste effekt af udredende afdelings volumen på resektionsraten ikke kan efterprøves i det nyere datasæt i DLCC. Analyserne har ikke, som det var tilfældet i den forudgående version af rapporten, i det nye datasæt kunnet efterprøve hypotesen om, at udredende afdelingsvolumen skulle have indvirkning på graden af misklassifikation og dermed på mulighederne for at tilbyde operation på et fejlagtigt grundlag. I lighed med indikator IV i den nationale indikatorrapport finder rapporten ret udtalt og statistisk signifikant variation på tværs af regioner hvad angår misklassifikation, men en høj grad af misklassifikation synes ikke umiddelbart forbundet med højere resektionsrate, idet forløb uden misklassifikation også udviser betydende geografisk heterogenitet.

Rapporten undersøger herefter mortaliteten i den totale patientpopulation, mortaliteten i den del af patientpopulationen, som ikke har fået foretaget resektion, og endelig mortaliteten i den del af patientpopulationen, der har fået foretaget resektion. I alle 3 analyser genfindes den tidligere konstaterede geografiske heterogenitet, men rapporten må nu ud fra de ovennævnte analyser konstatere, at man ikke med baggrund i data i DLCC (og vel fra andre kilder) kan fastslå konkrete årsager til de konstaterede forskelle. Region Hovedstaden fremstår generelt som den region, hvor patienterne har den bedste overlevelse. Dette findes i alle 3 analyser, men herudover kan analyserne ikke på entydig måde identificere behandlingssteder med specielt dårlig prognose. Specielt hvad angår overlevelse efter resektion synes der at være ensartet korttidsprognose efter resektion på tværs af opererende centre og regioner, hvorfor forskellene formentligt skal forklares ved forhold, der har med patienternes efterbehandling at gøre, først og fremmest i onkologisk regi.

Med denne baggrund kan det konkluderes

- at der findes betydende regionale forskelle i overlevelsen for danske lungecancerpatienter.
- at overlevelsen generelt er størst i Region Hovedstaden i sammenligning med landets øvrige regioner
- at det ikke med baggrund i tilgængelige data i DLCC er muligt at drage konklusioner vedrørende mulige årsager til disse forskelle
- at der findes betydende regionale forskelle i tilbuddet om operativ behandling for danske lungecancerpatienter
- at det umiddelbart ikke er muligt at forklare disse forskelle ud fra regionale forskelle med hensyn til validiteten af den kliniske stadiestemmelse for patienterne forud for operation

2. SAMMENFATNING

I mortalitetsanalyserne indgår alle lungecancerforløb med diagnose stillet i årene 2007, 2008, 2009 og 2010, i alt 16274 forløb. Ud over mortalitetsanalyse af det samlede materiale fremlægges analyser af mortalitet hos patienter uden registreret resektion samt en analyse for patientgruppen med registreret resektion. Der bringes endvidere analyser af grad af overensstemmelse i stadieklassifikation ved udredning versus efter operation, samt af resektionsforhold.

Denne opdaterede og reviderede version af rapporten er ny ved at begrænse sig til den seneste del af Dansk Lunge Cancer Register, som anses for at være mest komplet og som afspejler mest optimalt de seneste års bestræbelser på at optimere udredning og behandling af lungecancer inden for pakkeforløb.

Generelt tages der forbehold over for analysens resultater som følge af stor variabilitet på tværs af regioner og over kalendertid hvad angår de udrednings- og behandlingsaktiviteter, der er registreret i Dansk Lunge Cancer Register.

For en række forhold er resultaterne som forventet: Højere alder ved diagnose er associeret med stigende mortalitet. Stigende sygdomsstadie er associeret med stigende mortalitet. Analysen genfinder det velkendte fænomen, at mortaliteten er lavere for kvinder i forhold til for mænd med lungecancer; denne forskel er specielt udtalt hvad angår mortalitet efter operation.

I alle analyser påvises en udtalt variation i mortalitet på tværs af regioner, når der justeres for de faktorer, der indgår i DLCCR-NIP sammenhæng. Denne variation er til stede, uanset om der er registreret resektion eller ej i lungecancerforløbene. Det generelle billede er, at mortalitetsniveauet er lavest i region Hovedstaden. I analysen af mortalitetsforholdene efter resektion findes, at misklassifikation (her defineret som klinisk betydende uoverensstemmelse mellem stadie vurderet ved udredning i forhold til stadie vurderet i forbindelse med operation) resulterer i en markant overdødelighed. Sammenholdt hermed findes der en ret udtalt variation på tværs af regioner i forekomsten af misklassifikation. Regioner med relativt højt mortalitetsniveau efter operation tenderer til at have høj forekomst af misklassifikation og øget sandsynlighed for resektion. Høj resektionsrate i en region kan dog ikke alene forklares med misklassifikation, idet specielt region Midtjylland ligger højt i resektionsrate også for operationer, hvor der ikke foreligger misklassifikation. Dette kunne tolkes som om, der er regionale forskelle hvad angår at give tilbud til operation. Analyserne sandsynliggør, at når der tages højde for misklassifikation og klinisk stadie ud over alder og køn, refterer der en ret markant forskel på tværs af behandlingssteder og regioner hvad angår mortalitetsniveau efter resektion.

Analysen af mortalitet og prognose efter resektion synes at godtgøre, at den påviste forskel i prognose på tværs af afdelinger og regioner formentligt skal tilskrives ikke så meget den helbredsmæssige effekt af det kirurgiske indgreb, men snarere effekten af andre faktorer, som ikke har kunnet belyses inden for rammene af nærværende rapport. Der tænkes i denne sammenhæng først og fremmest på den onkologiske behandling initialt og senere i det enkelt lungecancerforløb.

Analysen har forsøgt systematisk at inddrage udredende afdelings volumen (vurderet som antal gennemførte udredningsforløb i 2009) som prognostisk faktor. Dette har imidlertid ikke været muligt, da de detaljerede analyser har demonstreret en høj grad af korrelation mellem inddeling på regionsniveau og udredningsvolumen. Det er således ikke muligt i en statistisk model at kunne skelne mellem effekt af regionale forhold og effekt af udredningsvolumens størrelse.

Resultaterne af de her fremlagte analyser skal ses i sammenhæng med årsrapporterne fra NIP-lungecancer. Det har i denne sammenhæng været et gennemgående fund, at 60-70 % af patienterne med primær lungecancer diagnosticeres på et så fremskredent tidspunkt, at der ikke med de nuværende behandlingsmuligheder findes tilbud om intenderet kurativ behandling. Ikke desto mindre er der igennem de seneste år registreret et støt fald i den generelle mortalitetsrate i patientpopulationen. Skal der for alvor ske en forbedring af prognosen, forudsættes en væsentligt styrket opsporende indsats, først og fremmest i primærsektoren, således at stadigt flere patienter kommer til udredning og diagnose på et tidspunkt, hvor der gives reel mulighed for intenderet kurativ terapi.

3. BAGGRUND

I forbindelse med den nationale audit d. 29. april 2008 for lungecancer i Det Nationale Indikatorprojekt (NIP) blev det besluttet, at anmode Kompetencecenter Syd for Landsdækkende Kliniske Databaser (KCS) om at udarbejde en supplerende analyse:

”Med henblik på yderligere belysning af resultaterne for indikatorer I og II besluttede indikatorgruppen at rekvirere supplerende analyser for overlevelse under ét med operation som prognostisk faktor. Gruppen besluttede ligeledes at rekvirere en supplerende analyse af resultaterne for indikator V med henblik på belysning af resektionsrate versus stadiefordeling”.

Den første version af den rekvirerede rapport blev udsendt i december 2008. Det besluttedes i formandskabet for indikatorgruppen at indhente en uafhængig ekspertvurdering af rapporten hos professor i biostatistik Werner Vach, Syddansk Universitet. Rapporten og professor Vachs gennemgang blev drøftet på møde i indikatorgruppens formandskab d. 15. januar 2009. Det besluttedes på mødet, at der skulle foretages en række justeringer af rapporten, som herefter skulle indgå i den nationale auditproces for NIP-lungecancer i foråret 2009.

Den opdaterede rapport tog udgangspunkt i samme datasæt fra Dansk Lunge Cancer Register (DLCR), som lå til grund for NIP-lungecancerrapporten for året 2008. Skæringsdato for indberetning var 9. februar 2009. Rapportens analyser blev dog begrænset til at inkludere forløb med diagnoseår fra og med 2003 til og med 2007 af hensyn til at opnå passende lang observationstid for alle patienter.

Den opdaterede rapport blev gennemdrøftet ved møde d. 29. april 2009 i den nationale indikatorgruppe for NIP-lungecancer. På grund af rapportens resultater blev det besluttet at anmode KCS om at foretage yderligere analyser. Først og fremmest blev det fundet påkrævet at søge belyst, hvorvidt inddragelse af data om patienternes komorbiditet kunne bidrage til forklaring af rapportens fund.

Den nye rapport blev drøftet i forbindelse med audit-processen for NIP-lungecancer for årsrapporten 2009 på møde d. 7. april 2010. På basis af de fremlagte resultater og under hensyntagen til de organisatoriske ændringer, der i de senere år har fundet sted inden for udredning og behandling af primær lungecancer i Danmark, blev det besluttet at anmode om at udarbejdet endnu en version – nu begrænset til de senere års forløb for at sikre et så homogent datasæt som muligt. Endvidere skulle der sættes fokus på den mulige prognostiske effekt af udredende afdelings størrelse. For at optimere datagrundlaget skal den endelige rapport medtage data for året 2010.

Hermed foreligger den (formentligt) endelige reviderede version af rapporten. I forhold til de foregående versioner adskiller den foreliggende version sig på følgende måde: Der er kun medtaget førstegangsførløb af primær lungecancer med diagnose i årene 2007, 2008, 2009 og 2010. Datasættet er i øvrigt i alle henseender identisk med det datasæt, der ligger til grund for årsrapportering fra DLCR for året 2010.

3. PROBLEMFORMULERING

Rapporten har til formål at vurdere, hvorvidt mortalitet og derved prognose for primær lungecancer varierer på tværs af behandlingsenheder og regionale inddelinger under de organisatoriske rammer, som aktuelt er etableret for udredning og behandling af primær lungecancer i Danmark. I analysen skal indgå en vurdering af den prognostiske effekt af udredende afdelings størrelse. Da det på forhånd er godtgjort, at operation i sig selv udvirker en væsentlig gunstig effekt på prognosen ved primær lungecancer, indgår ligeledes en vurdering af misklassifikation af stadie, det vil sige hvorvidt der i udredningsforløbet konstateres et klinisk stadie, som ved efterfølgende operation viser sig at være klinisk væsentligt afvigende fra det reelle sygdomsstadie på behandlingstidspunktet.

Rapporten falder herved i følgende delanalyser:

Resektionsrater og de faktorer, der indvirker herpå

Misklassifikation, herunder forekomst af misklassifikation i forhold til resektionsrate-niveau

Mortaliteten i den totale patientpopulation

Mortaliteten i den del af patientpopulationen, som ikke har fået foretaget resektion

Mortaliteten i den del af patientpopulationen, der har fået foretaget resektion

4. DATAGRUNDLAG

Oversigt over data

Det datasæt, der ligger til grund for mortalitetsanalyserne i nærværende rapport, repræsenterer alle forløb af primær lungecancer, som i DLCR er registreret med diagnosedato inden for årene 2007, 2008, 2009 og 2010 pr. 28. januar 2011 svarende til tidspunktet for udtrækket til årsrapporten for 2010.

Figur 4.1 illustrerer afgrænsningen af de lungecancerforløb, der indgår i nærværende analyse. Det bemærkes, at der kun er medtaget det første lungecancerforløb for de patienter, der er registreret med mere end et forløb. Efter eksklusionerne indgår totalt 16274 patientforløb med diagnosedato inden for den anførte periode. For yderligere detaljer henvises i øvrigt til den nationale rapport for året 2010.

Figur 4.1. Flowdiagram over udvælgelsen af lungecancerforløb til nærværende analyse

			Antal tilbageblevne	
			Patienter	Forløb
Totalt registreret i DLCR pr. 28. januar 2011			39256	37798
	Patienter	Forløb		
Testpatienter	4	4	39252	37794
Patienter uden forløb	1823	0	37429	37794
Patienter med forløb uden indhold	122	165	37307	37629
Forløbsnummer >1	0	322	37307	37307
Patologidiagnose mesoteliom	464	464	36843	36843
Bopælsdiagnose i Grønland	124	124	36719	36719
Diagnosedato før 1. januar 2007	20390	20390	16329	16329
Diagnosedato efter 31. december 2010	55	55	16274	16274

For de enkelte analyser vil et vist antal forløb blive ekskluderet på grund af manglende data. Dette gøres der rede for i de konkrete sammenhænge.

Tabel 4.1 viser lungecancerforløbene fordelt på diagnoseår og registreret bopælsregion. Antallet steg fra 3872 for diagnoseår 2007 til 4172 for diagnoseår 2010; denne stigning skal formentligt overvejende tilskrives stigende komplethedsniveau i DLCR.

Table 4.1. Lungecancerforløbene fordelt på diagnoseår og registreret bopælsregion

Diagnoseår	Hovedstaden	Sjælland	Syddanmark	Midtjylland	Nordjylland	Uoplyst	I alt
2007	1099	632	900	807	416	18	3872
2008	1108	713	948	837	447	21	4074
2009	1143	704	958	857	461	33	4156
2010	1077	725	965	918	464	23	4172
I alt	4427	2774	3771	3419	1788	95	16274

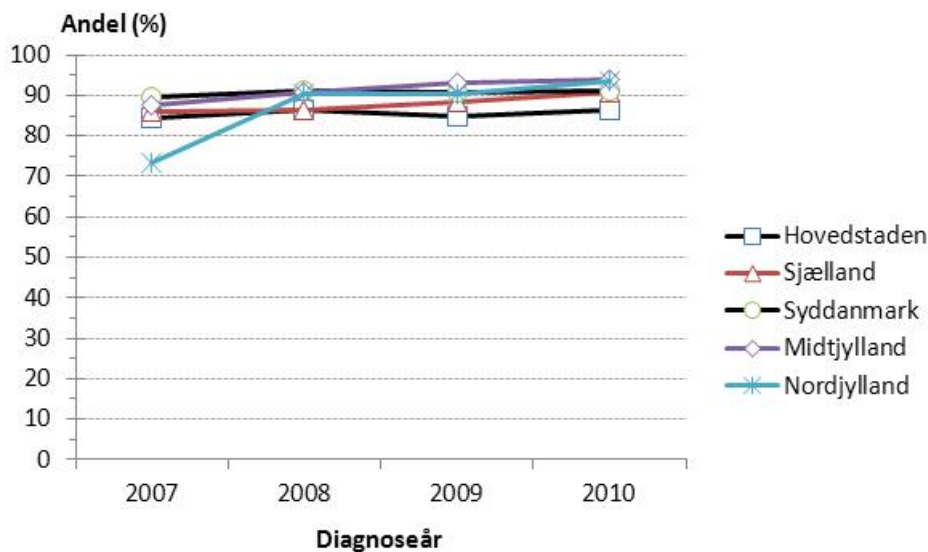
Lungecancerforløbene er forskelligt belyst hvad angår udrednings- og behandlingsforhold. Derfor er det af afgørende betydning for den overordnede vurdering af rapportens resultater at have indblik i, hvorledes forløbene er fordelt efter kalendertid (diagnoseår) og geografi (region i henhold til patienternes registrerede bopæl svarende til lungecancerforløbet).

Figur 4.2 viser for hver region udviklingen i andelen af lungecancerforløbene med indberettede data fra primært udredende afdelinger. Med undtagelse af år 2007 ligger andelen af forløb med primær udredning inden for et bånd på 85-95 %.

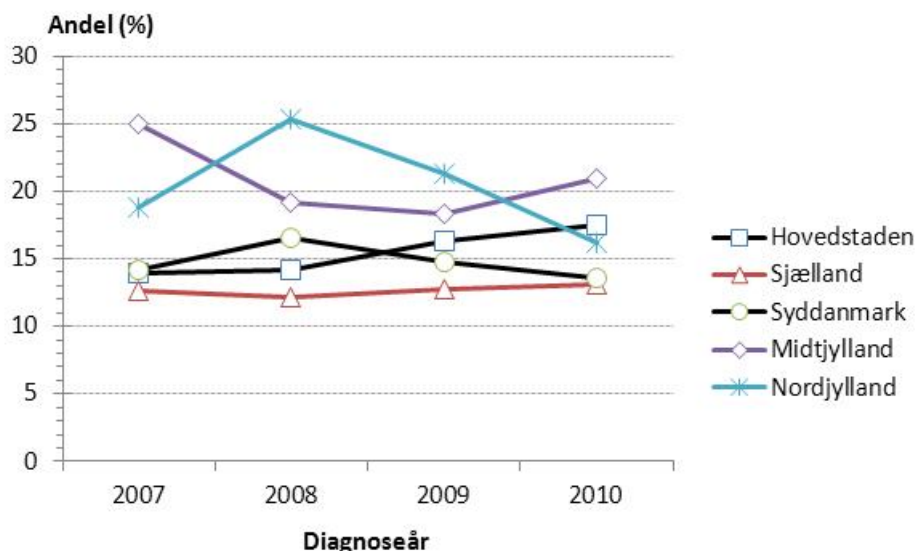
Figur 4.3 viser tilsvarende for hver region udviklingen over tid i andelen af de registrerede lungecancerforløb med indberettede operationsdata. Billedet afspejler betragtelig heterogenitet, både på tværs af regioner og som funktion af tiden.

Det er uvist, hvorvidt kurveforløbene i figur 4.2 og 4.3 afspejler de reelle forhold omkring udredning og operationsaktivitet, eller om kurveforløbene helt eller delvist skal forklares ved ændringer i praksis for indberetninger af udredning, henholdsvis operation. I DLCR-samarbejdet er det imidlertid en generel opfattelse, at specielt indberetning af operationsaktivitet må anses for praktisk taget komplet over den undersøgte periode.

Figur 4.2. Andel af lungecancerforløb med udredningsdata. Efter bopælsregion og diagnoseår. I alt 16179 forløb er medtaget, efter eksklusion af 95 forløb med uoplyst bopælsregion



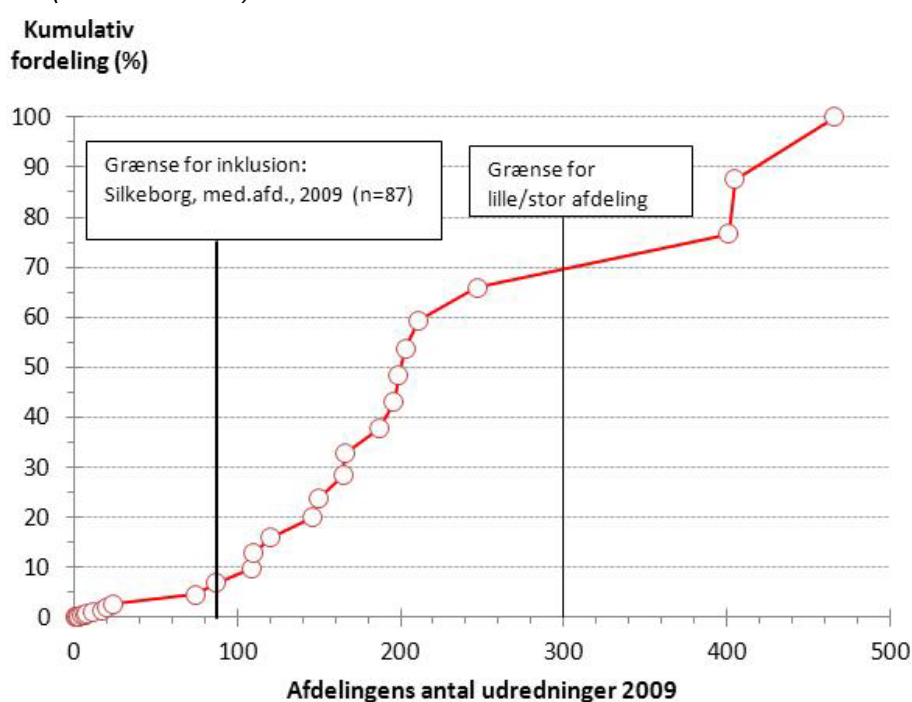
Figur 4.3. Andel af lungecancerforløb med operationsdata registreret. Efter bopælsregion og diagnoseår. I alt 16179 forløb er medtaget, efter eksklusion af 95 forløb med uoplyst bopælsregion



Som tidligere anført har det været hensigten at medtage de udredende afdelingers aktivitetsniveau i analyserne. Det er gjort ved at indeksere hver udredende afdeling efter antal registrerede udredninger i DLCR. Figur 4.4. viser den kumulative fordeling af lungecancerforløb med udredningsdata i forhold til udredende afdelings volumen. Det skal understreges, at datagrundlaget udelukkende omfatter de udredningsforløb, som har resulteret i diagnosen primær lungecancer. De tre største udredende afdelinger står for 34 % af forløbene. For ca. 5 % af forløbene er udredningen foretaget på afdelinger, der i 2009 havde et udredningsvolumen under den arbitrært fastsatte nedre grænse, sv.t. udredningsvolumen for medicinsk afdeling i Silkeborg (n=87). En tærskelværdi på 300 anses for rimelig til at adskille afdelinger efter lille, henholdsvis stort udredningsvolumen. Der er efterprøvet alternative grupperinger af udredningsvolumen, men den anførte afskæring mellem de tre største udredende afdelinger over for resten af de udredende afdelinger er fastholdt, fordi den resulterer i den mest markante prognostiske effekt.

Den kumulative fordelingskurve for samtlige diagnoseårgange 2007-2010 (inkl.) kombineret er stort set sammenfaldende med kurveforløbet for diagnoseårgang 2009 (figur 4.4).

Figur 4.4. Kumulativ andel af lungecancerforløb med udredningsdata efter udredende afdelings volumen, år 2009 (n=3741 forløb)



Tabel 4.2 grupperer alle forløb med udredningsdata efter udredende afdelings volumen i år 2009 over for afdelingernes regionale tilhørsforhold.

Tabel 4.2. Antal forløb fordelt efter primært udredende afdelings volumen (antal udredninger resulterende i diagnosen primær lungecancer) i år 2009 i forhold til afdelingernes regionale tilhørsforhold

Volumen	Hovedstaden	Sjælland	Syddanmark	Midtjylland	Nordjylland	I alt
<87 (*)	718	0	55	21	56	850
87 - 149	0	0	875	1337	0	2212
150 - 299	0	2405	2837	1557	0	6799
≥300	3182	0	0	0	1452	4634
Ail	3900	2405	3767	2915	1508	14495

(*): Grænsen 87 er fastlagt ud fra udredningsvolumen år 2009 ved medicinsk afdeling i Silkeborg

Der ses en meget høj grad af korrelation mellem udredende afdelingers volumen og regionalt tilhørsforhold. I både region Hovedstaden og region Nordjylland er der langt overvejende foretaget udredning på afdelinger med højt volumen. I region Sjælland er der udelukkende foretaget udredning på afdelinger med volumen i mellemklassen. I både region Syddanmark og region Midtjylland falder forløbene i de tre laveste klasser, mens der ikke foretages udredning på afdelinger med højt volumen i disse to regioner. Disse forhold umuliggør, at man i en statistisk analyse vil kunne adskille effekt af udredende afdelings volumen over for effekten af afdelingens regionale tilhørsforhold.

Selv om det således ikke er muligt at isolere effekten af udredende afdelings volumen over for effekten af udredende afdelings regionale tilhørsforhold, kan det ikke udelukkes, at lungecancerforløb udredt på afdelinger med specielt lavt volumen repræsenterer specielle tilfælde. Det skal endvidere tages i betragtning, at den foreliggende analyse så vidt muligt skal afspejle de seneste års bestræbelser på en centraliseret og koordineret udrednings- og behandlingsindsats for primær lungecancer. Af disse grunde er alle forløb (n 0 698), hvor der er foretaget udredning på en afdeling med volumen < 87 i året 2009 (se tabel 4.2) udeladt fra de efterfølgende statistiske analyser, mens de er medtaget i de ujusterede grafiske oversigter.

Tabel 4.3. viser, at det totale analyserede datasæts kompletthed for en række af de væsentligste variable er på meget højt niveau. Den mest betydende kilde til datadeficit kan henføres til manglende udredningsdata.

Tabel 4.3. *Oversigt over kompletthedsforholdene for væsentlige variable*

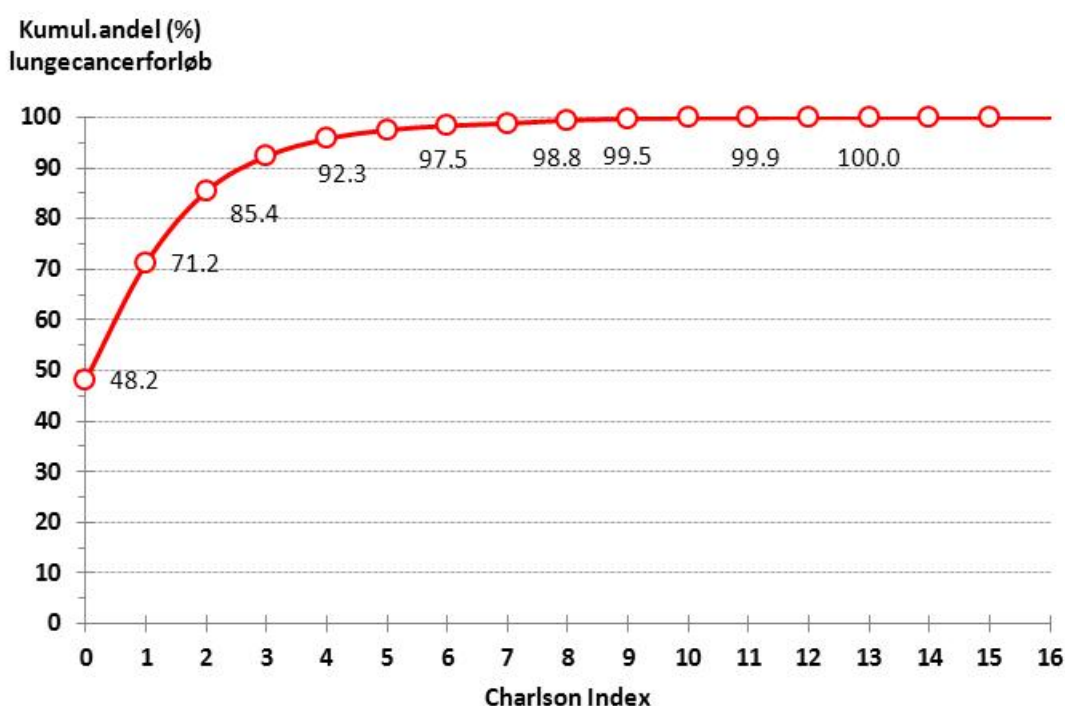
Variabel	Med komplette data	Med manglende data	Datakompletthed (%)	I alt
Alder	16274	0	100.0	16274
Køn	16274	0	100.0	16274
Klinisk stadie	14242	1851	88.6	16274
Operationsstatus	16274	0	100.0	16274
Region	16179	95	99.4	16274
Diagnoseår	16274	0	100.0	16274
Charlson Index	16211	63	99.6	16274

Justering for komorbiditet

Justering af komorbiditet er foretaget efter principperne oprindeligt rapporteret af Charlson et al.¹. Der er anvendt al tilgængelig information i LPR om de enkelte patienter op til diagnosticeringstidspunktet for lungecancerforløbet. Rapportens Appendiks redegør detaljeret for håndteringen af dataudtrækket fra datasættets modtagelse til fastlæggelse af den enkelte patients Charlson Index. Det fremgår også af Appendiks, hvilke bestræbelser der er gjort for at sikre, at lungecancerforløbet ikke i sig selv fejlagtigt vil belaste Charlson Index for den enkelte patient.

Figur 4.5 viser den kumulerede fordeling af registret Charlson Index for alle lungecancerforløbene. For 63 forløb har der ikke kunnet findes registreringer i LPR (ud over hvad der kan tilskrives lungecancerforløbet); disse forløb er allokeret til Charlson Index-værdien 0.

Figur 4.5. Kumulativ andel af lungecancerforløb med information om LPR-data efter Charlson Index (n=16274 forløb, hvor der for 63 forløb ikke er fundet registreringer i LPR)



For knap 50 % af forløbene er der ikke registreret forudgående LPR-aktiviteter, der er belastende for Charlson Index. For ca. 7.5 % af forløbene er der registreret forudgående aktiviteter, der udløser et Charlson Index på 4 eller højere.

¹ Charlson ME, Pompei P, Ales KL, MacKenzie CR: "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation". J. Chron. Dis. **40**: 373-383, 1987

Dataanalyse

Der henvises generelt til årsrapport 2010 (udgivet i april 2011) for NIP-lungecancer hvad angår den statistiske indfaldsvinkel til analyse af data.

I kapitel 5 vedrørende resektionsforhold er der anvendt logistisk regression for at estimere odds ratio for resektion for hver region, stratificeret for stadie og med justering for alder, køn og Charlsongruppe. Rent arbitrært er der som referencekategori anvendt region Hovedstaden og laveste stadie (stadie 0 og I kombineret).

I kapitel 6 er der til vurdering af validiteten af stadieklassifikation ved udredning anvendt kappastatistik med test af homogenitet af kappakoefficienter på tværs af regioner.

Dødelighedsanalyserne (afsnit 7-9) har benyttet Cox proportional hazard model til estimering af mortalitetsrateratio (hazard ratio). For at belyse heterogenitet med hensyn til kalendertid og geografi (på regionsniveau, henholdsvis sygehusniveau) er der suppleret med frailty modeller.

På regionsniveau er region Hovedstaden anvendt som reference, mens der på hospitalsniveau (vedr. mortalitet efter operation) er anvendt Rigshospitalet som reference. Med hensyn til udvikling over kalendertid er diagnoseåret 2007 anvendt som reference.

Der er ved visuel sammenligning af kurverne for estimererne på $-\ln(-\ln(\text{overlevelse}))$ over for Kaplan-Meier kurverne for hver Cox regressionsanalyse konstateret acceptabel opfyldning af analysernes forudsætninger.

Som anført ovenfor er der en meget høj grad af korrelation mellem udredende afdelings volumen og den regionale inddeling. Inklusion af udredningsvolumen som variabel har efterfølgende givet resultater, der ikke har kunnet fortolkes meningsfyldt såfremt også region er inkluderet i modellen som geografisk variabel. Der kan der med andre ord ikke foretages en adskillelse af effekt på regionsniveau over for effekt af udredende afdelings volumen. Der er derfor foretaget det valg at bibeholde region som geografisk variabel og udelade udredende afdelingsvolumen fra den statistiske analyse.

I overensstemmelse med gældende praksis for anvendelse af Charlson Index er det enkelte lungecancerforløbs beregnede Charlson Index grupperet som følger:

Charlson gruppe 0: Charlson Index værdi = 0

Charlson gruppe 1: $1 \leq \text{Charlson Index} \leq 3$

Charlson gruppe 2: Charlson Index ≥ 4 .

5. RESEKTIONSFORHOLD

Data og analysestrategi

Analysen omhandler den regionale variation med hensyn til resektion, når der tages højde for stadie og visse andre justeringsfaktorer. Ved resektion forstås operation *på nær eksplorativt indgreb*.

Faktor af primær interesse for vurdering af variation:

- Region (patientbopæl, med region Hovedstaden som reference)

Prognostiske faktorer:

- Sygdomsstadie (klinisk stadie, med laveste stadieklasse som reference))

Justeringsfaktorer:

- Alder ved diagnose (dikotomiseret med alder < 67 år som reference)
- Køn (mænd som reference)
- Diagnoseår (2007 som reference)
- Charlsongruppe (3 grupper, med gruppe 0 (ingen komorbiditet) som reference)

Som anført i kapitel 4 er der en udtalt korrelation mellem udredende afdelings volumen og behandlende afdelings regionale tilhørsforhold. Derfor er det umuligt at adskille prognostisk effekt af geografi over for udredningsvolumen. Da geografi er en faktor af primær interesse, er udredende afdelings volumen fra den statistiske analyse.

Justering af alder er ud fra forskellige statistiske afprøvninger bedømt at kunne foretages mest hensigtsmæssigt ved dikotomisering svarende til alder 67 år.

Ved vurdering af den prognostiske betydning af stadie er stadie 0 slået sammen med stadie I af hensyn til små tal, således at der opereres med følgende stadietkategorier: 0+I, II, IIIa, IIIb og IV.

Der anvendes logistisk regression for at estimere odds ratio for resektion for hver region, stratificeret for stadie og med justering for alder, køn og diagnoseår. Rent arbitrært er der som referencekategori anvendt region Hovedstaden og laveste stadie (stadie 0 og I kombineret).

Tabel 5.1. fremstiller lungecancerforløbene efter bopælsregion og registreret operation (samt specielt eksplorativt indgreb). Region Nordjylland og først og fremmest Region Midtjylland skiller sig ud med markant højere operationsrater end de øvrige regioner, og forholdet kan ikke forklares ved specielt høj rate af eksplorative indgreb i disse to regioner.

Tabel 5.1. Oversigt over forløb med registreret operation

Region	Med operation	Heraf eksplorativt indgreb	Uden operation	Andel opererede (%)	I alt
Hovedstaden	685	43	3742	15.5	4427
Sjælland	352	18	2422	12.7	2774
Syddanmark	558	31	3213	14.8	3771
Midtjylland	711	29	2708	20.8	3419
Nordjylland	364	9	1424	20.4	1788
Uoplyst	14	1	81	14.7	95
I alt	2684	131	13590	16.5	16274

Datakompletheden hvad angår de relevante variable fremgår af tabel 5.2 og dokumenterer acceptabel complethed. Største deficit (16-17 %) er manglende information om udredningsstadiet på grund af manglende udredningsdata. Hvad angår netop manglende information om udredningsstadiet er der ret udtalt forskel på regionerne (se også kapitel 6). Til beregningerne foreligger der fornøden information om 13494 forløb (82.3 % af det totale antal forløb i materialet).

Tabel 5.2. Oversigt over datakomplethed for analyse af resektionsforhold (n=2551 forløb med resektion)

Variabel	Med komplette data	Med manglende data	Datakomplethed (%)
Alder	2551	0	100.0
Køn	2551	0	100.0
Diagnoseår	2551	0	100.0
Klinisk stadium	2122	429	83.2
Regionsbopæl	2538	13	99.5
Data vedr. komorbiditet	2541	10	99.6

Resultater

Indledningsvis viser tabel 5.3, at resektionsraten fordeler sig som forventet, med faldende rate efter stigende udredningsstadium. For de laveste stadier ses en faldende resektionsrate med stigende komorbiditet.

Tabel 5.3. Resektionsrater (%) efter grupperinger i henhold til udredningsstadium og Charlsongruppe

Klinisk stadium	Charlsongruppe				Alle
	0	1	2	Ingen data	
0, I	931/1509 (61.7)	773/1543 (50.1)	102/267 (38.2)	6/8 (75.0)	1812/3327 (54.5)
II	185/426 (43.4)	126/395 (31.9)	10/50 (20.0)	0/1 (0.0)	321/872 (36.8)
IIIa	76/574 (13.2)	47/566 (8.3)	8/95 (8.4)	0/0 (-)	131/1235 (10.6)
IIIb	50/1365 (3.7)	46/1207 (3.8)	8/204 (3.9)	0/1 (0.0)	104/2777 (3.7)
IV	46/3512 (1.3)	17/3007 (0.6)	2/504 (0.4)	0/9 (0.0)	65/7032 (0.9)
Ingen data	189/804 (23.5)	199/829 (24.0)	37/173 (21.4)	4/45 (8.9)	429/1851 (23.2)
Alle	1297/7783 (16.7)	1085/7177 (15.1)	159/1251 (12.7)	10/63 (15.9)	2551/16274 (15.7)

Tabel 5.4 giver en oversigt over de estimerede odds ratio'er. Patienter fra region Hovedstaden med laveste kliniske stadiumkategori er anvendt som reference. Høj alder er (som forventet) statistisk signifikant associeret med reduceret odds for resektion, mens køn og diagnoseår ikke indvirker statistisk signifikant på odds for resektion. Odds ratio for resektion falder som forventet med stigende stadium. Specielt for stadierne IIIb og IV er der dog tale om stor usikkerhed på grund af et begrænset antal resektioner i disse stadiumkategorier. For specielt kategorien med stadium II falder region Midtjylland ud med markant højere odds ratio end de øvrige regioner. Det bemærkes at de højere Charlsongrupper under de givne modelantagelser (som forventet) ses at være associeret med lavere resektionsrate (se også tabel 5.3).

Tabel 5.4. Odds ratio for resektion i henhold til stadie og region med justering for alder, køn, Charlsongruppe, udredende afdelings volumen og diagnoseår. Reference: Stadie 0+I komb. for region Hovedstaden. Antal analyserede forløb: 14187 ud af 17490 (81,1 %)

	Variabel	Odds Ratio	C.L. (95%)	P-værdi
	Alder	0.41	0.36 - 0.47	0.000
	sex	0.90	0.79 - 1.02	0.108
	comorb 1,2 vs comorb 0	0.67	0.59 - 0.77	0.000
	comorb >2 vs comorb 0	0.40	0.30 - 0.52	0.000
	SCLC vs NSCLC	0.22	0.16 - 0.30	0.000
	Uoplyst vs NSCLC	0.08	0.05 - 0.11	0.000
	Diagnoseår	1.03	0.97 - 1.09	0.320
Stadie 0+I	Hovedstaden	(1.00)		
	Sjælland	0.49	0.36 - 0.68	0.000
	Syddanmark	0.52	0.40 - 0.69	0.000
	Midtjylland	0.87	0.67 - 1.14	0.314
	Nordjylland	0.76	0.55 - 1.05	0.100
Stadie II	Hovedstaden	0.30	0.21 - 0.43	0.000
	Sjælland	0.26	0.17 - 0.39	0.000
	Syddanmark	0.17	0.11 - 0.25	0.000
	Midtjylland	0.41	0.28 - 0.61	0.000
	Nordjylland	0.27	0.17 - 0.41	0.000
Stadie IIIa	Hovedstaden	0.10	0.06 - 0.15	0.000
	Sjælland	0.06	0.04 - 0.10	0.000
	Syddanmark	0.02	0.01 - 0.04	0.000
	Midtjylland	0.04	0.02 - 0.07	0.000
	Nordjylland	0.07	0.04 - 0.12	0.000
Stadie IIIb	Hovedstaden	0.02	0.02 - 0.04	0.000
	Sjælland	0.02	0.01 - 0.03	0.000
	Syddanmark	0.01	0.01 - 0.02	0.000
	Midtjylland	0.01	0.01 - 0.02	0.000
	Nordjylland	0.01	0.01 - 0.03	0.000
Stadie IV	Hovedstaden	0.00	0.00 - 0.01	0.000
	Sjælland	0.00	0.00 - 0.00	0.000
	Syddanmark	0.00	0.00 - 0.00	0.000
	Midtjylland	0.01	0.01 - 0.01	0.000
	Nordjylland	0.01	0.00 - 0.00	0.000

Konklusion

Analysen finder som forventet, at odds ratio – og dermed sandsynligheden – for resektion aftager med stigende stadie. Høj alder er forbundet med reduceret odds for resektion, mens der ingen association findes for køn og diagnoseår. Endvidere findes, at høj grad af komorbiditet er forbundet med lav sandsynlighed for resektion.

Specielt for stadie II ses en markant højere sandsynlighed for resektion i region Midtjylland. Disse forhold belyses yderligere i kapitel 6 vedrørende misklassifikation.

Der må i fortolkningen af resultaterne tages generelt forbehold, da dataindberetningen for operativ aktivitet såvel som for udredningsforhold og dermed klassifikation af klinisk stadie varierer mellem regioner og over kalendertid.

6. MISKLASSIFIKATION

Data og analysestrategi

I denne analyse belyses misklassifikation, udtrykt ved klinisk betydende uoverensstemmelse mellem stadietklassifikation ved udredning versus efter operation. Analysen er begrænset til forløb med registreret operation, herunder også eksplorative indgreb. Bortfald fra analysen betinges af manglende/ufuldstændig kirurgisk stadietklassifikation og/eller manglende/ufuldstændig registrering af stadiet ved udredning (klinisk stadiet).

Faktor af primær interesse for vurderinger af variation:

- Region (patientbopæl)

I denne analyse indgår ingen yderligere justeringsfaktorer eller prognostiske faktorer.

I analysen er anvendt kappastatistik med test af homogenitet af kappakoefficienter på tværs af regioner. Der er anvendt χ^2 -teknik til vurdering af heterogenitet på tværs af regioner.

Datakompletheden hvad angår information om udredningsstadiet for de opererede fremgår af nedenstående oversigtstabel og dokumenterer kompromitteret komplethed. Der er udtalt variation i kompletheden på tværs af regioner; således mangler region Nordjylland indberetninger om udredningsklassifikation af stadiet for næsten halvdelen af operationerne. Der er fuld information om stadiet for 2222 forløb (82.8%) af det totale antal forløb med registreret operation. Efter eksklusion af forløb med ukendt regionalt tilhørsforhold samt forløb med udredning på afdeling med for lavt udredningsvolumen indgår 2130 forløb (79.4%) i den statistiske analyse.

Tabel 6.1. Oversigt over datakomplethed for analyse af misklassifikation ($n=2684$ forløb med operation)

Region	Antal opererede	Med fuld information om stadiet	Datakomplethed (%)
Hovedstaden	685	567	82.8
Sjælland	352	283	80.9
Syddanmark	558	445	79.7
Midtjylland	711	635	89.3
Nordjylland	364	284	78.0
Uoplyst	14	8	57,1
I alt	2684	2222	82.8

Resultater

Tabel 6.2 sammenfatter resultaterne. Konventionelt anses kappaværdier ≥ 40 for at repræsentere 'acceptabel' overensstemmelse, mens værdier ≥ 75 anses for at repræsentere 'excellent' overensstemmelse. På tværs af regioner varierer kappaværdierne mellem 0.544 (region Midtjylland) og 0.559 (region Nordjylland) som det laveste (men inden for acceptniveaue) og 0.784 (region Hovedstaden) som er i det optimale område. Der er statistisk evidens for heterogenitet på tværs af regionerne ($P < 0.001$). Totalt for landet er kappaværdien 0.656 (95% sikkerhedsinterval: 0.640-0.697).

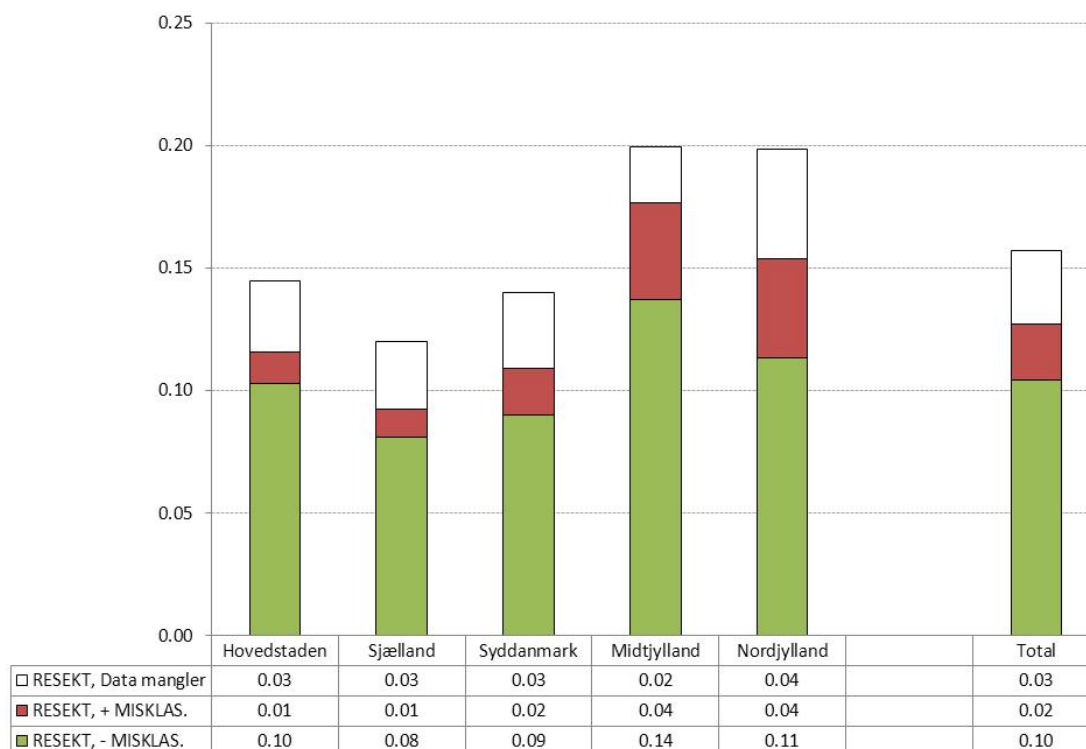
Tabel 6.2. Overensstemmelse (kappastatistik) mellem stadieklassifikation ved udredning, henholdsvis operation. Antal analyserede forløb: 2130 ud af 2684 (79.4%)

Region	Kappa	C.L. (95%)	Antal
Hovedstaden	0.784	(0.756 - 0.802)	509
Sjælland	0.756	(0.700 - 0.803)	274
Syddanmark	0.631	(0.588 - 0.673)	437
Midtjylland	0.544	(0.509 - 0.585)	627
Nordjylland	0.559	(0.499 - 0.597)	283
Hele landet	0.656	(0.640 - 0.697)	2130

Der er i kapitel 5 redegjort for, at region Midtjylland har en specielt høj resektionsrate. Analysen i nærværende kapitel viser, at samme region ligger relativt lavt hvad angår grad af overensstemmelse mellem udredningsstadiet og postoperativt stadium. Det er derfor nærliggende at belyse, hvorvidt der er en sammenhæng mellem høj resektionsrate og højt niveau af misklassifikation, således at højt niveau af misklassifikation medfører at flere patienter fejlagtigt klassificeres i et stadium foreneligt med operation.

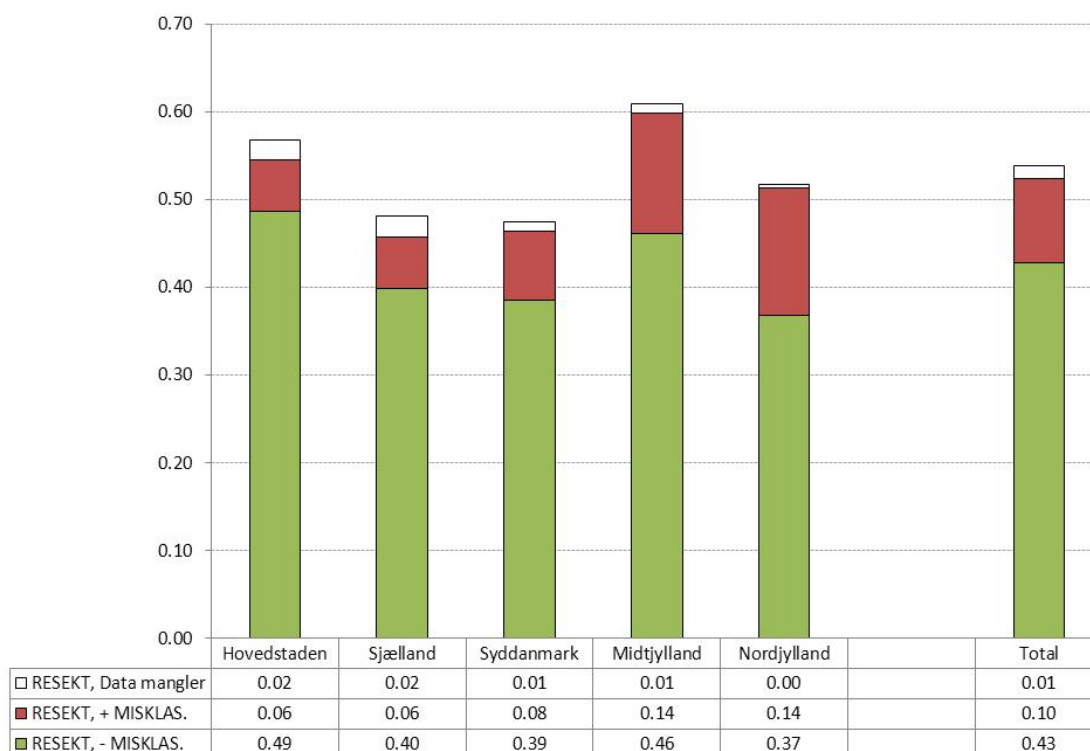
Figur 6.1A sammenfatter den regionale fordeling af resektionsrater med specielt fokus på misklassifikation, uanset niveau hvad angår udredningsstadiet. Der tegner sig et billede af ret udtalt heterogenitet i resektionsraterne på tværs af regioner. Endvidere viser figuren, at højere niveau for misklassifikation ikke umiddelbart kan forklare de regionale forskelle i resektionsraten: For forløb uden misklassifikation er resektionsraten i Midtjylland væsentligt større end resektionsraten i de øvrige regioner kombineret. Fortolkningen vanskeliggøres på grund af variation på tværs af regionerne med hensyn til manglende information om klassifikation af udredningsstadiet.

Figur 6.1A Andel resekerede patienter i det totale materiale, fordelt på regioner (efter patientbopæl) og misklassifikationsstatus. Der er IKKE taget hensyn til udredningsstadiet. Antal forløb: 16179



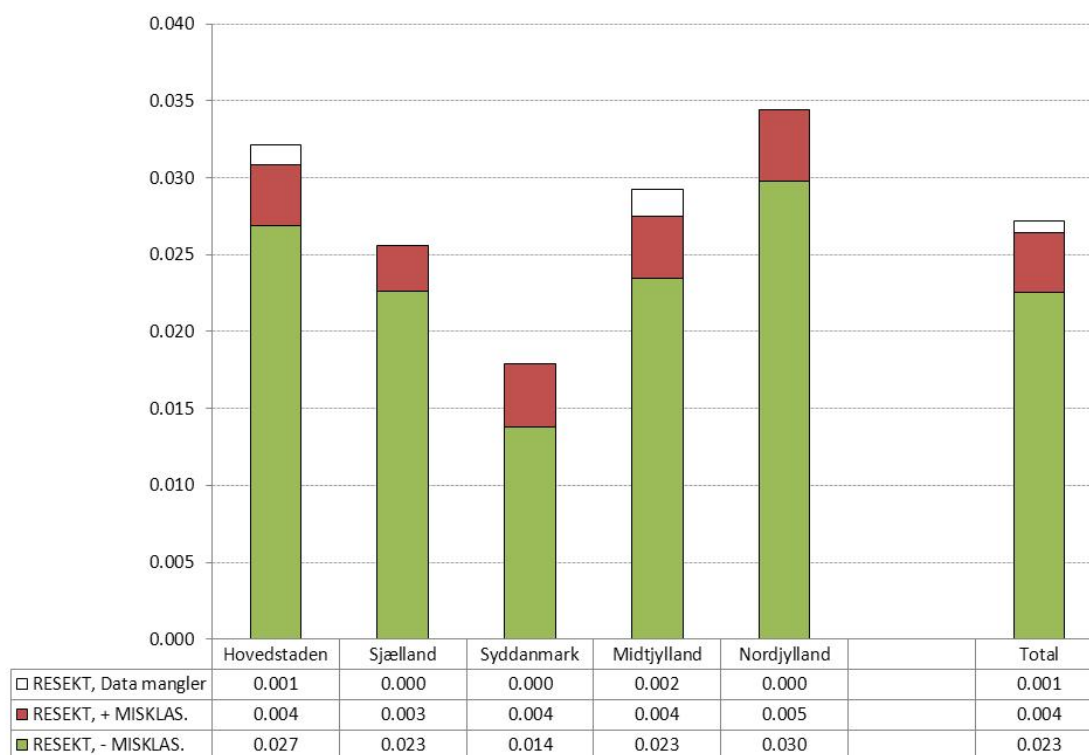
Figur 6.1B illustrerer forholdene for specielt for udredningsstadier med værdier mindre end III. Som forventet er resektionsraten for denne stadietkategori væsentligt højere end for totalpopulationen. Region Midtjylland ligger højest med en resektionsrate på 61%, men variationsbredden er ikke så udtalt som for resektionsraten i totalpopulationen. Andelen af forløb med resektion uden data til belysning af misklassifikation er som forventet beskedent og skyldes en manglende angivelse af stadiet påvist ved operation eller ufuldstændig angivelse af klinisk stadiet. Relativt set er misklassifikationsraten størst i region Midtjylland, efterfulgt af region Nordjylland.

Figur 6.1B. Andel resekerede patienter i det totale materiale, fordelt på regioner (efter patientbopæl) og misklassifikationsstatus. Kun forløb med udredningsstadier \leq II er medtaget. Antal forløb: 3370



For fuldstændighedens skyld er i figur 6.1C vist forholdene for specielt forløb med klinisk stadiet i kategorierne IIIa, IIIb og IV. Som forventet er resektionsraten generelt på meget lavt niveau og med kun meget få fuldt belyste forløb. Selv på det meget lave niveau ses der ret udtalt regional variation med en resektionsrate i region Syddanmark på 1.8% som det laveste, over for en resektionsrate på 3.5% i region Nordjylland som det højeste. Relativt er der ikke væsentlige forskelle mellem regionerne hvad angår forekomst af misklassifikation for forløb med resektion.

Figur 6.1C. Andel resecerede patienter i det totale materiale, fordelt på regioner (efter patientbopæl) og misklassifikationsstatus. Kun forløb med udredningsstadier >II er medtaget. Antal forløb: 11011



Det er nærliggende at søge belyst, hvorvidt der er sammenhæng mellem aktivitetsniveau (volumen) ved den udredende afdeling og forekomsten af misklassifikation for forløb med registreret operation (tabel 6.3.)

Tabel 6.3. Association mellem aktivitetsniveau (volumen) ved udredende afdeling og påvist misklassifikation ved efterfølgende operation. Analysen medtager kun fuldt belyste forløb

Udred.afd. Volumen	+ MISKLAS	- MISKLAS	TOTAL
Lille	284	1053	1337
Stort	141	652	793
TOTAL	425	1707	2130

Der er en 1.25 gange øget forekomst af misklassifikation ved forløb med operation udredt på afdeling med lille volumen i forhold til ved afdeling med stort volumen (odds ratio = 1.25; proportionsratio = 1.20). Forøgelsen er dog ikke statistisk signifikant; χ^2 (DF=1) = 3.733; P = 0.053).

Konklusion

Analysen har dokumenteret en ret udtalt og statistisk signifikant variation på tværs af regioner hvad angår overensstemmelse mellem stadieklassifikation ved udredning (klinisk stadie) over for stadieklassifikation ved efterfølgende operation. Region Midtjylland ud viser højest niveau hvad angår misklassifikation. Alle regioner ligger dog inden for accept-niveauet.

Høj grad af misklassifikation synes ikke umiddelbart forbundet med højere resektionsrate, idet forløb uden misklassifikation også udviser betydelig geografisk heterogenitet. I modsætning til den forudgående version af rapporten, findes der i analysen med det nyere og formodet mere homogene datasæt ikke statistisk signifikant evidens for, at udredende afdelings volumen skulle have indvirkning på misklassifikationsniveauet og dermed på mulighederne for at tilbyde operation på et fejlagtigt grundlag.

Det skal understreges, at den geografiske variabilitet i grad af overensstemmelse i stadieklassifikation ved udredning over for stadieklassifikation ved operation udelukkende har kunnet vurderes i det udsnit af forløb, hvor der er registreret et operativt indgreb. Ifølge sagens natur kan analysen ikke belyse variationen i misklassifikation for de patienter, der ikke får foretaget operation.

Der må i fortolkningen af resultaterne tages generelt forbehold, da dataindberetningen specielt hvad angår registrering af udredningsdata og herunder klinisk stadie varierer mellem regioner og over kalendertid.

7. MORTALITET I DET TOTALE MATERIALE

Data og analysestrategi

I denne analyse belyses, hvorledes mortaliteten for patienter med primær lungecancer ændres over tid og mellem regioner. Der tages udgangspunkt i det totale materiale.

Faktorer af primær interesse for vurderinger af variation:

- Region (patientbopæl, med region Hovedstaden som reference)
- Diagnoseår (2007 som reference)

Prognostiske faktorer:

- Sygdomsstadie (efter klinisk stadie)
- Patologi (SCLC/NSCLC/uoplyst, med NSCLC som reference)
- Operation (ja/nej med nej som reference)
- Charlsongruppe (3 grupper, med gruppe 0 (ingen komorbiditet) som reference)

Justeringsfaktorer:

- Alder ved diagnose (Alder < 67 år som reference)
- Køn (mænd som reference)

Justering af alder er ud fra forskellige statistiske afprøvninger bedømt at kunne foretages mest hensigtsmæssigt ved dikotomisering svarende til alder 67 år.

Ved vurdering af den prognostiske betydning af stadie er stadie 0 slået sammen med stadie I af hensyn til små tal, således at der opereres med følgende stadiekategorier: 0+I, II, IIIa, IIIb og IV.

Stadie 0+I er anvendt som reference for andre stadiekategorier.

Ved vurdering af faktorer af primær interesse er året 2007 anvendt som reference for de andre år, og region Hovedstaden er reference for regioner.

Der er anvendt Cox regressionsmodel med kovariater: alder, køn, operation, stadie, region og diagnosticeringsår. Til belysning af heterogenitet med hensyn til kalendertid og geografi på regionsniveau er der suppleret med frailty modeller.

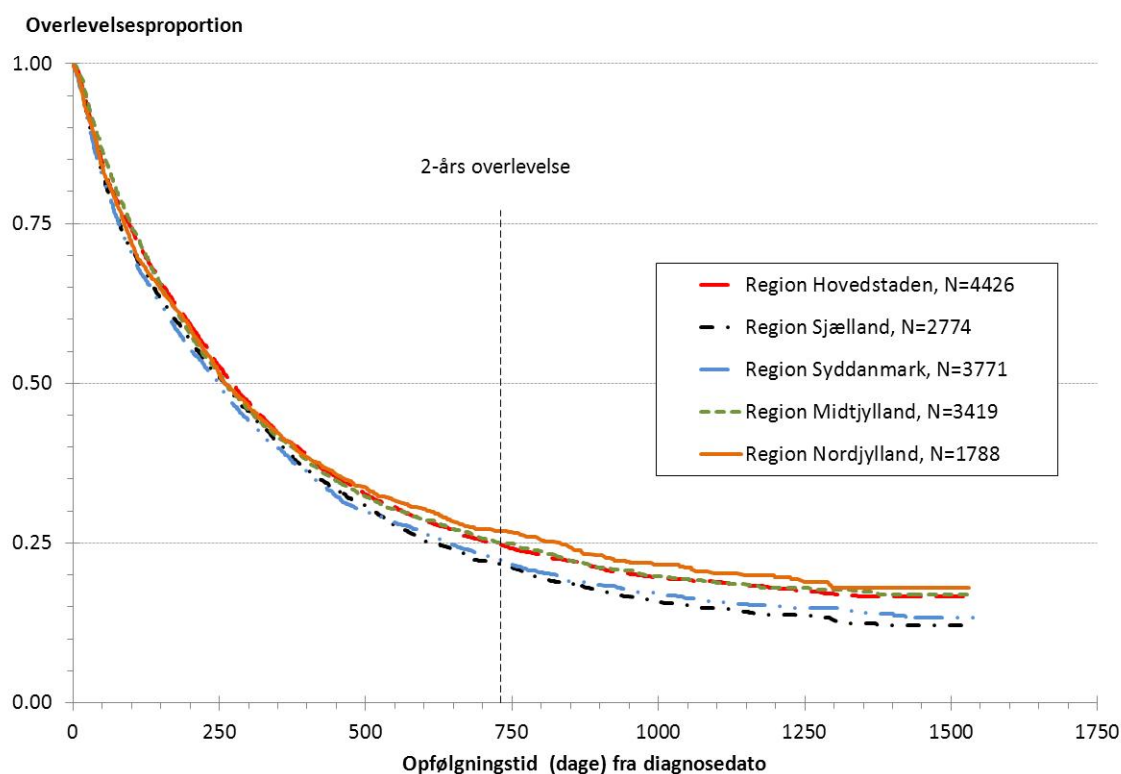
Hvad angår oversigt over datakomplethed henvises til tabel 4.2.

Resultater

Indledningsvist bringes *ujusterede* overlevelseskurver for det totale materiale efter patienternes registrerede bopælsregion (figur 7.1). Et beskedent antal forløb med inkonsistente opfølgningstider og/eller ukendt region er udeladt. I den ujusterede analyse har region Sjælland lavest 2-års overlevelse (21.8%), med region Syddanmark på næsten samme niveau (22.3%). De tre øvrige regioner har stort set sammenfaldende kurveforløb med region Nordjylland øverst (26.8%) efterfulgt af region Midtjylland (25.0%) og region Hovedstaden (24.8%).

I figur 7.2 vises de regionale og nationale (og fremdeles ujusterede) overlevelseskurver, men nu med stratifikation for Charlsongruppe. Den generelle tendens er for alle regioner som forventet, at højere kategori for komorbiditet er forbundet med lavere overlevelse.

Figur 7.1. Analyse af det totale materiale: Overlevelseskurver (uanset diagnoseår) for patienter efter registreret bopælsregion. Antal forløb: 16274 i den ujusterede analyse



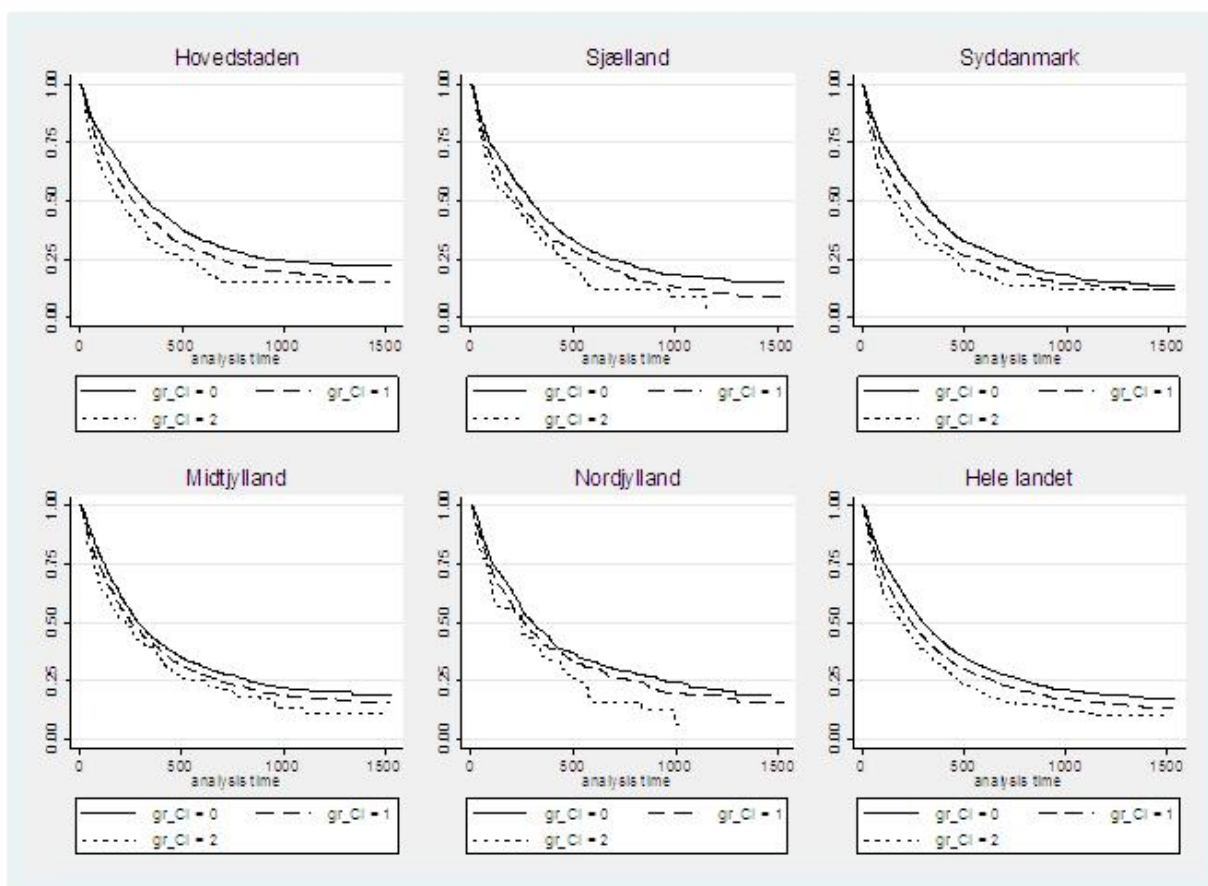
Tabel 7.1 indeholder resultaterne af den detaljerede analyse af den fulde statistiske model. I analysen indgår 13493 forløb (82.9% af samtlige 16274 forløb) med alle baggrundsvariable og prognostiske faktorer belyst.

For køn, alder, klinisk stadie og operationsstatus er resultaterne som forventet, dvs. at kvinder har lavere mortalitetsniveau end mænd, operation medfører en væsentlig mortalitetsreduktion, og høj alder såvel som højt klinisk stadie er forbundet med højere mortalitet; for de laveste stadier er billedet knapt så entydigt, idet stadie II umiddelbart ser ud til at være forbundet med marginalt højere mortalitet end stadie IIIa.

Øget komorbiditetsniveau (høj Charlsongruppe) er associeret med høj dødelighed. Det seneste diagnoseår (2010) er forbundet med væsentligt lavere mortalitetsniveau i forhold til de forudgående diagnoseår.

I den fulde model har regionerne Sjælland, Syddanmark, Midtjylland og Nordjylland hver især mortalitetsniveauer, der er statistisk signifikant forøget (op til 30%) i forhold til region Hovedstaden. Overordnet set er den geografiske heterogenitet statistisk signifikant ($P=0.000$). Dette billede adskiller sig derved ret markant fra mønstret set ved den ujusterede analyse (figur 7.1), hvor kun regionerne Sjælland og Syddanmark ligger lavere mens de øvrige regioner er ligget samlet på et lidt bedre niveau.

Figur 7.2. Analyse af det totale materiale: Overlevelseskurver (uanset diagnoseår) for patienter efter registreret bopælsregion og yderligere stratificeret efter Charlsongruppe. Antal forløb: 13493



Konklusion

Resultaterne er som forventet hvad angår mortalitetsforholdene for alder, køn, klinisk stadie og operationsstatus. Endvidere er høj grad af komorbiditet, her vurderet ved Charlson Index, forbundet med øget mortalitetsniveau. Med korrektion for de øvrige variable i den statistiske model er small cell lung cancer forbundet med statistisk signifikant lavere dødelighed i forhold til non small cell lung cancer.

Efter justering for de medtagne faktorer bekræfter den opdaterede analyse til en vis grad billedet af geografisk heterogenitet. Generelt fremstår region Hovedstaden med laveste mortalitetsniveau, over for de andre regioner som i større eller mindre grad ligger noget højere i mortalitetsniveau. På grund af den høje grad af korrelation mellem udredende afdelings volumen og regionale tilhørsforhold, kan analyseresultatet ikke forklares isoleret ved en volumeneffekt, men må søges forklaret i andre forhold, som varierer mellem regionerne, men som ikke umiddelbart kan belyses ved de data og rammer, der er tilgængelige for nærværende rapport. Hypotetisk kunne man i denne sammenhæng pege på effekten af eventuelle regionale forskelle i den onkologiske behandling af patienterne.

Analyseresultaterne understreger dels kompleksiteten ved at skulle søge forklarende årsager til variation i overlevelsesforholdene ved primær lungecancer, dels nødvendigheden af at inddrage en lang række faktorer, herunder alder, køn, klinisk stadie og operationsforhold.

Der må i fortolkningen af resultaterne tages generelt forbehold, da dataindsamlingen hvad angår udrednings- og behandlingsaktivitet varierer mellem regioner og over kalendertid.

Tabel 7.1. Analyse af det totale materiale. Statistiske resultater. Der indgår 13493 forløb (82.9%) af det totale antal forløb

Variabel	Hazard ratio	C.L. (95%)	P-værdi
Alder (ref.: <67 år) ≥67 år	1.39	1.33 - 1.45	0.000
Køn (ref.: Mand) Kvinde	0.88	0.85 - 0.92	0.000
Operation (ref.: Nej) Ja	0.32	0.29 - 0.35	0.000
Charlsongruppe (ref: Gruppe 0) Gruppe 1 Gruppe 2	1.17 1.38	1.12 - 1.22 1.28 - 1.49	0.000 0.000
Patologi (ref.: NSCLC) SCLC Uoplyst	0.87 1.25	0.82 - 0.93 1.18 - 1.32	0.000 0.000
Stadie (ref.: 0+I kombineret) Stadie II Stadie IIIa Stadie IIIb Stadie IV	1.62 1.51 2.09 3.34	1.45 - 1.82 1.36 - 1.68 1.91 - 2.28 3.07 - 3.63	0.000 0.000 0.000 0.000
Region (ref.: Region Hovedstaden) Sjælland Syddanmark Midtjylland Nordjylland	1.11 1.23 1.19 1.20	1.04 - 1.18 1.16 - 1.30 1.12 - 1.26 1.12 - 1.30	0.002 0.000 0.000 0.000
Diagnoseår (ref.: 2007) 2008 2009 2010	1.01 0.96 0.80	0.95 - 1.06 0.91 - 1.01 0.75 - 0.86	0.853 0.130 0.000

8. MORTALITET UDEN RESEKTION

Data og analysestrategi

I denne analyse belyses, hvorledes mortaliteten ændres over tid og mellem regioner for forløb, hvor der *ikke* er fået foretaget resektion. Analysen medtager de 131 forløb, hvor der er foretaget eksplorativt indgreb.

Faktorer af primær interesse for vurderinger af variation:

- Region (patientbopæl, med region Hovedstaden som reference)
- Diagnoseår (2007 som reference)

Prognostiske faktorer:

- Sygdomsstadie *i henhold til udredning* (klinisk stadie, med laveste gruppe som reference)
 - Patologi (SCLC/NSCLC/uoplyst, med NSCLC som reference)
- Charlsongruppe (3 grupper, med gruppe 0 (ingen komorbiditet) som reference)

Justeringsfaktorer:

- Alder ved diagnose (alder <67 år som reference)
- Køn (mænd som reference)

Justering af alder er ud fra forskellige statistiske afprøvninger bedømt at kunne foretages mest hensigtsmæssigt ved dikotomisering svarende til alder 67 år.

Ved vurdering af den prognostiske betydning af stadie er stadie 0 slået sammen med stadie I af hensyn til små tal, således at der opereres med følgende stadiekategorier: 0+I, II, IIIa, IIIb og IV.

Stadie 0+I er anvendt som reference for andre kategorier.

Ved vurdering af faktorer af primær interesse er året 2007 anvendt som reference for de andre år, og region Hovedstaden er reference for regioner.

Der er anvendt Cox regressionsmodel med kovariater: alder, køn, stadie, region og diagnosticeringsår. Til belysning af heterogenitet med hensyn til kalendertid og geografi (på regionsniveau, henholdsvis hospitalsniveau) er der suppleret med frailty modeller.

Datakompletheden hvad angår de relevante variable fremgår af tabel 8.1 og dokumenterer acceptabel complethed. Største deficit (ca. 10%) er manglende information om udredningsstadie på grund af manglende udredningsdata for de pågældende forløb. Der er fuld information for i alt 11460 (83.5%) af det totale relevante antal forløb.

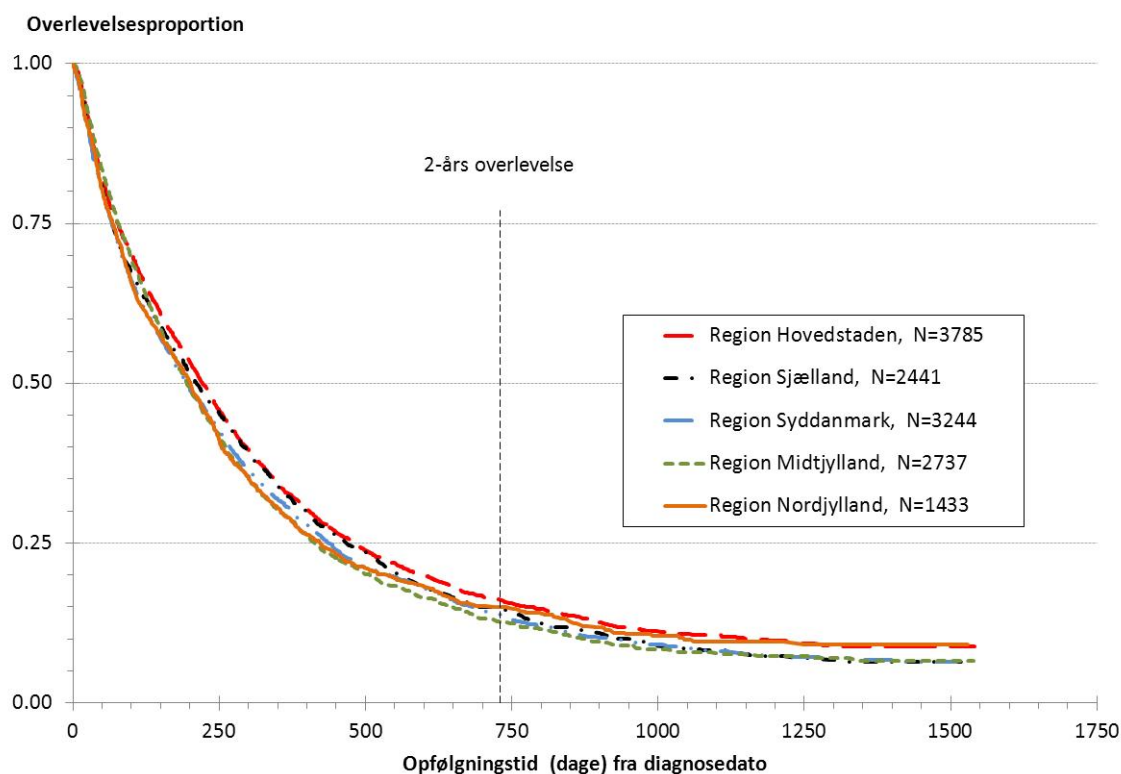
Tabel 8.1. Oversigt over datakomplethed for analyse mortalitet uden operation

Variabel	Med komplette data	Med manglende data	Datakomplethed (%)	I alt
Alder	13723	0	100.0	13723
Køn	13723	0	100.0	13723
Klinisk stadie	12301	1422	89.6	13723
Region	13641	82	99.4	13723
Diagnoseår	13723	0	100.0	13723
Charlsongruppe	13538	53	99.6	13723

Resultater

Indledningsvist bringes *ujusterede* overlevelseskurver for patienterne uden resektion efter patienternes registrerede bopælsregion (Figur 8.1). I den første del af kurveforløbet ligger regionerne ret tæt; hvorefter kurveforløbene begynder at afvige fra hinanden. I den estimerede 2-års overlevelse ligger region Midtjylland lavest (12.6%), mens region Hovedstaden ligger højest (16.1%). Region Nordjylland ligger i kurveforløbets afslutning (med kun begrænset antal observationer) på relativt højt niveau sammen med region Hovedstaden. Det skal dog anføres, at region Nordjyllands kurve ligger lavest i det første år efter diagnostetidspunktet.

Figur 8.1. Analyse af patientmaterialet uden resektion. Overlevelseskurver (uanset diagnoseår) for patienter efter registreret bopælsregion. Forløb med inkonsistent opfølgningstid og ukendt region er udelukket



Resultaterne af den statistiske analyse er vist i tabel 8.2. Der indgår 11460 forløb (83.5% af de mulige) i den statistiske analyse. Generelt findes et effektmønster for de inkluderede variable, der i mange henseender er sammenfaldende med resultaterne fra analysen af det totale materiale (tabel 7.1).

For køn, alder, klinisk stadie og operationsstatus er resultaterne som forventet, dvs. at kvinder har lavere mortalitetsniveau end mænd, operation medfører en væsentlig mortalitetsreduktion, og høj alder såvel som højt klinisk stadie er forbundet med højere mortalitet; for de laveste stadier er billedet knapt så entydigt, idet stadie II umiddelbart ser ud til at være forbundet med marginalt højere mortalitet end stadie IIIa.

Øget komorbiditetsniveau (høj Charlsongruppe) er associeret med høj dødelighed. Det seneste diagnoseår (2010) er forbundet med væsentligt lavere mortalitetsniveau i forhold til de forudgående diagnoseår.

I den fulde model har regionerne Syddanmark, Midtjylland og Nordjylland hver især mortalitetsniveauer, der er statistisk signifikant forøget (af størrelsesordenen 20%) i forhold til region Hovedstaden, mens region Sjælland ikke adskiller sig statistisk signifikant fra region Hovedstaden. Overordnet set er den geografiske heterogenitet statistisk signifikant ($P=0.000$). Dette billede adskiller sig derved ret markant fra mønsteret set ved den ujusterede analyse (figur 8.1), hvor regionerne Hovedstaden og Nordjylland tilsyneladende ligger på samme niveau med relativt bedre langtidsoverlevelse end de tre øvrige regioner.

Tabel 8.2. Patientmaterialet uden operation. Statistisk analyse efter REGION. Antal analyserede forløb: 11460 ud af 13273 (83.5 %)

Variabel	Hazard ratio	C.L. (95%)	P-værdi
Alder (ref.: <67 år)			
≥67 år	1.37	1.31- 1.43	0.000
Køn (ref.: Mand)			
Kvinde	0.89	0.85 - 0.93	0.000
Charlsongruppe (ref: Gruppe 0)			
Gruppe 1	1.16	1.11 - 1.21	0.000
Gruppe 2	1.35	1.24- 1.46	0.020
Stadie (ref.: 0+I kombineret)			
Stadie II	1.56	1.36 - 1.79	0.000
Stadie IIIa	1.53	1.37- 1.72	0.000
Stadie IIIb	2.11	1.91 – 2.33	0.000
Stadie IV	3.34	3.05 - 3.66	0.000
Region (ref.: Region Hovedstaden)			
Sjælland	1.10	1.03 - 1.17	0.004
Syddanmark	1.22	1.15 - 1.29	0.000
Midtjylland	1.17	1.10 - 1.25	0.000
Nordjylland	1.19	1.10 - 1.29	0.000
Diagnoseår (ref.: 2007)			
2008	1.00	0.94 - 1.06	0.927
2009	0.97	0.91 - 1.03	0.265
2010	0.80	0.75 – 0.86	0.009

Konklusion:

Resultaterne falder ud som forventet hvad angår mortalitetsforholdene for alder, køn, klinisk stadie og operationsstatus. Endvidere er høj grad af komorbiditet, her vurderet ved Charlson Index, forbundet med øget mortalitetsniveau. Med korrektion for de øvrige variable i den statistiske model er small cell lung cancer forbundet med statistisk signifikant lavere dødelighed i forhold til non small cell lung cancer.

Efter justering for de medtagne faktorer bekræfter den opdaterede analyse til en vis grad billedet af geografisk heterogenitet. Generelt fremstår region Hovedstaden med laveste mortalitetsniveau, og med region Sjælland på stort set samme niveau. Heroverfor ligger de andre regioner af størrelsesordenen 20% højere i mortalitetsniveau, hvilket hver især for regionerne er statistisk signifikant højere i forhold til region Hovedstaden. På grund af den høje grad af korrelation mellem udredende afdelings volumen og regionale tilhørsforhold, kan analyseresultatet ikke forklares isoleret ved en volumeneffekt, men må søges forklaret i andre forhold, som varierer mellem regionerne, men som ikke umiddelbart kan belyses ved de data og rammer, der er tilgængelige for nærværende rapport. Hypotetisk kunne man i denne sammenhæng pege på effekten af eventuelle regionale forskelle i den onkologiske behandling af patienterne.

Som ved analysen af det totale materiale understreges nødvendigheden af at inddrage en lang række faktorer, herunder alder, køn, klinisk stadie og operationsforhold, som alle indgår på kompleks måde i indflydelsen på overlevelsesforholdene ved primær lungecancer.

Der må i fortolkningen af resultaterne tages generelt forbehold, da dataindsamlingen hvad angår udrednings- og behandlingsaktivitet varierer mellem regioner og over kalendertid.

9. MORTALITET EFTER RESEKTION

Data og analysestrategi

I denne analyse belyses mortaliteten for patienter med primær lungecancer efter resektion, specielt hvorvidt prognosen ved resektion ændres over tid og mellem regioner. Det skal understreges, at 131 forløb med eksplorativt kirurgisk er udeladt af denne analyse.

Faktorer af primær interesse for vurderinger af variation:

- Opererende hospital (med Rigshospitalet som reference)
- Diagnoseår (2007 som reference)

Prognostiske faktorer:

- Sygdomsstadie *i henhold til udredning* (klinisk stadie, med laveste gruppe som reference)
- Misklassifikation (dikotomiseret, med "Nej" som reference)
- Charlsongruppe (3 grupper, med gruppe 0 (ingen komorbiditet) som reference)

Justeringsfaktorer:

- Alder ved diagnose (alder <67 år som reference)
- Køn (mænd som reference)

Justering af alder er ud fra forskellige statistiske afprøvninger bedømt at kunne foretages mest hensigtsmæssigt ved dikotomisering svarende til alder 67 år.

Ved vurdering af den prognostiske betydning af stadie er stadie 0 slået sammen med stadie I af hensyn til små tal, således at der opereres med følgende stadiekategorier: 0+I, II, IIIa, IIIb og IV.

Stadie 0+I er anvendt som reference for andre kategorier.

Ved vurdering af faktorer af primær interesse er året 2007 anvendt som reference for de andre år. I stedet for region er operationssygehus anvendt som geografisk variabel, med Rigshospitalet som reference for de øvrige behandlende hospitaler.

Der er anvendt Cox regressionsmodel med kovariater: alder, køn, stadie, misklassifikation, region/hospital og diagnosticeringsår. Til belysning af heterogenitet med hensyn til kalendertid og geografi er der suppleret med frailtymodeller.

Datakompletheden hvad angår de relevante variabler fremgår af tabel 9.1 og dokumenterer nedsat komplethed, som overvejende kan tilskrives manglende information om udredningsdata og dermed klinisk stadie. Der er fuld information for i alt 2019 (79.1%) af det totale relevante antal forløb (n=2551).

Tabel 9.1. Oversigt over datakomplethed for analyse af mortalitet med operation

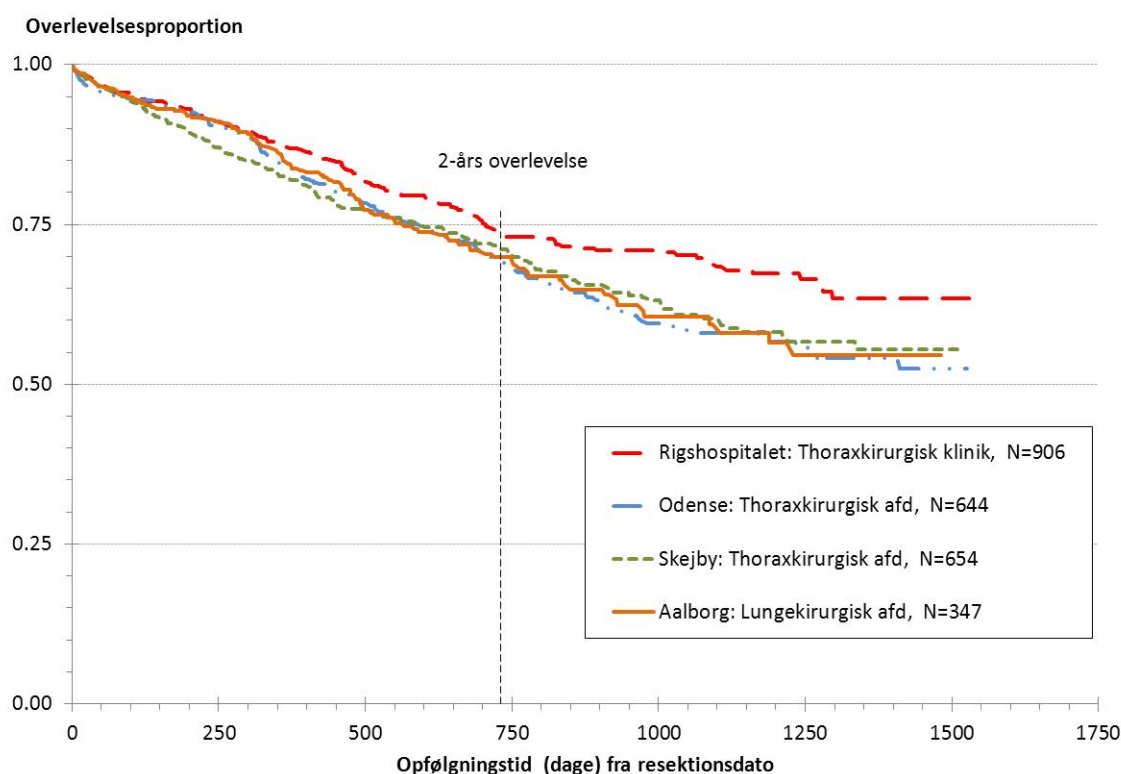
Variabel	Med komplette data	Med manglende data	Datakomplethed (%)	I alt
Alder	2551	0	100.0	2551
Køn	2551	0	100.0	2551
Stadie	2122	429	83.1	2551
Region	2551	0	100.0	2551
Diagnoseår	2551	0	100.0	2551
Misklassifikation	2060	491	80.7	2551
Charlsongruppe	2541	10	99.6	2551

Resultater

Indledningsvist bringes *ujusterede* overlevelseskurver for materialet med registreret operation efter opererende sygehus (Figur 9.1). Efter ca. 100 dage fra resektionstidspunktet skiller Skejby Sygehus sig ud ved at ligge lidt lavere i forhold til de tre øvrige operationssteder. Efter ca. et års opfølgningstid devierer kurveforløbene yderligere i forhold til hinanden, men sådan at Rigshospitalet ret konsekvent fremstår med den relativt højeste overlevelse.

I estimerne på den ujusterede 2-års overlevelse ligger Rigshospitalet højest (73.5%), mens værdierne ligger tæt for Odense Universitetshospital (69.0%), Skejby Sygehus (71.4%) og Aalborg Sygehus (69.9%).

Figur 9.1. Analyse af patientmaterialet med resektion. Overlevelseskurver (uanset diagnoseår) fra resektionsdato for patienter efter opererende sygehus.



Resultaterne af den statistiske analyse er sammenfattet i tabel 9.2 med den geografiske dimension repræsenteret ved behandlende sygehus. Der indgår 2019 forløb (79.1% af de mulige) i analysen.

For alder og køn findes effekter som forventet. Endvidere er der øget mortalitetsniveau i takt med øget komorbiditet. Der er ikke evidens for ændringer i mortaliteten i forhold til kalendertiden.

Ifølge modellen er misklassifikation en betydende prognostisk faktor med væsentligt øget mortalitetsniveau for forløb med misklassifikation. Det skal understreges, at misklassifikation i den foreliggende sammenhæng er dikotomiseret som en 'ja/nej'-variabel, hvor 'ja' til misklassifikation oftest vil betyde et *højere* stadie efter operation i forhold til udredningsstadiet. Dette medfører potentiel interaktion mellem misklassifikation og stadie. Med disse forbehold er modellens estimer i overensstemmelse med statistisk signifikant stigende mortalitetsniveau med stigende udredningsstadie.

Der er statistisk signifikant heterogenitet ($P=0.004$) mellem behandlende sygehuse, idet Odense Universitetshospital i henhold til modellen har en statistisk signifikant øgning i mortalitetsniveau på ca. 30% i forhold til Rigshospitalet, som ligger lavest. Ingen af de to øvrige behandlende sygehuse udskiller sig signifikant i forhold til Rigshospitalet.

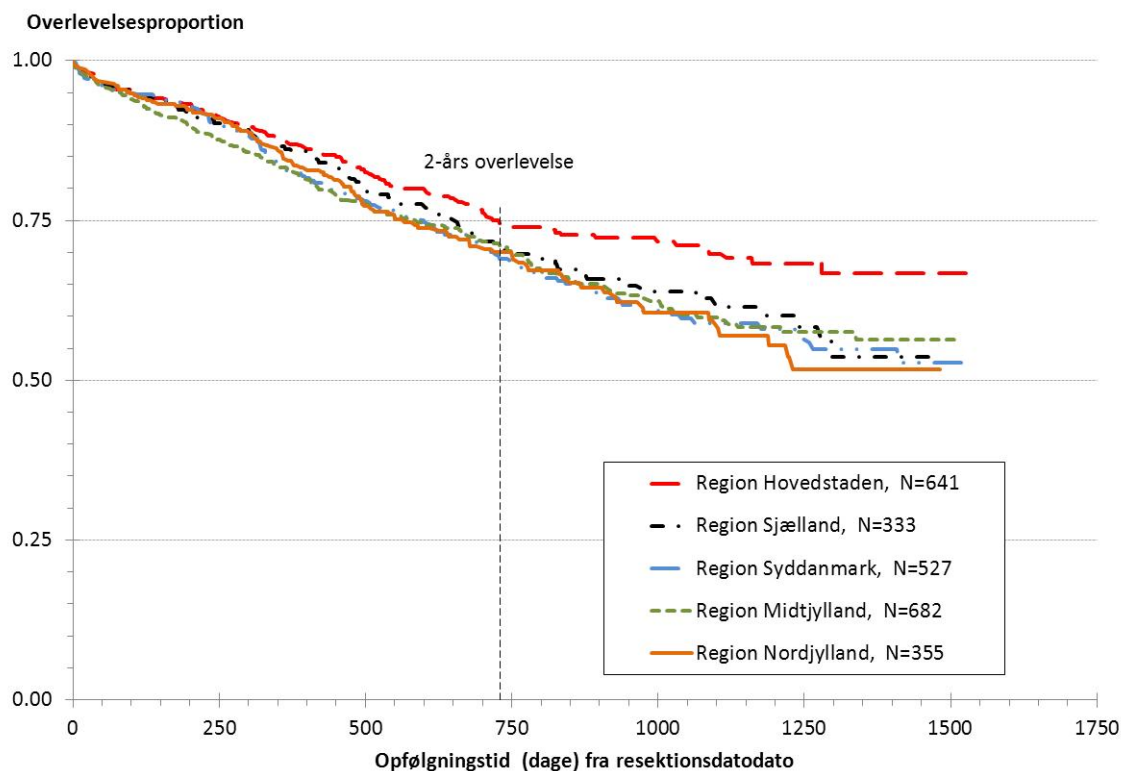
Tabel 9.2. Patientmaterialet med resektion. Statistisk analyse efter BEHANDLENDE SYGEHUS. Antal analyserede forløb: 2019 ud af 2551 (79.1%)

Variabel	Hazard ratio	C.L. (95%)	P-værdi
Alder (ref.: <67 år)			
≥67 år	1.59	1.33 - 1.89	0.000
Køn (ref.: Mand)			
Kvinde	0.80	0.67 - 0.96	0.014
Misklassifikation (ref.: Nej)			
Ja	2.42	2.00 - 2.92	0.000
Charlsongruppe (ref.: Gruppe 0)			
Gruppe 1	1.29	1.08 - 1.54	0.005
Gruppe 2	1.93	1.39 - 2.68	0.000
Stadie (ref.: 0+I kombineret)			
Stadie II	1.68	1.35 - 2.08	0.000
Stadie IIIa	1.40	0.97 - 2.01	0.069
Stadie IIIb	1.63	1.12 - 2.37	0.010
Stadie IV	2.94	2.01 - 4.30	0.000
Behandelende sygehus (ref.: Rigshospitalet)			
Odense Universitetshospital	1.31	1.04 - 1.66	0.021
Århus Universitetshospital, Skejby	1.11	0.88 - 1.41	0.370
Ålborg Sygehus	1.06	0.79 - 1.41	0.709
Diagnoseår (ref.: 2007)			
2008	1.08	0.88 - 1.34	0.451
2009	0.77	0.60 - 1.00	0.053
2010	0.79	0.56 - 1.12	0.187

Resultaterne af den statistiske analyse demonstrerer, hvor komplekst det er at skulle identificere mulige forklaringer bag den påviste variation i mortalitet på tværs af behandlingssteder. Kurveforløbene i figur 9.1. antyder, at prognosen efter resektion først begynder at variere flere måneder efter resektionstidspunktet, altså lang tid efter den umiddelbare helbredsmæssige effekt af operationen. Hertil kommer, at selv om der er justeret for komorbiditet ved hjælp af Charlson Index, vil der formentligt være andre konkrete kliniske vurderinger af patientoperabilitet, som der ikke kan tages højde for på det foreliggende datagrundlag. Det kan heller ikke vides, om der overordnet set er forskelle mellem de opererende centres måder at udvælge patienter til resektion. Endelig vil sammensætningen af patientgrundlaget for resektion ved et givet center ikke alene afhænge af profilen for patienterne fra egen region, men også af profilen af de patienter der bliver henvist til operation fra andre regioner. I forhold til lungecancer er dette en problemstilling af størst relevans for Rigshospitalet og Odense Universitetshospitalet på grund af samarbejdsaftaler vedrørende den operative behandling af patienter fra region Sjælland.

Som supplerende belysning af disse forhold er der foretaget en ujusteret analyse af overlevelsen for patienter efter resektion, men nu opgjort i forhold til patientbopæl (regionsniveau) som alternativ til opgørelse i forhold til behandelende sygehus (figur 9.2). Som forventet er der stort set sammenfaldende kurveforløb i de første 100 dage efter resektion, hvorefter Rigshospitalet fremdeles ligger med relativt højest overlevelsesniveau, mens de øvrige regioner er mere eller mindre sammenfaldende. Dette mønster bekræfter, at Rigshospitalet og region Hovedstaden generelt fremstår med den højeste overlevelse, hvorimod der ikke er nogen klar evidens for at identificere specielt dårlige behandlingsresultater for de øvrige behandlingssteder og regioners vedkommende.

Figur 9.2. Analyse af patientmaterialet med resektion. Overlevelseskurver (uanset diagnoseår) fra resektionsdato for patienter efter patientbopæl på regionsniveau



Konklusion

Resultaterne af den statistiske analyse understøtter, at der overordnet set forefindes heterogenitet på tværs af behandlende sygehuse. Ud over at Rigshospitalet fremstår som behandlingsstedet med den bedste prognose, kan der ikke ud fra resultatet af analysen identificeres behandlingssteder med specielt dårlig prognose. Det skal i denne sammenhæng understreges, at der synes at være ensartet korttidsprognose efter resektion på tværs af opererende centre og regioner, hvorfor forskellene formentligt skal forklares ved forhold, der har med patienternes efterbehandling at gøre, først og fremmest i onkologisk regi. Det er ikke muligt inden for rammerne af nærværende rapport og det forhåndenværende datamateriale at belyse sådanne forhold nærmere.

APPENDIX: BEREGNING AF CHARLSON INDEX

Dette Appendix dokumenterer, hvorledes de kumulerede data fra Landspatientregisteret (LPR) er blevet anvendt til at beregne Charlson Index for populationen af registrerede patienter i DLCR.

Grupperingen af diagnoser vises i bilag 1.

Bilag 2 fremstiller den SAS-algoritme (script), der er modtaget sammen med data og som anses for dokumentation af, hvorledes Charlson Index indtil nu har været håndteret i dansk registerforskning. I forbindelse med den aktuelle analyse har denne algoritme vist sig uhensigtsmæssig hvad angår bl.a. håndteringen af diabeteskoder. Endvidere er det udokumenteret, hvorledes SAS-algoritmen håndterer elimineringen af bidrag til Charlson Index fra den sygsom (her lungecancer), som patientgruppen repræsenterer.

Klargøring af data

Data fra LPR modtages på CD eller DVD fra Kræftens Bekæmpelse eller Forskerservice i Sundhedsstyrelsen. Vi bruger to tabeller fra LPR, adm og diag, som indeholder henholdsvis personnumre og diagnosekoder for de relevante patienter. Gruppen af relevante patienter leveres af (ellers hentes fra) DLCR's inddaterings-system, Topica.

Desuden bruger vi en tabel med diagnosedato for lungecancer. Der skal være én –og kun én– diagnosedato for hver patient. Diagnosedatoerne hentes i den såkaldte analysedatabase, som KCS/FKE har udviklet og driver.

Adm-tabellerne

adm7793 og adm9408		
Felt	Format	Note
cpr	tekst	<i>indeks</i>
DxDATE	dato	
V_RECNUM	heltal	<i>indeks</i>
D_INDDTO	dato	
D_UDDTO	dato	

Adm-tabellerne fra LPR valideres mht. om datoerne er gyldige og v_recnum er gyldigt som heltal. Desuden kan vores udsnit af adm-tabellen have strukturelle defekter vedr. recordnummer og manglende diagnoser.

Recordnummer

I adm-tabellen er recordnummeret (V_RECNUM) entydigt, dvs. hver registrering har sit eget recordnummer. Men det er ikke altid tilfældet i vores udtræk af adm. Her finder vi lejlighedsvis recordnumre, som hver især er tilknyttet 2 forskellige CPR-numre. Vi har dog ikke set recordnumre, som er tilknyttet flere end 2 CPR-numre. Det er desværre ikke muligt at reparere disse *dobbelte recordnumre*, men nogle af dem er ikke relevante her, fordi:

- diagnosekoden indgår ikke i Charlson Index eller
- diagnosekoden er en solid kræftform eller en metastaserende kræftform, og diagnosedato er højst 150 dage tidligere end lungekræftdiagnosen (herom senere).

Når man ser bort fra dem, bliver der ofte nogle dobbelte recordnumre tilbage. Disse registreringer har en relevant diagnosekode, men patienten kan ikke identificeres entydigt. Det er ikke muligt at gøre recordnumrene entydige med det udleverede materiale. Derfor er analysen foretaget *med* de dobbelte recordnumre.

Manglende diagnoser

Udtræk fra Adm-tabellen kan indeholde patienter, som ikke findes i Diag-tabellen. Patienterne er ellers forespurgt hos LPR, hvor de er registreret, men deres diagnoser ses ikke modtaget. Det betyder, at patienterne ikke kommer til at indgå i analysen. Vi ser typisk under 200 af disse patienter, hvis fødselsdato og inklusionsdato ellers betyder, at patienterne skulle være med i analysen. Fødselsåret skal være 1920 eller senere, og inklusionsdato skal være fra 2003 til seneste hele registreringsår (pt. 2010).

De udleverede data giver ikke mulighed for at indkredse årsagen til de manglende diagnoser, og data har typisk været igennem flere instanser før ankomsten til KCS/FKE.

Diag-tabellerne

diag7793 og diag9408		
Felt	Format	Note
cpr	tekst	<i>indeks</i>
V_RECNUM	heltal	<i>indeks</i>
C_DIAG	tekst	<i>indeks</i>
C_DIAGTYPE	tekst	
C_TILDIAG	tekst	

Diagnosekoderne opbevares i feltet C_DIAG, mens feltet V_RECNUM refererer til Adm-tabellen.

Sidste klargøring

Adm- og Diag-tabellerne er indekseret for at forkorte svartiden fra databasen. Tabellerne ovenfor viser hvilke felter, der er indekseret. Alle indices indeholder ét felt hver især.

Data er ikke ændret under klargøringen.

Metode

Charlson Index anvender 19 Charlsongrupper, som kan ses i bilag 1. Dette bilag viser også, hvilke diagnosekoder, som tælles med i de respektive grupper.

Beregning af Charlson Index er lavet i en række skridt, som beskrives nedenfor. De overordnede skridt (hovedpunkterne) hedder M1, M2 og M3, og de beskriver metoden i grove træk. Detaljerne i metoden er beskrevet med punkterne C1 – C11, som giver flere tekniske detaljer.

Charlsonvægtene bliver beregnet i M1, hvor der tages hensyn til, at lungekræftdiagnosen ikke skal indgå i beregningen. Men en gennemgang af data har vist, at man bør tage hensyn til flere diagnoser end lungekræft, når man beregner Charlsonvægtene. Det skyldes, at data indeholder en række kræftdiagnosekoder, som umiddelbart kan tilskrives lungekræft, selvom de ikke er registreret under lungekræft specifikt. Disse diagnosekoder bør skilles fra, når man beregner Charlson Index med lungekræft som indekssygdom, og det sker i M2. Det er Charlsongrupperne for solide og metastaserende kræftformer, som påvirkes af disse kræftdiagnosekoder.

Kort fortalt beregner M2 Charlsonvægtene for solide og metastaserende kræftformer, men beregningen ser bort fra registreringer, som med stor sandsynlighed kan relateres til lungekræft. De 150 dage er valgt efter en manuel gennemgang af data.

M1 – almindelige Charlsonvægte

Formål: beregn Charlsonvægte i alle Charlsongrupper for alle patienter. Grundlaget er registreringer, som ikke har en lungekræftdiagnosekode, og hvor diagnosedato er tidligere end inklusionsdato for den pågældende patient.

Design

Dette første skridt (M1), hvor vi finder vægte for alle Charlsongrupper, laves med følgende metode:

- C1. Find alle registreringer, hvor følgende er opfyldt.
 - a. diagnosedato (D_INDDTO) er tidligere end inklusionsdato (DxDATE),
 - b. diagnosekoden tilhører en Charlsongruppe og
 - c. diagnosekoden er ikke lungekræft, dvs. ICD10-koder som ikke begynder med DC34.
 Resultatet er en tabel med Cpr-nr, diagnosedato, diagnosekode og Charlsongruppe.
- C2. Resultatet fra C1 inddeles i grupper. En gruppe består af 1 eller flere registreringer, som har samme Cpr-nr og samme Charlsongruppe.
- C3. Find den tidligste diagnosedato i hver gruppe.
- C4. Rapportér følgende oplysninger fra hver gruppe: cpr-nummer, tidligste diagnosedato, Charlsongruppe samt Charlsonvægt. Sidstnævnte er let at finde, fordi Charlsongruppen er kendt.

Resultatet er en tabel med CPR-nummer, tidligste diagnosedato, Charlsongruppe og Charlsonvægt. Bilag 3 viser SQL-koden, som udfører trinene C1 – C4.

M2 – Charlsonvægte for kræftdiagnoser

Formål: beregn Charlsonvægte for solide kræftformer og metastaserende kræftformer. Brug registreringer, som opfylder flg:

- a. Diagnosekoden er en solid eller metastaserende kræftform.
- b. Diagnosekoden er ikke lungekræft specifikt.
- c. Diagnosedato er mindst 150 dage tidligere end patientens inklusionsdato i DLCR. De 150 dage er valgt efter en manuel gennemgang af data.
- d. Diagnosedatoen falder ikke sammen med de dage, hvor patienten har en lungekræftdiagnosekode.

Design

Dette skridt (M2) følger samme princip som M1, men her indgår flere søgekriterier. Derfor bliver søgeresultatet mere afgrænset.

M2 bruger det samme datasæt (input) som M1. Det vil sige, at alle patienter og diagnosekoder er input til M2. Metoden er følgende:

- C5. Find registreringer, hvor følgende er opfyldt.
 - a. diagnosedato falder ikke sammen med en dato, hvor patienten har en lungekræftdiagnose;
 - b. diagnosedato er senest 31.12.2007;
 - c. diagnosekoden solid eller metastaserende kræft, men ikke lungekræft;
 - d. diagnosedatoen er 150 dage før inklusionsdato eller tidligere; der er ud fra en arbitrær betragtning indlagt en 'karenstid' på 150 dage for at kompensere for det forhold, at første lungecancer-relevante diagnose i LPR kan forekomme før inklusionsdato (diagnosedato) registreret i DLCR; dette vil for eksempel kunne forekomme, hvor udredningsdata ikke er indberettet til DLCR.
- C6. Resultatet fra C1 inddeles i grupper. En gruppe består af 1 eller flere registreringer, som har samme Cpr-nr og samme Charlsongruppe. Find den tidligste diagnosedato i hver gruppe.
- C7. Rapportér følgende oplysninger fra hver gruppe: cpr-nummer, tidligste diagnosedato, Charlsongruppe og Charlsonvægt.

Strukturelt set ligner resultatet C4, men SQL-koden er anderledes her, se bilag 4. Dels indeholder koden en kolonne til patientens inklusionsdato, som bruges til at implementere C5d. Dels indgår kolonnen *Serial*, som ikke er relevant her, men bruges i anden sammenhæng.

M3 – det samlede Charlsonindeks

Formål: Kombiner resultaterne fra M1 og M2 med denne metode:

- a. Patienter, som ikke fik en Charlsonvægt under punkt M2, tildeles vægten 0 under solide og metastaserende kræftformer.
- b. Charlsonvægtene fra M1 og M2 lægges sammen for hver patient. Summen er lig med Charlsons indeks.

Design

Det sidste skridt, M3, går ud på at flette M1 og M2 sammen og summere vægterne til et samlet Charlson Index. Resultatet bliver en tabel med kolonner til CPR-nummer, inklusionsdato og Charlson Index samt en kolonne til hver Charlsongruppe. Sidstnævnte bruges til at opbevare vægterne for de respektive grupper. Vi skaber også en ny talkolonne, som hedder NonCharlson. Den skal bruges til at markere de patienter, som ikke har et Charlsonindeks. Resultatet bliver en tabel, som har én række pr. patient. Metoden er:

- C8. Opret en tabel med de nævnte kolonner. Indsæt cpr-numre og tilhørende inklusionsdatoer, så der dannes en række pr. patient. Opdater tabellen med Charlsonvægtene fra M1.
- C9. Opdater tabellen, så Charlsongrupperne for solid og metastaserende kræft får vægten 0 (nul).
- C10. Opdater tabellen med Charlsonvægtene fra M2. Opdater kun patienter og Charlsongrupper (solid og metastaserende kræft), som er med i resultatet fra M2.
- C11. Læg Charlsonvægtene sammen for hver patient. Dette giver patienternes respektive Charlsonindeks. Evt. patienter, som er optaget i LPR, men har Charlsonindeks lig med 0, tildeles et 1-tal i kolonnen NonCharlson.

SQL-koden til C8 fremgår af bilag 5. Koden indeholder bla. en SQL-sætning til hver Charlsongruppe, dvs. i alt 19 sætninger. De følger samme mønster, så derfor er kun 3 af dem vist. SQL-koden bruger forkortelserne fra bilag 1 til at identificere Charlsongrupperne, og koden bruger forkortelserne i samme rækkefølge som bilaget: AMI, HF, ... AIDS.

Koden til C9 – C10 (bilag 6) opretter en hjælpetabel, som hedder DiagnosesCollapsed. Vi bruger dog kun 3 af kolonnerne i denne sammenhæng: Cpr, CA_SOLID og CA_METAST. Koden opretter en række i DiagnosesCollapsed for hver patient, som har en Charlsonvægt for solid eller metastaserende kræft. Felterne CA_SOLID og CA_METAST bliver tildelt henholdsvis 2 og 6 afhængigt af, om patienten har diagnoser i en eller begge Charlsongrupper. Tallene 2 og 6 er Charlsonvægtene for de to grupper, se også bilag 1.

Det sidste punkt, C11, er implementeret som vist på bilag 7. Charlsonvægtene lægges sammen til en sum, som udgør Charlson Index. Efter denne beregning ser man, at nogle patienter har Charlson Index lig med nul, fordi patienterne kun har diagnoser, som er relateret til lungekræft. Disse patienter tildeles et 1-tal i kolonnen NonCharlson. Denne markering betyder, at patienten ikke har diagnoser, som er relevante i denne beregning af Charlson Index.

Den sidste blok SQL-kode (bilag 7) finder patienter, som har en diagnosekode, men ikke er med i resultatet indtil nu. Disse patienter har ingen diagnoser, som indgår i *dette* Charlson Index – dvs. med lungekræft som indekssygdom. Patienterne har naturligvis en lungekræftdiagnose, og de har måske nogle mindre vigtige diagnoser, fx springfinger. Men patienterne har ingen Charlson-relevante diagnosekoder i vores sammenhæng. Vi indsætter patienterne i resultatet og sætter NonCharlson lig med 1 som tegn på, at patienterne ikke har en diagnosekode af relevans for det aktuelle Charlson Index.

Resultat

Efter denne beregning fremkommer resultatet i tabellen CharlsonTransposed. Den har én række pr. patient, og hver række indeholder CPR-nummer, inklusionsdato, Charlsonindeks, alle 19 Charlsonvægte samt tallet NonCharlson. Tallet NonCharlson er medtaget for at kunne belyse, hvorvidt der foreligger diagnosekoder for en patient uden at en eneste af dem falder i en Charlsongruppe; i dette tilfælde er værdien 1 for NonCharlson. I modsat fald er NonCharlson lig med 0 (talværdien nul).

Bilag 1. Gruppering af diagnosekoder (ICD8 og ICD10) i Charlson Index

Charlson Index diagnosegruppe	Vægt	Forkortelse	ICD-8	ICD-10
Akut myokardie infarkt	1	AMI	410	I21-I23
Hjerteinsufficiens	1	HF	427.09-427.11; 427.19; 428.99; 782.49	I50; I11.0; I13.0; I13.2
Karsygdomme	1	VASC	440-445	I70-74; I77
Cerebrovaskulære sygdomme	1	CVD	430-438	I60-I69; G45; G46
Demens	1	DEMENS	290.09-290.19; 293.09	F00-F03; F05.1; G30
Kroniske lungesygdomme	1	COL	490-493; 515-518	J40-J47; J60-J67; J68.4; J70.1; J70.3; J84.1; J92.0; J96.1; J98.2; J98.3
Bindevævssygdomme	1	CONNECT	712; 716; 734; 446; 135.99	M05; M06; M08; M09; M30- M36; D86
Ulcussygdomme	1	ULCUS	530.91; 530.98; 531- 534	K22.1; K25-K28
Milde leversygdomme	1	LIVER_MILD	571; 573.01; 573.04	B18; K70.0-K70.3; K70.9; K71; K73; K74; K76.0
Diabetes, ukompliceret	1	DM_UNCOMP	249.00; 249.06; 249.07; 249.09; 250.00; 250.06; 250.07; 250.09	E10.0; E10.1; E10.9; E11.0; E11.1; E11.9; E14.0; E14.1; E14.9
Hemiplegi	2	HEMIPLEG	344	G81; G82
Nyresygdomme	2	KIDNEY	403; 404; 580-583; 584; 590.09; 593.19; 753.10- 753.19; 792	I12; I13; N00-N05; N07; N11; N14; N17-N19; Q61
Diabetes med komplikationer	2	DM_COMP	249.01-249.05; 249.08; 250.01-250.05; 250.08	E10.2-E10.8; E11.2-E11.8; E14.2-E14.8
Solide kræftformer	2	CA_SOLID	140-194	C00-C75
Leukæmi	2	LEUC	204-207	C91-C95
Lymfomer	2	LYMPH	200-203; 275.59	C81-C85; C88; C90; C96
Moderate til svære lever- sygdomme	3	LIVER_SEVERE	070.00; 070.02; 070.04; 070.06; 070.08; 573.00; 456.00-456.09	B15.0; B16.0; B16.2; B19.0; K70.4; K72; K76.6; I85
Metastaserende cancer	6	CA_METAST	195-199	C76-C80
AIDS	6	AIDS	079.83	B21-B24

Bilag 2. SAS-program til beregning af Charlson Index

```

libname lpr 'U:\CANULI\LPR';
libname opdt 'U:\CANULI\LPR\opdatering2009';
libname dlcr 'U:\CANULI II\Lung';
libname dlcr_c 'U:\CANULI\LPR\opdatering2009\DLCR_Charlson';

options mprint;

/*Cpr-numre fra DLCR, første gang de optræder*/

proc sort data=dlcr.raa_lunge;
  by cpr dxdate;
run;

data dlcr_c.cpr_lunge (keep=cpr dxdate);
  set dlcr.raa_lunge;
  by cpr dxdate;
  if first.cpr;
  if year(foddto) < 1920 then delete;
run;

/*Kobler med oplysninger fra LPR*/

%macro old;

proc delete data=dlcr_c.old_adm;
run;

%do x=1 %to 3;

proc sql;
  create table adm as
  select a.cpr, dxdate, k_recnum as v_recnum, d_inddto, d_uddto
  from dlcr_c.cpr_lunge a, lpr.adm&x b
  where a.cpr=b.v_cpr and a.dxdate > b.d_uddto;
quit;

proc append data=adm base=dlcr_c.old_adm force;
run;

%end;
%mend;

%old;

%macro old_diag;

proc delete data=dlcr_c.old_diag;
run;

%do x=1 %to 6;

proc sql;
  create table adm as
  select b.*
  from dlcr_c.old_adm a, lpr.diag&x b
  where a.v_recnum=b.v_recnum;
quit;

proc append data=adm base=dlcr_c.old_diag force;
run;

%end;
%mend;

```



```

%old_diag;

%macro new;

proc delete data=dlcr_c.new_adm;
run;

%do x=1 %to 2;

proc sql;
create table adm as
select a.cpr, dxdate, v_recnum, d_inddto, d_uddto
from dlcr_c.cpr_lunge a, opdt.gruppe&x._adm b
where a.cpr=b.v_cpr and a.dxdate > b.d_uddto;
quit;

proc append data=adm base=dlcr_c.new_adm force;
run;

%end;
%mend;

%new;

%macro new_diag;

proc delete data=dlcr_c.new_diag;
run;

%do x=1 %to 2;

proc sql;
create table adm as
select b.*
from dlcr_c.new_adm a, opdt.gruppe&x._diag b
where a.v_recnum=b.v_recnum;
quit;

proc append data=adm base=dlcr_c.new_diag force;
run;

%end;
%mend;

%new_diag;

data dlcr_c.adm7793 dlcr_c.adm9408;
set dlcr_c.old_adm dlcr_c.new_adm;
if d_uddto < mdy(1,1,1994) then output dlcr_c.adm7793;
else output dlcr_c.adm9408;
run;

data diag;
set dlcr_c.old_diag (drop=c_diagmod c_tildiag) dlcr_c.new_diag (drop=c_diagmod
c_tildiag);
if c_diagtype in ('H' '+') then delete;
run;

proc sql;
create table dlcr_c.diag7793 as
select cpr, b.v_recnum, c_diag, c_diagtype
from dlcr_c.adm7793 a, diag b
where a.v_recnum=b.v_recnum;
quit;

proc sql;
create table dlcr_c.diag9408 as
select cpr, b.v_recnum, c_diag, c_diagtype

```

```

from dlcrc_c.adm9408 a, diag b
where a.v_recnum=b.v_recnum;
quit;

```

```

data char7793;
set dlcrc_c.diag7793;
dkd=c_diag;
IF '41000'<=dkd<='41099' THEN dsd1=1;
ELSE dsd1=0;
IF dkd in('42709','42710','42711','42719','42899','78249') THEN dsd2=1;
ELSE dsd2=0;
IF '44000'<=dkd<='44599' THEN dsd3=1;
ELSE dsd3=0;
IF '43000'<=dkd<='43899' THEN dsd4=1;
ELSE dsd4=0;
IF '29009'<=dkd<='29019' OR dkd='29309' THEN dsd5=1;
ELSE dsd5=0 ;
IF '49000'<=dkd<='49399' OR '51500'<=dkd<='51899' THEN dsd6=1;
ELSE dsd6=0 ;
IF '71200'<=dkd<='71299' OR '71600'<=dkd<='71699'
OR '73400'<=dkd<='73499' OR '44600'<=dkd<='44699'
OR dkd='13599' THEN dsd7=1;
ELSE dsd7=0;
IF dkd in('53091','53098') OR '53100'<=dkd<='53499' THEN dsd8=1;
ELSE dsd8=0;
IF '57100'<=dkd<='57199' OR dkd in('57301','57304') THEN dsd9=1;
ELSE dsd9=0;
IF dkd in('24900','24906','24907','24909','25000',
'25006','25007','25009') THEN dsd10=1;
ELSE dsd10=0;
IF '34400'<=dkd<='34499' THEN dsd11=1;
ELSE dsd11=0;
IF '40300'<=dkd<='40499' OR '58000'<=dkd<='58499'
OR dkd in('59009','59319')
OR '75310'<=dkd<='75319' OR '79200'<=dkd<='79299' THEN dsd12=1;
ELSE dsd12=0;
IF '24901'<=dkd<='24905' OR dkd in('24908','25008')
OR '25001'<=dkd<='25005' THEN dsd13=1;
ELSE dsd13=0;
IF '14000'<=dkd<='17299' OR '18000'<=dkd<='19499' THEN dsd14=1;
ELSE dsd14=0;
IF '17400'<=dkd<='17499' OR dkd='174' THEN mc=1;
ELSE mc=0;
IF mc=1 AND dsd14=1 THEN dsd14=0 ; * c.m. eks ;
IF '20400'<=dkd<='20799' THEN dsd15=1;
ELSE dsd15=0;
IF '20000'<=dkd<='20399' OR dkd='27559' THEN dsd16=1;
ELSE dsd16=0;
IF dkd in('07000','07002','07004','07006','07008','57300')
OR '45600'<=dkd<='45609' THEN dsd17=1;
ELSE dsd17=0;
IF '19500'<=dkd<='19999' THEN dsd18=1;
ELSE dsd18=0;
IF dkd='07983' THEN dsd19=1;
ELSE dsd19=0;
RUN;

```

```

data char9408;
set dlcrc_c.diag9408;
dkd=c_diag;
IF 'DI210'<=dkd<='DI239' THEN dsd1=1 ; ELSE dsd1=0 ;

IF 'DI500'<=dkd<='DI509' OR
dkd in('DI110','DI130','DI132') THEN dsd2=1 ; ELSE dsd2=0 ;

IF 'DI700'<=dkd<='DI749' OR 'DI770'<=dkd<='DI779' THEN dsd3=1 ; ELSE dsd3=0 ;

```

```

IF 'DI600'<=dkd<='DI699' OR 'DG450'<=dkd<='DG469' THEN dsd4=1 ; ELSE dsd4=0 ;

IF 'DF000'<=dkd<='DF039' OR dkd='DF051' OR
'DG300'<=dkd<='DG309' THEN dsd5=1 ; ELSE dsd5=0 ;

IF 'DJ400'<=dkd<='DJ479' OR 'DJ600'<=dkd<='DJ679'
OR dkd in('DJ684','DJ701','DJ703','DJ841',
'DJ920','DJ961','DJ982','DJ983') THEN dsd6=1 ; ELSE dsd6=0 ;

IF 'DM050'<=dkd<='DM069'
OR 'DM080'<=dkd<='DM099' OR 'DM300'<=dkd<='DM369'
OR 'DD860'<=dkd<='DD869' THEN dsd7=1 ; ELSE dsd7=0 ;

IF dkd='DK221' OR 'DK250'<=dkd<='DK289' THEN dsd8=1 ; ELSE dsd8=0 ;

IF 'DB180'<=dkd<='DB189' OR 'DK700'<=dkd<='DK703'
OR dkd in('DK709','DK760')
OR 'DK710'<=dkd<='DK719' OR 'DK730'<=dkd<='DK749' THEN dsd9=1 ; ELSE dsd9=0 ;

IF dkd in('DE100','DE101','DE109','DE110','DE111','DE119') THEN dsd10=1 ; ELSE dsd10=0
;

IF 'DG810'<=dkd<='DG829' THEN dsd11=1 ; ELSE dsd11=0 ;

IF 'DI120'<=dkd<='DI129' OR dkd='DI131'
OR 'DI133'<=dkd<='DI139' OR 'DN000'<=dkd<='DN059'
OR 'DN070'<=dkd<='DN079' OR 'DN110'<=dkd<='DN119'
OR 'DN140'<=dkd<='DN149' OR 'DN170'<=dkd<='DN199'
OR 'DQ610'<=dkd<='DQ619' THEN dsd12=1 ; ELSE dsd12=0 ;

IF 'DE102'<=dkd<='DE108' OR 'DE112'<=dkd<='DE118' THEN dsd13=1 ; ELSE dsd13=0 ;

IF 'DC000'<=dkd<='DC439' OR 'DC450'<=dkd<='DC759' THEN dsd14=1 ; ELSE dsd14=0 ;

IF 'DC500'<=dkd<='DC509' OR dkd in('174','DC50') THEN mc=1 ; ELSE mc=0 ;
IF mc=1 AND dsd14=1 THEN dsd14=0 ; * c.m. eks ;

IF 'DC910'<=dkd<='DC959' THEN dsd15=1 ; ELSE dsd15=0 ;

IF 'DC810'<=dkd<='DC859' OR 'DC880'<=dkd<='DC889'
OR 'DC900'<=dkd<='DC909' OR 'DC960'<=dkd<='DC969' THEN dsd16=1 ; ELSE dsd16=0 ;

IF dkd in('DB150','DB160','DB162','DB190','DK704','DK766')
OR 'DK720'<=dkd<='DK729' OR 'DI850'<=dkd<='DI859' THEN dsd17=1 ; ELSE dsd17=0 ;

IF 'DC760'<=dkd<='DC809' THEN dsd18=1 ; ELSE dsd18=0 ;

IF 'DB210'<=dkd<='DB249' THEN dsd19=1 ; ELSE dsd19=0 ;
run;

data dlcr_c.char;
set char7793 char9408;
run;

PROC SORT data=dlcr_c.char;
BY cpr;
run;

PROC MEANS data=dlcr_c.char NOPRINT ;
VAR dsd1-dsd19 ;
BY cpr ;
OUTPUT OUT=ss sum=sum1-sum19 ;
RUN;

DATA dlcr_c.charlson7708;
SET ss (keep=cpr sum1-sum19) ;
IF sum1>1 THEN sum1=1 ; IF sum2>1 THEN sum2=1 ;

```

```
IF sum3>1 THEN sum3=1 ; IF sum4>1 THEN sum4=1 ;
IF sum5>1 THEN sum5=1 ; IF sum6>1 THEN sum6=1 ;
IF sum7>1 THEN sum7=1 ; IF sum8>1 THEN sum8=1 ;
IF sum9>1 THEN sum9=1 ; IF sum10>1 THEN sum10=1 ;
IF sum11=>1 THEN sum11=2 ; IF sum12=>1 THEN sum12=2 ;
IF sum13=>1 THEN sum13=2 ;
IF sum14=>1 THEN sum14=2 ; IF sum15=>1 THEN sum15=2 ;
IF sum16=>1 THEN sum16=2 ; IF sum17=>1 THEN sum17=3 ;
IF sum18=>1 THEN sum18=6 ; IF sum19=>1 THEN sum19=6 ;
suma=sum(of sum1-sum19) ;
IF suma<0 THEN suma=0 ;
rename suma=char7790;
run;
```

Bilag 3. SQL algoritme til beregning af Charlson Index, trin C1 – C4

```

CREATE TABLE CharlsonD0 (
  Cpr VARCHAR(15),
  MinDiagnosedato DATETIME,
  AntalDiagnoser INT,
  CharlsonKey VARCHAR(20)
);

INSERT INTO CharlsonD0 (Cpr, MinDiagnosedato, AntalDiagnoser, CharlsonKey)
SELECT L.cpr AS Cpr, MIN(A.d_inddto) AS MinDiagnosedato,
       COUNT (DISTINCT D.c_diag) AS AntalDiagnoser, I.CharlsonKey
FROM ((dbo.cpr_lunge L INNER JOIN dbo.Adm A ON (L.cpr = A.cpr))
      INNER JOIN dbo.Diag D ON (A.v_recnum = D.v_recnum))
      INNER JOIN Opslag.dbo.IcdCharlson I ON (D.c_diag LIKE I.IcdKey)
WHERE (D.c_diag NOT LIKE 'DC34%') AND (D.c_diag NOT LIKE '1621%')
      AND (A.d_inddto <= L.DxDATE)
GROUP BY L.cpr, I.CharlsonKey;

SELECT A.Cpr, A.MinDiagnosedato AS MinDiagnoseDate,
       A.AntalDiagnoser AS DiagnoseCount,
       O.CharlsonKey AS CharlsonGr, O.Weight as Weight,
       L.DxDATE AS Inclusion
INTO CharlsonDlcr
FROM (dbo.CharlsonD0 A INNER JOIN Opslag.dbo.Charlson O
      ON (A.CharlsonKey = O.CharlsonKey))
      INNER JOIN dbo.cpr_lunge L ON (A.cpr = L.cpr);

```

Bilag 4. SQL algoritme til beregning af Charlson Index, trin C5 – C7

```

SELECT DISTINCT A.cpr, A.d_inddto AS LCDate
INTO #LCDiagAll
FROM Adm A INNER JOIN Diag D ON (A.v_recnum = D.v_recnum)
WHERE (D.c_diag LIKE 'DC34%') OR (D.c_diag LIKE '1621%')
      OR (D.c_diag LIKE 'DC780%') OR (D.c_diag LIKE '1970%')
ORDER BY cpr;

-- Diagnosedatoer for solid eller metastaserende kræft,
-- men ikke lungekræft.
SELECT IDENTITY(INT, 1, 1) AS Serial, A.cpr,
       CAST(NULL AS DATETIME) AS Inclusion,
       MIN(A.d_inddto) AS MinDiagnoseDate, D.c_diag AS DiagnoseCode,
       MIN(I.Diagnosis) AS Diagnose,
       CAST(NULL AS INT) AS CA_METAST, CAST(NULL AS INT) AS CA_SOLID
INTO Diagnoses
FROM (Adm A INNER JOIN Diag D ON (A.v_recnum = D.v_recnum))
LEFT JOIN Opslag.dbo.IcdCharlson I ON (D.c_diag = I.IcdKey)
WHERE (I.CharlsonKey = 'CA_METAST' OR I.CharlsonKey = 'CA_SOLID')
      AND (D.c_diag NOT LIKE 'DC34%') AND (D.c_diag NOT LIKE '1621%')
      AND (D.c_diag NOT LIKE 'DC780%') AND (D.c_diag NOT LIKE '1970%')
      AND (A.d_inddto <= CONVERT(DATETIME, '2007-12-31', 120))
      AND A.d_inddto NOT IN (SELECT LCDate FROM #LCDiagAll DA
                           WHERE DA.cpr = A.cpr)

GROUP BY A.cpr, D.c_diag
ORDER BY cpr, MinDiagnoseDate;

CREATE INDEX Diagnoses_cpr ON Diagnoses (cpr);

-- Tilføj inklusionsdato:
UPDATE D SET D.Inclusion = L.DxDATE
FROM Diagnoses D LEFT JOIN cpr_lunge L ON (D.cpr = L.cpr);

-- Tilføj Charlsonvægt til diagnoser, som tilhører CA_METAST:
UPDATE D SET D.CA_METAST = I.Weight
FROM Diagnoses D INNER JOIN Opslag.dbo.IcdCharlson I ON (D.DiagnoseCode = I.IcdKey)
WHERE I.CharlsonKey = 'CA_METAST';

-- Tilføj Charlsonvægt til diagnoser, som tilhører CA_SOLID:
UPDATE D SET D.CA_SOLID = I.Weight
FROM Diagnoses D INNER JOIN Opslag.dbo.IcdCharlson I ON (D.DiagnoseCode = I.IcdKey)
WHERE I.CharlsonKey = 'CA_SOLID';

-- Slet rækker, hvor lprdato er højst 150 dage før dlcrdato.
-- Slet også rækker, hvor lprdato er senere end dlcrdato.
DELETE FROM Diagnoses
WHERE DATEDIFF(day, MinDiagnoseDate, Inclusion) <= 150;

```

Bilag 5. SQL algoritme til beregning af Charlson Index, trin C8

```

CREATE TABLE CharlsonTransposed (
  Cpr VARCHAR(15),
  Inklusion DATETIME,
  CharlsonIndex INT DEFAULT 0,
  AMI INT DEFAULT 0,
  HF INT DEFAULT 0,
  VASC INT DEFAULT 0,
  CVD INT DEFAULT 0,
  DEMENS INT DEFAULT 0,
  COL INT DEFAULT 0,
  CONNECT INT DEFAULT 0,
  ULCUS INT DEFAULT 0,
  LIVER_MILD INT DEFAULT 0,
  DM_UNCOMP INT DEFAULT 0,
  HEMIPLEG INT DEFAULT 0,
  KIDNEY INT DEFAULT 0,
  DM_COMP INT DEFAULT 0,
  CA_SOLID INT DEFAULT 0,
  LEUC INT DEFAULT 0,
  LYMPH INT DEFAULT 0,
  LIVER_SEVERE INT DEFAULT 0,
  CA_METAST INT DEFAULT 0,
  AIDS INT DEFAULT 0,
  NonCharlson INT DEFAULT 0
);

CREATE INDEX CharlsonTransposed_cpr ON CharlsonTransposed (cpr);

INSERT INTO CharlsonTransposed (Cpr, CharlsonIndex)
SELECT Cpr, SUM(Weight)
FROM CharlsonDlcr
GROUP BY cpr;

UPDATE T SET Inklusion = L.DxDATE
FROM CharlsonTransposed T INNER JOIN cpr_lunge L ON (T.cpr = L.cpr);

UPDATE T SET AMI = D.Weight
FROM CharlsonTransposed T INNER JOIN CharlsonDlcr D ON (T.cpr = D.cpr)
WHERE D.Charlsongr = 'AMI';

UPDATE T SET HF = D.Weight
FROM CharlsonTransposed T INNER JOIN CharlsonDlcr D ON (T.cpr = D.cpr)
WHERE D.Charlsongr = 'HF';
.
.
.

UPDATE T SET AIDS = D.Weight
FROM CharlsonTransposed T INNER JOIN CharlsonDlcr D ON (T.cpr = D.cpr)
WHERE D.Charlsongr = 'AIDS';

```

Bilag 6. SQL algoritme til beregning af Charlson Index, trin C9 – C10

```

-- Nulstil felterne CA_SOLID og CA_METAST i CharlsonTransposed:
UPDATE CharlsonTransposed
SET CA_SOLID = 0, CA_METAST = 0;

-- DiagnosesCollapsed er en hjælpetabel til Charlsonvægte for kræftdiagnoser:
CREATE TABLE DiagnosesCollapsed (
  Cpr VARCHAR(15),
  Inclusion DATETIME,
  CA_METAST INT DEFAULT NULL,
  CA_SOLID INT DEFAULT NULL,

  -- Resten af kolonnerne bruges ikke i denne sammenhæng:
  MinDiagnoseDate1 DATETIME,
  DiagnoseCode1 VARCHAR(6),
  Diagnose1 VARCHAR(100),

  MinDiagnoseDate2 DATETIME,
  DiagnoseCode2 VARCHAR(6),
  Diagnose2 VARCHAR(100),
  .
  .
  .

  MinDiagnoseDate10 DATETIME,
  DiagnoseCode10 VARCHAR(6),
  Diagnose10 VARCHAR(100)
);

CREATE INDEX DiagnosesCollapsed_cpr ON DiagnosesCollapsed (cpr);

-- Indsæt cpr-numre:
INSERT INTO DiagnosesCollapsed (Cpr)
SELECT D.cpr
FROM Diagnoses D
GROUP BY D.Cpr;

-- Opdater Charsongrupperne CA_METAST og CA_SOLID
UPDATE DC SET DC.CA_METAST = 6
FROM DiagnosesCollapsed DC INNER JOIN Diagnoses D ON (DC.cpr = D.cpr)
WHERE D.CA_METAST IS NOT NULL;

UPDATE DC SET DC.CA_SOLID = 2
FROM DiagnosesCollapsed DC INNER JOIN Diagnoses D ON (DC.cpr = D.cpr)
WHERE D.CA_SOLID IS NOT NULL;

-- Overfør CA_SOLID og CA_METAST til CharlsonTransposed:
UPDATE CT
SET CT.CA_SOLID = CASE WHEN DC.CA_SOLID IS NULL THEN 0 ELSE DC.CA_SOLID END,
  CT.CA_METAST = CASE WHEN DC.CA_METAST IS NULL THEN 0 ELSE DC.CA_METAST END
FROM CharlsonTransposed CT INNER JOIN DiagnosesCollapsed DC ON (CT.Cpr = DC.Cpr);

```


Bilag 7. SQL algoritme til beregning af Charlson Index, trin C11

```

-- Genberegner CharlsonTransposed.CharlsonIndex
UPDATE CharlsonTransposed
SET CharlsonIndex = AMI + HF + VASC + CVD + DEMENS + COL
    + CONNECT + ULCUS + LIVER_MILD + DM_UNCOMP + HEMIPLEG
    + KIDNEY + DM_COMP + CA_SOLID + LEUC + LYMPH + LIVER_SEVERE
    + CA_METAST + AIDS;

/*
 * Nogle rækker har CharlsonIndex = 0, fordi vi har nulstillet flere
 * CA_SOLID og CA_METAST, end vi har indsat fra DiagnosesCollapsed.
 * Disse rækker opdateres med NonCharlson = 1.
 */
UPDATE CharlsonTransposed
SET NonCharlson = 1
WHERE CharlsonIndex = 0;

-- Find alle patienter, som har en diagnose, ligemeget hvilken:
SELECT DISTINCT D.cpr
INTO #DiagnosePatients
FROM Diag D INNER JOIN Opslag.dbo.Icd I ON (D.C_DIAG = I.IcdKey)
WHERE D.C_DIAG IS NOT NULL AND D.C_DIAG != '';

-- Slet dem, der allerede er med i resultatet:
DELETE DP
FROM #DiagnosePatients DP INNER JOIN CharlsonTransposed CT ON (DP.cpr = CT.cpr);

-- Indsæt non-Charlsonpatienterne i CharlsonTransposed:
INSERT INTO CharlsonTransposed (Cpr, Inklusion, NonCharlson)
SELECT DP.cpr, L.DxDATE, 1
FROM #DiagnosePatients DP INNER JOIN cpr_lunge L ON (DP.cpr = L.cpr);

```