

Norsk mammografiscreening – mange selvmotsigelser i evalueringen

Offentlig mammografiscreening ble innført i Norge i 1996. Målet var 30 % reduksjon i brystkreftdødelighet. I 2006 fikk Norges forskningsråd i oppdrag å evaluere Mammografiprogrammet. Rapporten inneholder mange selvmotsigelser: Screening medfører ikke at færre får svulster med spredning, men reduserer likevel dødeligheten av brystkreft.

I evalueringsrapporten fra 2015 konkluderer man med at brystkreftdødeligheten er blitt redusert med 20–30 %, og at fem kvinner blir overdiagnostisert for hvert brystkreftdødsfall som forebygges (1).

Forekomsten av brystkreft er beregnet etter justering for hormonbruk og antatt underliggende insidensøkning, mens analyser av dødeligheten ikke justeres på samme måte. Videre antar man at det er ingen eller liten effekt av bedre behandling av brystkreftdødelighet.

RAMME 1

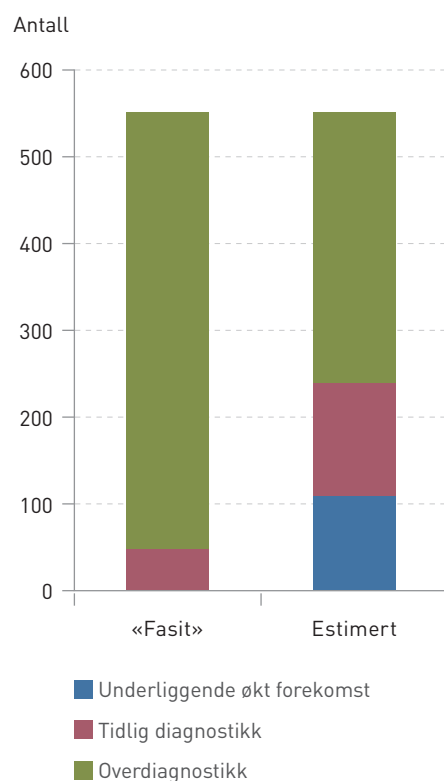
I dette tenkte scenarioet, som beskriver 20 år med mammografiscreening, blir 550 flere kvinner diagnostisert med brystkreft mellom 50 og 69 års alder med mammografiscreening – enn hvis de ikke var blitt screenet (ekstra tilfeller). Av de 550 er 50 kvinner reelt tidlig diagnostisert, mens 500 er overdiagnostiserte (fig 1).

Hvis man ved beregning av overdiagnostikk (søylen til høyre i figuren) antar at 20 % av de 550 ekstratilfellene skyldes underliggende økende forekomst (20 år med 1 % økning per år) (16), reduseres antallet ekstratilfeller fra 550 til 440.

Videre antar man at kvinner over screeningalder (70–79 år) også har 1 % årlig økning i forekomst av brystkreft. Dersom det var 400 tilfeller i denne alderen før screening ble introdusert, forventer man 480 tilfeller med 20 % økning. Når man så observerer 350 tilfeller i denne aldersgruppen (400 minus de 50 som var reelt tidlig diagnostisert), er dette 130 færre enn de 480 man forventet. Hvis disse 130 trekkes fra de 440 ekstratilfellene med begrunnelsen tidlig diagnostikk, konkluderer man med 310 overdiagnostiserte kvinner.

Ved å modellere en underliggende insidensøkning på 1 % per år og samtidig anta at det finnes mange svulster med lang fremskynningstid som krever at man skal følge opp individer i ti år etter at de ikke lenger blir invitert til screening, kan man underestimere nivået på overdiagnostikk med rundt 40 %.

Mammografi er en røntgenundersøkelse av brystene, og ved mammografiscreening ønsker man å påvise kreftsvulster mens de ennå er små og lokaliserte slik at kvinner kan behandles kurativt. Når man finner en svulst, tar man en biopsi, og rundt én av fem viser seg å være infiltrerende brystkreft (1). Ved mammografi oppdager man også mange svulster begrenset til melkekjertelgangene, såkalt duktalt carcinoma in situ (DCIS). Noen av disse svulstene utvikler seg til infiltrerende brystkreft, men flertallet utvikler seg ikke videre (2, 3). I dag kan vi ikke forutsi hvilke av disse svulstene som har utviklingspotensial, og hvilke som vil forbli uendret eller gå tilbake, og derfor er behandlingen som for infiltrerende brystkreft.



Figur 1 Scenario over ekstra krefttilfeller på grunn av mammografiscreening. Søylen til venstre: Av de 550 som ble diagnostisert med brystkreft, er 50 kvinner reelt tidlig diagnostisert, mens 500 er overdiagnostiserte. Søylen til høyre: Hvis man ved beregning av overdiagnostikk antar at 20 % av de 550 ekstratilfellene skyldes underliggende økende forekomst (20 år med 1 % økning per år) (16), reduseres antallet ekstra tilfeller fra 550 til 440

Rapporten til Norges forskningsråd baserer seg på ulike typer analyser av forekomst og dødelighet av brystkreft. Syv ulike forskningsgrupper ble etter konkurranse valgt til å gjøre analysene. Det skulle anvendes et felles datasett bestående av individdata fra Kreftregisteret og Dødsårsaksregisteret som var koblet til en rekke andre helseregistre.

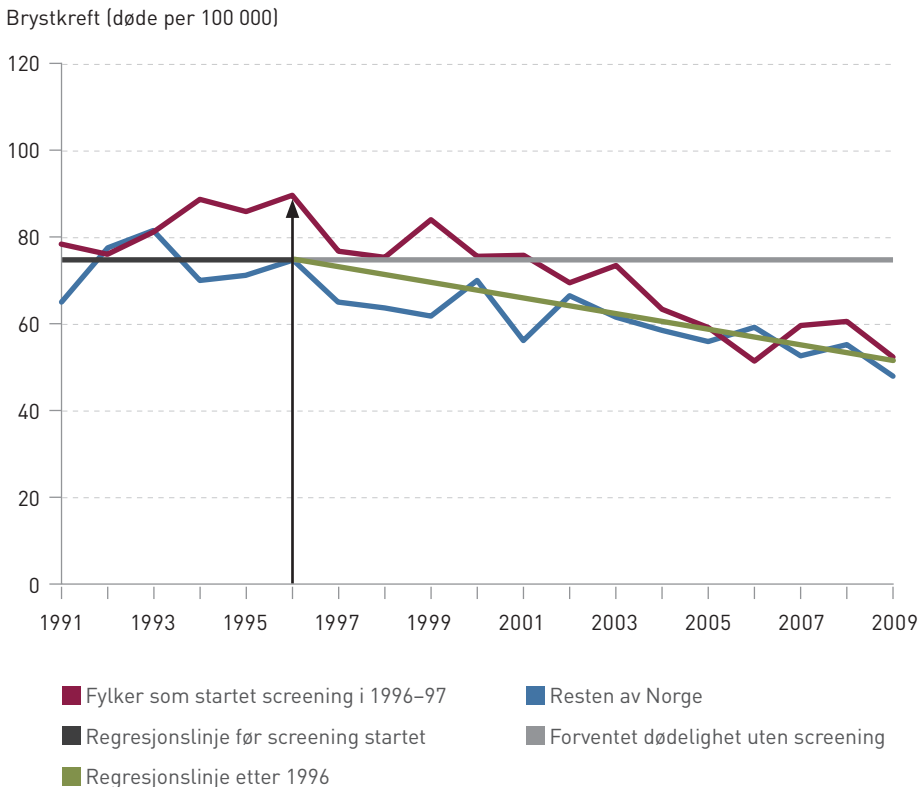
Overdiagnostikk

Overdiagnostikk er diagnostikk av kreftsvulster som ellers aldri ville gitt symptomer i pasientens levetid (4). Overdiagnostiserte svulster kan være svulster som vokser meget sakte, som ikke vokser og forblir subkliniske, eller som spontant forsvinner. Overdiagnostikk kan enkelt beregnes i randomiserte studier, men randomiserte studier av mammografiscreening kan ikke lenger gjøres. Det er også utviklet metoder for å beregne omfanget av overdiagnostikk i offentlige mammografi-programmer, og en vesentlig forskjell mellom metodene er hvilke svulster som inngår i nevneren. Enten man benytter alle tilfeller av brystkreft i aldersgruppen 50–69 år (screeningalder), 50–74 år, 50–84 år, eller man estimerer forventet antall svulster uten screening, så vil omfanget av overdiagnostikk variere mellom 10 og 50 % med bruk av det samme tallet i telleren (4). Mens denne variabiliteten skyldes ulike definisjoner, ikke skjevhet i analysene, gir vi nedenfor tre eksempler på statistisk justering i rapporten som faktisk øker risikoen for skjevhet og underestimering av overdiagnostikk.

Eksempel 1. Justering for bruk av hormoner mot plager i overgangsalderen

I rapporten fra Norges forskningsråd kan man lese at bruk av hormoner mot plager i overgangsalderen var en viktig årsak til den kraftige økningen i brystkreftforekomst i Norge i 1990-årene, og at det er svært viktig å justere for bruk av hormoner med individdata (1).

I år 2000 brukte rundt 40 % av postmenopausale kvinner hormoner mot plager i overgangsalderen (1, 5). I rapporten er det oversett at bare halvparten av disse brukte kombinasjonspreparater (5), og at det bare er kombinasjonspreparater som øker risikoen for brystkreft (5–7). Det er heller ikke vektlagt at sammenhengen mellom bruk av hormoner og risiko for brystkreft fremstår 2–4 ganger sterkere i observasjonsstudier enn



Figur 2 Brystkreftdødeligheten i aldersgruppen 55–74 år var rundt 76 per 100 000 kvinner før screening startet (svart pil), og i perioden 2005–9 var den rundt 55 per 100 000. Heltrukket rød kurve er de fire fylkene (Akershus, Oslo, Rogaland og Hordaland) som startet med screening i 1996–7, og blå kurve er resten av Norge. Heltrukket svart linje er regresjonslinjen før screening startet, og grå linje er forventet dødelighet uten screening. Grønn linje er regresjonslinjen etter 1996. Den viser en reduksjon på 28 %, fra 1991–95 til 2005–9

i randomiserte intervensjonsstudier (5–7). Årsaken er trolig at hormonbruk medfører mer røntgentette bryst med mindre kontrast mellom normalt og patologisk vev, og at dette forsinker diagnosetidspunktet (6, 8).

Hormonbruk er en tidsavhengig variabel, og derfor må man inkludere starttidspunkt og varighet av behandlingen når man skal justere. Slike data var dessverre ikke tilgjengelig for forskerne som evaluerte Mammografiprogrammet, og man måtte derfor kategorisere kvinnene enten som brukere, tidligere brukere eller ikke-brukere av hormoner mot plager i overgangsalderen (5). Ved å kategorisere kvinner på denne måten i observasjonsstudier, overestimerer man risikoen for brystkreft ved kvinners hormonbruk med mange 100 % (6, 7). Når en overdreven del av økningen i brystkreftforekomst forklares med hormonbruk, vil estimater av overdiagnostikk bli for lave.

Eksempel 2. Justering for underliggende økende forekomst

I rapporten justerer man for underliggende økende forekomst av brystkreft over tid. Forekomsten av brystkreft økte med rundt 1 % årlig for alle kvinner i perioden 1953–85. Økningen kan skyldes forandringer i fertilitet, kosthold og andre livsstilsfaktorer. Men økningen kan også skyldes økt oppmerksomhet og mer opportunistisk screening, og derav funn av flere små saktevoksende svulster

(1, 9–12). Etter 1985 økte forekomsten bare i aldersgruppen 50–69 år (13), og det er usannsynlig at ikke opportunistisk screening er vesentlig for å forklare denne aldersspesifikke økningen. Hvis en stor del av den observerte økningen i brystkreftforekomst forklares med en underliggende insidensøkning, så justerer man vekk mye av overdiagnostikken i analysene (ramme 1).

Eksempel 3. Justering for lang fremskyndingstid

I rapporten hevder man at kort oppfølgings-tid er en viktig feilkilde ved beregning av overdiagnostikk. Vi forstår det som at man mener antall år med oppfølging etter at kvinner ikke lenger innkalles til mammografiscreening. Det har betydning hvis det finnes mange svulster med lang fremskyndingstid (lead-time).

Fremskyndingstid er tidsintervallet fra en svulst oppdages ved screening og til svulsten ville ha blitt oppdaget klinisk. For eksempel må kvinner følges opp til de er minst 79 år hvis mange mammografidiagnostiserte svulster i aldersgruppen 65–69 år har ti års fremskyndingstid. Da er sammenligning av kumulative rater opptil 79 år i teorien en god metode for å beregne omfanget av overdiagnostikk – fordi muligheten for skjevhet er liten. Ulempen er at konfidensintervallet rundt kumulative rater vokser med lengden på oppfølging (14), så den statistiske usik-

kerheten blir høyere. Lang oppfølging er derfor ikke nødvendigvis bedre enn kort.

Mer alvorlig er det at kombinasjonen av å justere for lange fremskyndingstider og å justere for underliggende økende brystkreftforekomst utover det reelle, introduserer alvorlig skjevhet i resultatene. I studier av kvinner med mammografidiagnostiserte svulster som ikke er blitt operert, har gjennomsnittlig fremskyndingstid vært rundt ett år (4). Analyser av observasjonelle data viser det samme (4).

Forestillingen om at gjennomsnittlig fremskyndingstid er 2–7 år baserer seg på matematiske modeller som antar at alle svulster vokser, og at all økning i forekomst ved screening skyldes tidlig diagnose (altså antar man ingen overdiagnostikk) (4). Denne selvmotsigelsen – å anta at det ikke er overdiagnostikk når man skal estimere nivået av overdiagnostikk – er helt urimelig. Slike matematiske modeller er for øvrig falsifisert på ulike måter (4, 15). Hvis man antar for lange fremskyndingstider, underestimeres nivået av overdiagnostikk (ramme 1).

Regresjon av kreft

I en meget sitert artikkel fra det norske Mammografiprogrammet som ikke er nevnt i rapporten fra Norges forskningsråd, beregnet vi hvilken andel av screeningoppdagede kreftsvulster som ville forsvunnet spontant (15).

I denne studien er kvinner som er invitert tre ganger til mammografiscreening over seks år (testgruppen), sammenlignet med kvinner som først ikke screenes i fire år, og som så screenes én gang de neste to årene. Denne studien er designet slik at man justerer for nesten alle forskjeller i risiko mellom gruppene ved at de samme kvinnene brukes i både testgruppen og kontrollgruppen.

Uten overdiagnostikk burde det totale antallet svulster være det samme i begge gruppene etter seks år, men det var 22 % flere svulster i testgruppen. Studien kan tolkes som at mye av økningen i brystkreftforekomst ved screening er diagnostikk av svulster som ville gått i spontan regresjon hvis de ikke var diagnostisert ved mammografi. I tillegg må nesten alle dukalt carcinoma in situ -svulster forsvinne spontant eller forbli uendret fordi kirurgisk behandling ikke medfører færre brystkrefttilfeller. Poenget med å behandle og fjerne forstadier er at det ikke skal bli kreft etterpå. To randomiserte studier hvor kvinner tilbys aktiv oppfølging versus kirurgi er startet for å studere regresjon av dukalt carcinoma in situ (2, 3).

Dødelighet

Rundt halvparten av all dødelighet hos kvinner i aldersgruppen 50–74 år skyldes kreft, men bare 6 % av totaldødeligheten skyldes brystkreft (16). I rapporten er bare effekten av mammografiscreening på bryst-

kreftspesifikk dødelighet vurdert, og ikke screening på totaldødelighet eller total kreftdødelighet. De to siste endepunktene er også interessante da de fanger opp mulig økt dødelighet som følge av kreftbehandlingen (cellegift og strålebehandling øker dødeligheten av hjerte- og karsykdom og andre typer kreft) (9).

I rapporten fremhever man én studie som mest troverdig, og baserer sin konklusjon om effekt av mammografiscreening på brystkreftdødelighet på denne alene (17). Mens man i denne studien konkluderer med 28 % reduksjon av brystkreftdødelighet, viser to andre studier (18, 19) rundt 10 % ikke-signifikant reduksjon – store sprik i estimater kan også tolkes som usikkerhet om hva effekten egentlig er. Denne foretrukne studien justerer ikke for individuelle risikofaktorer, er basert på en statistisk metode som er vagt beskrevet og derfor vanskelig å forstå, og som ikke er vanlig i bruk, og analysen bygger på antagelser som det ikke er redegjort for, og som ikke er validert.

En av de viktigste antagelsene som det ikke er redegjort for, er hvordan effekten av bedre behandling er modellert. Dødelighet av brystkreft i Norge begynte å falle rundt 1993, altså like før offentlig mammografiscreening startet (fig 2) (20). Mye av dødelighetsreduksjonen etter 1993 skyldes trolig innføring av moderne brystkreftbehandling, som hormonbehandling, cellegift og trastuzumab (Herceptin). Bedre behandling er anslått å redusere brystkreftspesifikk dødelighet med 30 % (21).

Det sentrale spørsmålet er hvordan man har skilt mellom effekten av bedre behandling og effekten av offentlig mammografiscreening. Hvis man finner at nesten hele reduksjonen skyldes mammografiscreening, er dette et urimelig funn. Det er gjort en valideringsanalyse hvor man har lagt til pseudoinvitasjoner til screening av 50–69 år gamle kvinner i 1990–94, og hvor man finner relativ risiko rundt 1 % (17). Dessverre bidrar ikke denne øvelsen til å validere metoden.

Ingen mammografiscreeningprogrammer (10), heller ikke det norske (11), har redusert forekomsten av brystkreft med spredning. Mens man i rapporten trekker sine konklusjoner utelukkende fra statistisk modellering av brystkreftdødelighet, forteller en slik enkel observasjon en annen historie: Uten reduksjon av brystkreft i avansert stadium er det vanskelig å se for seg reduksjon i dødelighet – og nesten all insidensøkning er overdiagnostikk.

Konklusjon

I rapporten er konklusjonen at Mammografi-programmet har medført 20–30 % reduksjon i brystkreftspesifikk dødelighet, og at fem kvinner overdiagnostiseres for hvert brystkreftdødsfall som forebygges. Dette resultatet må tolkes med forsiktighet. Konklusjonen er basert på et snevert utvalg

av tilgjengelig kunnskap, og bare noen av endepunktene som anbefales ved evaluering av screeningprogrammer, er anvendt: De har ikke studert om man ved mammografiscreening oppdager kreft tidligere eller om totaldødeligheten synker. Det er lite tilfelle ved å justering for hormonbruk og antatt økende brystkreftforekomst er fundamentale for analysene av forekomst, men ikke i analyser av dødelighet.

Det mest tankevekkende er likevel hvordan man har håndtert effekten av bedre behandling i perioden etter at screening ble innført. Behandlingsendringer er anslått å redusere brystkreftdødeligheten med 30 %. Men det er urimelig å anta at både behandlingseffekten og screeningeffekten hver for seg skal være 30 % – det ville medført 60 % reduksjon i brystkreftdødeligheten.

Vi takker turnuslege og ph.d.-kandidat Henriette Jodal for verdifulle innspill til manuskriptet.

Per-Henrik Zahl
per-henrik.zahl@fhi.no
Øyvind Holme
Magnus Løberg

Per-Henrik Zahl (f. 1961) er dr.med. i biostatistikk og arbeider som statistiker ved Divisjon for psykisk og fysisk helse, Folkehelseinstituttet. Han har publisert flere artikler om mammografiscreening og brystkreftdødelighet. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Øyvind Holme (f.1970) er ph.d. og spesialist i generell indremedisin og i fordøyelsessykdommer ved Sørlandet sykehus, Kristiansand og Gruppe for klinisk effektforskning, Institutt for helse og samfunn, Det medisinske fakultet, Universitetet i Oslo. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Magnus Løberg (f. 1979) er lege, ph.d. og postdoktor ved Gruppe for klinisk effektforskning, Institutt for helse og samfunn, Det medisinske fakultet, Universitetet i Oslo og ved Avdeling for transplantasjonsmedisin, Oslo universitetssykehus, og ved K.G. Jebsen senter for tarmkreftforskning i Oslo. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Litteratur

1. Research-based evaluation of the Norwegian Breast Cancer Screening Program. Oslo: The Research Council of Norway, 2015.
2. Time Magazine. 2015. <http://time.com/4057310/breast-cancer-overtreatment/> [8.2.2016].
3. New York Times. 2015. <http://nytimes.com/2015/09/29/health/a-breast-cancer-surgeon-who-keeps-challenging-the-status-quo.html?smid=fb-share&r=0> [8.2.2016].
4. Zahl P-H, Jørgensen KJ, Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. Br J Cancer 2013; 109: 2014–9.

5. Suhrke P, Zahl P-H. Breast cancer incidence and menopausal hormone therapy in Norway from 2004 to 2009: a register-based cohort study. Cancer Med 2015; 4: 1303–8.
6. Zahl PH, Mæhlen J. Bias in observational studies of the association between menopausal hormone therapy and breast cancer. PLoS One 2015; 10: e0124076.
7. Ioannidis JPD. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005; 294: 218–28.
8. Banks E, Reeves G, Beral V et al. Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. BMJ 2004; 329: 477–82.
9. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. [review]. Cochrane Database Syst Rev 2013; 6: CD001877.
10. Autier P, Boniol M, Middleton R et al. Advanced breast cancer incidence following population-based mammographic screening. Ann Oncol 2011; 22: 1726–35.
11. Lousdal ML, Kristiansen IS, Møller B et al. Effect of organised mammography screening on stage-specific incidence in Norway: population study. Br J Cancer 2016; 114: 590–6.
12. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. BMJ 2015; 350: g7773.
13. Zahl PH, Strand BH, Mæhlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. BMJ 2004; 328: 921–4.
14. Aalen OO. Nonparametric inference for a family of counting processes. Ann Stat 1978; 6: 701–26.
15. Zahl PH, Mæhlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008; 168: 2311–6.
16. Statistisk sentralbyrå. Tabell 08880: Dødsfall, etter kjønn, alder og detaljert dødsårsak (avslutta serie). <https://ssb.no/statistikkbanken/SelectVarVal/Define.asp?MainTable=DødsfallDetaljAld&KortNavnWeb=dødsarsak&PLanguage=0&checked=true> [8.2.2016].
17. Weedon-Fekjær H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study. BMJ 2014; 348: g3701.
18. Kalager M, Zelen M, Langmark F et al. Effect of screening mammography on breast-cancer mortality in Norway. N Engl J Med 2010; 363: 1203–10.
19. Olsen AH, Lyng E, Njor SH et al. Breast cancer mortality in Norway after the introduction of mammography screening. Int J Cancer 2013; 132: 208–14.
20. Autier P, Boniol M, Gavin A et al. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. BMJ 2011; 343: d4411.
21. Davies C, Godwin J, Gray R et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 2011; 378: 771–84.

Mottatt 19.2.2016, første revisjon innsendt 16.3.2016, godkjent 23.9.2016. Redaktør: Kaveh Rashidi.

 Engelsk oversettelse på www.tidsskriftet.no

Publisert først på nett.