



BACHELOROPPGAVE, RAD 3911

**Forfattere:
Alexander Bergset og Rannveig Dagestad**

**Effekten av
mammografiscreening
- Faglige uenigheter**

Høgskolen i Gjøvik
Avdeling for helse, omsorg og sykepleie
Seksjon for helse, teknologi og samfunn
6. semester

Mai 2012

Antall ord: 9578

Forord

Dette er vår avsluttende oppgave ved radiografutdanningen ved Høgskolen i Gjøvik, som ble ferdigstilt i mai 2012. Oppgaven er i hovedsak rettet mot radiografer og annet interessert fagpersonell. Arbeidet med oppgaven har vært lærerikt og ga oss god innsikt i temaet.

Vi vil rette en stor takk til vår veileder Bjørn Morten Hofmann for god veiledning og konstruktiv tilbakemelding under hele arbeidet.

Gjøvik, våren 2012

Alexander Bergset og Rannveig Dagestad

Sammendrag

Tittel:	Effekten av mammografiscreening – faglige uenigheter	Dato : 16/05-12
Deltaker(e)/	Alexander Derås Bergset Rannveig Dagestad	
Veileder(e):	Bjørn Morten Hofmann	
Stikkord/nøkkel ord:	Mammografiscreening, redusert brystkreftdødelighet, faglige uenigheter	
Antall sider/ord:	Antall vedlegg: 1	Publiseringsavtale inngått: ja/nei
Kort beskrivelse av bacheloroppgaven:		
Problemstilling:		
Redusert brystkreftdødelighet som følge av mammografiscreening av kvinner: Hva er de faglige uenighetene?		
Hensikt:		
Belyse de faglige uenighetene omkring effekten av mammografiscreening. Hvordan kan forskerne komme til så ulike resultat? Reduksjonen i dødelighet varierer fra opp mot 50 % til liten eller ingen effekt.		
Metode:		
Vi benyttet en kvalitativ tilnærming og foretok en litteraturstudie. Der vi analyserte metodene i de ulike studiene. Studiene inkludert var alle originalstudier som benyttet ulik fremgangsmåte		
Resultat:		
Hvilken metode som er benyttet og hvilke variabler som er inkludert i studiene har svært mye å si for resultatene. De største uenighetene synes å ha to hovedkilder. Dette var lengden på oppfølgingstid og hvilke aldersgrupper som var inkludert i studiene.		
Konklusjon:		
Dette er et veldig komplekst tema der det er vanskelig å komme til enighet, spesielt så lenge metodene varierer i så stor grad.		

Abstract

Title:	The effect of mammography screening Scientific disagreements	Date : 16/5-12
Participants/	Alexander Derås Bergset Rannveig Dagestad	
Supervisor(s)	Bjørn Morten Hofmann	
Keywords (3-5)	Mammography screening, reduced mortality, scientific disagreements	
Number of pages/words:	Number of appendix: 1	Availability (open/confidential):
Short description of the bachelor thesis:		
Topic/Research question:		
Reduced breast cancer mortality among women participating in mammography screening. What are the scientific disagreements?		
Purpose:		
Shed light on the professional disagreements around the national mammography screening program. How can the results from the various studies vary by such large extent, from 25-50% till almost nothing?		
Method:		
We took a qualitative approach and underwent a literature review. Analyzing the methods used by the authors of the different studies. The main studies included were all original studies with different methods.		
Results:		
The results from the various studies depend strongly on what kind of method used and which factors included in the studies. The two main disagreements seem to originate from the amount of follow-up time used and which age groups that was included in the study.		
Conclusion		
This is a very complex topic and it is hard to reach consensus, especially when the factors included in each study are so different		

Innhold

Sammendrag.....	3
Abstract	4
1.0 Innledning.....	8
2.0 Metode.....	10
2.1 Vurdering av litteratur.....	10
2.1.1 Inkluderingskriterier.....	11
2.2 Søk.....	12
2.3 Litteratur.....	13
2.4 Metodekritikk.....	14
2.4.1 Litteraturstudie som metode.....	14
3.0 Teori	16
3.1 Bakgrunn	16
3.2 Absolutt og relativ risiko.....	17
3.3 Beslutninger under stor usikkerhet.....	18
3.3.1 Usikkerhet i forhold til konsekvensene av de ulike alternativene (2).....	18
3.3.2 Usikkerhet i forhold til hvilken informasjon en kan stole på (3).....	19
3.4 Bias.....	19
3.4.1 Self selection bias.....	20
3.4.2 Lead time bias	21
3.4.3 Lenght bias	22

3.5 Confounding.....	23
3.6 Studiedesign	23
4.0 Resultat.....	24
4.1 Hva er man uenig om metodisk?.....	27
4.1.1 Oppfølgingstid.....	27
4.1.2 Utvalg	27
4.1.3 Screeningdeltagelse	28
4.1.4 Villscreening	28
4.1.5 Bias.....	29
4.1.5.1 Lead time bias	29
4.1.5.2 Self selection bias.....	29
4.1.5.3 Confounding.....	30
4.2 Hva er man uenig om når det gjelder effekten av screening og tolkning av resultater?	30
4.3 Hva er man uenig om når det gjelder om screeningkriteriene (WHO) er oppfylt?.....	31
5.0 Diskusjon.....	32
5.1 Hva er man uenig om metodisk?.....	32
5.1.1 Oppfølgingstid.....	32
5.1.2 Utvalg	33
5.1.3 Screeningdeltagelse	34
5.1.3.1 Self selection	35
5.1.3.2 Length time bias	36

5.1.4 Villscreening	36
5.1.5 Studiedesign	37
5.2 Hva er man uenig om når det gjelder effekten av screening og tolkning av resultater?	37
5.2.1 Tolkning av resultater.....	37
5.2.1.1 Interessekonflikt	38
5.3 Konsekvenser for praksis	38
5.4 Kritikk	39
5.4.1 Søk.....	39
5.4.2 Reliabilitet og validitet	40
6.0 Konklusjon	41
7.0 Litteratur.....	42
Vedlegg 1: Sammendrag av de 6 hovedartiklene.....	45

Antall ord: 9578

1.0 Innledning

Vurdering av effekten av mammografiscreening har ført til en enorm faglig debatt. De store studiene spriker sterkt om hvorvidt screening reduserer brystkreftdødeligheten eller ikke, og i hvilken grad.

Det er av naturlige grunner meget vanskelig å orientere seg i et så bredt felt som mammografiscreening har utviklet seg til å bli. Til tross for ca. 50 års forskning, fra de første, store randomiserte forsøkene, og fram til dagens mange evalueringer av de offentlige screeningprogrammene i den vestlige verden, er den vitenskapelige usikkerheten større enn noen gang. I diverse fagtidsskrifter ser en etter artikler på emnet en engasjert, og til tider opphetet debatt omkring mammografiscreening generelt, metode og de enkelte funn. Det er ikke til å unngå å observere at meningene omkring emnet er delt, noen vil si polarisert. Denne debatten sees ikke kun i vitenskapelige kretser; også i nasjonal og internasjonal media er det stadig debatter, artikler og innlegg i diverse fora som kanskje gjør debatten ytterligere polarisert, da massemedia ofte fremstiller forskning på en helt annen måte enn tradisjonen er i vitenskapelige tidsskrifter.

I forkant av oppgaven gjorde vi søk etter lignende bacheloroppgaver. Etter det vi kjenner til er det ikke publisert bacheloroppgaver som tar sikte på å analysere debatten omkring mammografiscreening.

I denne oppgaven ønsker vi å se nærmere på den faglige debatten som gjør mammografiscreening til et så kontroversielt tema som det er i dag. Vi har valgt å utføre en litteraturstudie hvor vi tar utgangspunkt i et utvalg studier utført av, det vi mener, er noen av de toneangivende aktørene i den pågående debatten, spesielt sett fra et norsk perspektiv. Vi ønsker å finne ut hvorfor det er så vanskelig å bli enig om effekten av mammografiscreening på brystkreftdødelighet, og sammenfatte dette slik at de som er interessert i temaet, men ikke ønsker å fordype seg i forskningen, kan danne seg et bilde av hva de faglige uenighetene i debatten er. Vi mener dette er noe mange kan ha nytte av, og det er spesielt radiograffaglig relevant for de radiografer som daglig jobber med mammografi, i tillegg til radiografer og helsepersonell generelt som ønsker en bredere forståelse på et vanskelig område.

Med utgangspunkt i dette har vi valgt følgende problemstilling: ”Redusert brystkreftdødelighet som følge av mammografiscreening av kvinner: Hva er de faglige uenighetene?”

2.0 Metode

For å besvare problemstillingen og kartlegge de faglige uenighetene fant vi det hensiktsmessig å utføre en litteraturstudie med et egnet utvalg av studier som representerer to syn på saken; de som viser en vesentlig reduksjon i dødelighet og de som viser en liten eller ingen reduksjon.

Det finnes en stor mengde studier utført med det mål å analysere effekten av brystscreening på dødelighet som følge av brystkreft. Vi var opptatt av å få et utvalg som representerte de ulike synene som gjør det mulig å presentere og klargjøre uenighetene. Herunder er det ønskelig med studier hvor resultatene viser en reduksjon i dødelighet som følge av brystkreft, og studier som viser ingen eller liten utvikling i dødelighet.

Som sagt finnes det et stort antall studier gjort på effekten av mammografiscreening, noe som gir utfordringer ved begrensinger i utvalget. Etter å ha gått i gjennom et meget stort materiale, som et resultat av våre databasesøk, valgte vi å rette vårt fokus på studier utført av dem vi gjennom vårt arbeid oppfattet som de toneangivende aktørene i den pågående debatten. Dette anså vi som forfattere som publiseres mye og refereres til i vitenskapelige artikler. Dette så vi som hensiktsmessig på flere måter, da resultater og meninger gjennomgående er sammenlignbare innenfor de forskjellige sidene av debatten. Ofte fant vi også at de mest toneangivende henviser mye til annen forskning, noe som gir et tverrsnittsbilde av de to siders oppfatning. Disse forfatterne henvises også til massemedia, noe som bidrar til synligheten i den offentlige debatten.

2.1 Vurdering av litteratur

I denne studien, hvor dataene samles ved hjelp av databasesøk, er det viktig å ha en systematisk fremgangsmåte av flere grunner. Først og fremst er det for vår egen del viktig å dokumentere søkene og vurdere litteraturene opp mot de kriteriene som ligger til grunn, men det er også viktig for utenforstående, da våre funn skal kunne etterprøves og metoden vurderes.

For å sikre kvaliteten til studiene som ble inkludert har vi fulgt Cochrane collaborations retningslinjer. Herunder finner vi definisjon av forskningsspørsmål og inkluderingskriterier

for studiene, vurdering av risiko for bias i de inkluderte studiene, dataanalyse, rapportere bias, presentasjon av resultat, funn, tabeller og tolkning av resultater og konklusjon. Det er en streng protokoll som krever nøyaktighet i arbeidet, at man forholder seg kritisk til stoffet man behandler, at alt undersøkes og at det foregår over en lenger periode (Aveyard 2010)

Vårt søk var noe begrenset med tanke på de tidligere nevnte retningslinjene, da vi ikke fullt ut har fulgt disse. Blant annet er søket vårt begrenset noe i tid, grunnet tidsperspektivet på oppgaven, og søkene ble kun gjort i to databaser. Dette for å begrense datamengden vi var nødt til å gjennomgå, men også fordi vi mente datagrunnlaget vi fikk gjennom disse to databasene var godt nok for oppgavens omfang.

2.1.1 Inkluderingskriterier

Det ble gjort en systematisk gjennomgang av alle treffene fra søkene gjort i databasene. Artikler ble først valgt ut på grunnlag av tittel, deretter en ny runde ved gjennomlesing av abstrakt, og til slutt en gjennomlesing av de artikler som fortsatt var del av utvalget. Dette vises i figur 3. De artiklene vi mente var mest relevante til vår oppgave ble så valgt ut. Som et resultat av de systematiske søkene ble det også plukket ut artikler fra usystematiske søk. I dette legger vi blant annet artikler valgt ut fra referanser i artikler fra de systematiske søkene og artikler referert til i de faglige diskusjonene som følge av artiklene fra de systematiske søkene.

I vår oppgave var det viktig å få en oversikt over de forskjellige elementene i debatten, vi valgte derfor kun å inkludere originalartikler som hovedlitteratur. Studiene i de artiklene som er valgt bruker forskjellige metoder i sin forskning, noe vi hadde som et inkluderingskriterium. Vi ønsket også, i og med at utvalget måtte begrenses, å bruke artikler utgitt av dem vi anser som de mest toneangivende i debatten omkring mammografiscreening.

Vi valgte da vi gjorde søkene å begrense det tidsmessig fra 1.1.2007 – 9.3.2012. Ny forskning på området blir publisert fortløpende, og utviklingen på feltet skjer raskt. Vi ønsket derfor å begrense søket til de siste fem årene for å unngå forskning som kan ha mistet noe av sin relevans. Valget om å tidsbegrense til de siste fem år ble også gjort fordi søk gjort på et bredere tidsspekter resulterte i et meget stort antall treff, noe som, oppgavens omfang tatt i

betraktning, ville vært meget vanskelig å håndtere. Det at vi utførte usystematisk søk på bakgrunn av de systematiske har allikevel ført til at vi har inkludert én artikkel utgitt før 2007.

Alle artiklene som er inkludert i oppgaven er fagfellevurderte og de respektive tidsskriftene er sjekket opp mot database for statistikk og høgere utdannings (DBH) register over autoriserte publiseringskanaler, hvor alle tidsskrifter var oppført som nivå 1 eller 2-tidsskrifter.

2.2 Søk

Søkeord	Begrensninger	Database	Kombinasjoner	Antall treff
1. Mortality	All fields (no full text), peer reviewed 2007 – today	Proquest	1+3+6+7	76
2. Mortality rate			6+7+4+1	233
3. Effect			6+7+5+1	105
4. Breast cancer			6+2+3	45
5. Breast neoplasms			6+2	133
6. Mammography				
7. Screening	Abstract, title, keywords 2007 - today	Science Direct	6+7+4+1	69
8. Mass screening			6+3	145

Tabell 1: Oversikt over databasesøkene utført i perioden 16.1 - 9.3

Som tabellen over viser hadde vi en klar strategi for databasesøkene vi utførte i forbindelse med arbeidet på denne oppgaven. Vi definerte før databasesøkene ble gjort åtte søkeord/fraser som ble utarbeidet i samarbeid med bibliotekar på høgskolen i Gjøvik. Som en kan se ble ikke alle søkeordene brukt, da søkeresultatene var gode og ga et godt utvalg etter de første søkene.

En del av søkeordene listet opp fungerte som alternativer, dersom de andre ikke skulle gi tilfredsstillende resultater.

Man ser også at ikke alle søkene ble utført i begge databasene. I utgangspunktet hadde vi et godt utvalg artikler allerede etter søk i Proquest, men vi ønsket å gjøre søk i en alternativ database for å sikre at de første søkene var relevante for vår problemstilling, og for å se etter eventuelle andre funn. Vi valgte å begrense antall søk i Science Direct, da vi etter to søk hadde et godt utvalg av artikler og et godt grunnlag for videre usystematiske søk.

Det ble foretatt en innholdsanalyse av hver enkelt artikkel for å se etter argumenter, og for å besvare forskningsspørsmålene. Analysen gikk ut på å sammenligne de ulike studiene med fokus på metode og diskusjonsdelen. Relevant informasjon ble gjenkjent, plukket ut og sammenfattet.

For å få svar på vår problemstilling gikk vi ut i fra tre forskningsspørsmål når vi skulle analysere artiklene.

1. Hva er man uenig om metodisk?
2. Hva er man uenig om når det gjelder effekten av screening og tolkning av resultater?
3. Hva er man uenig om når det gjelder om screeningkriteriene (WHO) er oppfylt?

2.3 Litteratur

Her vil kort valg av litteratur og nettressurser brukt i oppgave gjøres rede for. Valg av kilder er fattet på bakgrunn av relevans og at vi på bakgrunn av informasjon om kildene finner de troverdige. Vi anser litteraturen vi har brukt som troverdig og anerkjent.

Av internettkilder har vi blant annet brukt Forskningsetisk bibliotek. Dette er en undervisningsressurs som har som formål å bidra til refleksjon og debatt omkring forskningsetiske spørsmål. Siden er skrevet av fagfolk og er et resultat av et prosjekt ledet av De nasjonale forskningsetiske komiteer i samarbeid med landets syv universiteter. Artiklene på nettstedet fungerer som en introduksjon til viktige forskningsetiske temaer. Vi har også

brukt Store norske leksikon på nett. Dette er en godkjent kilde ved skoler og universiteter, og artiklene er skrevet av forfattere med kompetanse på emnet.

Vi har også brukt et utvalg bøker vi mener har relevans for vår oppgave. Avyard et. al (2010) ble brukt som litteratur angående metode, etter anbefaling fra biblioteket på Høgskolen i Gjøvik. Denne boka er strukturert og konkret, og ga en god innføring i fremgangsmåten for en litteraturstudie. I tillegg har vi brukt Dalland (2007) som er en anerkjent forfatter innenfor metode, og som står i ressurslisten til radiografiutdanningen og pensumlisten til sykepleieutdanningene ved høgskolen i Gjøvik

Ved siden av bøker og internettkilder har vi brukt rapporter fra Helsedirektoratet og International Agency for Research on Cancer (IARC). Helsedirektoratet er et fagdirektorat og myndighetsorgan organisert under Helse- og omsorgsdepartementet, og IARC er en datterorganisasjon under World Health Organization (WHO) som har som hovedmål å fremme et internasjonalt samarbeid innenfor kreftforskning. Vi anser begge disse kildene som pålitelige.

I tillegg til overnevnte litteratur, og hovedartiklene, er det brukt en rekke vitenskapelige artikler i oppgaven. Disse er hentet fra de opprinnelige databasesøkene og usystematiske søk. Alle disse artiklene er fra vitenskapelige tidsskrifter sjekket i DBHs oversikt over publiseringskanaler.

2.4 Metodekritikk

2.4.1 Litteraturstudie som metode

En svak side ved litteraturstudium er muligheten for feiltolking av stoffet. Sekundærkilde vil si at noen andre enn forfatteren selv for eksempel presenterer en tekst, mens en primærkilde vil si at den opprinnelige forfatteren presenterer sin egen tekst. For å sikre en mest mulig korrekt fremstilling som mulig, er primærkilder alltid å foretrekke (Dalland 2007). På bakgrunn av dette og for å begrense mulighet for feiltolkninger i vår oppgave valgte vi hovedsakelig å holde oss til primærkilder. Man må også være oppmerksom på ikke å bli for opphengt i forfatterens synspunkter og meninger slik at en overser viktige poeng. Samtidig som det er viktig å plukke ut relevant stoff i forhold til problemstillingen. Ved en kvalitativ

tilnærming kan forforståelsen i stor grad påvirke søk og tolking av litteratur, og det kan være vanskelig å forholde seg objektiv. Forforståelsen ble i størst mulig grad lagt til side for ikke å påvirke oppgaven, men vi ser at dette var vanskelig å unngå.

3.0 Teori

Her presenteres teori som er relevant for å forstå oppgaven som helhet.

3.1 Bakgrunn

Forventningene til mammografiscreening har siden de første, store randomiserte forsøkene utført fra 60- til 90-tallet vært høye. IARC hevdet i en rapport fra 2008 at mammografiscreening er den eneste screeningmetoden som er beviselig effektiv, og at det hos kvinner over 50 år i høyinntektsland, ved en deltagelse på over 70 %, kan redusere dødeligheten som følge av brystkreft med 20-30 % (IARC 2008).

Kanskje er mammografiscreening blant det som er gjort mest forskning på i senere tid. Det er derfor et paradoks at det ikke er konsensus i forskningskretser for hvorvidt mammografiscreening har etterlevd resultatene man opprinnelig så i de randomiserte forsøkene. Samtidig er det naturlig at det forskes mye på området, da usikkerheten er stor med tanke på flere aspekter forbundet med mammografiscreening.

I Norge har Helse- og omsorgsdepartementet gitt Forskningsrådet i oppgave å koordinere en evaluering av det norske screeningprogrammet. Helsedirektoratet ga så sent som 22. mars i år ut rapporten "Prioriteringer i helsesektoren". Her skriver de at "det er meget viktig at evalueringen blir gjennomført slik at en kan vurdere hvorvidt den offentlige helsetjenesten i Norge fortsatt skal tilby mammografiscreening til kvinner i alderen 50-69 år" (Helsedirektoratet 2012).

En Cochrane-oversikt fra 2011 konkluderte med noe av det samme. I denne fant de en redusert dødelighet som følge av brystkreft, men samtidig er det uklart om screening gjør mer nytte enn skade, og at det derfor bør igangsettes evalueringer utført av uavhengige forskere (Gøtzsche & Nielsen 2011). Andre grupper finner på sin side resultater som samsvarer med, og overgår, de fra de randomiserte forsøkene, med redusert dødelighet opp mot 50 % (Schopper & de Wolf 2009; Duffy et al. 2010).

3.2 Absolutt og relativ risiko

Hvordan man presenterer data fra en studie kan være med på å forsterke eller bagatellisere et resultat. Når det kommer til studier på effekten på brystkreftdødelighet som følge av screening, og i stort sett alle kliniske studier, er begrepene absolutt og relativ risiko sentrale og viktige i framstillingen. Å forstå disse begrepene er essensielt for å forstå meningen bak tallene.

Relativ risiko (RR) representerer høyere eller lavere risiko for et utfall i én gruppe sammenlignet med en annen. Beregningen gjøres ved å dele prosentandelen av utfallet i én gruppe (interessegruppen) på prosentandelen i den andre gruppen (kontrollgruppen). Om resultatet blir 1.0 vil det si at risikoen i de to gruppene er lik. Om RR er 3.0, vil det si at risikoen er tre ganger høyere i interessegruppen enn i kontrollgruppen. Er RR 0,5 vil det si en halvert risiko.

Statistisk fremstilles RR oftest med et konfidensintervall (CI), for å kunne vurdere hvorvidt resultatet er signifikant.

RR kan kun regnes ut for grupper som følges over tid. Dette kan gjøres prospektivt, som er den mest brukte metoden i kliniske studier, eller i retrospekt, for eksempel kasus-kontrollstudier (Citrome 2010).

Absolutt risiko vil på sin side si sannsynligheten for en risikofaktor i en befolkning. Absolutt risiko angis som antall tilfeller delt på antall personer i en gitt undersøkelse. Denne enheten vil oftest gi den mest relevante informasjonen om risiko til den enkelte pasient, og bør derfor oppgis sammen med RR (Store medisinske leksikon 2011).

For eksempel kan man si at gjennomsnittskvinnen har 2 % sjanse for å dø av brystkreft. En kvinne som deltar i et screeningprogram vil på sin side ha en RR på 0,5 (50 %) i forhold til gjennomsnittskvinnen. Den absolute risikoen for å dø av brystkreft for en kvinne som deltar i et screeningprogram vil derfor være på 1 %. (Tallene er kun eksempler).

3.3 Beslutninger under stor usikkerhet

Å ta avgjørelser ved usikkerhet er ikke uvanlig verken i det daglige eller innenfor vitenskapen. I praksis er alle beslutninger tatt under en viss grad av usikkerhet. Med stor usikkerhet menes en situasjon hvor alle alternativer leder til en usikkerhet i forhold til eventuelle konsekvenser.

Usikkerhet kan deles inn i fire forskjellige komponenter:

1. Usikkerhet i forhold til hvilke alternativer en har å forholde seg til
 2. Usikkerhet i forhold til konsekvensene av de ulike alternativene
 3. Usikkerhet i forhold til hvilken informasjon en kan stole på
 4. Usikkerhet i forhold til verdier/personlige meninger hos beslutningstakeren(e)
- (Hansson 1996)

Under vil komponent to og tre presenteres videre med henblikk på relevans til vår oppgave.

3.3.1 Usikkerhet i forhold til konsekvensene av de ulike alternativene (2)

Usikkerhet forbundet med konsekvenser er den vanligste usikkerheten forbundet med beslutningstaking (Hansson 1996). Problemet med konsekvenser er at en beslutningstaker ikke alltid kan ta hensyn til dem, da det ofte er ekstremt vanskelig å konsekvensutrede og forutsi virkningen av det ene eller andre utfallet. I noen tilfeller vil det å sette konsekvensene av de forskjellige utfallene opp mot hverandre være til hjelp, andre ganger kan det motsatte være fordelmessig (Hansson 1996). For eksempel kan man si at å legge ned mammografiprogrammet er forbundet med mye større grad av usikkerhet enn ved å fortsette programmet slik det nå er.

Slik asymmetri i usikkerhet kan være til hjelp i en beslutningsprosess, men det er viktig å påpeke at det alene ikke kan være avgjørende.

3.3.2 Usikkerhet i forhold til hvilken informasjon en kan stole på (3)

En del av kunnskapen og informasjonen som trengs i en beslutningsprosess er tilgjengelig for allmennheten, men dette krever kunnskaper om hvor en skal lete. Full oversikt er på sin side nærmest fullstendig utilgjengelig for de fleste, men krever inngående teknisk kunnskap innenfor flere spesialområder. Dette gjelder også innenfor mammografiscreening, hvor kunnskap om epidemiologi, strålefysikk/biologi, sykdomsforløp etc. er viktig for å danne seg et fullstendig bilde. De som tar beslutninger må derfor sette sin lit til såkalte eksperter på området, noe som tilfører en annen dimensjon til usikkerheten; hvilke(n) ekspert(er)/informasjon skal en sette sin lit til (Hansson 1996).

Man kan skille mellom tre grader av usikkerhet til informasjon. Den første består i at en vet sikkert hvem som er ekspertene på området, men disse ekspertene er uenig. Den andre består i at vi ikke vet hvem som er de mest troverdige ekspertene på området, mens den siste går ut på om vi i det hele tatt skal stole på eksperter og hva de mener om saken (Hansson 1996).

Alle de tre gradene kan til en viss grad trekkes paralleller til innenfor debatten omkring mammografiscreening. Nummer en, i at debatten er polarisert omkring viktige spørsmål, nummer to, at det er vanskelig å vite hvilken av partene en skal stole på, og nummer tre, at en burde se bort i fra ekspertenes syn på saken og engasjere en tredjepart til å ta evaluere mammografiscreening. Alle tre er spørsmål som ofte stilles i forbindelse med mammografiscreening og kontroversene rundt.

Hva som gjør usikkerheten i forhold til mammografiscreening kanskje enda mer utfordrende er at det er snakk om eventuelt å avslutte noe som allerede er godt etablert i mange land, samtidig vet man ikke sikkert hva effekten i dag er, og derfor blir det ytterligere vanskelig, eller umulig, å konsekvensutrede for en eventuell nedleggelse eller fortsettelse.

3.4 Bias

Bias kommer av det franske ordet *biais* som betyr ”på snei”, på norsk oversettes det ofte til skjevhet. Ordet betyr ingenting alene, men må kobles sammen med et annet ord for å gi

mening. Bias er et uttrykk som brukes mye i forskning, og viser til systematiske feil i innsamling og tolkning av data. Forskjellige bias i metoden kan føre til at resultatene fra forskningen ikke samsvarer med virkeligheten. Bias fører til utfordringer i enhver forskning (Staff 2010).

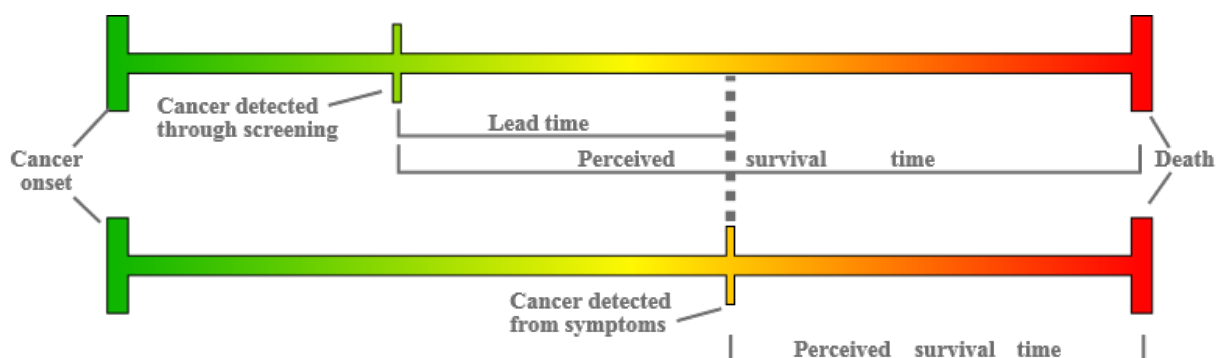
Det finnes mange bias innenfor forskning. I forskningen på effekten av mammografiscreening er det noen spesifikke bias som er spesielt viktig å vite betydningen av for å forstå hvilke utfordringer forskerne står overfor; self selection bias, lead time bias og lenght bias. Vi velger gjennomgående å bruke de engelske termene, da vi mener dette er mer dekkende enn å forsøke å oversette.

3.4.1 Self selection bias

Selection bias kan enkelt oversettes som utvalgsskjevhet. Valg av studiepopulasjon og studiedeltagelsen er avhengig av en rekke faktorer og kan potensielt ha mye å si for resultatene.

I mammografiscreening viser self selection bias til fenomenet at kvinner som velger å delta i screeningprogrammet har andre forutsetninger enn de som takker nei. For eksempel kan kvinner som screenes generelt ha en lavere risiko for brystkreft fordi de trener mer, spiser sunnere, er høyere utdannet etc., noe som vil gi et resultat som styrker effekten av screening, men som i virkeligheten ikke nødvendigvis er sannheten. Også kjent som "healthy screenee effect". En kan også si at self selection bias oppstår på grunn av at resultatene fra grupper som frivillig deltar i undersøkelser kan variere fra grupper som ikke takker ja til slike undersøkelser (Staff 2010).

3.4.2 Lead time bias

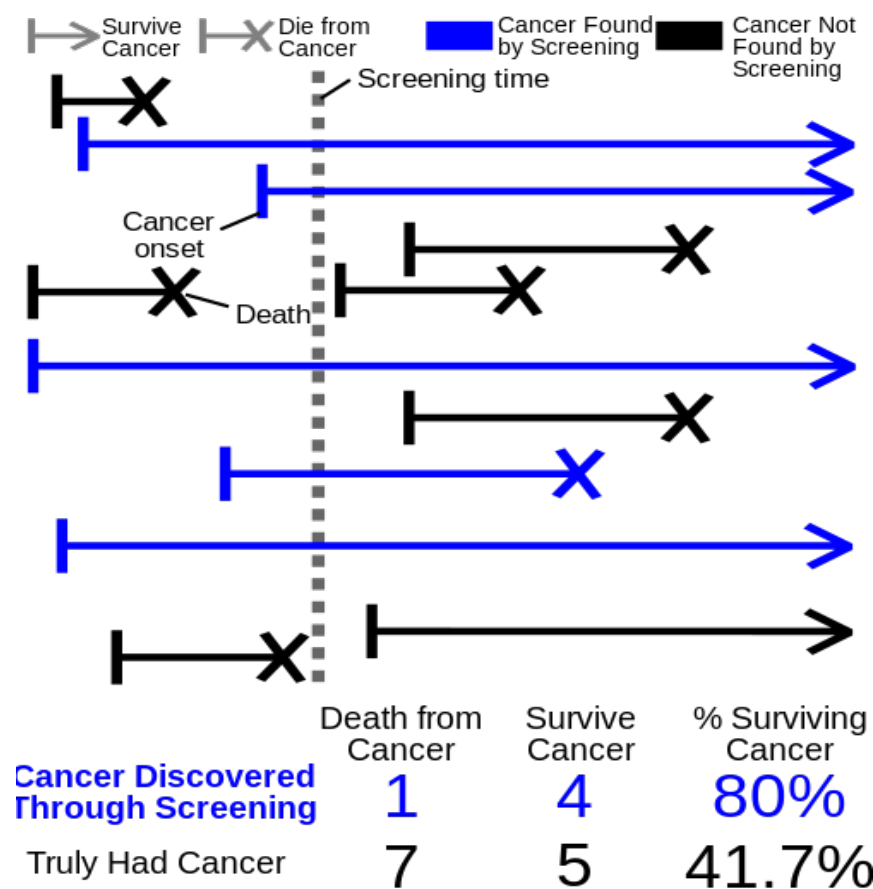


Figur 1: Illustrerer lead time bias, og hvordan den persiperte overlevelsestiden øker uten å endre sykdomsforløpet (Mcstrother 2011a)

Lead time er tidsperioden mellom oppdagelse av sykdom (med screening) og dens vanlige kliniske forløp og diagnose.

Lead time bias oppstår når to diagnostiske metoder for en sykdom blir sammenlignet og den ene metoden (ofte den nye eksperimentelle) diagnostiserer sykdommen før, men ikke har noen effekt på utfallet av sykdommen. Det fremstår da som om denne metoden forlenger overlevelsestiden, men det som egentlig skjer er at pasienten lever like lenge, men får diagnosen på et tidligere tidspunkt. For å oppnå valide sammenligningsgrupper må lead time bias trekkes i fra, noe som har vist seg vanskelig da man ikke har sikre tall på dette (Mahnken et al. 2008).

3.4.3 Length bias



Figur 2: Figuren viser hvordan length bias gir skjevhet i resultatet og dermed overestimerer effekten av screening (Mcstrother 2011b).

Length bias er en potensiell bias som oppstår ved at sakte voksende tumorer har like stor sannsynlighet for å bli oppdaget med mammografiscreening som raskt voksende tumorer (Mahnken et al. 2008). Dette kan bidra til at kreft som uten screening ikke ville manifestert klinisk i kvinnens levetid, diagnostiseres. Dette fører til overdiagnostisering, og videre en overvurdering av effekten av mammografiscreening.

3.5 Confounding

Confounding er en vanlig feilkilde i helsefaglig forskning. Det kan minne om, men er ikke det samme som bias. Mens bias involverer feil i måling av variabler er confounding feil i tolkning av hva som er en nøyaktig måling. Dette fører til ukorrekte estimater for assosiasjoner. Det er tilfeldige feil og ikke systematiske. Økning i studiepopulasjon vil redusere tilfeldige feil, i motsetning til systematiske feil som ikke kan reduseres for ved økt studiepopulasjon (Staff 2010).

3.6 Studiedesign

Randomiserte kontrollerte forsøk er en form for kliniske forsøk eller vitenskaplige prosedyrer brukt for å vurdere effektiviteten av en behandling eller medisin. Deltakerne er tilfeldig fordelt i tiltaks- og kontrollgruppe. Dersom deltakerne selv skulle fått velge grupper er det sannsynlig at de hadde valgt den mest populære gruppa og de to gruppene hadde ikke blitt like. Randomiserte kontrollerte forsøk kan brukes for å avgjøre om en behandlingsmetode er effektiv eller ikke.

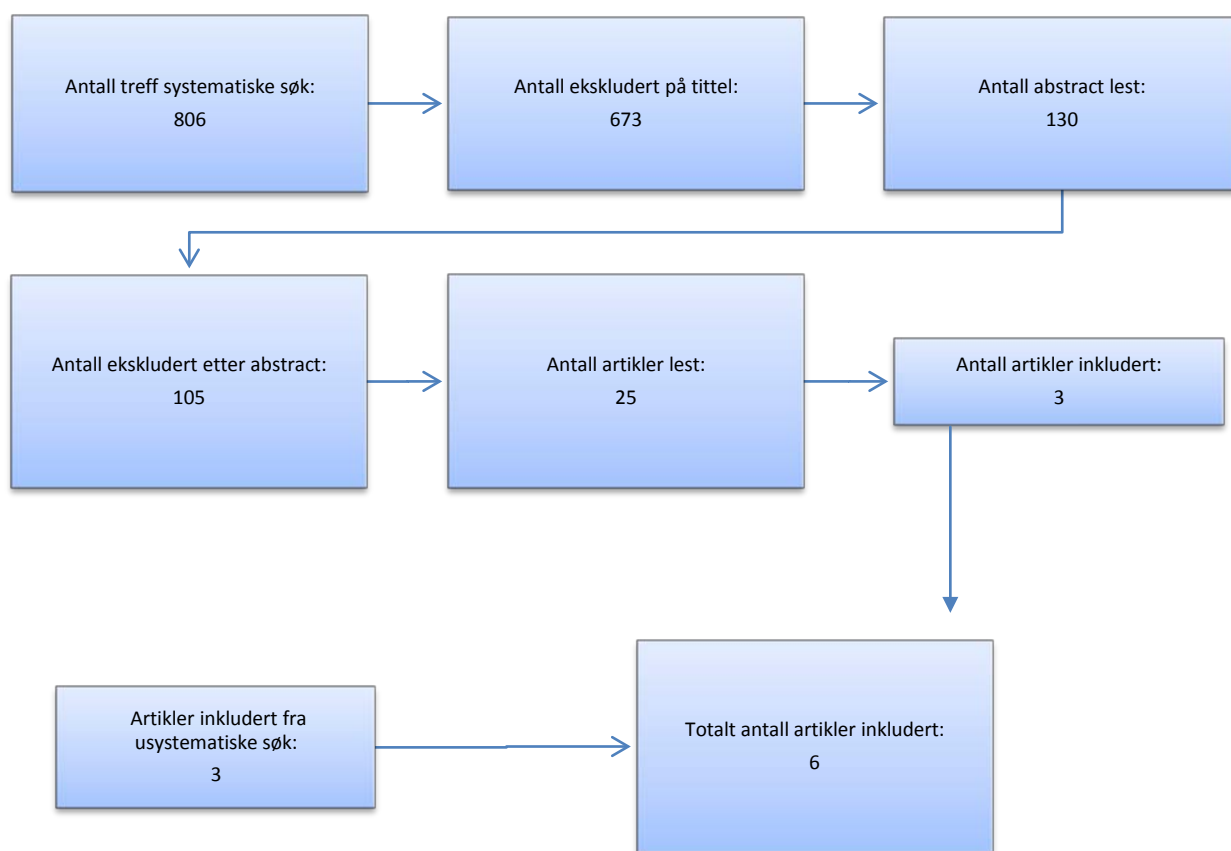
Kohorte- og kasus kontroll-studier er observasjonsstudier og blir ofte gjennomført når det ikke er mulig å gjennomføre et randomisert kontrollert forsøk. Slike observasjonsstudier kan brukes for å finne sammenhengen mellom ulike faktorer.

En kohortestudie er en studie av en gruppe mennesker der alle er utsatt for en spesiell hendelse eller livsstil, som følges opp over tid for å studere effekter.

Kasus-kontrollstudie er en studie av en gruppe som har blitt utsatt for en aktuell faktor eller har en bestemt sykdom, og som sammenlignes med en gruppe ikke har blitt utsatt for dette (Aveyard 2010).

4.0 Resultat

Her vil de ulike studien kort presanteres i en tabell, med henblikk på deres metode og resultater. Et mer utfyllende sammendrag hvor fremgangsmåten i de forskjellige studiene beskrives følger som vedlegg (vedlegg 1).



Figur 3: Fremgangsmåte for ekskludering og inkludering av artikler

Artikkel 1: Olsen, A. H.et al. (2005). Breast cancer mortality in Copenhagen after introduction of mammography screening

Artikkel 2: Allgood, P. C.et al. (2008). A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality

Artikkel 3: Duffy, S. W.et al. (2010). Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England

Artikkel 4: Kalager, M.et al. (2010). Effect of screening mammography on breast-cancer mortality in Norway

Artikkel 5: Jørgensen, K. J., P. H. Zahl og P. C. Gotzsche (2010). Breast cancer mortality in organised mammography screening in Denmark: comparative study

Artikkel 6: Autier, Philippe et al. (2011). Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database

	Design	Utvalg	Endepunkt	Tilhørighet
1 Olsen et al.	Kohorte-studie Oppfølgingstid: ca. 10 år Kun RR	40 000 kvinner: kvinner som var invitert til screening og kontrollgruppe med kvinner fra områder uten screening Aldersgruppe: 50-69år	Redusert dødelighet som følge av brystkreft på 25 %. 37 % blant de som faktisk ble screenet	
2 Allgood et al.	Kasus-kontrollstudie Oppfølgingstid: ca. 10 år Kun RR	300 kasuser, hver av dem med 2 kontroller (600) Aldersgruppe: 50-70år	Redusert dødelighet som følge av brystkreft på 30%	

3 Duffy et al.	Oppfølgingstid: 21,5 år RR og absolutt risiko	Svenske two county trial og tall fra mammografiprogrammet i England (UK). Aldersgrupper: <50, 50-69 og 70+	Screening reduserer 8,8 og 5,7 liv per 1000 screenet.	Tabár
4 Kalager et al.	Kohorte-studie Oppfølgingstid: ca. 10 år RR og absolutt risiko	4 grupper: Ikke screenet og screenet på grunnlag av ulik innførelse av mammografiprogrammet I Norge samt 2 historiske kontrollgrupper Aldersgrupper: 20-49, 50-69 og 70-84år	Reduksjon på 10% som følge av screening Samme reduksjon i aldersgruppen 70-84 år som i screeninggruppa. Størst reduksjon i den yngste aldersgruppa.	
5 Jørgensen et al.	Sammenlignende studie: Kohorte Oppfølgingstid: ca. 10 år Kun RR	København og Fyn sammenlignet med den ikke screenede populasjon. Aldersgrupper: 35-54, 55-74 og 75-84.	Størst reduksjon i den yngste aldersgruppen Ingen reduksjon som kunne tilskrives screening.	
6 Autier et al.	Økologisk studie/ geografisk sammenligning	Sammenligner 3 par naboland med ulikt tidspunkt for innføring av screening.	Liten variasjon mellom landene der screening var innført tidlig og sent. Tolket dette som at reduksjon kom av andre årsaker enn screening	

Tabell 2: Tabellen viser et kort sammendrag av hovedartiklene brukt i denne oppgaven

4.1 Hva er man uenig om metodisk?

4.1.1 Oppfølgingstid

Oppfølgingstiden varierer lite i de inkluderte studiene. Olsen et al., Allgood et al., Kalager et al. og Jørgensen et al. har alle ca. 10 års maksimal oppfølgingstid. Dette vil si observasjonstiden fra screeningsprogrammets start, og ikke medregnet den historiske kontrollperioden. Kun Duffy et al. med sine 21,5 år har en oppfølgingstid som er lengre¹. Det er ikke nødvendigvis maks oppfølgingstid som er av størst interesse, men den gjennomsnittlige. Dette er det flere av studiene som ikke informerer om, men er et tall som er interessant.

Duffy et al. mener at en for å måle effekten av mammografiscreening er nødt til å ha ”lang nok” oppfølgingstid. Kalager et al. har blitt kritisert for å bruke for kort oppfølgingstid i sin studie. I deres studie er gjennomsnittlig oppfølgingstid 2,2 år.

4.1.2 Utvalg

Duffy et al. (2010), Kalager et al. (2010), Jørgensen et al. (2010) og Autier et al. (2011) inkluderte kvinner i ulike aldersgrupper i sine studier. I tillegg til aldersgruppa 50-69(70) år hadde de med grupper, både yngre og eldre, som ikke ville dra nytte av mammografiscreening, og som derfor burde vise en annen trend i brystkreftdødelighet om mammografiscreening hadde en effekt. I alle studiene, med unntak av Duffy et al. (2010), fant de en større nedgang i aldersgruppen <50 år. I aldersgruppen over 70 år ble det observert tilsvarende nedgang som i aldersgruppen 50-69 år. Dette tolkes i studiene som at andre faktorer har bidratt mer enn mammografiscreening på reduksjon i brystkreftdødeligheten. Dog kan ingen av studiene si sikkert hvilke(n) faktor(er) som står for reduksjonen. Hofvind et al. (2011) gjorde en studie på sammenhengen mellom brystkreft-insidens og hormonterapi og mammografi i Norge. De fant en sammenheng mellom hormonterapi og brystkreftinsidens, men kunne ikke konkludere med noe.

¹ Autier et al. er ikke nevnt her, da de har et studiedesign hvor oppfølgingstid ikke er viktig på samme måte som i de andre inkluderte studiene.

4.1.3 Screeningdeltagelse

Allgood et al. og Olsen et al. har i sine studier oppgitt resultater for redusert brystkreftdødelighet for alle invitert til screening, i tillegg til kvinner invitert som faktisk ble screenet.

Kalager et al., Jørgensen et al. og Duffy et al. har i sine studier kun oppgitt redusert brystkreftdødelighet for alle kvinner invitert til mammografiscreening.

I Norge ble det parallelt med oppstart av mammografiscreening stilt krav til at det skulle være på plass multidisiplinære team, spesialisert på kreftbehandling. Dette gjorde det mulig for Kalager et al. å isolere effekten av forbedret behandling, fordi kvinner i aldersgrupper som ikke ble tilbudt screening allikevel ville dra nytte av de multidisiplinære teamene som ble implementert i fylkene hvor screening hadde startet.

4.1.4 Villscreening

Når vi her henviser til villscreening, også kjent som opportunistisk screening, sikter vi til kvinner som ikke er en del av et screeningprogram, som på eget initiativ får utført mammografi privat. Dette finnes det ingen måter å kontrollere per i dag, omfanget er derfor usikkert.

Olsen et al. hevder villscreening i Danmark er uvanlig og at dette ville hatt lite å si for resultatene. Det samme sier Jørgensen et al. i sin studie. Dersom for eksempel kvinner under 50 år benyttet seg av villscreening kunne dette vært en del av forklaringen på den store reduksjon i dødelighet i denne gruppen, som Olsen et al. fant.

Kalager et al. påpeker at noen av kvinnene i den ikke-screenede gruppen kan ha fått utført mammografi på eget initiativ, og at dette potensielt kan ha redusert effekten av mammografiscreening i studien. De sier de ikke har noen eksakte tall på omfanget, men mener villscreening ikke har utgjort en bias i deres studie.

Allgood et al. (2008) har i sin studie justert for villscreening ved kun å inkludere kasuser diagnostisert etter 1995 (seks år etter oppstart).

Duffy et al. sier ingenting om forekomst av villscreening i sin studie.

I Autier et al. argumenteres det med at villscreening ikke var vanlig i Norge etter at det nasjonale screeningprogrammet var innført. De andre landene inkludert i deres studie nevnes ikke spesifikt i forbindelse med villscreening.

4.1.5 Bias

4.1.5.1 Lead time bias

Lead time bias er vanskelig å justere for, da vitenskapen i dag ikke vet hvor lang lead time er, og videre hvor mye resultatene påvirkes av det. Flere av studiene oppgir å ha justert for lead time bias, men det finnes ikke konsensus på området om for eksempel hvor lang oppfølgingstid som er nødvendig for å negligjere effekten av denne biasen.

Duffy et al. har en lengre oppfølgingstid enn de andre, noe som kan bidra til at lead time påvirker resultatene i mindre grad enn ved kortere oppfølgingstid.

Kalager et al. sier at de har tatt høyde for lead time ved kun å telle dødsfall fra brystkreft der dødsfall og diagnose oppstod i en av de definerte gruppene. For eksempel ville dødsfall blant kvinner i screeninggruppen kun tilskrives brystkreft dersom sykdommen ble diagnostisert ved screening eller klinisk, mens kvinnen var en del av screeninggruppen. Av dette konkluderer de med at lead time ikke hadde noen innvirkning på deres resultater.

4.1.5.2 Self selection bias

Self selection bias vil kun være en aktuell bias i Allgood et al. og Olsen et al. Dette fordi det i deres studier er gjort beregninger for redusert brystkreftdødelighet blant de kvinnene som faktisk ble screenet. Allgood et al. har justert for self selection ved å ta utgangspunkt i kvinnenens sosioøkonomiske status, mens Olsen et al. har justert ved bruk av en matematisk modell som de forklarer i artikkelen.

4.1.5.3 Confounding

Autier et al. har et studiedesign som skiller seg fra de andre, og en rekke feilkilder kan forekomme når en velger et slikt design. Forskerne undersøkte flere faktorer som assosieres med brystkreftdødelighet, da forskjeller mellom landene kan påvirke resultatene. Av alle livsstilsfaktorer fant de at overvekt (målt i BMI) ved diagnose er den største risikoen for dårlig prognose. De hentet derfor inn data fra rundt år 2000 for overvekt blant kvinner over 18 år i de seks landene. De hentet også inn data over antall fødte barn og alder ved første fødsel, da dette også har noe å si for risikoen.

Autier et al. sier selv at en svakhet ved studiedesignet er at det er umulig å kontrollere forvekslinger/feilkilder, og at det er vanskelig å vite hva som forårsaket fraværet av forskjeller i nabolandene. De sier også at en ukjent faktor kan ha påvirket resultatet, men at dette er usannsynlig, da det mest sannsynlig ikke ville påvirket alle de tre parene.

De andre studiene har ikke nevnt noe angående confounding. Autier et al. skiller seg fra de andre studiedesignene og er av den grunn mer utsatt for denne feilkilden.

4.2 Hva er man uenig om når det gjelder effekten av screening og tolkning av resultater?

Hvordan resultatene presenteres og hvilke statistiske modeller som er brukt kan ha mye å si for hvordan man som leser oppfatter resultatene. Olsen et al., Allgood et al. og Jørgensen et al. presenterer sine funn kun ved bruk av relativ risiko. Duffy et al. og Kalager et al. presenterer sine funn ved bruk av både relativ og absolutt risiko. Autier et al. sammenligner kun trender i naboland, og har derfor ingen estimater på RR eller absolutt risiko.

For å komme fram til resultatene har de forskjellige studiene brukt ulike statistiske analysemodeller. Olsen et al., Duffy et al. og Jørgensen et al. har alle brukt såkalt Poissonregresjon, men med forskjellige inkluderingskriterier, noe som vil påvirke resultatene. Olsen et al. har brukt variablene femårige aldersgrupper, periode, region og eksponering, Duffy et al. kalkulerte absolutte tall på redusert antall døde ved å beregne forskjellen mellom de observerte tallene og de forventede tallene på basis av forandringer i dødeligheten i

aldergruppene <50 og 70+ år, mens Jørgensen et al. brukte Poissonregresjon for å kvantifisere forandringen i dødelighetstrender, sammenligne regioner og aldergrupper i regionene med og uten mammografiscreening. Allgood et al. har brukt logistisk regresjon, men nevner ingen av inklusjonskriteriene i denne. Kalager et al. nevner ikke hva slags statistisk analyse de har brukt, mens Autier et al. hadde et annet studiedesign hvor disse faktorene ikke er viktige slik som i de andre studiene.

4.3 Hva er man uenig om når det gjelder om screeningkriteriene (WHO) er oppfylt?

Ingen av de inkluderte studiene har vært opptatt av screeningkriteriene fra WHO. Disse oppfattes ikke lenger som viktige og er ikke en del av debatten.

5.0 Diskusjon

Når resultatene varierer i så stor grad er det viktig at metoden i de forskjellige studiene analyseres. Hvilke variabler er tatt med og hva slags fokus har forskerne hatt når de gjennomført studien. Det er tydelig at fremgangsmåten har hatt mye å si for resultatene. Metodiske konsekvenser og tolkning av data vil bli diskutert i dette kapittelet under hvert tema. I tillegg til å bruke de seks hovedartiklene har vi nyttet oss av kommentarer og respons publisert i de respektive tidsskriftene/nettsidene i etterkant av publisering av artiklene².

5.1 Hva er man uenig om metodisk?

5.1.1 Oppfølgingstid

Når en leser kritikken de forskjellige har til hverandre med tanke på oppfølgingstid er det ikke nødvendigvis maks oppfølgingstid som er av størst interesse, men den gjennomsnittlige. Dette er det flere av studiene som ikke informerer om, men er et tall som er interessant og kan ha stor innvirkning på resultatet til den enkelte studien, om en legger visse forutsetninger til grunn.

I følge de randomiserte studiene vil ikke en effekt som følge av mammografiscreening oppstå før etter 5 år. Flere, blant annet Duffy, mener at dersom Kalager et al. hadde hatt lengre oppfølgingstid ville effekten av mammografiscreening vist seg tydeligere, og reduksjonen i brystkreftdødelighet blitt betydelig høyere. Dette er kritikk vi kan forstå, men ser en på tallene til Olsen et al. vil en registrere at store deler av reduksjonen de fant i sin studie oppstod allerede etter 3 år, og forble konstant. Poenget med screening er å fremskynde diagnosetidspunkt, og effekten på brystkreftdødelighet er derfor ikke forventet å oppstå de første årene, men etter fem år, og deretter øke med økt oppfølgingstid. Det at full effekt oppstod kun 3 år etter oppstart av screening kan, i følge Jørgensen et al. (2010) tyde på at andre faktorer enn mammografiscreening forårsaket reduksjonen, herunder økt bevissthet omkring brystkreft og forbedret behandling.

² Såkalt rapid response (bmj) eller letter to the editor (nemj, j med screen). Dette var ikke tilgjengelig for alle artiklene, men er brukt der det fantes.

Kalager et al. (2011) sier i sitt tilsvarende til Duffy (2011) angående kritikk til kort oppfølgingstid at deres definisjon av oppfølgingstid skiller seg fra mange andre studier ved at den gjelder fra diagnosetidspunkt, og ikke fra tidspunkt for invitasjon, slik mange andre studier definerer det. De påpeker også at den gjennomsnittlige oppfølgingstiden, som i deres studie var på 2,2 år, alltid vil være begrenset, fordi det hvert år vil være nye kvinner som går inn i studiepopulasjonen, og på den måten holder gjennomsnittet lavt. De argumenterer også med at reduksjonen de fant begrenset seg til de første årene, og at de ikke så noen signifikant reduksjon etter dette, og videre at oppfølgingstiden derfor var tilstrekkelig.

Oppfølgingstid kan også, som tidligere nevnt, være med å påvirke i hvor stor grad lead time påvirker resultatet. Flere av studiene hevder å kompensere for lead time bias, men de opererer med meget store forskjeller i estimering av omfanget av disse, noe som er naturlig, da ingen med sikkerhet vet hvor mye man skal kompensere for. Samtidig svekker dette resultatene i forskningene generelt, og er en kilde til faglig uenighet.

5.1.2 Utvalg

Det virker å være noe uenighet om hvorvidt det er nødvendig med inkludering av andre aldersgrupper enn de som regnes å dra nytte av screening. Flere av de inkluderte studiene bruker andre aldersgrupper som kontrollgrupper for å se om trendene i én aldersgruppe skiller seg fra en annen, og på den måten danne seg et inntrykk av hvorvidt en eventuell reduksjon skyldes screening (som kun tilbys for kvinner i en bestemt alder) eller andre faktorer (som er gjeldende for alle aldersgrupper). Av de inkluderte studiene i denne oppgave har fire av dem inkludert andre aldersgrupper til sammenligning. De finnes dog noen viktige forskjeller blant disse fire. Jørgensen et al. skiller seg ut ved at de har ekskludert aldersgruppen 50-54 år, mens Duffy et al. skiller seg ut ved at de i aldersgruppen <50 år har inkludert kvinner helt ned i 18 års alder, og ikke begrenset til 30 år, noe som er gjennomgående i de andre studiene.

Duffy et al. sier i sin studie at de ikke ser den samme trenden hos kvinner <50 år som i aldersgruppen som tilbys screening. Dette blir av Jørgensen og Gøtzsche (2010) kritisert, fordi de mener Duffy et al. faktisk har funnet det samme som dem, men at deres måte å presentere dataene på ikke viser dette. Blant annet viser de til figur 1, hvor de har samlet data

for alle kvinner under 50 år, og at grafen derfor virker misledende. Dette begrunner de med at død som følge av brystkreft før fylte 30 år er ekstremt sjeldent, noe som fører til at grafen, slik den er utformet, ikke viser variasjoner. Den virker stabil, men i realiteten var det en stor nedgang i brystkreftdødeligheten i denne gruppen. Ser en på dataene bak, slik Jørgensen og Gøtzsche presenterer dem, kan en forstå at de mener Duffy et al. presenterer tallene på en måte som gjør det vanskelig for leseren å se utvikling i trenden, og som dermed virker fordelsmessig for mammografiscreening.

Dersom kvinner i andre aldersgruppene hadde vært inkludert og satt sammen på samme måte i alle studiene, ville dette vært med på å styrke resultatene i den enkelte studie. Spesielt ville dette styrket de studiene som fant en reduksjon i brystkreftdødeligheten som følge av screening, om de samtidig kunne vise at trenden blant de screenede ikke sammenfalt med trenden blant de aldersgruppene som ikke ville ha kunne hatt nytte av mammografiscreening.

5.1.3 Screeningdeltagelse

Både Olsen et al. og Jørgensen et al. har blitt kritisert for å bruke usikre tall, da screeningdeltagelsen har vært svært lav i Danmark, og det i følge IARC kreves at minimum 70 % av den aktuelle populasjonen deltar for at programmets utfall skal kunne evalueres (IARC 2009). Det at screeningdeltagelsen i Danmark er svært lav er ikke tatt i betraktning når screeninggruppene er definert.

Duffy et al. (2010) sier om sin egen studie at denne bygger på empiriske data som unngår feil, som å blande invitasjon til screening med faktisk å ha blitt screenet

Olsen et al. sin studie inkluderer både faktisk screenede kvinner og inviterte, men ikke screenede kvinner, mens Jørgensen et al. inkluderer alle i regionen, noe som kan svekke resultatene, da flere av kvinnene Jørgensen et al. regner som screenet aldri faktisk ble screenet. Samtidig unngår de på denne måten self selection bias. Uenigheten her står om hvorvidt en skal måle den generelle brystkreftdødeligheten, eller om man kun skal måle brystkreftdødelighetene blant kvinner som faktisk blir screenet. Et flertall av studiene inkludert i denne oppgaven har valgt førstnevnte. Slik vi ser det er det optimale kun å

inkludere de kvinnene som faktisk screenes, samtidig som man har nøyaktige estimater på self selection, noe som er meget vanskelig å oppnå.

Jørgensen et al. valgte i sin studie å utelukke aldersgruppen 50-54. Dette fordi de mener de fleste av krefttilfellene i denne gruppen ville ha blitt diagnostisert før fylte 50 år og ikke mens de var en del av screeningprogrammet. For å få et bredere utvalg og et mer korrekt bilde av effekten burde heller de tilfellene i aldersgruppen 50-54 år diagnostisert før screening startet vært ekskludert. Slik vi forstår det hadde ikke Jørgensen et al. tilgang på slik data, noe som svekker deres funn. Selv om dette ikke ville påvirket resultatene, slik Jørgensen et al. (2010) hevder, ville de stått sterkere mot kritikken som har kommet i etterkant til deres studie.

5.1.3.1 Self selection

Som med lead time og lenght time er det også vanskelig å vite hvor stor innvirkning på resultatene self selection har. Studier har vist at kvinner med lavere sosioøkonomisk status generelt har dårligere helse, og dermed kan være mer utsatt for brystkreft i utgangspunktet. Om disse kvinnene igjen er de samme som ikke deltar i mammografiscreening kan man ende opp med en overestimering av effekten av mammografiscreening. Olsen et al. (2005) estimerer i sin studie 37 % reduksjon i brystkreftdødelighet blant de som faktisk ble screenet. Dette mener Jørgensen et al. er et estimat som ikke er gyldig, da Olsen et al. ikke hadde noe grunnlag for å vite hvem i den ikke-screenede gruppen som faktisk ville sagt ja til screening om de hadde blitt invitert, jamført den tidligere nevnte "healthy screenee effect". Olsen et al. har i sin studie gjort beregninger for å justere for self selection bias, hvorvidt Jørgensen et al. ikke mener denne er god nok, eller om de har misforstått, er vanskelig å vite. Det vi konstaterer er at det på området er uenighet om hvordan man skal ta høyde for self selection.

Blant utvalget vårt er det kun Olsen et al. og Allgood et al. som har en metode som krever en justering for self selection. De andre studiene (med unntak av Autier et al.) har inkludert alle kvinner invitert. På en side er dette en mulighet for å utelukke self selection bias, samtidig svekker det resultatene, da en ikke kan si noe med sikkerhet. Spesielt gjelder dette for de mammografiprogrammene som har lav deltagelse, hvor man kan ende opp med et underestimat av effekten av mammografiscreening.

5.1.3.2 Length time bias

Dette er en viktig bias innenfor debatten omkring mammografiscreening, men den er mer fremtredende i forbindelse med overdiagnostisering. Dette er et tema som ikke blir diskutert i denne oppgaven, da det er meget omfattende, og nærmest kan regnes som en egen debatt i debatten. Det en kan se av den aller siste forskningen, er at overdiagnostisering har blitt et mer diskutert tema, og at length time bias derfor er viktigere i denne debatten.

5.1.4 Villscreening

Mye av uenigheten angående villscreening går på det faktum at flere av studiene mener villscreening ikke har påvirket deres resultater. Blant annet Olsen et al., Jørgensen et al. og Kalager et al. sier dette. Også Autier et al. trekker i sin studie fram at villscreening ikke er av stor betydning, og de nevner spesielt Norge i denne sammenheng. Kun Allgood oppgir å ha justert for villscreening. Dette gjorde de ved kun å inkludere kasuser diagnostisert seks år etter screeningoppstart. De gir ingen ytterligere forklaring på dette, men vi går ut i fra at de mener villscreening på dette tidspunktet hadde avtatt tilstrekkelig til at det ikke ville gi store utslag på resultatene.

Av de studiene som mener at villscreening ikke påvirket resultatene, får spesielt Autier et al. kritikk for denne påstanden. Lyng og Njor (2011) påpeker i sin respons til artikkelen at det er mer relevant å vite forekomst av opportunistisk screening før programmet ble innført for å kunne sammenligne det med Sverige. Med dette henviser de til at Autier et al. bruker estimer på villscreening fra etter innføringen av screening, mens de burde brukt estimer på villscreening blant kvinner før dette. Lyng og Njor mener at Norge hadde en betraktelig andel opportunistisk screening før introduksjon av mammografiprogrammet, noe som vil kunne forklare reduksjonen i brystkreftdødelighet startet før innføringen. Kalager et al. påpeker selv at villscreening kan ha minsket resultatet på effekten av screening. I hvilken grad er vanskelig

å si noe om, men i følge Lyng kan dette ha svekket resultatenes reliabilitet.

5.1.5 Studiedesign

Studier med forskjellige studiedesign ble bevisst valgt for å belyse flere sider av debatten. Utvalget viser også at det ikke er noe enighet om hva slags studiedesign som egner seg best ved evaluering av mammografiscreening. Hvorvidt ett studiedesign egner seg bedre enn et annet vil ikke vi ta stilling til, da vi ikke har grunnlag for å vurdere dette.

Autier et al. kritiseres Koning (2011) for valg av studiedesign. Han fastslår at Autier et al. ved å bruke geografisk sammenligning oppnår dårlige forskningsresultat. Autier et al. bruker naboland og hevder at disse er tilstrekkelig sammenlignbare. Koning påpeker at helsetilbudet varierer i stor grad fra land til land, til tross for at Autier et al. mener helsetilbudene er sammenlignbare. De baserer seg på salg av brystkreftmedisin for å avgjøre om helsetilbudet er lik i de to nabolandene, dette sier lite om resten av behandlingstilbudet, slik Koning vurderer det.

Med unntak av Autier et al. har vi ikke funnet noen konkret kritikk av studiedesignet som helhet av de andre studiene.

5.2 Hva er man uenig om når det gjelder effekten av screening og tolkning av resultater?

5.2.1 Tolkning av resultater

I følge Gøtzsche et al. (2010) varierer resultatene til Duffy et al. signifikant fra andre studier. For eksempel estimerer Duffy et al. med 2-2,5 liv reddet for hvert overdiagnostisert tilfelle, dette er 20-25 % mer enn cochranes oversikt fra de randomiserte forsøkene. Når resultatene fra UK presenteres motsir de seg selv. De vurderer dødelighetsratene i de ikke-screenede gruppene som stabile, samtidig som de rapporterer en signifikant reduksjon på 18 %. Cancer research UK fant i følge Gøtzsche et al. (2010) at kvinner i alderen 40-49 år hadde samme reduksjon i dødelighet på 44 % som aldersgruppen 50-64 år. I tillegg mener de grafene er villedende, da data for alle kvinner under 50 år er slått sammen. Dette gir et ukorrekt bilde, fordi dødsfall grunnet brystkreft før fylte 30 år er ekstremt sjeldent (Gøtzsche 2010). Som tidligere nevnt har Duffy et al. inkludert kvinner helt ned til 18 år i gruppen av kvinner under

50 år. Deres resultater er derfor annerledes enn andres, som kun inkluderer kvinner ned i 30 år. Ved å definere aldergruppene på forskjellige måter bidrar de forskjellige studiene til at resultatene ikke er sammenlignbare, og vi mener dette er med på å skape unødvendig uenighet blant de ulike gruppene.

Olsen et al. sitt valg av analysemodell kritiseres for å fremheve de positive sidene. Studien omtaler ikke negative sider ved mammografiscreening (Jørgensen et al. 2010). Ingen studie som viser positiv effekt av et tiltak burde presenteres uten å ta med de negative sidene, som i dette tilfellet er graden av overdiagnostikk. Flere av studiene kunne med fordel vært presentert på en mer objektiv måte, noe som kan være spesielt vanskelig da det foreligger en interessekonflikt. En ser at det er lett å tolke resultatene i den retningen en ønsker og at det kan være vanskelig å forholde seg subjektiv i slikt studie da en har en klar oppfatning på forhånd. Det er lett å velge den fremgangsmåten som er fordelaktig for sitt syn.

5.2.1.1 Interessekonflikt

I utgangspunktet har alle studiene erklært at det ikke foreligger noen interessekonflikt. Men etter en nærmere gjennomgang ble det funnet flere potensielle interessekonflikter. Dette gjelder Tabár, grunnlegger av Mammography Education inc. Arizona (Gøtzsche 2010), og to av forfatterne av Olsen et al., som begge er ledere for screening enheter (Gøtzsche 2005). Selv om de ikke skulle være påvirket av eventuelle egeninteresser, er det uheldig, da det åpner for kritikk av funn. Vi mener det er viktig å opplyse om all tilhørighet, da dette kan være med på å fjerne eventuell tvil og forhindre uenigheter om mulige interessekonflikter. Vi ser også at debatten til tider tar en noe uprofesjonell tone, noe flere også påpeker i sine responser. Her viser vi ikke til noen spesifikke tilfeller, men leser man en del av responsen til de ulike studiene ser en at tonen kan være anklagende, ofte uten noen gode begrunnelser.

5.3 Konsekvenser for praksis

Hensikten med denne rapporten var å kartlegge de faglige uenighetene slik at det skulle være lettere for radiografer og andre innefor fagfeltet å forstå konflikten i påvente av den offentlige evalueringen.

Anna Gavin, en av forfatterne av Autier et al., foretrekker i følge Michael Baum (2011) å fortsette med mammografiprogrammet til tross for at resultatene i studien ikke viser noen effekt på brystkreftdødeligheten som følge av mammografiscreening. Dette kan være på grunn av usikkerheten rundt effekten og det er lettere å fortsette som før enn å ta en beslutning om å legge ned et så omfattende tilbud når resultatene er såpass usikre. Det er vanskelig å vite hvilken informasjon en kan stole på ettersom metodene og utvalget varierer i så stor grad, og det kan da være lettere å ta beslutninger som har mindre umiddelbare konsekvenser som det å legge ned mammografiprogrammet ville ha ført til. Reaksjonene ville vært enorme, og det ville vært vanskelig å rettferdiggjøre uten konkrete beviser.

5.4 Kritikk

5.4.1 Søk

I oppgaver av denne art er det nødvendig å begrense seg med tanke på omfang av datamateriale og databasesøk. Våre resultater og funn kunne vært mer representative hadde vi utvidet søkene våre med flere databaser, søkeord og kombinasjoner. Allikevel valgte vi å gjøre de begrensinger vi gjorde, med tanke på oppgavens omfang og at vi følte vi hadde et godt utvalg etter de søkene vi gjorde. Vi så også at mange av de treffene vi fikk i våre søk gikk igjen i flere av søkene, og vi mente derfor det var begrenset hvor nyttig det ville være å gjøre ytterligere søk.

I de systematiske søkene gjorde vi en rekke begrensninger. Blant annet begrenset vi oss til kun nyere litteratur. Dette mente vi var viktig med tanke på at vi arbeider i et fagfelt som er i stadig endring, og at det derfor er viktig å inkludere den nyeste forskningen. Dette førte også til at vi fikk utelukket resultater som ikke lenger regnes som valide.

Ved siden av gjennomlesning av artikler og annen faglitteratur har vi også gjennomgått responsen på artiklene inkludert i vårt arbeid (der dette var tilgjengelig). Slik kunne vi kontrollere at vi tolket artiklene korrekt, i tillegg til at vi fikk et godt innsyn i den faglige debatten.

5.4.2 Reliabilitet og validitet

Søkestrategien i oppgaven er dokumentert og beskrevet i detalj slik at den kan etterprøves. Dette gir oppgaven reliabilitet, da andre kan komme til samme resultat ved å bruke samme premisser (Dalland 2007). Det meste av litteraturen består av artikler publisert i anerkjente tidsskrifter. Svært få internettkilder er brukt noe som styrker reliabiliteten siden vi ser at denne type kilder er vanskeligere å etterprøve, da innholdet på sidene stadig kan forandres.

For å sikre at vi fikk svar på vår problemstilling kontrollerte vi underveis at artiklene gikk direkte på vårt tema og kunne besvare forskningsspørsmålene.

Vi sikret validiteten ved å være saklige og pålitelige i vår innsamling og analyse av data. Når en konklusjon er basert på sanne premisser er den sann og dermed valid. Når en analyserer og siterer andres litteratur er det viktig å unngå feiltolkninger (Dalland 2007), dette vil vi si var den største utfordringen i gjennom arbeidet.

6.0 Konklusjon

De største uenighetene synes å ha to hovedkilder. Dette synes å være lengden på oppfølgingstid og hvilke aldersgrupper som skal inkluderes i studien. Oppfølgingstiden er vanskelig å gjøre noe med i en slik type kohortestudie som Kalager et al. utførte, siden de brukte ulik integrering av mammografiprogrammet som grunnlag for kohortegruppene.

Villscreening vil være vanskelig å kartlegge så lenge dette ikke blir registrert på samme måte som i de offentlige mammografiprogrammene.

Som vi ser i de ulike studiene er det store forskjeller i metoden. Studiedesign og hvordan studien har blitt utført har hatt mye å si for utfallet. Alle de inkluderte studiedesignene har sine svakheter. Spesielt siden gjennomførelsen av studiene varierer i stor grad, noe som kan gjøre at tilfeldigheter kan spille inn. Alle studiene kan kritiseres og den absolutt beste måten å måle effekten av et slikt tiltak vil være dobbelblindet randomiserte studier. Dog er ikke dette lenger etisk gjennomførbart, ettersom mammografiscreening nå er et offentlig tiltak og ingen kan nektes å delta. Ikke alle forskningsspørsmålene kunne besvares direkte i alle studiene.

Siden mammografiprogrammet ble innført samtidig med forbedringer innen helsevesenet vil effekten være vanskelig å anslå. Når det er så mange usikre elementer som spiller inn på forskningen er det vanskelig å konkludere med hva som er rett eller galt.

7.0 Litteratur

Allgood, P. C. et al. (2008). A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality.I: *The British Journal of Cancer* 98(1), s. 206-9.

Autier, Philippe et al. (2011). Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database.I: *Brit Med J* 343(7818), s. 300.

Aveyard, Helen (2010). *Doing a literature review in health and social care: a practical guide*. Maidenhead: McGraw-Hill/Open University Press.

Baum, Michael (2011). *The failure of breast cancer screening to meet its promise-yet again* [online]. [Response]: British Medical Journal. URL <http://www.bmj.com/content/343/bmj.d4411?tab=responses> (20.4.2012).

Citrome, L. (2010). *Relative vs. absolute measures of benefit and risk: what's the difference?*, 121: Wiley-Blackwell. 94-102 s.

Dalland, Olav (2007). *Metode og oppgaveskriving for studenter*. Oslo: Gyldendal akademisk.

Duffy, S. W. et al. (2010). Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England.I: *J Med Screen* 17(1), s. 25-30.

Duffy, S. W. (2011). More on screening mammography.I: *New England Journal of Medicine* 364 s. 281-286.

Gøtzsche, P C. K. Jørgensen (2005). *Benefits and harms of breast cancer screening* [online]. [Response]: British Medical Journal. URL <http://www.bmj.com/content/330/7485/220?page=1&tab=responses> (20.4.2012).

Gøtzsche, P. C. og M. Nielsen (2011). Screening for breast cancer with mammography.I: *Cochrane Database Syst Rev* (1), s. CD001877.

Gøtzsche, Peter C., Karsten Juhl Jørgensen, Per-Henrik Zahl (2010). Breast screening: why estimates differ by a factor of 20–25.I: *J Med Screen* 17(3), s. 158-159.

Hansson, S. O. (1996). Decision making under great uncertainty.I: *Philos Soc Sci* 26(3), s. 369-386.

Helsedirektoratet (2012). **Prioriteringer i helsesektoren - verdigrunnlag, status og utfordringer**: Helsedirektoratet.

Hofvind, S. et al. (2011). Breast cancer incidence trends in Norway-explained by hormone therapy or mammographic screening?I: *Int J Cancer*.

IARC (2008). **World Cancer Report 2008**. Lyon: International Agency for Research on Cancer.

IARC (2009). *Early Detection and Prevention - Screening Group* [online]: IARC. URL <http://www.iarc.fr/en/research-groups/SCR/index.php> (8.5.2012).

Jørgensen, K. J., P. H. Zahl og P. C. Gotzsche (2010). Breast cancer mortality in organised mammography screening in Denmark: comparative study.I: *Brit Med J* 340.

Jørgensen, K., Peter Gøtzsche (2010). *Mammography screening has'nt lived up to expectations* [online]: BMJ. URL <http://www.bmj.com/content/340/bmj.c1241?page=2&tab=responses> (20.04.2012).

Kalager, M. et al. (2010). Effect of screening mammography on breast-cancer mortality in Norway.I: *New England Journal of Medicine* 363(13), s. 1203-10.

Kalager, M., Marvin Zelen, Hans-Olav Adami (2011). More on screening mammography.I: *New England Journal of Medicine* 364 s. 281-286.

Koning, Harry J. de (2011). *Screen evaluation without evaluating screening?* [online]. [Response]: British Medical Journal. URL <http://www.bmj.com/content/343/bmj.d4411?tab=responses> (2.5.2012).

Lynge, E., Sisse Njor (2011). *Mammography screening and breast cancer mortality* [online]: British Medical Journal. URL <http://www.bmj.com/content/343/bmj.d4411?tab=responses> (1.5.2012).

Mahnken, J. D. et al. (2008). Reducing the effects of lead-time bias, length bias and over-detection in evaluating screening mammography: a censored bivariate data approach.I: *Stat Methods Med Res* 17(6), s. 643-63.

Mcstrother (2011a). *Lead time bias* [online]: Wikipedia. URL http://en.wikipedia.org/wiki/File:Lead_time_bias.svg (9.5.2012).

Mcstrother (2011b). *Lenght time bias* [online]: Wikipedia. URL http://en.wikipedia.org/wiki/File:Length_time_bias.svg (9.5.2012).

Olsen, A. H. et al. (2005). Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study.I: *Brit Med J* 330(7485), s. 220.

Schopper, Doris og Chris de Wolf (2009). How effective are breast cancer screening programmes by mammography? Review of the current evidence.I: *European Journal of Cancer* 45(11), s. 1916-1923.

Staff, Annetine (2010). *Bias* [online]: Forskningsetisk bibliotek. URL <http://www.etikkom.no/no/FBIB/Temaer/Spesielle-problemomrader/Bias/> (25.04.2012).

Store medisinske leksikon (2011). *Absolutt risiko* [online]: Store norske leksikon. URL http://snl.no/.sml_artikkel/absolutt_risiko (25.04.2012).

Vedlegg 1: Sammendrag av de 6 hovedartiklene

Sammendrag av de seks hovedartiklene

Olsen et al.

Olsen et al. er en kohortestudie gjort i Københavnregionen i Danmark. Studien inkluderer kvinner invitert til screening i København i de fem første screeningrundene (10 år). Resten av Danmark fungerte som kontrollgruppe, da resten av Danmark ikke har screening. Det ble også brukt historiske kontrollgrupper for de to. Studiepopulasjonen inkluderte 40000 kvinner i alderen 50-69 år. Studiepopulasjonen sammenlignes ikke med andre aldersgrupper i studien. Kvinner som ble diagnostisert med kreft før de mottok sin første invitasjon ble ekskludert fra studien. 71 % deltok i den innledende runden, et tall som falt noe de påfølgende rundene.

Alle resultater i studien oppgis kun ved bruk av relativ risiko (RR). Studien fant en redusert dødelighet som følge av brystkreft på 25 %. Reduksjonen som følge av å faktisk bli screenet var i følge studien 37 %.

Studien konkluderer med at mammografiscreening i København har redusert brystkreftdødeligheten med 25 %, og samtidig unngikk negative konsekvenser følger.

Inkluderingskriterier i Poisson-regresjon?

For å analysere effekten av invitasjon til screening sammenlignet de dødsratioen i studiepopulasjonen med kontrollgruppene, justert for alder, tidsperiode og region. Den statistiske analysen ble gjort ved bruk av Poisson-regresjon med variablene femårige aldersgrupper, periode, region og eksponering. Det var på den måten mulig å kontrollere for tidstrender og regionale forskjeller, men ikke å skille en eventuell effekt av interaksjon mellom de to og effekten av screening.

Allgood et al.

Allgood et al. er en kasus-kontrollstudie gjort i East Anglian-regionen i Storbritannia, hvor alle kvinner i alderen 50-70 år blir invitert til mammografiscreening hvert tredje år. Studien inkluderte 300 kasuser, hver av dem parett med to kontroller basert på fødselsdato, totalt 600, de aller fleste fra samme område. Studien ekskluderte ikke kvinner som ble invitert til screening, men som aldri faktisk ble screenet. Den sammenligner heller ikke forskjellige aldersgrupper og dødsratioen innenfor hver av disse gruppene, men ser utelukkende på målgruppen for mammografiscreening, kvinner i alderen 50-70 år.

Artikkelen konkluderer med at screeningprogrammet i East Anglian-regionen oppnår en reduksjon i brystkreftdødeligheten på minst 30 % blant kvinner som undergår mammografiscreening, og de anbefaler en utvidelse av kasus-kontrollstudien til andre deler UK og internasjonalt.

Som tidligere nevnt inkluderer studien kvinner invitert til screening. Kvinner invitert, men ikke screenet er derfor også inkludert i resultatene. Det er tatt høyde for dette, og resultater både for ”invitert” og ”invitert og screenet” er presentert hver for seg. I studien fremstilles reduksjonen i dødelighet som følge av brystkreft kun ved bruk av relativ risiko (RR).

Inkluderingskriterier i Poisson-regresjon?

Den statistiske dataanalysen ble gjort ved bruk av logistisk regresjon. Resultatene ble justert for self-selection bias ved å ta høyde for kvinnens sosioøkonomiske status (basert på postnummer), alternativt ved bruk av en modell av Duffy et al. (2002) som baserer seg på data fra de randomiserte studiene, hvor en sammenlignet den relative dødeligheten blant ikke-deltagende inviterte med kontrollgruppen. Studien ble også justert for villscreening ved kun å inkludere kasuser diagnostisert etter 1995 (seks år etter oppstart av screening). Artikkelen gir ingen god forklaring på dette, men vi går ut i fra at de mener villscreeningen hadde avtatt tilstrekkelig etter seks år med et screeningprogram, slik at dette ikke ga store utslag på resultatet.

Duffy et al.

Duffy et al. er en studie som tar sikte på å estimere absolutte tall på redusert brystkreftdødelighet basert på tall fra en randomisert studie og mammografiprogrammet i England. Den randomiserte studien er den såkalte two-county trial, en studie som ble gjort i Sverige. Det engelske mammografiprogrammet tilbyr screening hvert tredje år til kvinner i alderen 50-70 år (på tidspunktet for studien). Artikkelforfatterne hadde tilgang til data for femårige aldersgrupper og kalenderår for brystkreftforekomst i årene 1974-2003.

Den svenske randomiserte studien tilbydde screening til kvinner ned til 40 års alder, men kun kvinner i alderen 50-69 år ble inkludert av Duffy et al. (2010) for å sammenfalle med aldergruppen i det engelske programmet.

Dataanalysen ble gjort ved bruk av Poissonregresjon for å sammenligne brystkreftdødelighet blant kvinner i alderen 50-69 år i årene mellom 1995→ og før 1989 med det observert i andre aldersgrupper. De kalkulerte absolutte tall på redusert antall døde ved å beregne forskjellen mellom de observert tallene og de forventede tallene på basis av forandringer i dødeligheten i aldersgruppene <50 og 70+.

Duffy et al. sier tallene de kommer frem til er konservative fordi noen av dødsfallene fra 95→ vil være grunnet kreft diagnostisert før 1989 (før screening begynte), og fordi noen dødsfall i perioden 1989-1994 kan ha blitt forhindret av screening.

Studien presenterer resultatene både ved bruk av relativ risiko (RR) og absolutt risikoreduksjon. Den absolutte risikoen presenteres både med hvor mange liv som reddes og hvor mange som må screenes for å redde ett liv. Studien fant at mammografiscreening redder 8,8 og 5,7 liv per 1000 kvinne screenet i 20 år i hhv. two-county trial og det engelske mammografiprogrammet.

Duffy et al. fant at antall liv spart er ca. dobbelt så høyt som antall overdiagnostisert tilfeller, og at de positive konsekvensene derfor overgår de negative. Artikkelforfatterne sammenligner sine resultater med resultater fra en studie Gøtzsche tidligere har publisert (Duffy et al. 2010) som sier at overdiagnostisering er ti ganger vanligere enn redusert dødelighet. Duffy et al. (2010) tilbakeviser disse og mener de har for kort oppfølgingstid (10 år vs. 20 år). De sier også om sin egen studie at denne bygger på empiriske data som unngår feil, som å blande invitasjon til screening med å faktisk ha blitt screenet. Videre viser de til en studie av

Jørgensen og Gøtzsche som konkluderer med tilnærmet 100 % overdiagnostisering, noe de omtaler som en absurd påstand, og en feiltolkning av data.

Kalager et al.

Kalager et al. er en studie som tar sikte på å estimere effekten av det norske mammografiprogrammet. Dette ble startet opp i 1994, og gradvis implementert over hele landet de neste årene. I 2002 hadde Norge oppnådd et landsdekkende tilbud til alle kvinner i alderen 50-69 år.

Studien definerte fire grupper; en screenet og en ikke-screenet gruppe (nåværende), og to historiske grupper som speilet de nåværende gruppene over en tiårsperiode før implementering av screening. Parallelt med oppstart av screening ble det stilt krav til at det skulle være på plass et multidisiplinært team, spesialisert på kreftbehandling. Dette gjorde det mulig å isolere effekten av forbedret behandling, fordi kvinner i aldersgrupper som ikke ble tilbudt screening allikevel ville dra nytte av de multidisiplinære teamene som ble implementert i fylkene hvor screening hadde startet. De to nåværende gruppene var ikke direkte sammenlignbare, og de historiske kontrollgruppene ble derfor tatt i bruk for å justere for forskjeller i dødsstatistikken i de forskjellige delene av landet.

Studien brukte data fra 1996-2005, og historiske data fra 1986-1995.

Studien inkluderer kun invasiv kreft, ductalt carcinoma in situ (dcis) er derfor ekskludert.

Kun kreft som ble diagnostisert etter innføring av screening er inkludert i studien, men kvinner som ble diagnostisert etter innføring, men utenfor screening, såkalt intervallkreft, er inkludert.

Studien definerer effekten for aldersgruppene 20-49, 50-69 og 70-84 år. Ved å se på også andre aldersgrupper kan en se etter mønster i brystkreftdødeligheten, og vurdere hvorvidt det skyldes mammografiscreening eller andre faktorer. I sine resultater viser de tall på både relativ og absolutt risiko.

Det norske mammografiprogrammet alene blir i studien assosiert med en reduksjon på 10 % brystkreftdødeligheten. Total utgjorde mammografiscreening kun en tredjedel av den totale

reduksjonen man så i brystkreftdødeligheten, rest ble assosiert med forbedret behandling og andre faktorer.

Kalager et al. fant at reduksjonen i aldergruppen 70-84 år i stor grad var den samme som observert i screeninggruppen. Ingen av disse deltok i screening, men alle ble behandlet av de multidisiplinære teamene.

De påpeker svakheter ved egen studie. Blant annet hadde de en oppfølgingstid på maks 8,9 år, noe mange mener er for lite. De sier også at noen av kvinnene i den ikke-screenede gruppen kan ha fått utført mammografi privat, såkalt villscreening. Dette kan ha minsket resultatet på effekten av screening.

Kalager et al. sier deres funn er i tråd med resultater fra screeningprogrammet i England. Andre studier med lignende oppfølgingstid har tidligere konkludert med reduksjon på 6,4 % - 20 %, men at disse ikke har tatt trender over tid i betraktning, noe som kan føre til en overestimering av reduksjonen i brystkreftdødelighet.

Avslutningsvis konkluderer Kalager et al. med at deres funn støtter bevisene som sier mammografiscreening reduserer brystkreftdødeligheten, men at fordelene synes å være begrenset i programmet de evaluerte.

Jørgensen et al.

Jørgensen et al. er en sammenlignende studie som ønsker å se på hvorvidt reduksjonen på 25 % på brystkreftdødeligheten Olsen et al. (2005) fant i Københavnregionen faktisk kom som et resultat av screening, eller om andre faktorer har spilt inn. Sammenlignet med Olsen et al. hadde Jørgensen et al. fem år ekstra med datamateriale. Studien inkluderer også en ekstra screeningregion, Fyn fylke.

Studien sammenligner reduksjonen i brystkreftdødelighet i den screenede populasjonen med den ikke-screenede populasjonen, delt inn i femårige aldersgrupper. I motsetning til Olsen et al. (2005) valgte Jørgensen et al. å se på effekten i aldergruppen 55-74 år. Dette begrunner de med at de fleste i alderen 50-54 vil ha blitt diagnostisert før deres første screening, mens kvinner i alderen 70-74 år diagnostisert med kreft i stor grad vil ha blitt diagnostisert mens de enda var del av screeningprogrammet. Aldersgruppene 35-54 og 75-84 år vil i følge studien i

stor grad være uanfektet av screening, og fungerer i derfor i denne studien som kontrollgrupper, i tillegg til den nevnte ikke-screenede populasjonen i alderen 55-74 år.

For å definere en startdato for studien har de brukt resultater fra randomiserte studier som viste at en reduksjon viste seg først etter fem år, altså 1996 i København og 1998 i Fyn. De valgte derfor å bruke 1997 som startdato, noe som ga en tiårig observasjonsperiode. Til sammenligning brukte de en historisk kontrollgruppe over en tiårsperiode som endte når screeningen begynte, 1982-1991.

I sin statistiske analyse har de brukt Poissonregresjon for å kvantifisere forandringen i dødelighetstrender, sammenligne regioner og aldersgrupper i regionene med og uten mammografiscreening.

Jørgensen et al. sier deres funn er motstridende til de Olsen et al. fant i sin studie. De fant ikke eneffekt av det danske mammografiprogrammet. I tillegg fant de at i aldersgruppen som var for unge til å kunne ha dratt nytte av screening (35-54) fant man en større reduksjon i brystkreftdødeligheten enn i aldergruppen 55-74 år. De fant også at reduksjonen i dødeligheten fant sted for tidlig etter oppstart av screening til at dette kan ha vært en sannsynlig årsak. De påpeker at denne reduksjonen utgjorde basis for resultatene til Olsen et al. De påpeker også at reduksjonen i Fyn begynte allerede før innføringen av mammografiscreening. Dette tatt i betraktning mener de tilsier at andre faktorer, som økt bevissthet og forbedret behandling, er årsaken til reduksjonen. Retningslinjer for behandling av kreft ble innført parallelt over hele Danmark, men Jørgensen et al. mener at forbedringer i behandling kan ha kommet tidligere i noen regioner enn andre.

Studien ekskluderer ikke brystkreft diagnostisert før screeningen ble startet opp, men mener dette ikke har utgjort en særlig effekt på deres resultat. De tar heller ikke hensyn til at noen av kvinnene kan ha flyttet fra en region til en annen, men sier selv at dette ikke har hatt mye å si for resultatet, fordi det er lite flytting i de involverte områdene. Grunnlaget for dette utsagnet sies det dog ingenting om.

Jørgensen et al. kommenterer Olsen et al. sin statistiske modell og mener denne er i overkant komplisert, og at resultatet, avhengig av inkluderingskriterier i denne modellen, vil variere mye. Det vises blant annet til en artikkel som viser de forskjellige resultatene avhengig av modellvalg. Her ser man at resultatene spriker fra økt brystkreftdødelighet i studiepopulasjonen til de oppgitte resultatene på 25 % reduksjon (Olsen et al. 2007). Olsen et

al. valgte altså den modellen som ga det mest flatterende resultatet. Jørgensen et al. påpeker på sin side at deres modell er enkel, og kun presenterer rådata og enkle analyser.

Jørgensen et al. sier mange studier har nyttet såkalt populasjonsstatistikk til å estimere effekten av mammografiscreening, men at mange av disse kun inkluderer kvinner i aldersgruppen som kan dra nytte av screening, og ikke ser på utviklingen i andre aldersgrupper i samme tidsperiode. De sier også at mange av de samme studiene hevder å kompensere for lead time bias og lenght bias, men at de opererer med meget store forskjeller i estimering av omfanget av disse, noe som er naturlig da ingen med sikkerhet vet hvor mye man skal kompensere for. Samtidig svekker dette resultatene i disse studiene.

Jørgensen et al. konkluderer at de i sin studie ikke klarte å finne en effekt av det danske screeningprogrammet. Reduksjonen var lik i ikke-screenede områder og i aldersgrupper som ikke hadde dratt nytte av screening. De mener derfor at reduksjonen sannsynligvis har kommet som et resultat av andre faktorer, som forbedret behandling etc.

Autier et al.

Autier et al. er en studie som sammenligner trender i brystkreftdødelighet i tre par naboland hvor det ene landet startet opp mammografiscreening tidligere enn det andre, men de andre faktorene skal være av samme kvalitet. De tre inkluderte nabolandene i studien er Nord-Irland og Irland, Nederland og Belgia, og Norge og Sverige. Studien er en såkalt retrospektiv trendanalyse.

I arbeidet med å bestemme hvilke land som kunne sammenlignes vurderte forskerne flere faktorer, blant annet forventet levealder, antall kvinner i screeningalder, sosioøkonomiske forhold, kulturelt miljø, utdanningsnivå i populasjonen, kvalitet i helsetilbud og tilgang på dette. De tre tidligere nevnte parene møtte forskernes kriterier og ble således inkludert i studien.

En rekke feilkilder kan forekomme når en velger et slikt studiedesign. Forskerne undersøkte flere faktorer som assosieres med brystkreftdødelighet, da forskjeller mellom landene kan påvirke resultatene. Av alle livsstilsfaktorer fant de at overvekt (målt i BMI) ved diagnose er den største risikoen for dårlig prognose. De hentet derfor inn data fra rundt år 2000 for

overvekt blant kvinner over 18 år i de seks landene. De hentet også inn data over antall fødte barn og alder ved første fødsel, da dette også har noe å si for risikoen.

For å vurdere behandlingen tok de utgangspunktet i salgstall for kreftmedisiner. De opplyser om at de ikke fant noen data på tamixofenbruk som kunne brukes i sammenligning av landene.

Autier et al. fant at brystkreftdødeligheten varierer lite mellom landene hvor kvinner var blitt screenet over lengre tid sammenlignet med landene hvor kvinner i stor grad var uscreenet i samme periode. Av dette mener de det er mulig at mammografiscreening ikke har spilt en direkte rolle i reduseringen av brystkreftdødeligheten.

Autier et al. sier selv at en svakhet ved studiedesignet er at det er umulig å kontrollere forvekslinger/feilkilder, og at det er vanskelig å vite hva som forårsaker fraværet av forskjeller i nabolandene. De sier også at en ukjent faktor kan ha påvirket resultatet, men at dette er usannsynlig da det mest sannsynlig ikke ville påvirket alle de tre parene.

Autier et al. konkluderer med at mammografiscreening ikke kan forklare reduksjonen i brystkreftdødeligheten i landene i studien, da reduksjonen startet for tidlig etter implementering av screening, eller at screening fortsatt var uvanlig. Det synes mer sannsynlig at reduksjonen kom som et resultat av mer effektiv behandling. Den største reduksjonen så man hos kvinner under 50 år, uavhengig av screening. Reduksjonen hos kvinner over 70 år hos kvinner i Nederland har blitt tolket som at screening fungerer, men den samme trenden så man i Norge og Irland, og her kan ikke reduksjonen i samme periode tilskrives screening.