



10HBRAD

Delia Arbildo-Tverfjell

Kine Stangnes

BACHELOROPPGAVE, RAD 3911

”Hvor mye overdiagnostikk er det ved mammografi screening og hvorfor er det så stor uenighet?”

Høgskolen i Gjøvik

Avdeling for helse, omsorg, sykepleie Seksjon Radiografi

Mai 2013

Antall Ord: 7781

Forord

Dette er vår avsluttende bacheloroppgave ved radiografutdanningen ved høyskolen i Gjøvik. Oppgaven ble ferdigstilt i mai 2013. Oppgaven er av radiograffaglig relevans og er dermed rettet mot radiografer og annet interessert helsepersonell. Gjennomføringen av oppgaven har vært lærerikt og gitt oss god innsikt i temaet.

Vi vil takke vår veileder Bjørn Morten Hofmann for god veiledning. Han har vært en god støttespiller og gitt konstruktive tilbakemeldinger under veiledningstiden.

Gjøvik, våren 2013

Delia Arbildo-Tverfjell og Kine Stangnes

Sammendrag

Problemstilling:

Hvor mye overdiagnostikk er det ved mammografi screening og hvorfor er det så stor uenighet?

Hensikt:

Gi innsikt i hvor mye overdiagnostikk det er ved mammografiscreening. Hvordan har forskerne kommet frem til de forskjellige resultatene og hvorfor det er så stor uenighet ved estimeringen av overdiagnostikk?

Metode:

I oppgaven har vi benyttet en kvalitativ tilnærming og valgt litteraturstudie som metode. Litteraturen vi har benyttet i oppgaven er fra vitenskapelige artikler og enkelte bøker. Artikler inkludert i denne oppgaven var alle oversiktsartikler som benyttet forskjellige beregningsmetoder ved estimeringen av overdiagnostikk.

Resultat:

Hvilken metode som blir benyttet ved estimering av overdiagnostikk har mye å si for resultatene. Vi har benyttet totalt 6 artikler hvor alle artiklene bruker forskjellige metoder for å estimere overdiagnostikk. Det er to metoder som er mest foretrukket, den kumulative insidens metoden og insidens-rate metoden. De største uenighetene er hvilken metode og studiedesign som skal brukes for å gjennomføre studien.

Konklusjon:

Estimering av overdiagnostikk ved mammografiscreening er et kompleks tema grunnet variasjonene av metodene som benyttes for å beregne overdiagnostikk. Estimerer varierer fra – 4 til 69 % i de forskjellige studiene. Det er vanskelig å komme til enighet så lenge forskerne ikke kommer frem til en felles metode.

Stikkord/nøkkelord:

Mammografiscreening, brystkreft screening, overdiagnostikk, konsekvenser av overdiagnostikk.

Abstract

Research question:

How much overdiagnosis is there in mammographic screening and why are there so much disagreement?

Purpose:

Provide insight in to how much overdiagnosis there is in mammographic screening. How have the scientists come up with different results and why there is so much disagreement on the estimation of overdiagnosis?

Method:

In this thesis we have used a qualitative approach and selected literature study method. The literatures we have used in this thesis are scientific articles and some books. Articles included in this study were all review articles that used different methods to estimate overdiagnosis.

Result:

Which method is used in the estimation of overdiagnosis has a lot to say for the results. We have used a total of 6 articles where all these articles have used different methods to estimate overdiagnosis. There are two methods that are most preferred, the cumulative incidence method and incidence-rate method. The major disagreement is which method and study design that should be used to conduct the study.

Conclusion:

Overdiagnosis estimates in mammographic screening is a complex topic due to the variations of the methods used to estimate overdiagnosis. Overdiagnosis estimates vary from -4 to 69 % in the different studies. It is difficult to reach an agreement as long as the scientists have not come up with a common approach.

Keywords:

Mammography screening, breast cancer screening, overdiagnosis, consequence of overdiagnosis

INNHOLDFORTEGNELSE

Forord.....	1
Sammendrag.....	2
Abstract.....	3
1 INNLEDNING.....	6
2 METODE.....	7
2.1 Vurdering av litteratur.....	7
2.2 Inkluderings- og ekskluderingskriterier.....	8
2.3 Søk.....	9
2.4 Litteratur.....	9
2.5 Metodekritikk.....	10
2.5.1 Litteraturstudie som metode.....	10
3 TEORI.....	11
3.1 Hva er mammografiscreening?.....	11
3.2 Hva er overdiagnostikk?.....	12
3.2.1 Hva er konsekvensen av overdiagnostikk.....	12
3.3 Hvilke metoder brukes for å påvise overdiagnostikk.....	13
3.3.1 Insidensrate metode.....	13
3.3.2 Kumulativ innsidens metode.....	14
3.3.3 Tilnæringsmetode.....	15
3.4 Estimering av overdiagnostikk.....	15
3.5 Bias.....	15
3.5.1 Lead time bias.....	16
3.5.2 Self-selection bias.....	17
3.5.3 Length time bias.....	17
4 RESULTAT.....	18

4.1	Hva sier forskingsresultatene om omfanget av overdiagnostikk?.....	18
4.1.1	Effects of study methods and biases on estimates of invasive breast cancer over-detection with mammography screening: a systematic review, Biesheuvel et al. (2007)	18
4.1.2	Interpreting Overdiagnosis Estimates in Population-based Mammography Screening, Gelder et al. (2011)	20
4.1.3	The other side of technology: risk of overdiagnosis of breast cancer with mammography screening, Puliti and Paci (2009)	21
4.1.4	Complexities in the estimation of overdiagnosis in breast cancer screening, Duffy et al. (2008)	21
4.1.5	The benefits and harms of breast cancer screening: an independent review, Independent UK panel on breast cancer screening (2012).....	22
4.1.6	Overdiagnosis in cancer, Welch and Black (2010)	23
4.2	Hva er man uenige om når det gjelder overdiagnostikkestimater?	24
4.3	Oppsummering av resultater	24
5	DISKUSJON	25
5.1	Hvor stor er overdiagnostikk.....	25
5.2	Hvorfor er det så stor uenighet.....	27
5.3	Hvilken metode er mest foretrukket.....	27
5.4	Kritikk	28
5.4.1	Søk.....	28
5.4.2	Reliabilitet og validitet	29
6	KONKLUSJON	30
7	LITTERATURLISTE	31

1 INNLEDNING

Mammografiscreening gjennomføres for å detektere brystkreft i tidlig stadium og starte behandlingstiltak tidligst mulig for å øke prognosen. I Norge startet det offentlige mammografiscreeningsprogrammet i 1996. Det forskes mye på de positive og negative sidene ved mammografiscreening. Hovedfokuset er hvor stor nytteverdien i form av mortalitets reduksjon ved brystkreft er i forhold til hvor mye skade den gjør i form av overdiagnostikk. Overdiagnostikk betraktes som det mest negative utfallet ved mammografiscreening.

Dette er et aktuelt, spennende og interessant område samt et omdiskutert tema i media. Hensikten med oppgaven er å finne ut hvor mye overdiagnostikk det er ved mammografiscreening og hvorfor det er så stor uenighet. Uenighetene ved overdiagnostikk er et omdiskutert tema og ses ikke bare i vitenskapelige artikler, men også i internasjonal og nasjonal media er det stadig debatter om hvor stor overdiagnostikk ved mammografiscreening faktisk er. Overdiagnostikk er kreftsvulster som detekteres i et tidlig stadium ved mammografiscreening, ofte utvikler de seg så langsomt at de ikke vil bli merkbar i kvinnens levetid. Siden man ikke kan forutse om svulster er benigne eller maligne, vil disse kvinnene bli behandlet selv om de ikke ville trengt det.

Mammografi er et område radiografer jobber med daglig og det er relevant for oss som snart skal ut i arbeidslivet å ha kunnskap innen dette området. På bakgrunn av dette er vår valgte problemstilling som følger:

”Hvor mye overdiagnostikk er det ved mammografi screening og hvorfor er det så stor uenighet?”

For å svare på problemstillingen har vi valgt og utføre en litteraturstudie. Ved å søke igjennom databaser som var tilgjengelige via høgskolebiblioteket på Gjøvik har vi funnet artikler som er relevante for vår valgte problemstilling.

2 METODE

For å svare på oppgavens problemstilling har vi valgt og utføre en litteraturstudie. Det har blitt utført store mengder av studier som handler om både positive og negative sider ved mammografiscreening. Vi har valgt å skrive om de negative sidene ved mammografiscreening og vi vil avgrense oppgaven til kun overdiagnostikk.

Som sagt er det et stort antall studier som blir utført innen mammografiscreening, derfor har vi valgt å begrense søket vårt til studier av nyere dato. Dette både med tanke på utviklingen og fremgang i forskningen og for å begrense antall treff. Artikkene vi skal benytte i oppgaven er fra 2007 frem til dags dato 12.04.2013. I oppgaven skal det benyttes 6 hovedartikler.

En av oss (Delia Arbildo-Tverfjell) har deltatt på seminar den 31. januar 2013 som omhandlet fordeler og ulemper ved mammografi screening og hva som var de faglige uenighetene.

Hovedfokuset på seminaret var hvilke forskjellige metoder forskere benytter og hvordan de har kommet fram til de forskjellige resultatene når det gjelder mortalitet og overdiagnostikk. Det var over 200 mammografi eksperter og interesserte fagfolk som deltok på seminaret fra både innland og utland. Bakgrunnen for deltakelsen var å samle inn mest mulig informasjon og lære mer om fagområdet. Foredragsholderne i seminaret fokuserte mye på overdiagnostikk som var relevant for vår oppgave.

2.1 Vurdering av litteratur

For å vurdere kvaliteten på artiklene har vi gått igjennom innholdet, lest abstrakt og metode for å forsikre oss om at artiklene inneholder det som er relevant for vår oppgave. Ved hjelp av systematisk databasesøk har vi samlet relevante artikler i forhold til problemstillingen. Dette er viktig fordi litteraturstudier handler om å være kritisk til artiklens innhold. Det er først og fremst viktig for vår egen del og dokumentere søkene og vurdere litteraturen mot kriteriene som er satt for vurdering av litteratur.

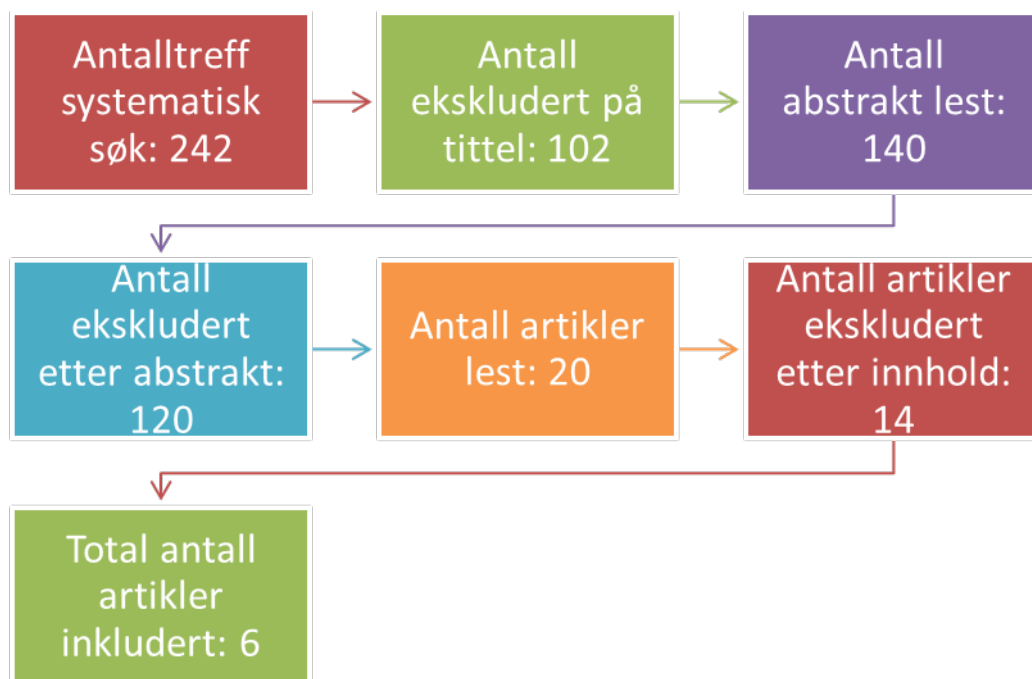
For å sikre artiklens kvalitet og vurdere hvilket vitenskapelig nivå de forskjellige artiklene har, har vi brukt publiseringskanaler via Norsk samfunnsvitenskapelig datatjeneste (NSD). Vi benytter artikler som har både vitenskapelig nivå 1 og 2. I tillegg til å sjekke vitenskapelignivå på artiklene har vi benyttet kunnskapssenterets retningslinjer for og kvalitetssikre artiklene. Her har vi gått igjennom om artiklene har formålet med oversikten

klart formulert, om det er klare inkluderingskriterier, om relevante studier er funnet, om bias er vurdert, om analysen er gjennomført på en tilfredsstillende måte, hva forteller resultatene, er de presise og kan de være til hjelp i praksis.

Et av hovedpoengene med systematisk vurdering er at man følger en streng protokoll for å sikre at prosessen som brukes er systematisk ved å bruke nøyaktige og strenge metoder for å identifisere og kritisk vurdere relevante studier for å kunne svare på et forhåndsdefinert spørsmål (Aveyard, 2010).

2.2 Inkluderings- og ekskluderingskriterier

For å vurdere hvilke artikler som skulle inkluderes eller ekskluderes i oppgaven har vi gjort en systematisk gjennomgang av alle treffene fra søkene gjort i databasene. Først valgte vi artiklene ut i fra tittelen, deretter leste vi hver av artiklenes abstrakt hvor vi valgte de artiklene vi mente var relevante til vår oppgave og til slutt gjennomlesing av de relevante artiklene. Deretter ekskluderte vi artikler ut i fra innhold og ble sittende igjen med de artiklene som vi mente ga svar på vår valgte problemstilling. Nedenfor har vi laget en tabell for å illustrere fremgangsmåten for inkluderings- og ekskluderingskriteriene.



Tabell 1: fremgangsmåte for ekskludering og inkludering av artikler.

2.3 Søk

Tabellen under viser hvilken strategi vi hadde for gjennomføring av databasesøk i forbindelse med oppgaven. Vi har benyttet totalt fire søkeord i forbindelse med databasesøk i de forskjellige databasene. Artikkene som vi har benyttet i oppgaven er fra databasene PubMed, Science Direct og Proquest. Ved databasene PubMed og Proquest benyttet vi søkeordene ”mammografi screening AND breast cancer screening AND overdiagnosis”, mens ved Science Direct benyttet vi søkeordene ”mammografi screening AND breast cancer screening AND consequence of overdiagnosis”. De fleste artiklene ble hentet fra PubMed og Science Direct siden disse databasene ga et godt utvalg av artikler i forhold til oppgaven.

Søkeord	Begrensninger	Database	Kombinasjoner	Antall treff	Antall artikler benyttet
1 Mammography screening	All fields, Peer reviewed, 2007- today	PubMed	1+2+3	12	3
2 Breast cancer screening	All fields, Peer reviewed, 2007- today	Science Direct	1+2+4	35	2
3 Overdiagnosis	All fields, Peer reviewed, 2007- today	Proquest	1+2+3	195	1
4 Consequence of overdiagnosis	All fields, Peer reviewed, 2007- today	Proquest	1+2+3	195	1

Tabell 2: oversikt over databasesøk gjennomført i perioden 20.11.12- 12.04.13.

2.4 Litteratur

Her gjør vi rede for valg av litteratur vi har brukt i oppgaven. Kildene vi har valgt å benytte i oppgaven har vi valgt på bakgrunn av relevans og at vi finner innholdet troverdig. Vi har benyttet et utvalg bøker som vi mener er relevante i forhold til oppgavens problemstilling. I metoddelen har vi benyttet Aveyard (2010) og Dalland (2007) som litteratur siden disse bøkene gir en god innføring av fremgangsmåten for en litteraturstudie og er strukturert og konkret. I tillegg har vi brukt Gøtzsche (2012), Bjørndal (2004) og Ertzaas (2003) i teoridelen. Gøtzsche (2012) og Ertzaas (2003) er relevante bøker som gir en god innføring i

mammografiscreening, mens Bjørndal (2004) er en god veiledning innen statistikk for helsepersonell. I tillegg til overnevnt litteratur har vi brukt Norsk samfunnsvitenskapelig datatjeneste. Dette er en side som har ansvar for registret over autoriserte publiseringskanaler som for eksempel tidsskrifter, serier og forlag. Denne siden har vi brukt for å kvalitetssikre artiklene vi har brukt i oppgaven.

2.5 Metodekritikk

Vi hadde i første omgang tenkt å oppveie fordeler og ulemper ved mammografiscreening, men vi fant ut at dette ble for stort for en bacheloroppgave og valgte derfor å avgrense oppgaven til risikofaktorer. Vi vurderte så å skrive om overdiagnostikk og falsk negativ/positiv svar men selv om dette var to interessante risikofaktorer som kan avdekkes valgte vi heller å fokusere på overdiagnostikk.

2.5.1 Litteraturstudie som metode

Hovedprinsippet ved en litteraturstudie er å oppdage nye innblikk innen forsknings temaer ved å gjennomføre en systematisk vurdering av litteraturen som benyttes (Aveyard 2010). En ulempe ved litteraturstudie er at fagstoff og pensum ofte er sekundærlitteratur og kan føre til feiltolkning av stoffet og at det opprinnelige perspektivet kan ha endret seg. Derfor er det viktig at man bør tilstrebe og bruke primærkilder for å sikre seg fra å arve andres feiltolkning eller unøyaktigheter (Dalland 2007). Dette ser vi at kan bli vanskelig å gjennomføre siden mye av forskningen er basert på andre studier. Vi har i denne oppgaven benyttet oversiktsartikler fordi disse samler kunnskap fra mange andre artikler og er med på å kvalitetssikre artiklene for oss.

3 TEORI

Her presenterer vi relevant teori til vår valgte problemstilling.

3.1 Hva er mammografiscreening?

Mammografiscreening er en organisert masseundersøkelse som inviterer kvinner i en bestemt aldersgruppe til mammografiundersøkelse. Mammografiprogrammet ble startet i Norge i 1996. Den hyppigste kreftformen hos kvinner er brystkreft og hvert år rammes mer enn 2400 kvinner av brystkreft i Norge. Av disse kvinnene er omtrent 1000 i aldersgruppen 50 til 69 år, men forekomsten av brystkreft øker i alle aldersgruppene (Ertzaas 2003).

Hensikten med mammografiscreening er at den reduserer dødeligheten ved brystkreft i alderen 50 til 69 år ved en 15 års tid eller lenger med 15 til 30 %. Et screeningprogram bør kun iverksettes når fordelene overstiger ulemperne (Harris, Yeatts et al. 2011). Screening har ikke bare fordeler, den har også viktige risiko faktorer. Derfor er det viktig at kvinnene informeres om mulig fordeler og ulemper, slik at de har forståelse av hva mammografiscreening innebærer (Brodersen, Jorgensen et al. 2010). Informasjonen bør inkludere overdiagnostikk, slik at kvinnen er klar over mulige risikofaktorer ved screening før de tar et valg om å delta i screeningprogrammet. Det diskuteres om det er hensiktsmessig å informere kvinner om risiko for overdiagnostikk før de velger å delta i screening. Mange mener at det er hensiktsmessig, mens andre er skeptisk til dette fordi de mener at det er mulig at slik informasjon kan fraråde kvinner fra å delta i screening (Hersch, Jansen et al. 2013).

Det er umulig å skille mellom lesjoner som vil utvikle seg til invasiv kreft og de som ikke utvikler seg under pasientens levetid og derfor blir disse pasientene behandlet etter protokollene. I Nyere studier kommer det frem at oppdagede lesjoner som er blitt behandlet unødvendig er så høy som 10 til 30 %, avhengig av alder og intensitet av screeningen. Det er reist viktig bekymring i disse studiene om de relative fordelene ved mammografiscreening. Det ble beregnet at screening av 1000 kvinner fra 50 år årlig i ti år vil resultere i at en kvinne vil unngå død av brystkreft, mens 2 til 10 kvinner blir overdiagnostisert og behandlet unødvendig, blir ca. 10 til 15 kvinner fortalt at de har brystkreft tidligere enn hva som ville vært tilfelle og 100 til 500 kvinner med minst en falsk alarm, hvor omtrent halvparten vil gjennomgå biopsi (Corbex, Burton et al. 2012).

3.2 Hva er overdiagnostikk?

Overdiagnostikk er svulster som oppfyller de histologiske kriteriene for kreft, men som aldri ville blitt klinisk tydelig i pasientens levetid. Selv om svulsten blir oppdaget tidlig, kan det gå flere år før symptomene utvikler seg og kvinnen vil mest sannsynlig dø av andre årsaker før hun vil dø av kreften. Kvinnen vil da bare oppleve de skadelige effektene av tidlig diagnose og behandling. Siden det ikke er mulig å gjenkjenne hvilke individuelle tilfeller av brystkreft som faktisk resulterer fra overdiagnostikk vil kvinnen bli behandlet etter protokollene (Puliti and Paci 2009, Puliti, Duffy et al. 2012)

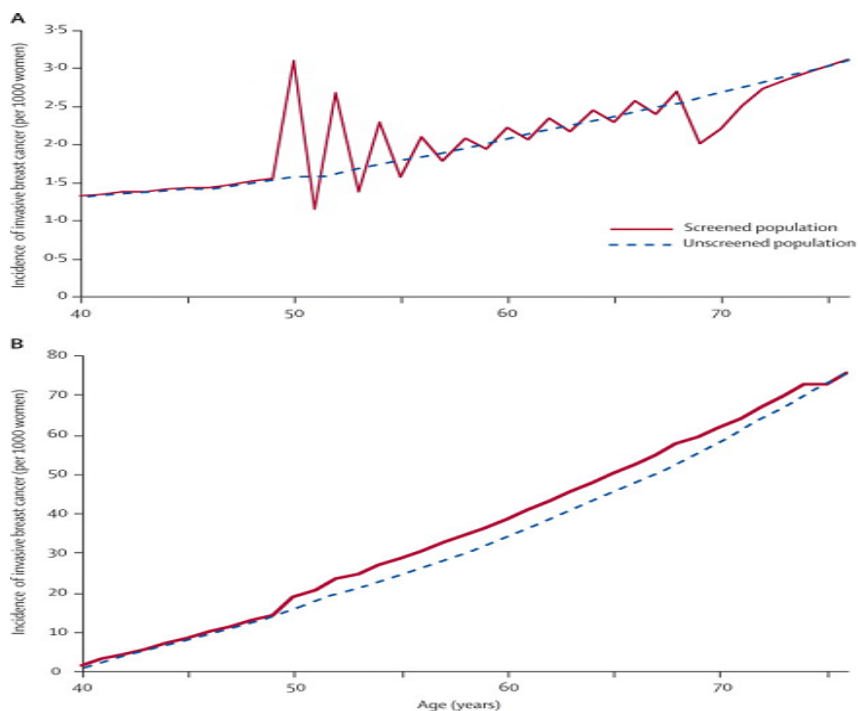
Overdiagnostikk kan ha en av to forklaringer. Kreften utvikles aldri eller kreften utvikler seg langsomt nok til at pasienten dør av andre årsaker. Overdiagnostikk bør ikke bli forvekslet med falsk positiv resultater, da dette er en positiv test i et individ som senere ved etterfølgende undersøkelser ikke har kreft (Welch and Black 2010). Hensikten med screening er å senke dødeligheten av kreft i populasjonen. Screening forårsaker alltid skade og det viktige spørsmålet da er om screening reduserer dødeligheten av kreft og om reduksjonen er stor nok til å rettferdiggjøre de skadene den påfører den friske befolkningen (Gøtzsche 2012).

3.2.1 Hva er konsekvensen av overdiagnostikk

Overdiagnostikk resulterer ofte i overbehandling av harmløse lesjoner (Hersch, Jansen et al. 2013). Behandlingen kan innebære kirurgi og enkelte tilfeller med strålebehandling og medisiner. Overdiagnostikk og overbehandling er ansett som den mest alvorlige skaden assosiert med tidlig påvisning av kreft via screening (Independent 2012, Hersch, Jansen et al. 2013). En overdiagnostisert person vil lite trolig få nytte av screeningen og kan i stedet få både psykologiske og emosjonelle plager i følge av diagnostisk og behandling (Hersch, Jansen et al. 2013).

3.3 Hvilke metoder brukes for å påvise overdiagnostikk

I følge Biesheuvel et al. (2007) er det tre brede metoder for å estimere overdiagnostikk: insidens-rate metoden (insidence–rate method), den kumulative insidens metoden (cumulative incidence method), og tilnæringsmetode (modelling approaches). Teoretisk sett er kumulativ insidens metode den beste metoden for å påvise overdiagnostikk (Biesheuvel, Barratt et al. 2007).

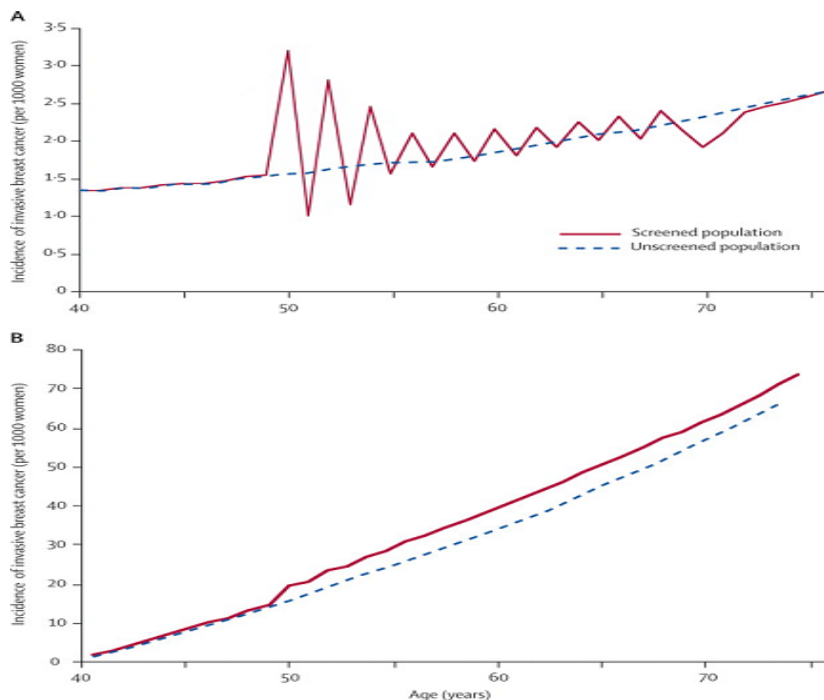


Figur1: Illustrasjon viser effekten av årlig screening av kvinner i alderen 50-68 år på forekomsten av invasiv brystkreft i fravær av overdiagnostikk. Alle kvinnene som gjennomgikk screening gjorde det i samme år. Figur A fremstiller forekomst-rate metoden, mens figur B fremstiller Kumulativ insidens metode. Kumulativ insidensen var høyere for den screenede populasjonen enn for den uscreenede populasjonen pga. lead time effekten (Biesheuvel, Barratt et al. 2007).

3.3.1 Insidensrate metode

Forskerne kan sammenligne den gjennomsnittlige årlige insidensen av brystkreft i en bestemt aldersgruppe over en viss periode i screenede og uscreenede kvinner på samme alder. Når overdiagnostikk ikke er til stede, vil forskjellen i insidensen være utelukkende i følge av lead

time (forskyvning av diagnosetidspunkt). Økt forekomst i screenet gruppe etter justering for lead time viser overdiagnostikk (Biesheuvel, Barratt et al. 2007).



Figur2: Illustrasjon viser effekten av annethvert år screening av kvinner i alderen 50-68 år på forekomsten av invasiv brystkreft ved tilstedeværelse av overdiagnostikk. Alle kvinnene som gjennomgikk screening gjorde det i samme år. Figur A fremstiller forekomstrate metoden, mens figur B fremstiller Kumulativ insidens metode. Kumulativ insidensen var høyere for den screenede populasjonen enn for den uscreenede populasjonen pga. lead time effekten og overdiagnostikk (Biesheuvel, Barratt et al. 2007).

3.3.2 Kumulativ insidens metode

Med denne metoden sammenligner man den kumulative insidens av en gruppe screenede kvinner mellom en gruppe som ikke er screenet over samme tidsperiode. Alderen mellom de to gruppene kan for eksempel være etter fylte 50 år til 68 år. Insidensen av brystkreft reduseres flere år etter at screening har stoppet i følge av lead time, derfor bør begge gruppene overvåkes i flere år etter at screening har stoppet. Hvis overdiagnostikk ikke er tilstede vil den kumulative insidensen etter denne perioden være identisk i begge gruppene. Hvis overdiagnostikk er tilstede i følge av screening vil det være et overskudd av brystkreft tilfeller i den screenede gruppen sammenlignet med den uscreenede gruppen (Biesheuvel, Barratt et al. 2007).

3.3.3 Tilnæringsmetode

Denne metoden kan rangeres fra enkle modeller som omfatter færre sykdomstilstander og komplekse modeller med mange sykdomstilstander. I tilnæringsmetoden er det en avveining mellom kompleksiteten, åpenheten, og validiteten av modellen og dens resultat. For eksempel, den enkle modellen krever færre forutsetninger, den har færre data krav og kan være mer transparente, men deres enkelhet kan føre til utilstrekkelige representasjoner av de virkelige data. Komplekse modeller kan mer nøyaktig oppfatte virkelige sykdoms prosesser og er generelt mindre transparente. På grunn av denne mangelen av transparens er det vanskelig å vurdere om de er utsatt for bias (Biesheuvel, Barratt et al. 2007). Tilnæringsmetoden er lite omtalt i de artiklene vi bruker og vil derfor ikke bli tatt med i diskusjonsdelen.

3.4 Estimering av overdiagnostikk

Kompleksiteten av tolkning av kreft insidens data i screeningperioden ved primærforskning og oversiktsartikler, er viktig å ta i betraktning ved estimering av overdiagnostikk. I tillegg til dette er det viktig at man tar i betraktning lead time fenomenet, overdiagnostikk og endring av insidensdata (Duffy, Lynge et al. 2008). Overdiagnostikk kan estimeres fra randomiserte kontroll studier eller fra observasjonsstudier. Valide estimater er avhengig av tilsvarende bakgrunn av brystkreft risiko for screenede og uscreenede kvinner og at effekten av lead time er tatt i betraktning. Det finnes ikke en enkel optimal måte for å estimere overdiagnostikk på, men de to nyttige estimatene er: fra befolkningsperspektiv, andelen av krefttilfeller noensinne diagnostisert hos kvinner invitert til screening som er overdiagnostisert og fra perspektivet til en kvinne invitert til screening hvor sannsynligheten for diagnostisert kreft i løpet av screening perioden blir presentert som overdiagnostikk (Independent 2012).

3.5 Bias

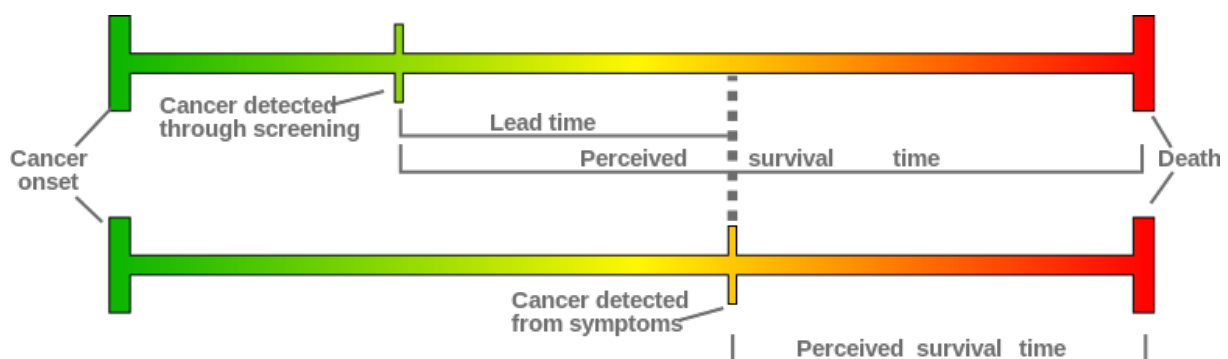
Bias har opphav av det franske ordet bias, som betyr skjevhet. Ordet bias har ingen betydning alene, derfor må det kobles med ett annet ord for at det skal ha allmenn akseptert mening (Staff, 2010). Bias innen forskning omtales som en systematisk feil som kan oppstå ved at undersøkelses design eller gjennomføring har mangler. Forskere har en måte å gå frem for å

gjennomføre studier innen et gitt tema. Det vil si at forskeren må ha forskningsdesign, målemetoder, datainnsamling, og oppfølging av det som skal forskes på. Den systematiske feilen kan oppstå i alle de overnevnte leddene, dette vil si at feilen går igjen hele tiden og ødelegger sammenlikningsgrunnlaget (Bjørndal and Hofoss 2004).

Innen forskning benyttes det mange forskjellige bias, men når det gjelder forskning innen mammografiscreening er det tre bestemte bias typer som er av stor betydning for forskerne. De tre typene bias som brukes er lead time bias, self selection bias, og lenght time bias. Nedenfor skal vi definere alle de tre typene bias slik at man kan få en forståelse av hva forskerne må tilstrebe med.

3.5.1 Lead time bias

Lead time bias refererer til tidsintervallet mellom diagnose av kreft og død. Siden screening fremskynder tidspunktet for diagnose vil pasienter med screeningoppdaget kreft ha bedre overlevelse enn klinisk påviste tilfeller, selv om screeningen ikke gjør noe for å utsette tidspunktet for dødsfallet. Bias henviser til intervallet mellom diagnosen av kreft ved screening og ved vanlig klinisk påvisning. Som et resultat av lead time bias, kan det se ut som screeningen forlenger livet, når det i teorien bare forlenger diagnostidspunktet. Screening av brystkreft viser bedre overlevelse enn klinisk påvisning, men screeningen viser ingen reduksjon av dødeligheten. Så å sammenlikne brystkreft overlevelse mellom screenede og uscreenede grupper er meningsløst. Derfor må man i stedet avgjøre om screening virkelig reduserer dødeligheten av brystkreft (Jatoi 1999).



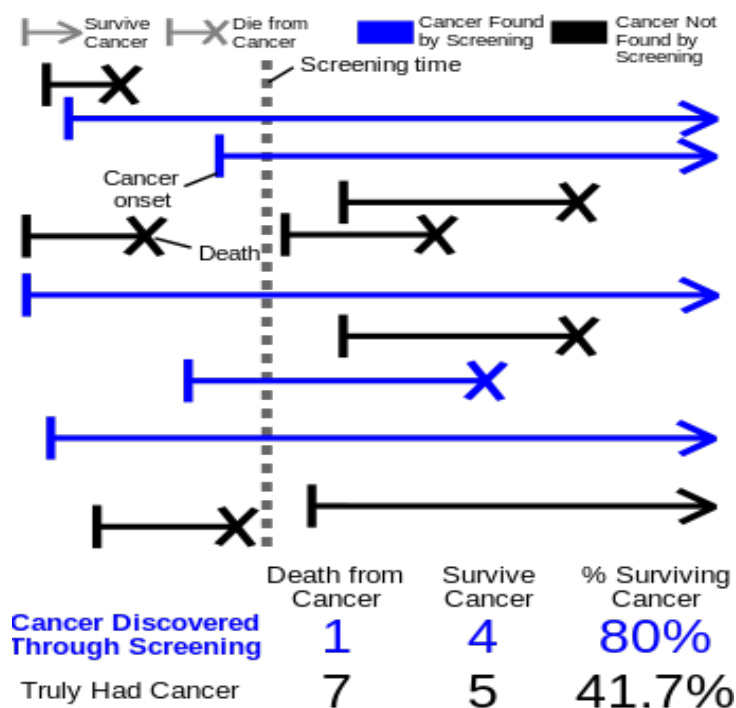
Figur3: Illustrasjon av lead time bias (Mcstrother2011a).

3.5.2 Self-selection bias

Self selection bias er et fenomen hvor kvinnen som velger å delta i screeningprogrammet er helsebevisst og har dermed en lavere mortalitet rate. Det gjøres en sammenligning mellom de kvinnene som velger å delta i screeningprogrammet og de som takker nei til screening. Ved sammenligning var brystkreft dødeligheten høyere for de som takket nei til screening enn for de som deltok i programmet. En sammenligning av screenede og uscreenede kvinner kan føre til utvalgsskjevhet. Derfor må man ta i betraktning en rekke faktorer som kan føre til skjevhet i studiets resultat, som for eksempel genetiske risiko faktorer, om kvinnen har genetisk risiko for å få brystkreft eller ikke (Jatoi 1999).

3.5.3 Length time bias

Screening har en tendens til å oppdage kreft som vokser langsomt og er derfor biologisk gunstig. Dette fenomenet kalles for lengt time bias. Saktevoksende krefttyper eksisterer i en lengre periode i den prekliniske fasen og det er dermed mer sannsynlig at den blir detektert ved screening. Hurtigvoksende krefttyper eksisterer kun i en kort periode i preklinisk fase, dermed er det mer sannsynlig at disse blir diagnostisert i intervallene mellom to screeninger. Ved å sammenligne de biologiske egenskapene mellom brystkreft som blir detektert ved screening og brystkreft som blir detektert klinisk, fant man ut at screening detektert tumor var mer gunstig (Jatoi 1999).



Figur4: Illustrasjon av lengt time bias (Mcstrother2011b)

4 RESULTAT

Her skal det presenteres et kort sammendrag av forskningsresultatene fra de forskjellige artiklene som benyttes i oppgaven. Artiklenes resultater vil bli presentert hver for seg, slik at det blir mer oversiktlig.

4.1 Hva sier forskningsresultatene om omfanget av overdiagnostikk?

Alle kildene vi benytter sier at det finnes overdiagnostikk ved mammografiscreening, men omfanget varierer fra studie til studie. Derfor har vi i slutten av resultat delen valgt å oppsummere de forskjellige resultatene fra hver artikkel.

4.1.1 Effects of study methods and biases on estimates of invasive breast cancer over-detection with mammography screening: a systematic review, Biesheuvel et al. (2007)

Denne artikkelen er skrevet av Biesheuvel et al.(2007). Biesheuvel et al. gjennomfører en systematisk vurdering av totalt 8 artikler for å finne ut omfanget av overdiagnostikkestimer ved invasiv brystkreft. For hver artikkel vurderte de studiedesignen og metoden for å kartlegge omfanget og effekten av bias. De har funnet totalt 22 estimer av overdiagnostikk fra de forskjellige kildene de har benyttet. Av de 8 artiklene de har brukt, var det 6 artikler som hadde benyttet kumulativ innsidens metode og de resterende 2 artiklene hadde benyttet insidens-rate metoden. Biesheuvel et al. presenterer følgende overdiagnostikkestimer som de beste. Disse estimatene rangerer fra -4 til 7,1 % for kvinner i alderen 40 til 49 år, 1,7 til 54 % for kvinner i alderen 50 til 59 år og 7 til 21 % for kvinner i alderen 60 til 69 år. I tillegg til dette presenterer vi en forenklet tabell fra Biesheuvel et al. (2007) som viser en oversikt av overdiagnostikkestimer fra metodene presentert over.

Studie	lignende risiko for utvikling av brystkreft for screenede og uscreenede grupper	Kontroll gruppen ble ikke screenet på slutten av studiet	Justering for lead time	Alders grupper	Overdiagnostikk estimatet (%) etter beregning av relativ risiko eller absolutt risiko
--------	---	--	-------------------------	----------------	---

Malmø	Ja	Ja	Nei ↑	45-69	23,6
Malmø	Ja	Ja	Nei ↑	45-69	32,0*
Canada 1	Ja	Ja	Ikke akt.	40-49	7,1
Canada 1	Ja	Ja	Nei ↑	40-49	30,0*
Canada 2	Ja	Ja	Ikke akt.	50-59	1,7
Canada 2	Ja	Ja	Nei ↑	50-59	25,9*
Stockholm	Ja	Nei ↓	Nei ↑	40-64	-4,7
Stockholm	Ja	Ja	Nei ↑	40-64	49,0*
Gothenburg	Ja	Nei ↓	Nei ↑	39-59	-10,0
Gothenburg	Ja	Ja	Nei ↑	39-49	13,0*
Two county	Ja	Nei ↓	Nei ↑	40-74	-5,1
Two county	Ja	Ja	Nei ↑	40-74	32,8*
Healt insurance plan	Ja	Ja	Nei ↑	40-64	4,6*
Edinburgh	Nei ↓	Ja	Nei ↑	45-64	5,7
Zackarisson et al. (sweden)	Ja	Ja	Ikke akt.	45-54 55-69 45-69	8,0 7,0 7,0
Peeters et al. (Netherlands)	Ja	Ja	Nei ↑	35-44 45-54 55-64 >65 ≥ 35	8,0* 41,0* 19,0* -13,0* 11,0*
Paci et al. (Italy)	Nei ↑	Ja	Ja	50-84	2,0
Paci et al. (Italy)	Nei ↑	Ja	Ja	50-54 55-59 60-64 65-69 70-74 50-74	7,4 -0,6 0,7 5,7 9,7 4,6

Tatt i betraktning (Ja), Ikke tatt i betraktning (Nei), Ikke aktuelt (ikke akt.), pilene viser underestimert (↓) og overestimert (↑). * viser beregning for alle brystkrefttyper, hvor de har inkludert Ductal carcinoma insitu (DCIS).

Tabell 3: viser vurderinger av metodologisk og skjevhets faktorer som kan påvirke estimering av overdiagnostikk i studier som bruker den kumulative insidens metode (Biesheuvel et al. 2007).

Studie	lignende risiko for utvikling av brystkreft for screenede og uscreenede grupper	Kontroll gruppen ble ikke screenet på slutten av studiet	Justering for lead time	Alders grupper	Overdiagnostikk estimatet (%) etter beregning av relativ risiko
Zahl et al. (sweden)	Nei ↑	Nei ↓	Nei ↑	50-69 70-74 75-78	45 1 -12
Zahl et al. (Norway)	Nei ↑	Nei ↓	Nei ↑	50-69 70-74	56 -11

Johnsson et al. (sweden)	Nei ↑	Ja	Ja	40-49	22
				50-59	54
				60-69	62
				70-74	84
Johnsson et al. (Sweden)	Nei ↑	ja	Ja	40-49	-4
				50-59	54
				60-69	21
				70-74	3
Tatt i betraktning (Ja), Ikke tatt i betraktning (Nei), pilene viser underestimert (↓) og overestimert (↑).					

Tabell 4: viser vurderinger av metodologisk og skjevhets faktorer som kan påvirke estimering av overdiagnostikk i studier som bruker insidens rate metoden. (Biesheuvel et al. 2007).

4.1.2 Interpreting Overdiagnosis Estimates in Population-based Mammography Screening, Gelder et al. (2011)

Denne artikkelen er skrevet av Gelder et al.(2011). Gelder mener at overdiagnostikkestimater ved mammografiscreening rangerer fra 1 til 54 %. Artikkelen forklarer hvorfor overdiagnostikkestimater varierer så mye. Dette viser Gelder ved å benytte et eksempel av gradvis implementering av mammografiprogrammet i Nederland. Overdiagnostikkestimater ble gjennomført årlig fra 1990 til 2006, hvor de inkluderte karsinom in situ og invasiv brystkreft.

Gelder beregnet overdiagnostikk ved å benytte forskjellige metoder identifisert i litteraturen de har brukt. Overdiagnostikkestimater i denne artikkelen er basert på modellert innsidens av brystkreft og den forventede innsidens uten screening. Når modellert brystkreft innsidens i en screening situasjon blir sammenlignet med den forventete innsidens uten screening, vil estimert overdiagnostikk i den totale befolkningen i gjennomføringsfasen av screening øke fra 1,0 % av all forventet brystkrefttilfeller i 1990 til 11,4 % i 1993. I 1993 toppet den modellerte overskudd av brystkrefttilfeller seg til 17,1 % av all forventet kreft hos kvinner i alderen 50 til 69 år, mens den modellerte underskudd i innsidens blant kvinner ikke lenger screenet var 0,8 % av all forventet kreft i denne aldersgruppen.

Overdiagnostikkestimatet ble redusert til 5,6 % i 1997, dette fordi det var flere kvinner som hadde etterfulgt screening. Under utvidelsesfasen, økte overdiagnostikkestimatet til 10,0 % i 1999, hvor det i 2001 ble redusert til 4,7 %. I løpet av stabil stand fase (steady-state phase) av screening økte estimatet først til 4,9 % i 2003, men falt til 2,8 % av all forventet brystkreft i 2006. I 2006 var overskuddet av brystkreft i aldersgruppen 50 til 69 år på 7,0 %. Avhengig av

hvilken nevner som brukes for å definere populasjonsrisikoen kan overdiagnostikk øke til 8,9 % hvis raten er beregnet som en fraksjon av alle screenede krefttilfeller. Hvis det er beregnet som en fraksjon av alle diagnostiserte tumorer mellom kvinner i screening alder og i en screening situasjon, blir estimatet 4,6 %. De estimerte ratene av overdiagnostikk ble beregnet som en relativ økning blant kvinner i screening alder og eldre og for kvinner som kun er i screening alder var estimer henholdsvis 3,6 % og 5,0 %. Hvis man ikke tar i betraktning for lead time vil beregningen av overdiagnostikk raten for kvinner i screening alder bli 9,7 %.

4.1.3 The other side of technology: risk of overdiagnosis of breast cancer with mammography screening, Puliti and Paci (2009)

Artikkelen er skrevet av Puliti og Paci (2009). I denne artikkelen diskuterer de konseptene og definisjonen av overdiagnostikkestimater, og presenterer ulike studie metoder for å estimere overdiagnostikk av brystkreft ved mammografiscreening. Overdiagnostikkestimater de kommer frem til i artikkelen er fra randomiserte kontroll studier og observasjonsstudier. Forfatterne har analysert 8 randomiserte kontroll studier, hvor de skiller mellom to typer studie metoder. I de tre svenske studiene (Two-country study, Stockholm studie og Gøteborg studie) har kontrollgruppen blitt tilbudt screening. I de resterende studiene (HIP, Malmø studie, Canadian 1 og 2 og Edinburg studie) har de ikke blitt tilbudt screening av kontrollgruppen. Ved randomiserte kontroll studier har de benyttet den kumulative insidens metoden og kommer frem til at overdiagnostikkestimater rangerer fra 10 til 14 %. I Observasjonsstudiene har de benyttet statistisk justering for lead time for å estimere overdiagnostikk. I observasjonsstudiene varierte estimatene i de forskjellige studiene. Her ble estimatet lavere enn 5 % i to studier, mens det ble mye høyere i en annen studie hvor estimatet rangerte fra 21 til 54 %.

4.1.4 Complexities in the estimation of overdiagnosis in breast cancer screening, Duffy et al. (2008)

Artikkelen er skrevet av Duffy et al. (2008). Her viser de hvordan screening presenterer kompleksiteter i en analyse og tolkning av innsidens data. For eksempel bringer lead time frem tid og aldersrelatert økning i insidens. I tillegg til dette kan risikofaktorer som hormonterapi være med på å påvirke estimeringen av overdiagnostikk. De sier ikke noe om hvordan en nøyaktig kan korrigere slik kompleksitet, men de bruker noen enkle

uformelle justeringer for å vise hvordan de kan utgjøre en betydelig økt insidens. Dette illustrerer de ved å bruke eksempler av vurderinger ved brystkreft insidens i Sverige.

Her viser de brystkreft insidensen av kvinner i Sverige fra alderen 50 til 69 år. Det er tatt med 463 405 deltakere fra elleve fylker. Hvor det ble antatt at brystkreft insidensen hos kvinner i alderen 50 til 69 år i en 7 års periode eller mer i forhold til forventet prescreening tendens var på 1.69. Altså 69 % overdiagnostikk. Etter justering av lead time involverte dette flytting av enkelte av aldersgruppene hvor 40-49 ble flyttet til 50-59 og noen fra 50-59 ble flyttet til 60-69 aldersgruppene, etter justeringen kom de fram til et relativ insidens på 1.54 (54 %). For å vurdere mulige overnevnte kompleksiteter som kan ha hatt innvirkning på rå data på 1.69, har de fjernet effekten av screening prevalens hos kvinner som nærmer seg 50 år. I prevalensfasen var overdiagnostikk insidensen på 84 %. Ved start av screening i alderen 50 omfattet dette rundt 26 % av populasjonen og innen disse fylkene ble 20 % av screenede kvinner i alderen 50 til 59 estimert til å være screening prevalens. Ved justering av lead time ved en spesifikk aldersgruppe økte overdiagnostikk insidensen til ca. 64 % i fravær av screening mellom aldersgruppene 50 til 59 og 60 til 69. Hvis en justerer for faktorer som hormonrelatert terapi har man en relativ overdiagnostikk insidens på 39 %.

4.1.5 The benefits and harms of breast cancer screening: an independent review, Independent UK panel on breast cancer screening (2012)

Artikkelen er utarbeidet av Independent UK panel on breast cancer screening (2012). Hensikten med artikkelen var å finne ut nytteverdien og ulemper ved mammografiscreening. Panelet har i hovedsak fokusert på randomiserte kontrollerte studier av brystkreft screening hvor de sammenligner kvinner som er invitert til screening med kontrollgruppen som ikke var invitert til screening. Her har de også analysert estimer fra observasjonsstudier. Når det gjelder estimering av overdiagnostikk har randomiserte kontroll studier en stor fordel ved design, hvor de sammenligner en gruppe kvinner med samme gjennomsnittlig prognose.

Panelet estimerte at diagnostisert kreft hos kvinner som har blitt invitert til screening i deres levetid og diagnostisering av kvinner i perioden de faktisk er under screeningprogrammet utgjør en overdiagnostikk estimat på 11 og 19 %. Når dataene kombineres for å estimere

nytteverdien og overdiagnostikk, anslår panelet at for 10 000 kvinner invitert til screening ved fylte 50 år i en 20 års periode, vil ca. 681 kreftformer bli oppdaget, hvorav 129 vil representere overdiagnostikk, og 43 dødsfall i følge av brystkreft vil bli forhindret. Det vil si at for hver brystkreft død forhindret vil omtrent tre overdiagnostikk tilfeller bli identifisert og behandlet. Av ca. 307 000 kvinner i alderen 50 til 52 år som blir invitert til screening hvert år vil over 1 % ha en overdiagnostisert kreft i løpet av en 20 års periode.

Panelet mener at data fra tre randomiserte kontroll studier (Malmö, Canada 1 og 2) uten screening ved endt studie gir mest pålitelig estimer av omfanget av overdiagnostikk, men det er en liten andel data og beregningsmetodene er underlagt flere usikkerheter i liket med estimer av mortalitets fordeler. Resultater fra observasjonsstudier viser at overdiagnostikk eksisterer, men estimer av sitt omfang er upålitelig. Panelet sier også at det ikke finnes noen klare bevis som tilsier at nåværende risiko for overdiagnostikk vil være lavere eller høyere enn de originale studiene.

4.1.6 Overdiagnosis in cancer, Welch and Black (2010)

Artikkelen er skrevet av Welch og Black (2010). Denne artikkelen oppsummerer fenomenet av kreft overdiagnostikk. De estimerer størrelsen av overdiagnostikk fra randomiserte kontroll studier, men de benytter også data fra observasjonsstudier. Welch og Black sier at det sterkeste beviset for overdiagnostikk kommer fra langtids oppfølging etter en randomisert kontroll studie av screening. På slutten av studien er det forventet at screening gruppen vil ha et større antall detekterte kreftformer enn i kontroll gruppen. Dette på grunn av at screening fremskyver tidspunktet for diagnose og fremskyver deteksjonen av enkelte krefttyper fremover i tid.

Av de 9 randomiserte studiene Welch og Black har analysert har bare en artikkel rapportert langsiktig oppfølging data av kreft forekomst. Rapporten på 15 år med utvidet oppfølging etter endt studie ga bevis for brystkreft overdiagnostikk. Ved slutten av 10års studiet ble 741 brystkreft tilfeller detektert i screeninggruppen i sammenligning med 591 i kontroll gruppen. I løpet av de etterfølgende 15 årene reduserte denne forskjellen på 150 krefttilfeller til 115, noe som tyder på 35 etterfølgende krefttilfeller. Det vedvarende overskudd av 115 krefttilfeller tyder på overdiagnostikk. Funnene ved endt studie med 35 etterfølgende krefttilfeller viser en kompleksitet i beregningen av overdiagnostikk. Med dette kan man si at 16 % (115 av 741) av

oppdaget kreft i screening gruppen var overdiagnostikk. Alternativt kan man avgrense nevneren kun til screening oppdaget kreft dette fordi overdiagnostikk bare kan skje i denne undergruppen. En tidligere publikasjon av studien viste at 64,4 % av de detekterte kreftilfellene i screeninggruppen, var en konsekvens av screening. Dette tyder på at ca. 477 var detektert ved screening. Ved bruk av denne nevneren er risikoen for overdiagnostikk ca. 24 % (115 av 477).

4.2 Hva er man uenige om når det gjelder overdiagnostikkestimerer?

Uenigheten er om hvor mye overdiagnostikk det er ved mammografiscreening. Dette fordi overdiagnostikkestimerer varierer fra studie til studie. Metodene som benyttes for å estimere overdiagnostikk varierer fordi forskere benytter forskjellige metoder, studiedesign og oppfølgingstid. Det er flere faktorer enn de overnevnte punktene som kan påvirke overdiagnostikkestimeringen, som for eksempel hormonterapi, lead time og alder. I følge Independent UK panel (2012), må minimum oppfølgingstid etter endt studie være fra 5 til 10 år.

4.3 Oppsummering av resultater

Her presenterer vi en kort oppsummering av resultatene til de forskjellige artiklene. Tabellen er blitt utarbeidet ut i fra Biesheuvel et al. (2007) sin tabell i artikkelen og vi har tatt utgangspunkt i denne for å utvide med resultater fra de andre artiklene som er benyttet i oppgaven.

Ref.	Metode	forutsetninger	Alder	Overdiagnostikk estimatet (%)
Biesheuvel et al.(2007)	<i>Kumulativ insidens og insidensrate</i>	<i>Invasiv brystkreft</i>	40-49 50-59 60-69	-4-7,1 % 1,7-54 % 7-21 %
<i>Gelder et al. (2011)</i>	<i>7 forskjellige metoder</i>	<i>Ductal carsinom in situ og invasive brystkreft mellom 1990-2006</i>	0-100 49-69 eller eldre 49-69	11,4 % 3,6 % 9,7 %
<i>Puliti og Paci (2009)</i>	<i>Kumulativ insidens og lead</i>	<i>Randomiserte studier og</i>	40-69 50-69	10-14 % 5 %

	<i>time</i>	<i>observasjonsstudier</i>	<i>50-59</i>	<i>21-54 %</i>
<i>Duffy et al. (2008)</i>	<i>Lead time</i>	<i>Lead time og hormonterapi</i>	<i>50-59</i> <i>50-59(etter justering av lead time)</i> <i>50-59 (hormonterapi)</i>	<i>69 %</i> <i>54 %</i> <i>39 %</i>
<i>The independent, U.K. Panel (2012)</i>	<i>Ukjent</i>	<i>Randomiserte studier og observasjonsstudier</i>	<i>0-100</i> <i>50-69</i>	<i>11 %</i> <i>19 %</i>
<i>Welch og Black (2010)</i>	<i>Ukjent</i>	<i>Randomiserte studier og observasjonsstudier</i>	<i>Ukjent</i> <i>Ukjent</i>	<i>24 %</i> <i>16 %</i>

Tabell 5: oppsummering av resultater fra de forskjellige artiklene som benyttes i oppgaven.

5 DISKUSJON

Her skal vi diskutere omfanget av overdiagnostikk, hvilken metode som egner seg best ut i fra de metodene vi har tatt med i denne oppgaven og hva som er uenighetene ved overdiagnostikk ved mammografiscreening.

5.1 Hvor stor er overdiagnostikk

Omfanget av overdiagnostikk varierte fra -4 til 69 % i de forskjellige studiene. Dette kommer av at hver studie har brukt forskjellige metoder for å estimere overdiagnostikk. Hvis man går ut i fra de forskjellige alderne vil overdiagnostikkestimatet variere fra de forskjellige aldersgruppene. I alderen 40 til 49 ligger overdiagnostikkestimatet mellom -4 til 7,1 % (Biesheuvel et al 2007). I alderen 40 til 69 ligger estimatet på 3,6 til 14 % (Gelder et al 2011, Puliti og Paci 2009). I alderen 50 til 59 var estimatet på 1,7 til 54% (Biesheuvel et al. 2007, Puliti og Paci 2009), men i følge Duffy et al. (2008) ligger overdiagnostikkestimeringen for alderen 50 til 59 på 69 %, men etter justering av lead time ble den liggende på 54 % og etter justering for hormonterapi på 39 %. I alderen 50 til 69 var estimatet på 5 til 19 % (Puliti og Paci 2009, The independent U. K 2012). I alderen 60 til 69 var estimatet på 7 til 21 %

(Biesheuvel et al. 2007) og i alderen 0 til 100 var estimatet på 11 til 11,4 % (Gelder et al. 2011, The independent U. K 2012). I en av studiene (Welch og Black 2010) kommer det fram et overdiagnostikkestimat på 16 til 24 %, men det er ikke spesifisert noen spesiell alder til disse tallene.

Her ser vi at estimatene varierer fra aldersgruppe til aldersgruppe. Dette kan komme av at alder, metode, lead time og hormonterapi spiller en rolle i estimeringen og om tallene er fra randomiserte studier eller observasjonsstudier. Som det kommer frem i resultat delen har noen av studiene inkludert bare invasiv brystkreft, mens andre har inkludert både invasiv brystkreft og karsinom in situ. Dette kan også være en faktor som er med på å påvirke overdiagnostikkestimatene. Så inkludering av invasiv brystkreft kan gi et lavere overdiagnostikkestimat, mens inkludering av karsinom in situ kan gi et høyere estimat. Dette fordi invasiv brystkreft ofte vil bli oppdaget mellom to screeninger, mens karsinom in situs ofte oppdages ved screening på et tidlig stadium. For eksempel inkluderer Biesheuvel et al. (2007) invasiv brystkreft, mens Gelder et al. (2011) inkluderer både invasiv brystkreft og karsinom in situ. Biesheuvel et al. (2007) sine estimater rangerer fra -4 til 54 %, mens Gelder et al. (2011) sine estimater rangerer fra 3,6 til 11,4 %. Her ser vi at det er stor variasjon av omfanget av overdiagnostikk, men det må her tas i betraktning at forskerne har benyttet forskjellige aldersgrupper og metoder. Dette kan være grunnen til at variasjonene er så store. Ut i fra dette kan det se ut som at det å inkludere invasiv brystkreft og karsinom in situ gir et lavere estimat av overdiagnostikk, mens inkludering av kun invasiv brystkreft kan gi et høyere estimat av overdiagnostikk.

Det er vanskelig å si nøyaktig hvor stor overdiagnostikkestimatet er fordi den varierer fra studie til studie, men tallene nevnt over kan gi en indikasjon på hvor stor overdiagnostikk estimatet kan være i hver aldersgruppe. Dette er ingen sikker indikasjon siden noen av studiene har sammenlignet kvinner som er invitert til screening med kontrollgruppen som ikke var invitert til screening mens de i andre studier har gjort de samme sammenligningene men kontrollgruppen har her blitt screenet på slutten av studien og dette kan føre til skjevheter i resultatene.

5.2 Hvorfor er det så stor uenighet

Mange forskere er uenige om hvor stor overdiagnostikk ved mammografiscreening faktisk er. Noen forskere mener overdiagnostikkestimatene ikke er så store mens andre mener overdiagnostikkestimatene er for høye. Mye av grunnlaget for uenighetene kommer nok av at forskerne bruker forskjellige metoder, inkluderingskriterier, studie design, aldersgrupper og om tallene er fra randomiserte kontroll studier eller observasjonsstudier. De forskjellige metodene forskerne bruker vil gi ulike estimater av overdiagnostikk. For eksempel viser Puliti og Paci (2009) at ved randomiserte studier benyttet de den kumulative insidens metoden og kom frem til at overdiagnostikkestimater rangerer fra 10 til 14 % og ved observasjonsstudier hvor de hadde benyttet statistisk justering for lead time varierte estimatet fra så mye som 5 til 54 %. Dette viser at valg av metode og om tallene er fra randomiserte kontroll studier eller observasjonsstudier har mye å si på resultatene av overdiagnostikk og det er dette som gjør det vanskelig for forskerne å kunne komme til enighet om omfanget av overdiagnostikk.

I tillegg til dette kan aldersgruppe, studiedesign og inkluderingskriterier ha innvirkning på overdiagnostikkestimatene. Dette fordi det kan være forskjellige estimater av overdiagnostikk innen de forskjellige aldersgruppene, om studiedesignen forskerne bruker har tatt hensyn til når diagnosen ble satt i forhold til invitasjon til screening eller om de har det som et krav og hvilke inkluderingskriterier de har valgt å ta med i studien. Dette er faktorer som kan gjøre at overdiagnostikkestimatene varierer så mye fra studie til studie og er med på og skape uenigheter mellom forskerne. Det ideelle alternativet ville være og kun bruke de estimatene som gir mest pålitelig overdiagnostikkestimat og bruke de samme inkluderingskriteriene og studiedesign for å få resultater som er tilnærmet like. Dette ser vi kan bli vanskelig å gjennomføre siden det er stor uenighet om hvilken metode, studiedesign og målinger som bør brukes. Selv om uenigheten er stor og resultatene varierer er det viktig at resultatene ikke bør ses hver for seg men som en helhet. Dette fordi hver enkelt studie kan bidra til å komme nærmere sannheten.

5.3 Hvilken metode er mest foretrukket

I to av artiklene (Puliti og Paci 2009, Biesheuvel et al. 2007) konkluderer de med at kumulativ insidens metoden i forhold til insidens-rate metoden er den metoden som kan gi mer valide resultater ved estimering av overdiagnostikk. For eksempel i artikkelen til Biesheuvel et al. (2007) benytter de den kumulative insidens metoden ved de randomiserte studiene og

insidens-rate metoden ved observasjonsstudiene. Ut ifra tabellen til Biesheuvel et al. (2007) kan det se ut som at insidens-rate metoden viser noe høyere overdiagnostikkestimater enn ved kumulativ insidens metoden. Dette kan komme av at ved kumulativ insidens metoden sammenligner man screeningsgruppen og kontrollgruppen uten screening, hvor disse gruppene bør bli overvåket i en lengre periode etter at screening har stoppet.

Ut ifra det vi har lest i artiklene kan det virke som det er vanskelig å gjennomføre overvåkning i lengre perioder på grunn av kostnader og tid. For eksempel The Independent U. K panel (2012) sier at minimum oppfølgingstid bør være fra 5 til 10 år. Vi mener at dette kan føre til skjevheter i overdiagnostikkestimatene. Dette fordi de som har en oppfølgingstid på 2 eller 3 år kan få forskjellige estimater enn de som har hatt en oppfølgingstid på 10 år. Insidens-rate metoden med justering for lead time vil også kunne gi valide overdiagnostikkestimater, men estimatene kan variere mye. For eksempel viser Biesheuvel et al. (2007) at estimeringene ved insidens-rate metoden varierer med så mye som -4 til 54 % i aldersgruppene 40-49,50-59 og 60 til 69. Denne variasjonen kan komme fra faktorer som alder, lead time og selve metoden som kan gi de varierende resultatene. Ut ifra dette mener vi at kumulativ insidens metoden er den metoden som gir mest valide overdiagnostikk estimater hvis oppfølgingstiden etter endt studie er lang nok, men hvis det er kort eller ingen oppfølgingstid bør lead time tas i betraktning

5.4 Kritikk

5.4.1 Søk

Det er viktig i en slik oppgave og foreta begrensninger i forhold til datamateriale og databasesøk. I vår oppgave ble det kun benyttet 4 forskjellige søkeord og 3 forskjellige databaser. For å få mer representative resultater og funn kunne vi utvidet søkene med flere søkeord, databaser og kombinasjoner. Dette ble ikke gjort siden vi følte vi hadde et godt utvalg av artikler ut ifra de søkeordene og databasene vi valgte å bruke. Det er også viktig å begrense søket med tanke på antall treff på artikler. Dette fordi vi arbeider i et fagfelt som er i stadig endring og er derfor viktig å inkludere nyere forskning. Vi har allikevel valgt å ta med en artikkel fra 2007 siden vi mener denne artikkelen gir oss en god innføring i de forskjellige metodene som brukes ved estimering av overdiagnostikk. Denne artikkelen blir også benyttet som referanse i noen av de andre artiklene vi har med i oppgaven. Samtidig ser vi at det er

viktig å inkludere enkelte artikler fra eldre studier som for eksempel artikkelen fra 2007 hvor vi kan få et innblikk av endringene av overdiagnostikk estimatene og om de eventuelt har forandret seg over årene.

I vår oppgave valgte vi og kun å inkludere oversiktsartikler. Dette fordi oversiktsartikler samler kunnskap fra flere artikler og som sagt tidligere er med på og kvalitetssikre artiklene. Det har vært vanskelig å tolke resultatene til noen av artiklene siden i enkelte av artiklene har det vært vanskelig å få konkrete tall, metode og fordeling av aldersgrupper. Ut ifra dette tenkte vi at vi kanskje burde ha inkludert original artikler siden disse gir et mer konkret valg av metode, design, utvalg av populasjon og et resultat som kan være noe enklere å tolke siden dette er forfatteren av artikkelens egne data. Selv om de har mer konkret valg av metode, design og utvalg av populasjon er det ikke nødvendigvis sikkert at dataene vil være mer konkrete etter som disse dataene kan ha vært påvirket av forfatterens egne synspunkter. Derfor valgte vi og kun å benytte oversiktsartikler siden disse virket mer troverdige.

5.4.2 Reliabilitet og validitet

I metode kapittelet har vi dokumentert og beskrevet søkeordene slik at de kan etterprøves. Dette er med på å gi oppgaven reliabilitet, slik at andre kan komme frem til det samme resultatet. Artiklene vi benytter i oppgaven er hentet fra anerkjente tidsskrifter som for eksempel The Lancet og British journal of cancer. Internettkilder er blitt lite brukt i denne oppgaven. De internettkildene som er brukt er kun brukt for å finne illustrasjoner. Liten bruk av internettkilder er med på å styrke reliabiliteten siden innholdet på sidene stadig kan endres og de blir dermed vanskeligere å kunne etterprøves.

For å få svar på problemstillingen valgte vi artikler som kunne besvare vårt valgte tema. Vi har i stede for å bruke bare bøker i teori delen også valgt å bruke vitenskapelige og peer review artikler. Dette fordi artiklene inneholdt den informasjonen vi trengte for å besvare de valgte punktene i teoridelen. For å sikre validiteten til oppgaven har vi vært nøye ved innsamling og analyse av data. Ved at en konklusjon er basert på sanne betingelser vil den dermed bli valid. Det er viktig å unngå feiltolkning når en analyserer og siterer andres litteratur (Dalland 2007). Vi mener at den største utfordringen ved utarbeidningen av denne oppgaven har vært å prøve å unngå feiltolkning av materialet.

6 KONKLUSJON

Ut ifra artiklene vi har benyttet ser vi at det finnes overdiagnostikk ved mammografiscreening. Overdiagnostikk kan føre til overbehandling av harmløse lesjoner og det er vanskelig å vurdere om lesjonene er maligne eller benigne og derfor vil kvinnene bli behandlet enten ved kirurgi eller i enkelte tilfeller med strålebehandling. Dette kan igjen føre til at kvinnene som blir behandlet unødvendig kan få psykologiske eller emosjonelle plager som følge av diagnose og behandling. Samtidig er det vanskelig å få et nøyaktig omfang av overdiagnostikk. Dette på grunn av alle de forskjellige faktorene som kan påvirke overdiagnostikkestimeringen, som for eksempel hvilken metode som benyttes, studiedesign, hvor lang oppfølgingstid de har brukt og hvilke inkluderingskriterier som er tatt med.

Overdiagnostikkestimatene i artiklene vi har brukt varierte fra -4 til 69 %, men disse tallene er et resultat av flere forskjellige metoder, studiedesign og utvalg av populasjon fra de forskjellige studiene. Så dette blir bare en sammensetting av resultatene for å vise hvor mye overdiagnostikkestimatene varierer uten å ta i betraktning hvilken metode og studiedesign som blir benyttet. Vi kan ikke nøyaktig si hvor stor overdiagnostikkestimeringen er men vi kan anta at for aldersgruppen 40 til 49 er estimatet mellom -4 til 7,1 %, 40 til 69 mellom 3,6 til 14 %, 50 til 59 mellom 1,7 til 54 %, 50 til 69 mellom 5 til 19 %, 60 til 69 mellom 7 til 21 % og 0 til 100 mellom 11 til 11,4 %. Det var også et estimat på 16 til 24 % men siden det ikke var spesifisert noe spesiell alder ved disse tallene er det vanskelig å vite hvor valide disse estimatene er. Det er disse forskjellene i estimatene som gjør at forskerne er uenige. Det er stor uenighet om hvilken metode og studiedesign som bør benyttes for å få et pålitelig estimat av overdiagnostikk. Vi mener at forskerne bør komme til enighet om hvilken metode, studiedesign og målinger som skal benyttes for å få et pålitelig resultat av overdiagnostikk.

7 LITTERATURLISTE

Bøker

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

Bjørndal, A. and D. Hofoss (2004). Statistikk for helse- og sosialfagene. Oslo, Gyldendal akademisk.

Dalland, O. (2007). Metode og oppgaveskriving for studenter. Oslo, Gyldendal akademisk.

Ertzaas, A. K. O. (2003). Mammografiprogrammet: kvalitetsmanual. Oslo, Kreftregisteret.

Gøtzsche, P. C. (2012). Mammography screening: truth, lies and controversy. London, Radcliffe Publishing.

Artikler

Biesheuvel, C., et al. (2007). "Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review." Lancet Oncol **8**(12): 1129-1138.

Brodersen, J., et al. (2010). "The benefits and harms of screening for cancer with a focus on breast screening." Pol Arch Med Wewn **120**(3): 89-94.

Corbex, M., et al. (2012). "Breast cancer early detection methods for low and middle income countries, a review of the evidence." The Breast **21**(4): 428-434.

Duffy, S. W., et al. (2008). "Complexities in the estimation of overdiagnosis in breast cancer screening." The British Journal of Cancer **99**(7): 1176-1178.

Harris, R., et al. (2011). "Breast cancer screening for women ages 50 to 69 years a systematic review of observational evidence." Prev Med **53**(3): 108-114.

Hersch, J., et al. (2013). "Women's views on overdiagnosis in breast cancer screening: a qualitative study." BMJ **346**: f158.

Independent, U. K. P. o. B. C. S. (2012). "The benefits and harms of breast cancer screening: an independent review." Lancet **380**(9855): 1778-1786.

Jatoi, I. (1999). "Breast cancer screening." The American Journal of Surgery **177**(6): 518-524.

Puliti, D., et al. (2012). "Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review." J Med Screen **19** Suppl 1: 42-56.

Puliti, D. and E. Paci (2009). "The other side of technology: risk of overdiagnosis of breast cancer with mammography screening." Future Oncol **5**(4): 481-491.

Welch, H. G. and W. C. Black (2010). "Overdiagnosis in cancer." J Natl Cancer Inst **102**(9): 605-613.

Internettkilder

Mcstrother (2011a). *Lead time bias* [online]: Wikipedia. URL
http://en.wikipedia.org/wiki/File:Lead_time_bias.svg (13.05.2013)

Mcstrother (2011b). *length time bias* [online]: Wikipedia. URL
http://en.wikipedia.org/wiki/File:Length_time_bias.svg (14.05.2013)

Staff, Annetine (2010). *Bias* [online]: Forskningsetisk bibliotek. URL
<http://www.etikkom.no/FBIB/Temaer/Spesielle-problemomrader/Bias/> (24.03.2013)