

Epidemiologiske metoder til vurdering af screeningsprogrammer

Jørn Olsen

STATUSARTIKEL

Sektion for Epidemiologi,
Institut For Folkesund-
hed, Aarhus Universitet

Ugeskr Læger
2014;176:V01140012

Screening synes at være en indlysende rigtig ide. »Kræft, der behandles i tide, kan helbredes« er et tautologisk slogan, for hvad vil det sige at blive behandlet i tide? Det forekommer intuitivt rigtigt, at vi skal behandle ikkereversible syge så tidligt som muligt. Kravene til vurdering af screeningsprogrammer var derfor beskedne tilbage i tiden. Havde man en brugbar og billig test uden svære bivirkninger, kunne man gå i gang.

Undertiden støttet af simple og fejlagtige mål, som viste, at overlevelsestiden efter tidlig cancerdiagnose med screening var længere end overlevelsestiden efter klinisk diagnose, selv om denne »længere« levetid måske bedst afspejler *lead time bias* [1]; tidligere diagnostik uden bedre overlevelse.

Kravene til vurdering af screeningsprogrammer blev derfor større med tiden, bl.a. fordi screeningsprogrammer kan have svære og alvorlige bivirkninger. Screeningstesten i sig selv er måske ikke ufarlig, og mange vil i de fleste screeningsprogrammer få falsk positive testresultater, der kan gøre dem urolige, og der er risiko for overbehandling. Det er f.eks. unødvendigt at gennemgå en stor og omfattende cancerbehandling, hvis man vil dø af andre sygdomme, inden cancersygdommen ville afslutte livet. Risikoen for denne funktionelle overdiagnosticering er ikke ubetydelig og må medtages i vore vurderinger af screeningsprogrammer, noget der fortsat forsømmes ganske ofte. I de følgende afsnit gennemgås kortfattet fordele og begrænsninger ved nogle af de klassiske epidemiologiske design. Formålet er at påpege, hvad man kan opnå med gode, traditionelle epidemiologiske studier.

DET RANDOMISEREDE FORSØG

Nu vil man næppe gennemføre større screeningsprojekter, uden at de har været afprøvet i randomiserede forsøg (RCT), fordi RCT opfattes som den forskningsmodel, der giver de færreste muligheder for fejlslutninger.

Man undgår konfounding i et stort forsøg, fordi den tilfældige allokering til screening også forventes at medføre en tilfældig fordeling af de kendte og ukendte potentielle konfoundere, der vil have betydning for udfaldene.

Det vigtigste metodeproblem ved dette design er manglende kompliance til screeningsprogrammet. Et problem, der blot øges over tid, og som betyder mindre sammenlignelighed i de randomiserede grupper over tid. *Intention to treat* eller *as randomized so analyzed* gør ikke sammenligningerne *unbiased*, og de vil blive tiltagende meningsløse, hvis man gerne vil vide, hvad fordelene er for den enkelte, der skal beslutte sig til at blive screenet eller ej.

Fra et folkesundhedssynspunkt er manglende compliance ikke noget metodeproblem, men en vigtig del af det, der skal studeres. Hvad er den forventede deltagelse i et screeningsprogram – og hvor meget vil det påvirke effekten på befolkningsniveau?

RCT'et skal ofte være stort og strække sig over en lang periode. Der er derfor risiko for, at screeningsundersøgelsen er baseret på en screeningstest, der er forældet, når studiet er slut, måske ti år eller mere efter dets start.

Et sådant RCT skal ikke alene omfatte det slutmål, screeningen adresserer, f.eks. dødelighed af brystcancer eller kolorektal cancer, men også brud på opfølgingsprocedurer etc. Men et RCT kan ikke stå alene. For det første omfatter et RCT en vurdering af et screeningsprogram under optimale betingelser, *efficacy*, selv om hovedinteresser knytter sig til *effectiveness*, hvordan programmet virker under normale, rutinemæssige forhold. For det andet er et RCT sjældent knyttet til den egentlige præventive handling, f.eks. fjernelse af polypper eller afficeret cervixepitel, men til mere distale processer som effekten af at tilbyde et bestemt program. For det tredje skal man gøre det klart, hvilket problem, man forsøger at løse. Hvis man, som i kolorektalcancerscreeningsforsøget i det daværende Fyns Amt, randomiserer befolkningen a priori til enten at få tilbudt screening eller ingenting, uden at informere dem på forhånd, får man svar på, hvad man kan forvente ved at tilbyde et sådant program til en befolkningsgruppe [2]. Hvor mange vil deltage, og hvad får den gruppe ud af det, som blev randomiseret til screening? Hvis man gerne vil vide, hvilket benefit den enkelte kan forvente at få ved at sige ja til screeningen, må man randomisere individuelt blandt dem, der har sagt ja til at acceptere tilbuddet.

OBSERVATIONELLE STUDIER

Da langt de fleste screeningsprogrammer vil have både positive og negative effekter, som formentlig vil skifte over tid, er det vigtigt at bruge vores almindelige monitoreringsmetoder i målrettede makroepidemiologiske eller økologiske studier [3-5].

Screeningen har oftest et omfang, der bør sætte aftryk i rutinebefolkningsdata [5]. Vigtigst er det at følge sygdommens årsagsspecifikke mortalitet over tid. I et screeningsprogram, der omfatter sekundær prævention, vil man ikke forvente, at incidensen ændres, når man ser bort fra tiden kort efter screening. Når *steady state* er genoprettet, forventes den aldersjusterede incidensrate at være som før, men den årsagsspecifikke mortalitet skal falde efter en vis latens-tid. Efter brystcancercreeningen i Frankrig [6] har man set stigende brystcancerincidens, men uændret brystcancerdødelighed, og det kan skyldes et effektivt screeningsprogram, der modvirker konsekvenserne af en stigende hyppighed af brystkræft. Denne udvikling kan dog også skyldes funktionel overbehandling og ingen eller næsten ingen overlevelseseffekt af screeningsprogrammet.

I områder, hvor screeninger er igangsat på forskellige tidspunkter, kan økologiske studier bruges til at vurdere deres effekt ved før-efter-sammenligninger og sammenligninger mellem forskellige områder [7, 8]. Disse studier er vigtige, men siger ikke nødvendigvis noget om de enkelte deltagere i screening, bl.a. på grund af den økologiske fejl. Vi ved ikke, om det er de screenede, der dør eller bliver helbredt, kun hvordan forholdene er på befolkningsniveau. Man vil derfor ofte supplere med andre observationelle studier på individniveau. Case-kontrol-studier er her en designmulighed til vurdering af både helbredsgevinster ved screeninger og bivirkninger ved screening [7].

CASE-KONTROL-STUDIER

Case-kontrol-studier er ofte billige og hurtige at gennemføre, men de kræver metodologisk indsigt at udføre korrekt [8].

I et case-kontrol-studie vil man f.eks. vælge alle med det caseudfald man vil studere, f.eks. død af colon- eller rectumcancer. Man vil identificere alle med disse dødsfald i en given periode og i en given befolkning. Man vil derefter udvælge en kontrolgruppe, som er tilfældigt udtrukket fra den population, de pågældende personer kom fra, indtil man når den undersøgelsesstørrelse, man ønsker. Ved at indhente oplysninger om screeningsaktivitet for både casepersoner og kontrolpersoner og sikre, at man ikke fejlklassificerer screeninger med diagnostiske udredninger, kan man få et estimat af, hvor meget screeninger evt. mindsker dødelighed af kolorektalcancer.



Mammografitest.

Problemet er her, som i et RCT, at få valide dødsårsager, der ikke blev påvirket af evt. screeningsprogrammer (kun et studie over den totale dødelighed er ikke ramt af dette problem). Kan man indhente oplysninger om faktorer, der påvirker både deltagelse i et screeningsprogram og/eller prognoser for cancer, kan man rette op på denne konfounding i analysen.

Vi vil forvente, at sådanne studier, især hvis de baseres på registre uden bortfald, vil kunne give resultater, som er meget lig et RCT. Et RCT kan være den eneste model, der kan rette op på konfounding af ukendte faktorer. Der er ikke nogen grundlæggende forskel på logikken i et case-kontrol-studie og et followupstudie [8]. Hvis man vælger at lave alle disse observationelle studier på registre alene, kan man selvfølgelig gøre det i follow-upmodellen – data er allerede til rådighed, man skal blot sikre, at de ikke er systematisk fejlkodet.

KOHORTESTUDIER

RCT'et er et eksperimentelt kohortestudie, fordi man manipulerer/eksperimenterer med »eksponeringen« for at studere dets »effekt«. I det observationelle kohortestudie af et screeningsprogram lader man folk selv vælge, om de vil modtage tilbuddet om screening eller ej (i et randomiseret studie bliver de spurgt, om de vil lade denne beslutning afhænge af en lodtrækning). Det betyder, at alle forhold af betydning for »effekten« – f.eks. brystcancerdødelighed, som korreleret med deltagelsen i programmet har potentiel konfunderkarakter og muligvis skal kontrolleres. Det vil sige, at ganske mange oplysninger om kohortedeltagerne skal indhentes, og rigtig mange skal rekrutteres (ofte +50.000), hvilket gør kohortestudiet til et lidet attraktivt design, hvis det skal baseres på



FAKTABOKS

Screeningsprogrammer skal evalueres, før de sættes i gang, og de skal løbende evalueres.

Både eksperimentelle og observationelle metoder skal anvendes.

primære data. Det ser anderledes ud, hvis man kan anvende sekundære elektroniske data (data indsamlet med et andet formål).

Sådan data kan anvendes til at se, om man når de forventede mål for befolkningen for deltagelse, sygelighed og dødelighed for alle og fordelt på køn, alder, bopæl, socialgruppe etc. En sådan followupmonitorering vil være en naturlig del af et screeningsprogram i Danmark, hvor disse muligheder findes. Ofte vil det dog være nødvendigt at kvalitetskontrollere data i de landsdækkende registre.

Disse analysemuligheder bør bruges, ikke blot fordi vi har dem, men for at sikre, at vi får mest ud af de resurser, vi vil anvende, og for at sikre, at vi ikke gør mere skade end gavn og får afsluttet programmer, der ikke virker godt nok. Især er det vigtigt at kontrollere, hvordan udviklingen i sensitivitet og specificitet er over tid. De mange forventede falsk positive testresultater kan betyde, at grænsen for klassifikationen af et positivt testresultat skrider, så man får en højere specificitet på bekostning af flere falsk negative resultater og mindre sensitivitet.

ANDRE STUDIER

I vurderingen af et screeningsprogram indgår naturligvis også andre komponenter [9-11], og man vil blive nødt til at indsamle oplysninger direkte fra deltagerne selv, deres erfaringer og meninger om screeningsprogrammet. Kvalitative studier har en vigtig rolle i screeningsforskningen.

LØBENDE MONITORERING AF SCREENING

Et screeningsprogram vil ofte lægge beslag på betydelige kliniske og samfundsmæssige resurser og skal derfor løbende monitoreres. Bivirkningerne kan blive hyppigere over tid, når den daglige rutine sætter ind, måske med svigtende kvalitetskontrol til følge.

Værdien af screeningen vil også afhænge af befolkningens lyst og villighed til at deltage i screeningsprogrammet og naturligvis også af kvaliteten af den kliniske behandling. Jo bedre behandlingsresultater, der opnås i sene stadier af sygdom, jo mindre screeningsgevinster vil man forvente, givet den tidlige behandling forbliver uændret. Hvis sen behandling giver lige så gode behandlingsresultater som tidlig behandling vil den eneste »effekt«, der så bliver tilbage være *lead time bias*.

EPILOG

Screening er en omkostningstung del af sundhedsvæsenet, og både den organiserede screening og den opportunistiske screening skal vurderes [9], fordi sundhedsvæsenet påtager sig et etisk ansvar for bivirkningerne. Deltagerne inviteres ind til prøver, test og undersøgelser på sundhedsvæsenets opfordring. De har ikke bedt om at blive undersøgt, og de bliver ikke undersøgt på grund af symptomer.

Screening har længe været en kamplads for begejstrede tilhængere og lige så arge modstandere. Sundhedsstyrelsen besluttede helt tilbage i 1991, at læger derfor skulle have adgang til en introduktionstekst om screening [12, 13]. Uenigheden gik så vidt i England, at man besluttede at nedsætte en gruppe af personer, som havde kendskab til epidemiologi og ikke havde deltaget i debatten eller været involveret i screeningsforskningen, med *M. Marmot* som formand [14].

Forskning i screening bør derfor styrkes for at give os et bedre beslutningsgrundlag, både til at igangsætte nye screeningsprogrammer og til at lukke programmer, der ikke længere har en værdi, der svarer til deres omkostninger [15].

KORRESPONDANCE: Jørn Olsen, Sektion for Epidemiologi, Institut For Folkesundhed, Aarhus Universitet, Bartholins Alle 2, 8000 Aarhus C. E-mail: jo@soci.au.dk

ANTAGET: 28. marts 2014

INTERESSEKONFLIKTER: Forfatterens ICMJE-formular er tilgængelig sammen med artiklen på Ugeskriftet.dk

LITTERATUR

- Day NE. The assessment of lead time and length bias in the evaluation of screening programmes. *Maturitas* 1985;7:51-8.
- Kronborg O, Fenger C, Olsen J et al. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996;348:1467-71.
- Day NE, Williams DR, Khaw KT. Breast cancer screening programmes: the development of a monitoring and evaluation system. *Br J Cancer* 1989;59:954-8.
- Day NE. Quantitative approaches to the evaluation of screening programs. *World J Surg* 1989;13:3-8.
- von Euler-Chelpin M, Lynge E, Rebolj M. Register-based studies of cancer screening effects. *Scand J Public Health* 2011;39(suppl 7):158-64.
- Junod B, Zahl PH, Kaplan RM et al. An investigation of the apparent breast cancer epidemic in France: screening and incidence trends in birth cohorts. *BMC Cancer* 2011;11:401.
- Weiss NS. Commentary: case-control studies of screening for colorectal cancer: tailoring the design and analysis to the specific research question. *Epidemiology* 2013;24:894-7.
- Olsen J, Christensen K, Murray J et al. An introduction to epidemiology for health professionals. New York: Springer, New York Dordrecht Heidelberg London, 2010.
- Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organized mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
- Nyström L, Andersson I, Bjurstam N et al. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002;359:909-19.
- Paci E, Euroscreen Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012;19(suppl 1):5-13.
- Hugod C, Fog J. Screening. Why, when, and how? København: Sundhedsstyrelsen, 1992.
- Arbejdsgruppe under Sundhedsstyrelsen. Screening. Hvorfor, hvornår, hvordan? Forebyggelse og Hygiejne 1990;13.
- Marmot MG. Sorting through the arguments on breast screening. *JAMA* 2013;309:2553-4.
- Miller AB, Wall C, Baines CJ et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* 2014;348:g366.