

# **Forskning i kliniske kræftdatabaser**

Region Hovedstaden  
Forskningscenter for Forebyggelse og Sundhed

Titel           Forskning i kliniske kræftdatabaser

Forfattere    Birgitte Lidegaard Frederiksen  
                  Merete Osler  
                  Torben Jørgensen

Udgiver       Region Hovedstaden  
                  Koncern Plan & Udvikling  
                  Forskningscenter for Forebyggelse og Sundhed  
                  Glostrup Hospital, afsnit 84/85  
                  Nordre Ringvej 57  
                  2600 Glostrup  
                  Telefon: 38 63 32 60  
                  Telefax: 38 63 39 77  
                  [www.fcfs.dk](http://www.fcfs.dk)

Forskningscenter for Forebyggelse og Sundhed, Region Hovedstaden, 2010

ISBN 87-987414-8-9

Denne rapport er udarbejdet af Forskningscenter for Forebyggelse og Sundhed, Region Hovedstaden, som et led i Indenrigs- og Sundhedsministeriets initiativ til styrkelse af infrastrukturen for klinisk kræftforskning i Danmark. Forskningscenter for Forebyggelse og Sundhed er en del af Kompetencecenter Øst, der, i samarbejde med Enheden for Klinisk Kvalitet, varetager epidemiologisk, statistisk og IT-mæssig bistand til en stor del af de Landsdækkende Kliniske Kvalitetsdatabaser i Danmark, herunder flere kræftdatabaser.

## Indholdsfortegnelse

1	Indledning	4
2	Forskning i de kliniske kræftdatabaser	5
3	En fælles analysemodel	6
3.1	Hvilke faktorer påvirker patientforløbet?	6
3.2	Model for analyser i kliniske kræftdatabaser	8
4	Afgrænsning af forskningsspørgsmålet	10
4.1	Årsagsdiagrammer	10
4.2	Confounding og mediering	10
4.4	Interaktion	12
5	Analyseprocessen	13
5.1	Forberedelse af data	13
5.2	Deskriptive analyser	13
5.3	Regressionsanalyser	13
6	Håndtering af nedsat datakvalitet	15
6.1	Dækningsgrad (selektionsbias)	15
6.2	Datakomplethedegrad	15
6.3	Validitet (informationsbias)	16
7	Kliniske udfald	17
7.1	Stadie	17
7.2	Diagnostiske undersøgelser	17
7.3	Ventetid	17
7.4	Behandling	17
7.5	Komplikationer	17
7.6	Overlevelse	17
7.7	Recidiv	18
7.8	Tilbagevenden til arbejdsmarkedet	18
7.9	Psykosociale følger	18
8	Data fra centrale registre	19
8.1	Comorbiditet	19
8.2	Socioøkonomisk position	19
8.3	Recidiv	20
9	Praktisk om opkobling til Danmarks Statistik	20
10	Hvad kan kompetencecentrene hjælpe med?	20
11	Udvalgte studier baseret på kliniske kræftdatabaser	21
12	Nyttige hjemmesider	22
	Appendix 1 - årsagsdiagrammer	23
	Appendix 2 – Charlsons comorbiditetsindex	25
	Referencer	26

## 1 Indledning

De danske kræftdatabaser udgør et unikt forskningsmateriale, både nationalt og internationalt set. Danmark var blandt de første nationer der etablerede kliniske databaser til monitorering af cancerpatienternes behandlingsforløb med brystkræftdatabasen, der blev etableret i 1977, som den første (1). Kræftdatabaserne blev dannet som en konsekvens af nationale sammenligninger af mortalitetsrater, der viste at danske cancerpatienter havde en dårligere overlevelse sammenlignet med de øvrige nordiske lande (2;3). Grundet den tidlige opstart har mange danske kræftdatabaser observationer fra en lang årrække, inkluderende et stort antal patienter. Det er unikt at data indsamles på nationalt plan, og ikke, som i de fleste udenlandske databaser, er begrænset til bestemte geografiske regioner, hospitaler eller sygesikringsordninger, som alle rummer risiko for selektion. Det danske CPR-nummer og muligheden for opkobling til talrige centrale registre giver endvidere enestående muligheder for opfølgning af patienterne og indhentning af supplerende oplysninger om fx social status, forbrug af receptpligtig medicin eller tidligere kontakt til sygehusvæsenet.

Den store og detaljerede mængde af oplysninger gør det til en forpligtelse, lige så vel som en mulighed, at udnytte disse data til dybdegående forskning i årsagerne til gode og dårlige patientforløb.

De mange fælles træk der er mellem de forskelle cancersygdomme med hensyn til diagnostik, stadietklassifikation og behandling bevirker at kræftdatabaserne til en vis grad rummer de samme typer af data. Dette gør det oplagt at systematisere tilgangen til anvendelse af kræftdatabaserne til forskning.

Hovedformålet med denne rapport er at opstille en model for udnyttelse af kræftdatabaserne til forskningsformål, der kan inspirere klinikere og forskere til initiering af nye forskningsprojekter. Et andet væsentligt formål er at dele ud af de epidemiologiske og praktiske erfaringer forfatterne har med forskning i de kliniske kræftdatabaser. Herunder hvilke oplysninger der bør inkluderes i de statistiske analyser for at opnå valide resultater, hvordan man bør forholde sig til dårlig dækningsgrad i databasen, hvordan man kan tage højde for manglende data, hvordan man får oprettet en kobling til centrale registre via Danmarks Statistik, hvordan man danner et comorbiditetsindex ud fra registerdata mm.

De ældste af databaserne har allerede et større antal publikationer bag sig, og det er relevant at trække på erfaringer opnået i forbindelse hermed, når nye projekter skal initieres. Det er vort håb, at rapporten kan inspirere til initiering og gennemførelse af mange nye forskningsprojekter med udgangspunkt i kliniske kræftdatabaser, samt at rapporten kan bidrage til at højne kvaliteten af sådanne projekter.

Kompetencecentrene i Danmark besidder epidemiologisk og statistisk ekspertise, og vil kunne råde og vejlede mhp initiering og udførelse af nye forskningsprojekter.

## 2 Forskning i de kliniske kræftdatabaser

Det overordnede mål for sundhedsvæsenets indsats overfor kræftpatienter er at øge patienternes overlevelse og livskvalitet på kort og på lang sigt. Hvis en sådan indsats skal være evidensbaseret kræver det forskning i, hvad der påvirker patientforløbet. De kliniske databaser rummer information om patientforløb, som kan bruges til at analysere årsagerne til gode og dårlige patientforløb. Der kan benyttes både en deskriptiv og en analytisk tilgang.

Med deskriptive analyser ses på udvikling i behandlingsresultater over tid og sted samt på fordelinger af prognostiske faktorer. Deskriptive analyser er velegnede til at danne sig et overblik over sundhedsvæsenets indsats og effekten af nye behandlingstiltag, som fx implementeringen af kræftpakkeforløb. De deskriptive analyser tager dog ikke højde for forskelle i patientgrundlag.

En udforskning af hvilke faktorer, der har betydning for patientforløbet, kræver en analytisk tilgang, hvor der tages højde for forskelle i patientunderlaget. Typisk sammenlignes udfald mellem forskellige grupper af patienter eller forskellige afdelinger ved hjælp af avancerede statistiske analyser (e.g. multiple regressionsanalyser). Resultater fra sådanne analyser kan give vigtig information til brug for den overordnede sundhedsplanlægning. Mere dybdegående analyse af patientspecifikke determinanter er også mulig i mange databaser, og vil, eventuelt i kombination med data fra centrale registre, kunne bruges til at identificere patientgrupper, der skal have særlige hensyn under behandlingen. Resultaterne fra sådanne observationelle studier vil dog ofte skulle afprøves i interventionsstudier, for at sikre, at der er tale om en reel effekt af de forskellige initiativer. Epidemiologisk forskning i de kliniske databaser er således hypotesegenererende.

Forskning i de kliniske kræftdatabaser kan endvidere bruges til at understøtte resultater opnået i kliniske randomiserede undersøgelser (RCT), som ofte vil være udført i stærkt selekterede patientpopulationer og under specielle omstændigheder, som ikke svarer til den daglige behandling af den generelle patientpopulation. Data fra de kliniske databaser repræsenterer således 'den virkelige verden', som kan benyttes til at vurdere, hvorvidt resultater opnået i RCT kan implementeres i den daglige klinik, og hvorvidt der opnås samme effekt.

Endelig kan data benyttes til at udvikle prædiktionsmodeller. Sådanne modeller kan bruges til at vurdere individuelle patienters absolutte risiko for en bestemt hændelse (e.g. komplikationer, død). Forskningscenter for Forebyggelse og Sundhed arbejder i øjeblikket (2010) på at udvikle en prædiktionsmodel for kolorektalcancerpatienter, der, på baggrund af alder, køn, operationstype, ASA score og livsstil, kan estimere den enkelte patients absolutte risiko for at dø inden for 30 dage efter operation. Dette arbejde bygger på data fra Kolorektalcancer databasen. Sådanne modeller vil kunne hjælpe klinikerne i vurderingen af operationsindikation, ligesom de vil kunne bruges til at allokere patienter fra forskellige risikogrupper til differentieret behandling.

### 3 En fælles analysemodel

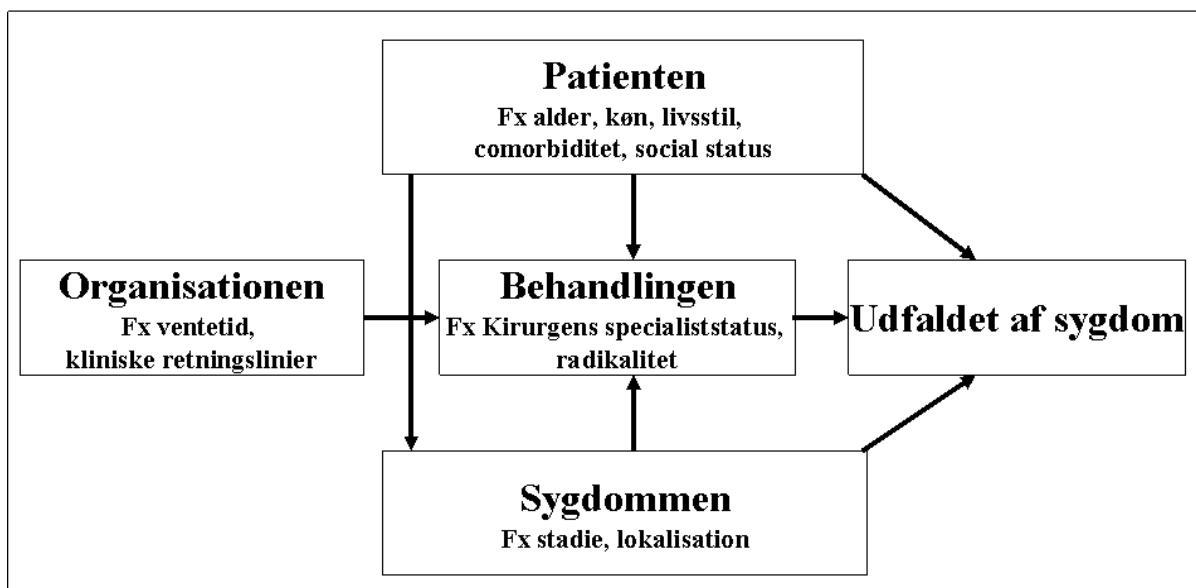
Kræftdatabaserne rummer muligheder for en systematisk tilgangsvinkel til nye forskningsprojekter. Mange af de samme problemstillinger vil være relevante at undersøge for de forskellige kræftsygdomme. En fælles model for analyser i kræftdatabaserne synes derfor gunstig, mhp en systematisk udnyttelse af de tilstedeværende data.

Det er dog vigtigt at pointere, at en systematisk tilgang til forskningsmulighederne på ingen måde nedsætter kravene til anvendelse af velgennemtænkte epidemiologiske og statistiske metoder i udførelsen af de enkelte analyser.

#### 3.1 Hvilke faktorer påvirker patientforløbet?

Figur 1 fremstiller en model over faktorer, der har betydning for udfaldet af en sygdom. Figuren er en videreudvikling af Sackett's model fra 1980'erne (4). Modellen illustrerer at udfaldet kan være et resultat af faktorer med relation til både organisationen omkring behandlingsforløbet, selve behandlingen, patienten selv samt sygdommens sværhedsgrad. Udfaldet kan altså påvirkes af faktorer både indenfor og udenfor klinikerens kontrol. Man risikerer at drage forkerte konklusioner, hvis man ikke tager højde for disse faktors indbyrdes afhængighed.

Figur 1:



De kliniske kræftdatabaser søger at indsamle data med relation til alle disse faktorer. Nedenfor bringes en oversigt med eksempler på variable, der afrapporteres i henholdsvis Kolorektalcancer Databasen og Lymfom Databasen (Tabel 1). Tabellen er suppleret med eksempler på data, der kan indhentes fra centrale registre. Landspatientregisteret, Sygesikringsregisteret, Lægemiddelregisteret og adskillige databaser med socio-demografiske informationer kan give relevante oplysninger om patientfaktorer (e.g. patienternes comorbiditet (se også 8.1), uddannelse, samlivsstatus, erhvervstilknytning og indkomst) og sygdomsudvikling.

Visse typer data er sværere at måle end andre, hvilket afspejler sig i, hvilke variable de kliniske kræftdatabaser har valgt at inkludere. Især sygdomsfaktorer og behandlingsfaktorer er meget grundigt afrapporteret. Sidstnævnte dog fortrinsvist hvad angår diagnostik og selve terapien, mens der ikke er mange oplysninger om kontrol, pleje og rehabilitering. Visse patientfaktorer er sparsomt afrapporteret, fx registrerer kun få kræftdatabaser livsstilsfaktorer, og det er ingen oplysninger om compliance. Organisationsfaktorerne fokuserer fortrinsvis på forskellige mål af ventetider, mens registrering af arbejdstilrettelæggelse, teknisk udstyr og samarbejde er vanskeligere at registrere. Opfølgingsvariable, som fx oplysninger om recidiv, registreres i princippet i visse databaser, men med meget lav datadækningsgrad, hvorfor opfølgning via centrale registre formentlig vil være mere valide (se 8.3).

Table 1: Faktorer af betydning for udfaldet af sygdom.

Faktorer	Generelle eksempler på variable	Eksempler på variable i Kolorrektalcancerdatabasen	Eksempler på variable i Lymfomdatabasen
<b>Patienten</b>	<ul style="list-style-type: none"> <li>▪ Demografiske faktorer (alder, køn, højde, vægt)</li> <li>▪ Comorbiditet</li> <li>▪ Livsstilsfaktorer</li> <li>▪ Compliance</li> </ul>	<ul style="list-style-type: none"> <li>▪ Alder, køn, højde, vægt</li> <li>▪ ASA score</li> <li>▪ Selvrapporterede konkurrerende sygdomme</li> <li>▪ Selvrapporteret tobaks- og alkoholforbrug, fysisk form</li> </ul>	<ul style="list-style-type: none"> <li>▪ Alder, køn, højde, vægt</li> <li>▪ WHO performance status</li> </ul>
<b>Sygdommen</b>	<ul style="list-style-type: none"> <li>▪ Lokalisation</li> <li>▪ Sværhedsgrad</li> </ul>	<ul style="list-style-type: none"> <li>▪ Lokalisation af tumor</li> <li>▪ Stadienddeling</li> <li>▪ Antal fjernede lymfeknuder m cancer</li> </ul>	<ul style="list-style-type: none"> <li>▪ Stadium</li> <li>▪ Forhøjelse af lactatdehydrogenase</li> <li>▪ Ekstranodal lokalisation</li> <li>▪ B-symptomer</li> <li>▪ IPI</li> </ul>
<b>Behandlingen</b>	<ul style="list-style-type: none"> <li>▪ Diagnostik</li> <li>▪ Terapi</li> <li>▪ Kontrol</li> <li>▪ Pleje</li> <li>▪ Rehabilitering</li> </ul>	<ul style="list-style-type: none"> <li>▪ Koloskopi</li> <li>▪ Billeddiagnostik</li> <li>▪ Operationstype</li> <li>▪ Komplikationer</li> <li>▪ Radikalitet</li> </ul>	<ul style="list-style-type: none"> <li>▪ Kemoterapiregime</li> <li>▪ Stråleterapi</li> <li>▪ Immunterapi</li> <li>▪ Toxicitet</li> </ul>
<b>Organisationen</b>	<ul style="list-style-type: none"> <li>▪ Sundhedspersonalets kompetence</li> <li>▪ Ventetid</li> <li>▪ Brug af kliniske retningslinjer</li> <li>▪ Arbejdstilrettelæggelse</li> <li>▪ Teknisk udstyr</li> <li>▪ Samarbejde</li> </ul>	<ul style="list-style-type: none"> <li>▪ Operatøruddannelse</li> <li>▪ Ventetid fra henvisning til forundersøgelse</li> <li>▪ Ventetid til operation</li> <li>▪ Indlæggelsestid</li> </ul>	<ul style="list-style-type: none"> <li>▪ Indlæggelsestid</li> <li>▪ Ventetid fra diagnose til behandlingsstart</li> <li>▪ Forsinkelse i behandlingen</li> <li>▪ Protokolleret behandling</li> </ul>
<b>Data fra centrale registre</b>	<ul style="list-style-type: none"> <li>▪ Opfølgning (fx dødsdato)</li> <li>▪ Recidiv (fx LPR)</li> <li>▪ Socioøkonomisk status (fx IDA-registeret)</li> <li>▪ Comorbiditet (fx LPR og Lægemedeldatabasen)</li> </ul>		

### 3.2 Model for analyser i kliniske kræftdatabaser

Nedenfor vises en oversigt over, hvordan informationerne i de kliniske kræftdatabaser systematisk kan anvendes til at få en indsigt i årsager til gode og dårlige patientforløb ved at analysere de opstillede sammenhænge. Som det vil fremgå tager modellen (punkt a-d) udgangspunkt i de faktorer der indgår i Figur 1.

Modellen er ikke udtømmende, og alle analyserne vil ikke kunne udføres i alle kræftdatabaserne.

#### Model for analyser i kliniske kræftdatabaser

- a. Organisatoriske faktorer (region / hospitalafdeling) betydning for
  - i. Sygdomsstadie
  - ii. Behandling
  - iii. Ventetid, indlæggelsestid, forsinkelse i behandling
  - iv. Sygdomsudfald (komplikationer, 30-dages mortalitet, overlevelse, remission, tilbagefald)
- b. Patientrelaterede (livsstil, comorbiditet, socioøkonomisk status, aldersgrupper etc.) faktorer betydning for
  - i. Sygdomsstadie
  - ii. Behandling
  - iii. Ventetid, indlæggelsestid, forsinkelse i behandling
  - iv. Sygdomsudfald (komplikationer, 30-dages mortalitet, overlevelse, remission, tilbagefald)
- c. Sygdomsrelaterede faktorer betydning for
  - i. Behandling
  - ii. Ventetid, indlæggelsestid, forsinkelse i behandling
  - iii. Sygdomsudfald (komplikationer, 30-dages mortalitet, overlevelse, remission, tilbagefald)
- d. Behandlingsrelaterede faktorer betydning for
  - i. Sygdomsudfald (komplikationer, 30-dages mortalitet, overlevelse, remission, tilbagefald).
- e. Forløb efter operationen – fx tilbagevenden til arbejdsmarkedet.
- f. Psykosociale følger af cancer – fx forbrug af psykofarmaka før og efter behandling for cancer.

#### *Betydningen af organisatoriske faktorer*

Disse analyser kan give indblik i forskelle i kvaliteten af behandlingen imellem afdelinger / regioner. Regionale forskelle i sygdomsstadie på diagnosetidspunktet kan være bestemt af forskelle i patienternes henvendelsesmønster, forskelle i udredningspraksis, samt et udtryk for henvisning mellem afdelinger/regioner af særlige patienttyper. Regionale forskelle i behandling, diverse ventetider og overlevelse kan også være et udtryk for kvalitetsforskelle i behandlingen. Fundne associationer påvirkes af patientsammensætningen, hvorfor justering for dette er af stor betydning.



#### *Betydningen af patientrelaterede faktorer*

Disse analyser kan blandt andet vise betydningen af sociale eller livsstilsrelaterede prognostiske faktorer og om der er en utilsigtet forskelsbehandling af bestemte patientgrupper (fx personer med lav socioøkonomisk status). Desuden er det muligt at eksplorere årsagerne til eventuel social ulighed.

#### *Betydningen af sygdomsrelaterede faktorer*

Disse analyser (fx stadie på diagnosetidspunktet) vil især være relevante i relation til estimering af prognose, samt til monitorering af om patienter med et givet stadie bliver allokeret til den foreskrevne behandling.

#### *Betydningen af behandlingsrelaterede faktorer*

Sammenhængen mellem behandling og sygdomsudfald vil ofte være studeret i protokollerede, eventuelt randomiserede studier. Analyser af behandlingsrelaterede faktorer på baggrund af data fra kræftdatabaserne vil typisk kunne anvendes til mere langsigtet monitorering af behandlingseffekter, samt til at kontrollere om de fund man har gjort i de randomiserede undersøgelser holder i 'den virkelige verden'.

#### *Forløb efter operationen*

Der er kun begrænset viden om, hvordan kræftpatienter klarer sig på arbejdsmarkedet efter end behandling. Analyse af data fra kræftdatabaserne koblet med data fra centrale registre kan vise, hvorvidt forskellige behandlingstyper har betydning for tilbagevenden til arbejdsmarkedet, ændringer i indkomst og ægteskabelige position. Det kan endvidere undersøges hvorvidt patienternes socioøkonomiske status på diagnosetidspunktet har betydning herfor.

#### *Psykosociale følger af cancer*

Psykosociale følger af en cancerdiagnose kan undersøges ved kobling til centrale registre, fx i form af forbrug af benzodiazepiner og antidepressiva fra Lægemedeldatabasen, depressionsudredning i almen praksis via ydelseskoder i Sygesikringsregisteret eller oplysning om indlæggelser med depression fra det Centrale Psykiatriregister. Data fra kræftdatabaserne kan anvendes til at analysere om patientrelaterede, sygdomsrelaterede og behandlingsrelaterede faktorer har betydning herfor.

## 4 Afgrænsning af forskningsspørgsmålet

Ovenstående model for analyser i de kliniske kræftdatabaser viser, at det er muligt at få svar på mangfoldige forskningsspørgsmål ud fra disse data. Dette øger vigtigheden af en klart defineret forskningsproblemstilling ved opstart af et studie.

En vigtig skelnen er, hvorvidt der ønskes svar på ætiologiske spørgsmål (altså at drage slutninger om årsagssammenhænge) eller det ønskes at prædiktere et givent klinisk udfald, hvor det ikke er størrelsen af eventuelle sammenhænge som er i fokus.

Ovenstående model for analyser i de kliniske databaser tager udgangspunkt i ætiologiske forskningsspørgsmål. Her er fokus at bestemme tilstedeværelse eller fravær af formodede årsagssammenhænge mellem en given eksponering (fx rygning) og et vist udfald (fx sårinfektion). Kontrol for confounding og brug af viden om sygdommens patofysiologi og behandlingsforløb er fundamentalt for sådanne analyser. En veldefineret problemstilling, hvor det klart defineres, hvad der er eksponering og udfald, gør det lettere at få overblik over hvilke andre faktorer, der vil kunne influere på den studerede sammenhæng – og altså lettere at foretage kontrol for mulige confoundere.

I prædikerende studier inkluderes ofte en del risikofaktorer i en statistisk model, og det analyseres hvor godt disse faktorer prædikerer udfaldet. Da der i sådanne studier typisk ikke tages højde for de forskellige faktorerers indbyrdes afhængighed, er det ikke muligt at udtale sig om *størrelsen* af effekten af *de enkelte* risikofaktorer.

Det vil ofte være hensigtsmæssigt for den klinisk funderede forsker at samarbejde med en epidemiolog og en statistiker med henblik på opstilling af de korrekte analysemodeller og -metoder.

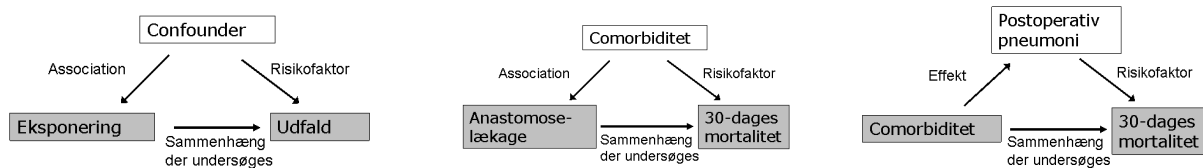
### 4.1 Årsagsdiagrammer

Ethvert forskningsprojekt bør starte med, at forskeren danner sig et overblik over den sammenhæng, der skal studeres, ved at tegne et årsagsdiagram. Årsagsdiagrammet fremstiller hvilke faktorer som forskeren, på baggrund af sin kliniske viden og faglitteraturen, mener, vil kunne påvirke sammenhængen mellem eksponering og udfald. Der kan tages udgangspunkt i de grupper af faktorer, der er afbildet i Figur 1, når problemstillingen tænkes igennem. Årsagsdiagrammet hjælper med at identificere confoundere, og skelne dem fra mediatorer.

I Appendix 1 gives en detaljeret beskrivelse af, hvordan et årsagsdiagram optegnes ud fra en forskningsproblemstilling, der undersøger sammenhængen mellem comorbiditet og 30-dages mortalitet for kolorektalcancer patienter.

### 4.2 Confounding og mediering

Confounding beskrives ofte som en sammenblanding af effekter. Det opstår, når forskeren prøver at bestemme betydningen af en eksponering for et bestemt udfald, men i stedet får målt effekten af en anden faktor, en confounder. Confounding er uønsket, fordi den skjuler den 'sande' effekt af eksponeringen.



Figur 2,3 og 4

Der er tre egenskaber, der skal være opfyldt, for at en variabel kan være en potentiel confounder (figur 2): (i) variabelen skal være en risikofaktor for udfaldet, (ii) den skal være associeret med eksponeringen, dvs fordelingen af variabelen skal være forskellig blandt de eksponerede og de ikke-eksponerede, og (iii) den må **ikke** være en følge af eksponeringen, og må således ikke være en del af årsagskæden mellem eksponering og udfald.

Confounding kan illustreres ud fra et studie vedrørende sammenhængen mellem anastomoselækage (sammensyningen af tarmen brister) og 30-dages mortalitet efter operation for coloncancer (figur 3). I en analyse af denne problemstilling vil det være korrekt at justere for comorbiditet, idet litteraturen og den kliniske erfaring viser, at høj comorbiditet er en risikofaktor for at dø i forbindelse med en operation (kriterium i). Det er desuden sandsynligt at hyppigheden af anastomoselækage er forskellige mellem patienter med høj og lav comorbiditet – formentlig højere frekvens hos patienter med lav comorbiditet; der er altså en association mellem anastomoselækage og comorbiditet (kriterium ii). Endelig er det ikke sandsynligt at en komplikation til operationen i form af anastomoselækage påvirker patientens comorbiditet, der jo er målt før operationen. Det tredje krav til en confounder er således også opfyldt, og det kan konkluderes, at comorbiditet er en potentiel confounder for sammenhængen mellem anastomoselækage og 30-dages mortalitet.

Et andet eksempel illustrerer sammenhængen mellem comorbiditet og 30-dages mortalitet efter operation for cancer (figur 4). Postoperativ pneumoni er en risikofaktor for postoperativ død (kriterium i). Ydermere er comorbiditet associeret med postoperativ pneumoni, idet en høj comorbiditet øger risikoen for postoperativ pneumoni (kriterium ii). Det mest sandsynlige er, at postoperativ pneumoni er en *effekt* af høj comorbiditet, og dermed en del af årsagskæden, der fører til højere 30-dages mortalitet. Postoperativ pneumoni opfylder således ikke kriterium iii – og altså ikke alle kriterierne for en confounder. Der skal således ikke justeres for denne variabel i analysen. Postoperativ pneumoni er derimod en *mediator* - en del af årsagskæden. Justeres analyserne for en mediator, vil noget af den sande effekt af sammenhængen fjernes, hvorved effekten underestimeres. På den anden side, selvom justering for en mediator er ukorrekt i relation til confounder kontrol, kan det være nyttigt at inkludere mediatorer i sine analyser for at eksplorere mulige årsdagsmekanismer. For en mere omfattende beskrivelse af confounding kan henvises til følgende reference (5).

Det er patofysiologisk viden og viden om mulige årsagssammenhænge fra klinikken, der hjælper forskeren med at afgøre, hvilke faktorer der opfylder kriterierne for at være en potentiel confounder eller om der eventuelt er tale om en mediator.

Da confounding skjuler den sande sammenhæng, der studeres, bør den forebygges eller fjernes. Der benyttes typisk multivariate regressions analyser. En forudsætning er selvfølgelig at man i databasen har tilgængelig information vedrørende confounderne, eller at man kan indhente oplysningerne fra centrale registre.

#### **4.4 Interaktion**

Når styrken af sammenhængen mellem to variable varierer med niveauet af en tredje variabel, tales der om statistisk interaktion eller effektmodifikation. Det kan for eksempel være hvis effekten af en behandling er bedre blandt mænd end blandt kvinder. Der er også tale om interaktion, når flere årsager virker sammen og den resulterende risiko er større (eller mindre) end man ville forvente ved at kombinere effekterne af de enkelte årsager. Klinikere taler her ofte om synergisme hhv. antagonisme.

Undersøgelse for interaktion kan foretages i multivariate regressionsanalyser. Der kan ikke gives specifikke retningslinjer for hvilke interaktioner, der bør undersøges for, men som tommelfingerregel bør der undersøges for interaktion mellem variable, som ud fra klinisk viden og litteraturen, kan formodes at interagere. Desuden anbefales det at teste for interaktioner mellem centrale variable og køn og alder.

## 5 Analyseprocessen

I det følgende vil analyseprocessen blive gennemgået ganske kortfattet. Der henvises til statistiske lærebøger for en grundig gennemgang, fx DG Altmans Practical statistics for medical research (6).

Analyseprocessen falder groft set i tre faser:

- Forberedelse af data
- Deskriptive analyser
- Multivariate regressionsanalyser

### 5.1 Forberedelse af data

Data kontrolleres for oplagte fejl, logiske fejl og datofejl. Kategoriske variable kontrolleres for om de registrerede data har værdier, der falder uden for de mulige kategorier. Det kontrolleres om kontinuerte variable falder inden for *rimelige* grænser. Der kontrolleres for outliers (værdier af data, der ikke er kompatible med resten af data), og der undersøges for missing data og mulige årsager hertil. Fordelingen af kontinuerte variable undersøges. Dette er specielt vigtigt, hvis de planlagte analyser forudsætter normalfordelte data. I disse tilfælde kan det komme på tale at transformere data for at opnå normalfordeling.

### 5.2 Deskriptive analyser

Næste trin er at gennemføre simple, deskriptive analyser. I univariate analyser undersøges en variabel ad gangen. For kategoriske variabel dannes frekvenstabeller, mens kontinuerte variable undersøges i form af gennemsnit, median og eventuelt percentiler. Grafisk afbildning kan være en fordel. I bivariate analyser undersøges variablene parvis overfor hinanden. De forskellige variable testes overfor henholdsvis eksponeringsvariabel og outcomevariabel. Kategoriske variable undersøges i krydstabeller, med udregning af chi-i-anden test. Gennemsnit af kontinuerte variable kan sammenlignes mellem forskellige niveauer af eksponerings- eller outcomevariablen og testes ved hjælp af en parametriske eller en non-parametriske test, afhængig af normalfordeling eller ej. Grafisk afbildning kan også her være til stor hjælp. Overlevelsedata undersøges ved hjælp af Kaplan-Meier plots og log-rank test, se evt (7).

Ovennævnte analyser giver en ide om størrelsen af association mellem eksponering og udfald, og hvorvidt og i hvor høj grad de potentielle confoundere er associeret med eksponering og udfald. Disse analyser er nødvendige inden der igangsættes mere avancerede analyser i form af multivariate regressionsmodeller.

### 5.3 Regressionsanalyser

Her følger kortfattet oversigt over de statistiske analysemetoder, der hyppigst finder anvendelse i relation til kliniske databaser. Med mindre forskeren har stort kendskab til denne type analyser, anbefales det at søge epidemiologisk/statistisk vejledning (fx i et kompetencecenter).

### **Logistisk regression**

Anvendes ved dikotome udfald, fx ja/nej udfald. Eksempler: Død inden for 30 dage ja/nej, recidiv ja/nej. Ofte vælger forskeren at gruppere udfald med flere niveauer i to udfald, fordi det er let at arbejde med og fortolke resultater fra en logistisk regression. Fx gruppering af Dukes' stadietinddeling for kolorektalcancere i Dukes A+B og Dukes C+fjernmetastaser.

For nærmere beskrivelse af logistisk regression se fx (8)

### **Ordinal regression**

Anvendes når udfaldet udgøres af ordinale data, dvs data hvor der er en logisk rangordning af kategorierne.

Eksempel: Dukes A/Dukes B/Dukes C/fjernmetastaser.

Resultaterne af en ordinal regressionsanalyse kan være svære at fortolke.

### **Lineær regression**

Anvendes når udfaldet er et kontinuert mål. Eksempel: Indlæggelsestid, ventetid.

For nærmere beskrivelse af lineær regression se fx (8).

### **Cox regression**

Cox regression, bruges til overlevelsesanalyser, og analyserer den tid det tager til en begivenhed indtræffer. Prototypen på en sådan begivenhed er død, heraf navnet 'overlevelsesanalyse'.

Overlevelsesanalyser kan dog anvendes i en meget bredere sammenhæng, hvori der indgår et tidsperspektiv, fx tid til recidiv.

Analyser af 30-dages mortalitet analyseres typisk ved brug af logistisk regression, som ovenfor anført, men kan også analyseres ved hjælp af Cox regression.

Der henvises endvidere til afsnit 7.6 for en beskrivelse af forskellige metoder til overlevelsesanalyser. For nærmere beskrivelse af Cox regression se fx (6).

### **Multilevel analyser**

Multilevel analyser bruges til at analysere data med flere i hinanden indlejrede niveauer. Fx *patienter* behandlet på *hospitaller*, eller som et eksempel på data med tre niveauer fx *patienter* behandlet af *læger* på *hospitaller*. Hvorvidt multilevel analyser er det rette analysevalg afhænger i høj grad af, hvad det er for en sammenhæng der skal undersøges. Hvis ønsket er at undersøge effekten af en variabel på ét niveau (fx kirurg volumen (mål for kirurgens erfaring)) på et udfald på et andet niveau (fx død af patienten) vil multilevelanalyser være relevante. Ligeledes benyttes multilevel analyser, hvis man fx ønsker at undersøge *variationen* i dødelighed mellem afdelinger, og ikke er interesseret i de enkelte afdelingers estimater. Ønskes derimod en direkte sammenligning af afdelinger, med et estimat for hver afdeling, kan de ovenfor beskrevne, mere gængse metoder være mere relevante.

Er eksponering og udfald på samme niveau, men er man bekymret for korrelation – dvs at patienterne ligner hinanden mere indenfor den samme afdeling end mellem afdelinger – kan der benyttes multi-level analyser. Men der kan også benyttes GEE (generalized estimating equations) eller sandwich estimators, som kan kontrollere både estimater og især varianser og p-værdier. Det vil være tilrådeligt at konsultere et af kompetencecentrene for disse analyser

## 6 Håndtering af nedsat datakvalitet

Datakvaliteten kan beskrives ved databasens dækningsgrad, datakomplethedensgraden og validiteten af de enkelte variable. Problemer med dækningsgrad og datakomplethed kan give anledning til selektionsbias, mens lav validitet kan give anledning til informationsbias.

### 6.1 Dækningsgrad (selektionsbias)

Det er databasernes målsætning at mindst 90 % af patienterne med den pågældende cancersygdom er registreret i databasen (jf. Basiskrav for landsdækkende kliniske kvalitetsdatabaser, afsnit 12). For at undersøge dette valideres databasen op mod data fra Landspatientregisteret og evt. Cancerregisteret. Ikke alle databaser kan overholde kravet til dækningsgrad. Lav dækningsgrad kan nedsætte generaliserbarheden af resultaterne. Bias opstår når patienterne i databasen er systematisk forskellige fra dem, som resultaterne skal generaliseres til. Det er fx sandsynligt, at der i databaserne er en underrapportering af patienter, der er døde undervejs i behandlingsforløbet, fordi disse patienters journaler bliver fjernet fra afdelingen, når patienten dør, og klinikerne så har sværere ved at få fat i data om disse patienter.

Forskeren bør beskrive mulige konsekvenser for resultaterne af lav dækningsgrad. Ved hjælp af oplysninger fra LPR og andre centrale registre er det muligt at indhente oplysninger om fx alder, køn, mortalitet, comorbiditet og socioøkonomiske forhold på de patienter, der ikke er registreret i databasen, hvorved disse kan karakteriseres og sammenlignes med patienterne i databasen.

### 6.2 Datakomplethedegrad

Nogle variable er mere komplet afrapporteret i databaserne end andre. Fx mangler ca. 40 % af livsstilsvariablene i DCCG databasen, og 8 % af patienterne i samme database er ikke stadietildelt i henhold til Dukes klassifikation. Normale statistiske procedurer ekskluderer patienter med manglende data i en eller flere indgående variable fra regressionsanalysen. Dette kan medføre både upræcise og biased resultater (9). Der bør redegøres for datakomplethedensgraden og mulige konsekvenser af lav komplethed for resultaterne. Dette kan gøres ved hjælp af sensitivitetsanalyser. Her estimeres betydningen af manglende data i studiet ved at antage forskellige grader af maldistribution af variabelen mellem grupperne der sammenlignes, hvorefter det iagttages hvordan dette ville påvirke resultaterne. I ovenfor nævnte eksempel med patienterne, der ikke var klassificeret i henhold til Dukes, kan variationen vurderes under en antagelse af at alle 8 % havde enten fjernmetastaser eller Dukes A.

Multipel imputation er en metode, der gør det muligt at inkludere alle de observerede patienter på trods af at nogle har manglende data i nogle af variablene. Metoden kan fx bruges hvis mange patienter mangler oplysninger om livsstil (10). Ideen bag multiple imputation er at bruge information fra patienter *uden* manglende livsstilsdata til at lære om sammenhængen mellem livsstilsvariablene og andre variable (fx køn, alder, socioøkonomisk status, comorbiditet, stadiet mm). Efterfølgende bruges kendskabet til denne sammenhæng til at estimere livsstilsdata hos patienter *med* manglende livsstilsdata. For også at tage højde for variansen skal proceduren gentages et vist antal gange, typisk 5-10 gange (heraf navnet 'multiple' imputation), hvilket resulterer i x datasæt, som analyseres på normal vis, hvorefter resultaterne kombineres.

### 6.3 Validitet (informationsbias)

Validitet udtrykker i hvor høj grad data måler det, de er tilsigtet at måle. Nogle variable, som fx ASA score, er kendt for at have forholdsvis lav validitet (11;12), mens mange andre kliniske variable, som fx tumorlokalisering og Dukes klassifikation, har høj validitet i databaserne.

Kvaliteten af data er valideret i fx DCCG og Danish Breast Cancer Groups database (DBC)(13;14). Også data fra de centrale registre har forskellig validitet; høj i fx de sociodemografiske registre og Lægemedelstatistik registeret (15), og mere varierende i Landspatientregisteret (16).

Hvis de indsamlede data er behæftet med fejl, kan det medføre fejl i analyserne, informationsbias. Man taler om misklassifikation, hvis den pågældende variabel måles på en kategorisk skala, og fejl fører til, at en person bliver placeret i den forkerte kategori. Det kan fx være en ikke-ryger, der bliver kategoriseret som en ryger. Misklassifikationen kan være differentiell eller non-differentiell. For misklassifikation af eksponeringsvariablen, vil misklassifikation være non-differentiell hvis den ikke er relateret til det udfald der studeres, men den vil være differentiell, hvis den *er* relateret til udfaldet. Vice versa med misklassifikation af udfaldsvariablen. Se fx følgende reference for en kort gennemgang af bias i kliniske forskning (17).



## 7 Kliniske udfald

### 7.1 Stadie

De fleste cancere klassificeres i henhold til TNM klassifikationen. For kolorektalcancer er der tradition for Dukes' klassifikation. Oftest dichotomiseres stadie, så den beskriver cancer med eller uden lymfeknudemetastaser, alternativt med eller uden fjernmetastaser (18;19). Hæmatologiske cancere udgør en speciel cancertype, hvor canceren ofte er lokaliseret til netop lymfeknuderne. Her vil det ikke udelukkende være Ann Arbor stadiet, der vedrører lokaliseringen af de afficerede lymfeknuder, der alene siger noget om hvor fremskreden canceren er. Der vil med fordel også kunne analyseres på følgende parametre som sygdomsudfald: Ann Arbor  $\geq$  II ja/nej, ekstranodel lokalisering ja/nej, LDH forhøjelse ja/nej, performance status  $>2$  ja/nej, IPI score  $\geq 3$  ja/nej, samt evt B-symptomer ja/nej.

### 7.2 Diagnostiske undersøgelser

Information om diagnostiske undersøgelser varierer mellem databaserne. Mange diagnostiske undersøgelser udføres i speciallægepraksis eller på en anden afdeling end den behandlende, hvorfor oplysninger herom kan være begrænsede. Et muligt diagnostisk udfald kunne fx være om patienten er blevet udredt mhp fjernmetastaser. I kolorektalcancer databasen vil det indebære, om der er foretaget billeddiagnostik af lunger og lever. I lungecancerdatabasen er en hel række diagnostiske procedurer registreret, hvor et muligt udfald vil være om patienterne har fået udført PET-scanning.

### 7.3 Ventetid

Forskellige ventetider vil være relevante:

Tid mellem henvisning og første besøg på behandlende afdeling.

Tid mellem henvisning og behandlingsstart (operation / kemoterapi / strålebehandling), alternativt tid mellem første besøg på behandlende afdeling og behandlingsstart (operation / kemoterapi / strålebehandling).

For patienter behandlet med kemoterapi kan ses på forsinkelse af behandlingen, svarende til om den observerede tid, som bruges på at gennemløbe det planlagte antal cyklusser, svarer til den forventede tid.

### 7.4 Behandling

Behandlingsregimer ændrer sig over tid, kan inkludere forskellige subgrupper af patienter over tid og kan være implementeret på forskellige tidspunkter i de forskellige regioner/sygehuse, hvorfor det kan være svært at komme med gode mål for 'den optimale behandling' på et givet tidspunkt. Af disse, og af analyse-mæssige årsager, vælges ofte i stedet forholdsvis simple målinger af behandlingskvaliteten fx sphincterbevarende kirurgi ja/nej (rectumcancerpatienter) og immunterapi ja/nej (lymfompatienter).

Har det videnskabelige selskab udstukket en rekommandation vil det kunne analyseres i hvor høj grad denne bliver fulgt.

### 7.5 Komplikationer

Behandles ofte som dichotome variable (ja/nej). Fx sårinfektion efter kirurgi.

### 7.6 Overlevelse

Overlevelse kan studeres ved hjælp af tre metoder: observeret overlevelse, cancer-specifik overlevelse og relativ overlevelse. Hver metode har sine fordele og ulemper.

*Observeret overlevelse* tæller alle dødsfald uanset årsag med, og er specielt anvendelig når sammenhængen mellem sygdoms- eller behandlingsrelaterede faktorer og overlevelse analyseres. Observeret overlevelse har derimod sine begrænsninger, når betydningen af patientrelaterede faktorer, fx socioøkonomisk status eller livsstilsfaktorer, studeres. Her vil det ikke kunne afgøres om en eventuel overlevelsesforskel imellem eksponeringsgrupperne er en konsekvens af canceren eller er en konsekvens af den i forvejen kendte forskel i overlevelsen blandt lavt vs. højt uddannede henholdsvis rygere vs. ikke-rygere.

*Cancer-specifik overlevelse* tæller kun dødsfald på grund af den aktuelle cancer med, mens alle andre dødsfald behandles som censurerede events. Med cancer-specifik overlevelse undgås ovennævnte sammenblanding af effekter, men metoden kræver information om dødsårsager fra dødsattester, som ofte er misklassificeret (20-22).

*Relativ overlevelse* er defineret som den observerede overlevelse i patientgruppen divideret med den forventede overlevelse i en sammenlignelig gruppe af baggrundsbefolkningen. Ofte estimeres den forventede overlevelse fra nationale overlevelsestabeller, der er stratificeret efter alder, køn og årstal. I tilfælde hvor der analyseres på socioøkonomiske determinanter for overlevelse bør de forventede overlevelsestabeller også være stratificeret på socioøkonomisk status, hvilket ikke er tilgængelig som standard fra Danmarks Statistik, men som dog kan beregnes(23;24). Der findes ingen nationale overlevelsestabeller, som er stratificeret på livsstil, som er en væsentlig dødsårsag.

Analyser af observeret og cancer-specifik overlevelse foretages typisk i en Cox regressionsmodel, mens analyser af relativ overlevelse er noget mere kompliceret (25).

## **7.7 Recidiv**

Se nedenfor 8.3 (data fra centrale registre)

## **7.8 Tilbagevenden til arbejdsmarkedet**

Fra Danmarks Statistik kan indhentes oplysninger om ansættelsesforhold. Et jobskifte vil kunne registreres som en ændring i arbejdsgiveridentifikationsnummer. Oplysninger om eventuel arbejdsløshed, førtidspensionering, efterløn eller alderspensionering er ligeledes let tilgængelige. Se endvidere nedenfor 'data fra centrale registre', punkt 8.2..

## **7.9 Psykosociale følger**

Data fra centrale registre vil kunne give visse oplysninger om mulige psykosociale følger af canceren, se fx (26). Fra Lægemiddelstatistik registeret kan man få oplysninger om patienternes forbrug af benzodiazepiner og antidepressiva før og efter kræftdiagnosen. Fra det Centrale Psykiatri Register kan indhentes oplysninger om eventuelle psykiatriske indlæggelser. Fra Sygesikringsregisteret kan indhentes oplysninger om depressionsudredning i almen praksis via ydelseskoder.

## 8 Data fra centrale registre

### 8.1 Comorbiditet

Comorbiditet er en vigtig confounder i mange studier af patientforløb. I DCCG databasen afrapporteres ASA-score og i Lymfomdatabasen rapporteres WHO performance status, som begge er meget grove mål for comorbiditet og har en høj interobservator variabilitet (11;12). Som supplement til de kliniske mål for comorbiditet, eller, i de databaser som ikke har comorbiditetsmål, kan Charlsons comorbiditets index benyttes. Det kan dannes på basis af udskrivningsdiagnoser fra Landspatientregisteret (27) og inkluderer 19 større sygdomskategorier (se appendix 2). Det er blevet adapteret og valideret til brug med hospitalsudskrivningsdiagnoser mhp prædiktions af kort- og langtidsoverlevelse (28). Dannelsen af comorbiditetsvariable og Charlson index kan endvidere ske på baggrund af sygdomsspecifik medicin, hvorom der kan indhentes oplysninger fra Lægemiddelstatistik registeret.

### 8.2 Socioøkonomisk position

Socioøkonomisk position er en vigtig confounder i mange studier, fx i analyser af livsstils betydning for prognosen. Danmarks Statistik råder over mange forskellige variable, der kan beskrive forskellige dimensioner af socioøkonomisk position. De hyppigst anvendte er højest fuldførte uddannelse, indkomst (brutto eller disponibel) og erhverv, der med fordel kan bruges i kombination. Der er fordele og ulemper ved de forskellige mål, se fx (29;30).

Hyppigt brugte socioøkonomiske variable:

Variabelbetegnelse hos Danmarks Statistik	Beskrivelse
<b>Familie</b>	
Civilst	Civilstand
ctype	C-familie type
<b>Uddannelse</b>	
ALMFSP	Højeste udførte almenudd
HFUDD	Højest fuldførte uddannelse
<b>Arbejdstilknytning</b>	
arledgr	andel af året ledig
sumgrad	Sum af ledighedsgrader
<b>Stilling</b>	
pstill2	Primær arbejdsstilling
Nystgr	Stillingskode
socio	Socioøkonomisk gruppering med vægt på jobrelevante færdigheder, t.o.m. 2001
Socio02	Socioøkonomisk klassifikation, fra 2002
DISCO	Lønmodtagerens arbejdsfunktion
<b>Indkomster</b>	
brutto	Bruttoindkomst er den indkomst, der kommer til almindelig indkomstbeskatning
QSPLINDK	Skattepligtig indkomst
DISPON_NY	Indkomst efter skat og renter (Samlet indkomst fratrukket renteudgifter, skat mm.)

### 8.3 Recidiv

Data fra LPR og Cancerregisteret kan give et mål for om patienterne har fået recidiv af deres cancer ved at definere ICD-10 koder, der er specifikke for recidiv af den pågældende cancersygdom. Sensitiviteten af denne metode er ikke optimal, idet nogle patienter kan have recidiv uden at være i kontakt med hospitalet herfor. Men det er aktuelt nok det bedste mål.

## 9 Praktisk om opkobling til Danmarks Statistik

Kobling af data fra en klinisk kræftdatabase til data fra centrale registre foretages af Danmarks Statistik (DS). DS etablerer et projekt-drev på deres server, som man kan logge sig på fra sin egen computer via internettet. Danmarks Statistik giver kun adgang til sådanne opkoblinger til autoriserede forsknings- og analysemiljøer (fx kompetencecentre).

Første trin er at sende en kort, foreløbig projektbeskrivelse til Forskningsservice hos Danmarks Statistik, der herefter i dialog med forskeren giver et overblik over datamulighederne og en beregning af omkostningerne. Der vil ofte være en ventetid på op til nogle få måneder før end data kan leveres. Der henvises i øvrigt til skrivelsen 'Introduktion til Forskningsservice i Danmarks Statistik', der findes på DS' hjemmeside (link: <http://www.dst.dk/TilSalg/Forskningservice/Intro.aspx>). Her beskrives også i hvilke tilfælde Datatilsynet og andre instanser skal søges om tilladelse, samt hvordan arbejde via fjerncomputer foregår. Danmarks Statistik giver mulighed for at arbejde i statistik softwareprodukterne SAS, STATA, SPSS, GAUSS.

## 10 Hvad kan kompetencecentre hjælpe med?

Kompetencecentre er knyttet til videnskabelige institutioner eller institutter under universiteterne og har ansat forskere med stor epidemiologisk erfaring samt fagstatistikere med stor kendskab til de komplekse epidemiologiske problemstillinger. Kompetencecentre har således stor erfaring i analyse af komplekse datasæt og håndtering og analyse af data fra centrale registre. Da der er tale om forskningsmiljøer, vil kompetencecentre løbende være på forkant med den metodologiske udvikling inden for den kliniske epidemiologi.

Der er ikke afsat økonomiske centrale midler til kompetencecentre til at hjælpe kliniske forskere med ovennævnte analyser, men kompetencecentre står til rådighed med initial vejledning og kan indgå i samarbejde med de kliniske databaser om specifikke analyser og projekter. En forudsætning for at kompetencecentre kan indgå i sådanne samarbejder, er der kan skaffes de fornødne økonomiske midler til aflønning af it-funktioner, statistik og epidemiologi.

## 11 Udvalgte studier baseret på kliniske kræftdatabaser

C. A. Bertelsen, A. H. Andreasen, T. Jorgensen, and H. Harling. Anastomotic leakage after anterior resection for rectal cancer: risk factors. *Colorectal Dis.*, 2008.

C. A. Bertelsen, A. H. Andreasen, T. Jorgensen, and H. Harling. Anastomotic leakage after curative anterior resection for rectal cancer: short and long term outcome. *Colorectal Dis.*, 2009.

S. O. Dalton, M. During, L. Ross, K. Carlsen, P. B. Mortensen, J. Lynch, and C. Johansen. The relation between socioeconomic and demographic factors and tumour stage in women diagnosed with breast cancer in Denmark, 1983-1999. *Br.J Cancer* 95 (5):653-659, 2006.

B. L. Frederiksen, M. Osler, H. Harling, S. Ladelund, and T. Jorgensen. Do patient characteristics, disease, or treatment explain social inequality in survival from colorectal cancer? *Soc.Sci.Med.* 69 (7):1107-1115, 2009.

B. L. Frederiksen, M. Osler, H. Harling, and T. Jorgensen. Social inequalities in stage at diagnosis of rectal but not in colonic cancer: a nationwide study. *Br J Cancer* 98 (3):668-673, 2008.

B. L. Frederiksen, M. Osler, H. Harling, S. Ladelund, and T. Jorgensen. The impact of socioeconomic factors on 30-day mortality following elective colorectal cancer surgery: A nationwide study. *Eur.J.Cancer* 45 (7):1248-1256, 2009.

H. Harling, S. Bulow, L. N. Moller, and T. Jorgensen. Hospital volume and outcome of rectal cancer surgery in Denmark 1994-99. *Colorectal Dis* 7 (1):90-95, 2005.

L. H. Iversen, S. Bulow, I. J. Christensen, S. Laurberg, and H. Harling. Postoperative medical complications are the main cause of early death after emergency surgery for colonic cancer. *Br J Surg* 95 (8):1012-1019, 2008.

T. N. Nickelsen, T. Jorgensen, and O. Kronborg. Lifestyle and 30-day complications to surgery for colorectal cancer. *Acta Oncol.* 44 (3):218-223, 2005.

T. N. Nickelsen, T. Jorgensen, and O. Kronborg. Thirty-day mortality after surgery for colorectal cancer in Denmark. *Colorectal Dis.* 7 (5):500-506, 2005.

V. Peuckmann, O. Ekholm, P. Sjogren, N. K. Rasmussen, P. Christiansen, S. Moller, and M. Groenvold. Health care utilisation and characteristics of long-term breast cancer survivors: nationwide survey in Denmark. *Eur.J.Cancer* 45 (4):625-633, 2009.

## 12 Nyttige hjemmesider

### Kliniske databaser

- Hjemmeside for Kompetencecenter Øst  
[www.kliniskedatabaser.dk](http://www.kliniskedatabaser.dk)
- Basiskrav for landsdækkende kliniske kvalitetsdatabaser  
[http://www.regioner.dk/ServiceOgAdministration/Kliniske%20kvalitetsdatabaser/~media/migration%20folder/upload/filer/sundhed/it%20og%20kvalitet/11.05.2007\\_basiskrav\\_for\\_landsdaek\\_kliniske\\_kvalitetsdatabaser.pdf.ashx](http://www.regioner.dk/ServiceOgAdministration/Kliniske%20kvalitetsdatabaser/~media/migration%20folder/upload/filer/sundhed/it%20og%20kvalitet/11.05.2007_basiskrav_for_landsdaek_kliniske_kvalitetsdatabaser.pdf.ashx)

### Danmarks statistik

- TIMES - detaljeret dokumentation af variable  
<http://dst.dk/HomeDK/Vejviser/dokumentation/TIMES.aspx>
- Dokumentation af høj kvalitetsvariable (dybdegående dokumentation af udvalgte variabler)  
<http://dst.dk/TilSalg/Forskningsservice/hkt4Forsker.aspx>
- Introduktion til Forskningsservice i Danmarks Statistik  
<http://www.dst.dk/TilSalg/Forskningsservice/Intro.aspx>.

### Sundhedsstyrelsen

- Registre – oversigt over Sundhedsstyrelsens registre  
<http://www.sst.dk/Indberetning%20og%20statistik/Sundhedsstyrelsens%20registre.aspx>
- Vejledning i udtræk fra Sundhedsstyrelsens registre  
[http://www.sst.dk/~media/Indberetning%20og%20statistik/Forskerservice/vejledning%20i%20registerudtrk\\_210408.ashx](http://www.sst.dk/~media/Indberetning%20og%20statistik/Forskerservice/vejledning%20i%20registerudtrk_210408.ashx)
- Vejledning om godkendelse af videregivelse af oplysninger fra patientjournaler mv. til brug for konkrete forskningsprojekter  
<https://www.retsinformation.dk/Forms/R0710.aspx?id=123933>
- Sundhedsvæsenets Klassifikationssystem (SKS) til klassifikation af diagnoser og procedurer  
[http://www.sst.dk/Indberetning%20og%20statistik/Klassifikation/SKS\\_vaerktoejer.aspx](http://www.sst.dk/Indberetning%20og%20statistik/Klassifikation/SKS_vaerktoejer.aspx)

### Lægemiddelstatistik registeret

- <http://www.laegemiddelstyrelsen.dk/1024/visLSArtikel.asp?artikelID=8479>

## Appendix 1 - årsagsdiagrammer

### Årsagsdiagrammer

I årsagsdiagrammer angiver pile retningen af kausaliteten mellem de forskellige faktorer. Statistiske undersøgelser kan *ikke* guide hvilken vej pilene skal vende. Det er klinikerens, der på baggrund af sin faglige viden, afgør hvilken vej pilen skal vende.

Eksempel:

Vi ønsker at undersøge sammenhængen mellem vores eksponering: comorbiditet, målt ved ASA-score umiddelbart før operation, og udfaldet: død inden for 30 dage efter en tarmkræft operation (indtegnet i diagrammet nedenfor i grå felter).

På baggrund af nedenstående overvejelser vedrørende kriterierne for en confounder hhv. mediator (afsnit 4.2) indtegnes variable og pile i årsagsdiagrammet.

- Alder er en risikofaktor for postoperativ død (pil indtegnet i diagrammet). Alder påvirker comorbiditet (pil indtegnet i diagrammet). Alder er ikke en del af årsagskæden mellem comorbiditet og postoperativ død.
- Køn er en risikofaktor for postoperativ død (pil indtegnet i diagrammet). Køn påvirker comorbiditet (pil indtegnet i diagrammet). I øvrigt er køn er ikke en del af årsagskæden mellem comorbiditet og postoperativ død.
- Årstallet vil muligvis kunne påvirke den postoperative mortalitet, fx hvis der er kommet nye mere skånsomme operations- og anæstesimetoder til over årene; årstal er risikofaktor for postoperativ død (pil indtegnet i diagrammet). Er observationerne indsamlet over en længere årrække, kan det tænkes at årstallet har betydning for comorbiditet, hvis generelle niveau i befolkningen kan have ændret sig over tid; årstal kan muligvis påvirke comorbiditeten (pil indtegnet i diagrammet). I øvrigt er årstal ikke en del af årsagskæden mellem comorbiditet og postoperativ død.
- Lav socioøkonomisk status er en risikofaktor for postoperativ død (pil indtegnet i diagrammet). Men er det socioøkonomisk position (SEP) der påvirker comorbiditeten eller omvendt? Personer med kort uddannelse har oftere livsstilsbetingede kroniske sygdomme, som diabetes, KOL og hjertekarsygdom. Måler vi SEP ved hjælp af uddannelse, som typiske er opnået i ungdommen, er vi således ikke i tvivl om, at det er uddannelsen, der påvirker comorbiditeten. Kort uddannede har ofte en lavere løn, hvorfor indkomst også kan påvirke comorbiditeten. Men omvendt vil comorbiditet også kunne føre til arbejdsløshed og dermed lavere indkomst – altså at comorbiditet påvirker indkomsten. Retningen på pilen må sættes i den retning, som vi tror mest på. Her sættes pilen fra SEP til comorbiditet, og vi mener således ikke at SEP er en del af årsagskæden. Da comorbiditet desuden i dette eksempel er målt umiddelbart før operationen, har den svært ved at påvirke SEP, som er opnået tidligere.
- Livsstil er en risikofaktor for postoperativ død (pil indtegnet i diagrammet). Spørgsmålet er om det er livsstil (fx rygning) der påvirker comorbiditeten eller omvendt? Det er veldokumenteret at en dårlig livsstil kan give KOL og hjertekarsygdom, som vil give patienterne en dårligere ASA-score. Vi tror således på at pilen går fra livsstil til comorbiditet (pil indtegnet i diagrammet). Livsstil er altså ikke en del af årsagskæden fra ASA til død.
- Komplikationer: Medicinske og kirurgiske komplikationer øger risikoen for postoperativ død (pil indtegnet i diagrammet). En høj comorbiditet giver en øget risiko for medicinske og formentlig også kirurgiske komplikationer til operationen (denne gang peger pilen fra

comorbiditet til komplikationer). Dette er et eksempel på en mediator, hvor komplikationerne er en del af årsagskæden mellem eksponering og udfald.

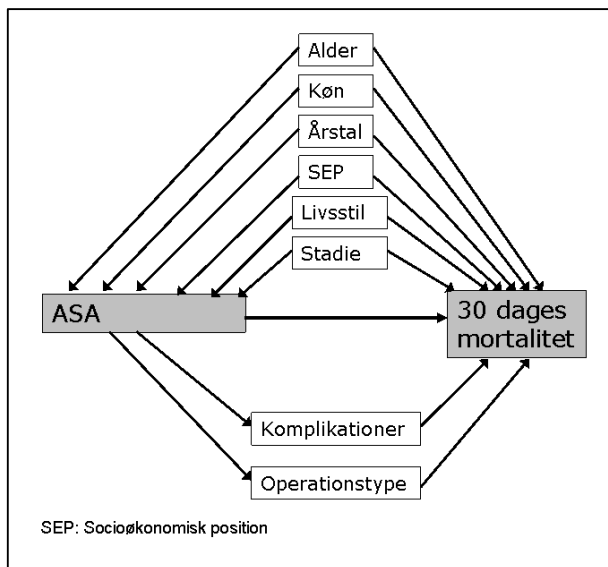
- Operationstype er en risikofaktor for postoperativ død (pil indtegnet i diagrammet). Hvis der er en sammenhæng mellem comorbiditet og operationstype, må det være, at comorbiditeten påvirker operationstypen (fx vil en høj comorbiditet kunne medføre at et mere skånsomt indgreb vælges), mens det ikke er sandsynligt at operationstypen påvirker comorbiditeten, der i øvrigt også er målt før operationen (pil indtegnet i diagrammet). Igen et eksempel på en mediator – en del af årsagskæden.
- Sygdomsstadie påvirker den postoperative død (pil indtegnet i diagrammet). Er det comorbiditet der påvirker sygdomsstadiet eller omvendt? Personer med høj comorbiditet går hyppigere til lægen og har måske derfor større chance for at få opdaget canceren i et tidligt stadie. Omvendt vil et højt sygdomsstadie måske i sig selv kunne påvirke anæstesiologens vurdering af patientens comorbiditet. Her er det svært at afgøre pilens retning, og forskeren må beslutte hvilken retning af pilen der er meste sandsynlig – eller vælge at analysere i både en model hvor pilen peger den ene og den anden vej. Vi har i figuren valgt at lade stadie indgå som en confounder (pil indtegnet i diagrammet).

Ved hjælp af årsagsdiagrammet har vi nu et overblik over forskningsspørgsmålets komplekse struktur. Confounderne, som skal inkluderes i regressionsanalysen, er de faktorer der er karakteriseret ved at to pile peger væk fra dem:  $x \leftarrow \text{'faktor (confounder)} \rightarrow y$ . Mediatorerne, som **ikke** skal inkluderes i regressionsanalysen, er de faktorer, hvor pilene peger igennem faktoren:  $x \rightarrow \text{'faktor' (mediator)} \rightarrow y$ . Mediatorerne kan dog inkluderes i separate analyser for at undersøge mulige mekanismer.

Vi har valgt at placere confounderne ovenfor den vandrette pil og mediatorerne under den vandrette pil for at øge overblikket.

Der kan meget vel være confoundere, som forskeren tror, er betydningsfulde, men som ikke er målt. Disse bør i princippet også indtegnes i diagrammet og mulig betydning heraf bør diskuteres, når resultaterne fremstilles.

Ovenstående er en let tilgængelig, men stærkt forenklet måde at anvende causale diagrammer på. I realiteten kan de forskellige confoundere og mediatorer også påvirke hinanden, hvilket også burde angives med pile. Der henvises til *Modern Epidemiology* for en nærmere gennemgang af årsagsdiagrammer (31). Desuden henvises til studie af Frederiksen et al., hvor årsagsdiagrammer er praktisk anvendt (32).





## Appendix 2 – Charlsons comorbiditetsindex

### Charlsons comorbiditetsindex

Definitions of disorders to be used in the comorbidity index as developed by Charlson (27)

<b>Disorder</b>	<b>Weight</b>	<b>ICD8</b>	<b>ICD10</b>
Myocardial infarction	1	410	I21;I22;I23
Congestive heart failure	1	427.09; 427.10; 427.11; 427.19; 428.99; 782.49	I50; I11.0; I13.0; I13.2
Peripheral vascular disease	1	440; 441; 442; 443; 444; 445	I70; I71; I72; I73; I74; I77
Cerebrovascular disease	1	430-438	I60-I69; G45; G46
Dementia	1	290.09-290.19; 293.09	F00-F03; F05.1; G30
Chronic pulmonary disease	1	490-493; 515-518	J40-J47; J60-J67; J68.4; J70.1; J70.3; J84.1; J92.0; J96.1; J98.2; J98.3
Connective tissue disease	1	712; 716; 734; 446; 135.99	M05; M06; M08; M09;M30;M31; M32; M33; M34; M35; M36; D86
Ulcer disease	1	530.91; 530.98; 531-534	K22.1; K25-K28
Mild liver disease	1	571; 573.01; 573.04	B18; K70.0-K70.3; K70.9; K71; K73; K74; K76.0
Diabetes type1	1	249.00; 249.06; 249.07; 249.09	E10.0, E10.1; E10.9
Diabetes type2		250.00; 250.06; 250.07; 250.09	E11.0; E11.1; E11.9
Hemiplegia	2	344	G81; G82
Moderate to severe renal disease	2	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 with end organ damage type1 type2	2	249.01-249.05; 249.08 250.01-250.05; 250.08	E10.2-E10.8 E11.2-E11.8
Any tumor	2	140-172; 174-194	C00-C43; C45-C75
Leukemia	2	204-207	C91-C95
Lymphoma	2	200-203; 275.59	C81-C85; C88; C90; C96
Moderate to severe liver disease	3	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
Metastatic solid tumor	6	195-198; 199	C76-C80
AIDS	6	079.83	B21-B24

Weights for each condition of the patient. The total equals the score: Example: chronic pulmonary (1) and lymphoma (2) = total score (3)

## Referencer

- (1) Blichert-Toft M, Christiansen P, Mouridsen HT. Danish Breast Cancer Cooperative Group-- DBCG: History, organization, and status of scientific achievements at 30-year anniversary. *Acta Oncol* 2008;47(4):497-505.
- (2) Engeland A, Haldorsen T, Dickman PW, Hakulinen T, Moller TR, Storm HH, et al. Relative survival of cancer patients--a comparison between Denmark and the other Nordic countries. *Acta Oncol* 1998;37(1):49-59.
- (3) Sant M, Capocaccia R, Verdecchia A, Gatta G, Micheli A, Mariotto A, et al. Comparisons of colon-cancer survival among European countries: The Eurocare Study. *Int J Cancer* 1995 Sep 27;63(1):43-8.
- (4) Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine.* Second ed. Boston/Toronto/London: Little, Brown and Company; 1991.
- (5) Jager KJ, Zoccali C, Macleod A, Dekker FW. Confounding: what it is and how to deal with it. *Kidney Int* 2008 Feb;73(3):256-60.
- (6) Altman DG. *Practical statistics for medical research.* London: 1999.
- (7) Jager KJ, van Dijk PC, Zoccali C, Dekker FW. The analysis of survival data: the Kaplan-Meier method. *Kidney Int* 2008 Sep;74(5):560-5.
- (8) Tripepi G, Jager KJ, Dekker FW, Zoccali C. Linear and logistic regression analysis. *Kidney Int* 2008 Apr;73(7):806-10.
- (9) Harrell JrFE. *Missing Data. Regression Modeling Strategies - with applications to linear models, logistic regression, and survival analysis.* Springer-Verlag New York, Inc.; 2001. p. 41-52.
- (10) Frederiksen BL, Osler M, Harling H, Ladelund S, Jorgensen T. The impact of socioeconomic factors on 30-day mortality following elective colorectal cancer surgery: A nationwide study. *Eur J Cancer* 2009 Jan 9;45(7):1248-56.
- (11) Ranta S, Hynynen M, Tammisto T. A survey of the ASA physical status classification: significant variation in allocation among Finnish anaesthesiologists. *Acta Anaesthesiol Scand* 1997 May;41(5):629-32.
- (12) Mak PH, Campbell RC, Irwin MG. The ASA Physical Status Classification: inter-observer consistency. *American Society of Anesthesiologists. Anaesth Intensive Care* 2002 Oct;30(5):633-40.
- (13) Jensen AR, Storm HH, Moller S, Overgaard J. Validity and representativity in the Danish Breast Cancer Cooperative Group--a study on protocol allocation and data validity from one county to a multi-centre database. *Acta Oncol* 2003;42(3):179-85.

- (14) Nickelsen TN, Harling H, Kronborg O, Bulow S, Jorgensen T. [The completeness and quality of the Danish Colorectal Cancer clinical database on colorectal cancer]. *Ugeskr Laeger* 2004 Aug 30;166(36):3092-5.
- (15) Gaist D, Sorensen HT, Hallas J. The Danish prescription registries. *Dan Med Bull* 1997 Sep;44(4):445-8.
- (16) Sørensen HT. Regional administrative health registries as a resource in clinical epidemiology. *International Journal of Risk & Safety in Medicine* 1997;(10):1-22.
- (17) Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C. Bias in clinical research. *Kidney Int* 2008 Jan;73(2):148-53.
- (18) Dalton SO, Doring M, Ross L, Carlsen K, Mortensen PB, Lynch J, et al. The relation between socioeconomic and demographic factors and tumour stage in women diagnosed with breast cancer in Denmark, 1983-1999. *Br J Cancer* 2006 Aug 8;95(5):653-9.
- (19) Frederiksen BL, Osler M, Harling H, Jorgensen T. Social inequalities in stage at diagnosis of rectal but not in colonic cancer: a nationwide study. *Br J Cancer* 2008 Feb 12;98(3):668-73.
- (20) Goldacre MJ. Cause-specific mortality: understanding uncertain tips of the disease iceberg. *J Epidemiol Community Health* 1993 Dec;47(6):491-6.
- (21) Percy C, Stanek E, III, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *Am J Public Health* 1981 Mar;71(3):242-50.
- (22) van der Laan MJ, Petersen ML. Estimation of direct and indirect causal effects in longitudinal studies. Technical report, University of California, Berkeley, Division of Biostatistics August 23, 2004 2008.
- (23) Dalton SO, Steding-Jessen M, Gislum M, Frederiksen K, Engholm G, Schuz J. Social inequality and incidence of and survival from cancer in a population-based study in Denmark, 1994-2003: Background, aims, material and methods. *Eur J Cancer* 2008 Sep;44(14):1938-49.
- (24) Dickman PW, Auvinen A, Voutilainen ET, Hakulinen T. Measuring social class differences in cancer patient survival: is it necessary to control for social class differences in general population mortality? A Finnish population-based study. *J Epidemiol Community Health* 1998 Nov;52(11):727-34.
- (25) Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Stat Med* 2004 Jan 15;23(1):51-64.
- (26) Hjerl K, Andersen EW, Keiding N, Mouridsen HT, Mortensen PB, Jorgensen T. Depression as a prognostic factor for breast cancer mortality. *Psychosomatics* 2003 Jan;44(1):24-30.
- (27) Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373-83.

- (28) de G, V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol* 2003 Mar;56(3):221-9.
- (29) Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey SG. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health* 2006 Jan;60(1):7-12.
- (30) Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey SG. Indicators of socioeconomic position (part 2). *J Epidemiol Community Health* 2006 Feb;60(2):95-101.
- (31) Glymour MM, Greenland S. Causal Diagrams. In: Rothmann KJ, Greenland S, Lash TL, editors. *Modern Epidemiology*. Third ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 183-212.
- (32) Frederiksen BL, Osler M, Harling H, Ladelund S, Jorgensen T. Do patient characteristics, disease, or treatment explain social inequality in survival from colorectal cancer? *Soc Sci Med* 2009 Aug 18;69(7):1107-15.