

# Ist es berechtigt, die Wirksamkeit der mammographischen Früherkennungsuntersuchung in Frage zu stellen?

Eine kritische Zusammenfassung des Artikels von Gøtzsche und Olsen

C. Bouchardy<sup>a</sup>, L. Raymond<sup>a</sup>, F. Levi<sup>b</sup>, N. Probst-Hensch<sup>c</sup>,  
 M. Tubiana<sup>d</sup>, C. de Wolf<sup>e</sup>

- <sup>a</sup> Institut de médecine sociale et préventive, Genf  
<sup>b</sup> Institut universitaire de médecine sociale et préventive, Lausanne  
<sup>c</sup> Institut für Sozial- und Präventivmedizin der Universität Basel, Basel  
<sup>d</sup> Centre Antoine-Béclère, Faculté de Médecine, Paris, Frankreich  
<sup>e</sup> Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland, Köln, Deutschland

## Zusammenfassung

Im Januar 2000 publizierte «The Lancet» einen von Gøtzsche und Olsen verfassten Artikel [1], der die Wirksamkeit des Mammographiescreenings anzweifelte. Gøtzsche und Olsen kritisierten darin die Methodik der prospektiven, randomisierten Studien. Sie beurteilten die Randomisierung in Studien, die eine Senkung der Brustkrebsmortalität zeigten, als suboptimal. Sie zogen daraus die Schlussfolgerung, dass Mammographiescreening ineffizient und deshalb nicht gerechtfertigt sei.

Die Schweizerische Krebsliga, in deren Aufgabebereich die Ausarbeitung nationaler Krebsbekämpfungsprogramme fällt, beauftragte eine Experten-Gruppe mit einer kritischen Analyse dieses Artikels. Unter Berücksichtigung der publizierten Repliken auf diesen Artikel und nach eingehender Prüfung der von Gøtzsche und Olsen vorgebrachten Argumente drängt sich die Schlussfolgerung auf, dass sie die Wirksamkeit des Mammographiescreenings nicht in Frage stellen können. Die von Gøtzsche und Olsen vorgenommene statistische Auswertung erfüllt nicht

die methodischen Anforderungen, die an die Evaluation von Studien solch praktischer Tragweite gestellt werden dürfen.

## Einführung

Ein von Gøtzsche und Olsen im Januar 2000 in «The Lancet» publizierter Artikel [1] zweifelt sogar die Wirksamkeit der Mammographie an: Die Autoren kritisierten die methodische Qualität jener randomisierten Studien, auf die sich die wichtigsten Wirksamkeitsbelege des Screenings stützen. Gøtzsche und Olsen behaupten, dass die «positiven» Studien (die eine signifikante Abnahme der Mortalität gezeigt haben) auf einer nicht «optimalen» Randomisierung beruhten, im Gegensatz zu den Studien mit negativen Ergebnissen. Daraus ziehen sie den Schluss, dass das Screening unwirksam und deshalb ungerechtfertigt sei. Die Argumente und Schlussfolgerungen dieses Artikels wurden unverzüglich von zahlreichen Fachleuten widerlegt [2–11].

Die Schweizerische Krebsliga, in deren Aufgabebereich die Ausarbeitung nationaler Krebsbekämpfungsprogramme fällt, beauftragte eine Experten-Gruppe mit einer kritischen Analyse dieses Artikels. Ziel der vorliegenden Arbeit ist es, die von Gøtzsche und Olsen vorgebrachten Argumente unter Einbezug der bereits publizierten Repliken auf den Artikel kritisch zu durchleuchten, und die Schlussfolgerungen der Autoren zu diskutieren.

## Artikel von Gøtzsche und Olsen

Nach dem Erscheinen eines Artikels mit der Schlussfolgerung, die Brustkrebsmortalität in Schweden, wo das Brustkrebs-Screening mittels Mammographie seit 1985 empfohlen wird, sei nicht zurückgegangen [12], werteten Gøtzsche und Olsen die 8 randomisierten Studien (Malmö, Kanada, Göteborg, Stockholm, Two-County [einzeln behandelt], New York und Edinburg) zur Wirksamkeit des Mammographiescreenings neu aus; die Autoren konzentrierten sich auf drei mögliche Ursachen der Effektverfälschung, die typischerweise bei therapeutischen Studien diskutiert werden, nämlich: a) nicht «optimale» Randomisierung, b) Fehlen der «blinden» Beurteilung der Todesursache, und c) Ausschlüsse von Frauen nach erfolgter Randomisierung. Parallel dazu versuchten die Autoren, die screeningbedingte Morbidität zu bestimmen. Ihre wichtigsten Schlussfolgerungen, die später kommentiert werden, können wie folgt zusammengefasst werden:

- Die Studien, welche eine Senkung der Brustkrebsmortalität zeigten, nämlich Göteborg, Stockholm, Two-County, New York, Edinburg und Phase II der Malmö-Studie, sind laut der Bewertung von Gøtzsche und Olsen mit einer Effektverfälschung, vor allem wegen Unterschieden in Alter und Sozialschicht zwischen der Interventionsgruppe und der Kontrollgruppe betroffen.

Korrespondenz:  
 Dr. Christine Bouchardy  
 Centre médical universitaire  
 Institut de médecine sociale et préventive  
 CH-1211 Genf 4

Diese Verzerrungen könnten möglicherweise erklären, weshalb die Gesamtmortalität in den schwedischen Studien erhöht war.

- Die Studien, die keine Senkung des Brustkrebsrisikos ergaben, nämlich die Malmö-Studie und die kanadische Studie, werden von Gøtzsche und Olsen auf Basis der gleichen Kriterien als nicht «Bias»-behaftet bewertet. Gemäss den Ergebnissen dieser beiden Studien wird die Screening-Mammographie als unwirksam und damit un gerechtfertigt betrachtet.

### Spezifische wissenschaftliche Kritikpunkte

#### Studie von Sjönell und Ståhle

Es sei zunächst daran erinnert, dass die Analyse von Sjönell und Ståhle [12], die der Auslöser der Studie von Gøtzsche und Olsen war, stark kritisiert und als unzulänglich eingestuft wurde, und zwar hinsichtlich ihres statistischen Powers, der Interpretation von Zeiteffekten und aufgrund mangelhafter Beschreibung der Rekrutierungs-Screening-Kohorte. [3, 13–19].

#### Schwedische Two-County-Studie

Die schwedische Two-County-Studie ist bei Gøtzsche und Olsen auf die schärfste Kritik gestossen. Es handelt sich um die grösste randomisierte Studie, und die Ergebnisse von Metaanalysen werden deshalb stark durch deren Resultate geprägt. Die Mortalitätsabnahme um 30% zwischen Interventions- und Kontrollgruppe blieb auch 20 Jahre nach der Randomisierung bestehen [20]. Die Kritik von Gøtzsche und Olsen lautet wie folgt:

- Die Interventionsgruppe ist durchschnittlich 5 Monate älter als die Kontrollgruppe (bei einer Population von 133 000 Frauen im Alter von 40 bis 74 Jahren); Gøtzsche und Olsen machen die Randomisierungsmodalitäten für diesen Altersunterschied verantwortlich.
- Die Analyse der Rohdaten ergibt für die Interventionsgruppe einen leichten Überschuss der Gesamtmortalität, bedingt durch den oben erwähnten Altersunterschied.
- Die Zahl der in die Studie aufgenommenen Frauen variiert von einer Publikation zur anderen.

Diese Kritikpunkte waren schon 1989 von den Autoren dieser Studie beantwortet worden. Sie kommentierten die Altersunterschiede und anerkannten, dass das Durchschnittsalter in der Interventionsgruppe leicht höher lag als in der Kontrollgruppe, was die leicht (nicht signifikant) höhere Mortalitätsrate an anderen Todesursachen erklärt [21]. Dieser Altersunterschied wurde übrigens von den Autoren adäquat korrigiert [21]. Die deutliche Senkung der Brustkrebsmortalität wurde bestätigt (relatives, nicht korrigiertes Risiko: 0,70; relatives, alterskorrigiertes Risiko: 0,69). Der Unterschied in der Gesamtmortalität wurde ausgeglichen (relatives, nicht korrigiertes Risiko: 1,02; relatives, korrigiertes Risiko: 0,99). Ausserdem kann – wie de Koning in seinem gleichzeitig

erschienenen Kommentar zum Artikel von Gøtzsche und Olsen festhält – dieser minimale Altersunterschied als «konservativ» betrachtet werden, weil dadurch die Vorteile des Screenings unterschätzt werden [3].

Die Inkonsistenz betreffend publizierte Teilnehmerzahlen war ebenfalls schon früher Gegenstand einer Erklärung der Autoren [21]. In der 1985 publizierten Originalanalyse [22] waren Frauen, bei denen vor Eintritt in die Studie ein Brusttumor diagnostiziert worden war, von den in Interventions- und Kontrollgruppe auftretenden Brustkrebsfällen und Brustkrebstodesfällen ausgeschlossen worden. In der Folge konnte durch eine Zusammenlegung mit den Daten der Krebsregister dieses Ausschlusskriterium auf die Gesamtheit der randomisierten Frauen erweitert werden. Deshalb umfasst die Studie seit 1989 eine «bereinigte» Population, die einige hundert Frauen weniger umfasst als die ursprüngliche Studienpopulation.

Cluster-Randomisierung ist gerechtfertigt und valid, wenn sie auf Gruppen von Frauen aus sozioökonomisch durchmischten Regionen angewandt wird, auch wenn es stimmt, dass diese weniger wirkungsvoll ist als die individuelle Randomisierung [23].

#### Göteborg-Studie

Bei den für die Frauen zwischen 39 und 49 Jahren publizierten Ergebnissen [24] finden Gøtzsche und Olsen einen (als statistisch signifikant beurteilten) Unterschied von einem Monat zwischen dem Durchschnittsalter der Frauen der Interventions- und den Kontrollgruppen, was ihrer Ansicht nach den Ausschluss der Studie aus den Metaanalysen rechtfertigt. In diesem Fall war die Randomisierung in Kohorten von aufeinanderfolgenden Geburtsjahren erfolgt. Weil der Randomisierungskoeffizient in Abhängigkeit der Kapazität zur Durchführung des Screenings variierte, waren die Altersunterschiede erwartet und die Analyse zur Altersanpassung bereits im voraus geplant worden und nicht erst im nachhinein, wie von Gøtzsche und Olsen behauptet. Idealerweise hätte die Korrektur nach Geburtsjahr und nicht nach Alter erfolgen sollen. In diesem Fall ergibt sich ein relatives Brustkrebsmortalitätsrisiko von 0,54 (CI 95%: 0,30 bis 0,96); dieser Effekt stimmt mit dem alterskorrigierten Effekt von 0,55 (CI 95%: 0,31 bis 0,96) [25] überein.

#### Malmö-Studie

Gøtzsche und Olsen berücksichtigten nur die ersten, negativen Ergebnisse der Malmö-Studie (Malmö I) und schlossen die später publizierten Ergebnisse von ihrer Reanalyse aus (Malmö II), die dank Vergrößerung der Studienpopulation und einem längeren Follow-up eine Senkung des relativen Brustkrebsmortalitätsrisikos auf 0,81 zeigten. Gøtzsche und Olsen rechtfertigten diesen Ausschluss damit, dass bei Malmö II alle aufgenommenen Frauenkohorten vermischt wurden, und dass wegen eines administrativen Fehlers die 1934 geborenen Frauen nicht rando-

**Tabelle 1**

Wichtigste Unterschiede zwischen Interventionsstudien in der Bevölkerung und klinischen Studien.

Klinische Studien	Interventionsstudien in der Bevölkerung
<i>Beispiel</i>	
Wirksamkeit der Einnahme eines Moleküls mit angenommener Wirkung versus Placebo	Wirksamkeit der systematischen Einladung zur Mammographie versus gängige medizinische Praxis
<i>Art der betroffenen Personen</i>	
Patienten	Theoretisch gesunde Bevölkerung
<i>Anzahl betroffener Personen</i>	
Ungefähr zehn bis hundert, selten tausend	Ungefähr tausend bis mehrere zehntausend
<i>Vergleichbarkeit der Gruppen</i>	
Unterschiede bei den prognostischen Kriterien sind nicht akzeptabel	Auch kleinste Unterschiede werden wegen der grossen Zahl beteiligter Personen signifikant
<i>Compliance in der Gruppe der Behandelten/Interventionsgruppe</i>	
Wichtige oder vollumfängliche Bedeutung	Teilweise Bedeutung
<i>Protokollabweichungen/Kontamination der Kontrollgruppe</i>	
Keine oder extrem kleine	Unvermeidlich und manchmal bedeutend, müssen dokumentiert werden
<i>Ausscheiden oder Ausschluss nach Randomisierung</i>	
Ausnahmefälle	Unvermeidlich

misiert, sondern alle in die Screeninggruppe aufgenommen worden waren. Weil Gøtzsche und Olsen nicht über Daten zur Altersverteilung verfügten, schätzten sie das Durchschnittsalter der Interventionsgruppe der Malmö-Studie auf sehr komplizierte und approximative Weise, ausgehend von den Meta-analysedaten der schwedischen Studien: «We used the number of women as reported in the meta-analysis and the mean ages as estimated above. We took account of the fact that women in Göteborg were randomly allocated to study and control groups in the approximate ratio of 1.2 in the 39–49-year age-group and 1.6 in the 50–59-year age-group. We had no data on age for the 50–59-year group, but since the imbalance in age in the 39–49-year group was numerically small, we used a mean age of 54 for both study and control groups. For Malmö, we used 57 years as estimated mean age in the study group, similar to the Koppberg and Östergötland trials. This approach yielded a mean age in the study groups of 54.93 years, very close to the 55.05 years reported in the meta-analysis. Since the mean age in the control groups was 0.51 years lower, that in the Malmö control group was estimated to be 56.85 years. The difference of 0.15 years is not significant [ $z = 1.53$ ,  $p = 0.13$ ] which suggests that the randomisation method in Malmö was adequate» [1]. Sie leiteten dar-

aus ab, dass der mit 0,15 Jahren bezifferte und auf Gruppen von jeweils 21 000 Frauen bezogene Altersunterschied unbedeutend sei. Bei der Göteborg-Studie beurteilten sie die Altersdifferenz von nur 0,09 Jahren, bezogen auf 12 000 Frauen, als signifikant für eine mögliche Effektverfälschung.

### Kanadische Studie

Die Untersuchung der Daten der kanadischen Studie bleibt erstaunlich oberflächlich im Vergleich zur harschen Kritik an den positiven Studien. Gøtzsche und Olsen akzeptieren die Korrektur der Randomisierung in der kanadischen Studie vorbehaltlos, trotz offenkundiger Effektverfälschungen, die weitaus schwerer wiegen als die Altersunterschiede, insbesondere die hohe Prävalenz fortgeschrittener Tumoren bei Studieneintritt [26–33]. Diese Studie kann auch kaum als bevölkerungsbezogene Studie betrachtet werden, da die Beteiligung in der Interventionsgruppe lediglich 25% betrug. Ausserdem erstaunt, dass Gøtzsche und Olsen behaupten, die kanadische Studie sei mit Abstand die am besten dokumentierte, aber gleichzeitig selbst aussagen, sie hätten keine Informationen über die Altersverteilung in den Vergleichsgruppen gefunden. Diese Information war ihnen erst nach der Publikation ihres Artikels zugänglich [34].

### Wissenschaftliche Kritikpunkte allgemeiner Art

#### Methodische Mängel

Gøtzsche und Olsen hatten keinen Zugang zu den gesamten individuellen Daten der randomisierten Studien und mussten sich deshalb mit den publizierten Daten begnügen, die häufig unvollständig, schon gruppiert oder aufgegliedert sind. Es scheint, dass sie auf wenig systematische Weise Kontakt mit den Studienverantwortlichen aufgenommen hatten, um weitere Details in Erfahrung zu bringen, insbesondere betreffend Randomisierungsmethoden. Sie geben an, die gewünschten Informationen nur sehr lückenhaft erhalten zu haben. So sollen in 2 von 8 Studien die Altersdaten und in 5 von 8 Studien die Daten zu sozioökonomischen Charakteristiken angeblich fehlen. Deshalb hätten sie eigentlich zum vornherein anerkennen müssen, dass ihre Review nicht machbar war. Zu dieser Einsicht sind Gøtzsche und Olsen aber erst später gelangt [35]. Zunächst haben sie versucht, dieses Informationsdefizit in mehreren Fällen durch Extrapolationen wettzumachen, die zumindest als gewagt bezeichnet werden müssen.

Die beobachteten Altersunterschiede sind entgegen den Behauptungen von Gøtzsche und Olsen nicht zwingend das Resultat eines suboptimalen Randomisierungsprozesses. Die Gesetze der Statistik lehren uns, dass selbst Unterschiede, deren Auftretenswahrscheinlichkeit sehr klein ist, eben doch ab und zu auftreten können. Es ist offensichtlich, dass die Durchführung von bevölkerungsbezogenen Studien, bei denen die Stichprobenziehung normalerweise auf Einwohnerregistern basiert, nicht zum primären Tätigkeitsbereich von Gøtzsche und Olsen gehören.

Einwohnerregister sind nie hundertprozentig à jour und enthalten immer eine gewisse Anzahl Personen, die in der Zwischenzeit nicht mehr in der Region wohnen (Auswanderung oder Rückwanderung). Folglich weiss man es nur von Teilnehmern der Studie, wenn sie weggezogen sind (zum Beispiel wenn die Post Einladungsbriefe für die Studienteilnahme retourniert). Deshalb ist ein gewisser Unterschied zwischen der Interventionsgruppe und der Kontrollgruppe unvermeidlich, insbesondere in bezug auf die Altersstruktur. Die Grösse des Unterschiedes hängt dabei vom Einfluss des Alters auf die Bevölkerungsbewegungen ab. Diese Unterschiede geben oft Anlass zu einer nachträglichen Korrektur bei der statistischen Auswertung.

Zumindest hätten Gøtzsche und Olsen daran denken müssen, dass Interventionsstudien in der Bevölkerung nur teilweise mit klinischen Studien vergleichbar sind. Die spezifischen Besonderheiten der beiden Methoden, wie sie in Tabelle 1 wiedergegeben sind, hätten unbedingt erwähnt werden müssen. So hätten sie besonders der Tatsache Rechnung tragen müssen, dass wegen der grossen Zahl der an Interventionsstudien beteiligten Frauen auch vernachlässigbare Altersunterschiede (von einigen Monaten) statistisch signifikant werden können, während in klinischen Studien mit zirka hundert Personen auch Unterschiede von mehreren Jahren oft nicht statistisch signifikant sind. Die Vergleichbarkeit der Gruppen kann in Bevölkerungsstudien nicht immer aufrechterhalten werden, vor allem wegen des unumgänglichen «loss to follow-up»; deshalb ist es üblich, bei der Auswertung der Resultate die notwendigen Anpassungen vorzunehmen, während dies in klinischen Studien nicht gerechtfertigt ist. Inzidenz- und Mortalitätsraten, die nicht altersstandardisiert sind, verunmöglichen jeden regionalen oder internationalen Vergleich. Gøtzsche und Olsen stellen jedoch dieses in der Epidemiologie geläufige Vorgehen stark in Frage. Ihre Überlegungen basieren auf den für klinische Studien üblichen Rahmenbedingungen, die darauf abzielen, die tatsächliche Wirkung der Behandlung zu messen, während bei Interventionsstudien in der Bevölkerung ein Verdünnungseffekt (zum Beispiel durch Nichtteilnahme eingeladener Frauen und Teilnahme von Frauen aus der Kontrollgruppe) unvermeidlich ist. Ausserdem ist es bekanntlich üblich, die Resultate einer einzigen klinischen Studie als schlüssig zu betrachten, während Schlussfolgerungen aus Interventionsstudien in der Bevölkerung im allgemeinen weniger endgültig sind und wiederholt werden müssen, um die Kausalität zwischen Intervention und Wirkung zu belegen.

Gøtzsche und Olsen betonen zu Recht den Vorwurf an einzelne randomisierte Studien, dass die Todesursachen nicht «blind» erhoben wurden. Dieses Problem wurde bereits beschrieben und diskutiert (siehe [36]). So wurden die Todesfälle bei der Metaanalyse der schwedischen Studien nach «Blinding» betreffend Screeningstatus ausgewertet [37]. Ausserdem stellt sich dieses Problem bei den Fallkontrollstudien nicht, weil diese per definitionem von der

Todesursache ausgehen und im nachhinein das Screeningverhalten erfassen; diese Studien finden jedoch ebenfalls eine deutliche Senkung der Brustkrebsmortalität bei Frauen, die an Mammographie-screening teilgenommen hatten.

Die statistische Methode, mit der die Heterogenität zwischen den Studien evaluiert werden sollte, wurde ungenügend beschrieben. Die Autoren verwendeten ein «Fix effect model», das heisst ein Modell, bei welchem die Variablen im voraus festgelegt wurden, und das nur vorgängig definierte Störvariablen berücksichtigt, ohne deren mögliche Interaktionen zu berücksichtigen. Die im Modell berücksichtigten Variablen sind nicht beschrieben. Während in epidemiologischen und klinischen Studien der Grenzwert für statistische Signifikanz traditionell auf 5% festgesetzt wird, erhöhten Gøtzsche und Olsen diesen Grenzwert in gewissen Fällen auf 10%, was es ihnen ermöglichte, einen statistisch signifikanten P-Wert für die Heterogenität zwischen den eingeschlossenen und den ausgeschlossenen Studien zu erhalten ( $p = 0,08$ ). Diese Heterogenität hätte mit der konventionellen Signifikanzgrenze nicht festgestellt werden können. Die Methode, mit der die Unterschiede zwischen der Interventionsgruppe und der Kontrollgruppe untersucht wurden, vor allem bezüglich Alter, ist nicht beschrieben. Man weiss nicht, ob das Alter bei Studieneintritt in Jahren und Monaten (als kontinuierliche Variable), in gerundeten Jahren oder in Form von Altersgruppen berücksichtigt wurde. Je nach gewähltem Vorgehen verändern sich zwingend die Schätzungen vom Durchschnittsalter und von Standardabweichungen. Die Schätzungen der Autoren können deshalb nur sehr annäherungsweise nachvollzogen werden. Es bleibt damit erstaunlich, dass kleine Altersdifferenzen (zum Beispiel 0,09 in der Göteborg-Studie) als signifikant beurteilt werden, während grössere Unterschiede bei einer grösseren Anzahl Frauen es nicht sind (zum Beispiel 0,15 in der Malmö-Studie). Zudem haben besonders bei der Göteborg-Studie unsere eigenen Berechnungen das Vorhandensein eines statistisch signifikanten Unterschieds nicht bestätigt.

Anerkennt man schliesslich, dass die jeder dieser Studien inhärenten, unvermeidlichen potenziellen Verzerrungen den Schätzwert für die tatsächliche Wirksamkeit des Screenings entweder reduzieren oder erhöhen, ist gerade eine Metaanalyse über sämtliche Studien unverzichtbar. Schlussfolgerungen, die auf nur 20% der Daten aus den randomisierten Studien basieren, dürfen deshalb angezweifelt werden. Es ist offenkundig, dass ein entsprechendes Weglassen eines Teils von Resultaten in einer Metaanalyse oder in einer Literaturreview über die Wirksamkeit einer therapeutischen Intervention mit Skepsis aufgenommen worden wäre.

Gøtzsche und Olsen diskutieren das Ausmass der möglichen Effektverfälschung auf die Schätzung der Wirksamkeit der Mammographie nicht. Sie begnügen sich mit der Feststellung, dass unvollständige Randomisierung in klinischen Studien theoretisch einen grossen Teil oder gar den gesamten beobachteten

**Tabelle 2**

Wichtigste Kommentare und Schlussfolgerungen von Gøtzsche und Olsen zu den randomisierten Studien und zum relativen Risiko betreffend Brustkrebsmortalität.

Studie	Schlussfolgerungen von Gøtzsche und Olsen	Wichtigste Kommentare betreffend beobachtete Unterschiede zwischen Interventions- und Kontrollgruppe	Relatives Risiko der Brustkrebsmortalität
Malmö I	Adäquat	Altersverteilung nicht verfügbar. Approximativ errechneter Unterschied von 0,15 Jahren*, wird als nicht signifikant beurteilt	0,96
Kanada	Adäquat	Vergleichbares Bildungsniveau; Altersverteilung nicht verfügbar**	1,08
Malmö II	Inadäquat	Altersverteilung nicht verfügbar. Frauen mit Geburtsjahr 1934 wurden nicht randomisiert, sondern alle in die Interventionsgruppe eingeschlossen	0,81
Göteborg	Inadäquat	Als relevant beurteilter Altersunterschied: Frauen in der Interventionsgruppe 0,09 Jahre jünger	0,55 (Alter: 39 – 49)
Stockholm	Inadäquat	Als relevant beurteilter Altersunterschied: Frauen in der Kontrollgruppe 0,18 Jahre jünger	0,73
Two-County (Kopparberg und Östergötland)	Inadäquat (beide Studien)	Als relevant beurteilter Altersunterschied: Frauen in der Kontrollgruppe Kopparberg 0,45 Jahre älter, in der Kontrollgruppe Östergötland 0,27 Jahre älter	0,70
New York	Inadäquat	Unterschiede bei den Risikofaktoren, insbesondere bei der Schulbildung	0,79
Edinburg	Inadäquat	Unterschiede bei der sozioökonomischen Schicht	0,87

\* Ungefähres Alter, Schätzung von Gøtzsche und Olsen ausgehend von den Daten der Metaanalyse der schwedischen Studien.

\*\* Alter von Gøtzsche und Olsen nicht gefunden.

Effekt einer therapeutischen Intervention erklären kann. Im konkreten Fall ist aber klar, dass die zwischen der Interventions- und Kontrollgruppe festgestellten Altersunterschiede eine Mortalitätsabnahme in der Grössenordnung von 25% nicht erklären können. Nach Berechnungen, die wir ausgehend von den schwedischen Daten zur Brustkrebsmortalität von 1995 (WHO-Daten) durchgeführt haben, würde ein Unterschied von einem Jahr im Durchschnittsalter zwischen zwei Bevölkerungsgruppen zwischen 40 und 70 Jahren nur einen relativen Unterschied bezüglich Brustkrebsmortalität in der Grössenordnung von 6% erklären. Die grösste Altersdifferenz, die jedoch von Gøtzsche und Olsen festgestellt wurde (Kopparberg-Studie), beträgt 0,45 Altersjahre, was einen Unterschied in der Brustkrebsmortalität von weniger als 3% ergibt. Zudem diskutieren Gøtzsche und Olsen die Tatsache nicht, dass selbst Resultate von Studien mit potenziell entgegengesetzten Verzerrungen in bezug auf die Senkung der Brustkrebsmortalität übereinstimmen. Tatsächlich waren in der schwedischen Two-County-Studie die Frauen der Interventionsgruppe älter als die der Kontrollgruppe (um 0,45 Jahre in der Kopparberg-Studie und um 0,27 Jahre in der Östergötland-Studie), während in der Stockholm-Studie die eingeladenen Frauen jünger waren als in der Kontrollgruppe (um 0,18 Jahre). Diese Altersdifferenzen sollten zur Folge haben, dass die Wirksamkeit des Screenings im ersten Fall zu tief und im zweiten Fall zu hoch eingeschätzt wird. Die Autoren berücksichtigen auch die Wirksamkeitsun-

terschiede des Mammographiescreenings in Abhängigkeit des Alters nicht. Schliesslich gibt es auch keinen Hinweis dafür, dass durch die Randomisierung systematisch ein Bias zugunsten der Interventionsgruppe eingeführt worden wären. Wenn zudem die in den schwedischen Studien beobachtete Senkung der Brustkrebsmortalität nur durch diesen Bias bedingt gewesen wäre, hätte es keinen Grund gegeben, dass diese Verzerrung nicht auch bei den unter 50jährigen gleich gross gewesen wäre, in einem Altersbereich, wo die geringere Wirksamkeit des Screenings nur schwer zu bestreiten ist [38–40]. Auch hätte die Tatsache, dass die eingeladenen Frauen einer begünstigteren sozioökonomischen Schicht angehört haben sollen (wie dies anscheinend bei den Studien aus New York und Edinburg der Fall war), normalerweise eine zu tiefe Schätzung der Wirkung des Screenings zur Folge gehabt, weil das Brustkrebsrisiko in sozial besser gestellten Kreisen höher ist.

**Offensichtlicher Mangel an Unvoreingenommenheit**  
 Keine Interventionsstudie mit Hunderten oder Tausenden von Personen kann perfekt sein; Abweichungen vom Protokoll sind unvermeidlich. Als erstes muss in Frage gestellt werden, ob eine Altersdifferenz dichotomes und definitives Kriterium zur Beurteilung einer Randomisierungsverzerrung sein kann. Unabhängig von der Stichprobengrösse dürfen nämlich Altersunterschiede nicht ausschliesslich der Randomisierung zugeschrieben werden. Ausserdem werden bei so grossen Teilnehmerzahlen auch kleinste Altersunter-

schiede statistisch signifikant. Gemäss Wahrscheinlichkeitsrechnung sind bei mindestens einer von 8 Studien statistisch signifikante Unterschiede vorhanden. Zudem wurde das Alterskriterium nicht konsequent angewendet, indem die kanadische Studie und die erste Phase der Malmö-Studie trotz fehlender Altersangaben akzeptiert wurden, während die schwedischen Studien auf der Basis minimaler, aber als relevant beurteilter Unterschiede ausgeschlossen wurden (Tab. 2).

Gøtzsche kennt den «Investigator-Bias» im Zusammenhang mit nicht blind durchgeführten Beurteilungen genau, er hatte diese Gefahr im leicht provokativen Artikel «Blinding during data analysis and writing of manuscripts» beschrieben [41]. Seine Wahl des Beurteilungskriteriums für den Randomisierungserfolg und die Intensität, mit der die Autoren nach möglichen Verzerrungen des Schätzwertes suchen, scheint vielmehr durch das Resultat der entsprechenden Studien als durch offensichtliche Randomisierungsfehler oder objektive Kriterien geleitet zu sein. Wie zufällig führt die Anwendung dieses Selektionskriteriums dazu, das Gøtzsche und Olsen alle Studien verwerfen, deren Ergebnisse zugunsten des Screenings lauten. Ein Mangel an Unvoreingenommenheit ist umso wahrscheinlicher, als man den Zusammenhang kennt, der den Impuls für den Bericht von Gøtzsche und Olsen gegeben hat, sowie die a priori negative Einstellung des ersten Autors gegenüber Vorsorgeuntersuchungen im allgemeinen: «Do we wish to turn the world's healthy citizens into fearful patients-to-be who, in the not too distant future, might be asked to deliver, for example, annual samples of faeces, urine, sputum, vaginal smear, and blood, and undergo X-ray and ultrasound examination with all they entails in term of psychological morbidity and the potential for harm because of further testing and interventions due to false positive findings?» [42]. Diese a priori negative Haltung gegenüber dem Screening hätte eine kritische Analyse der Studien ohne Kenntnis ihrer Ergebnisse erfordert.

#### Die Studie ist keine Metaanalyse der Cochrane-Organisation

Gøtzsche und Olsen erwecken den falschen Eindruck, ihr Lancet-Artikel stelle einen offiziellen Bericht der internationalen Forschungsorganisation Cochrane Collaboration dar, die sich aber öffentlich von dieser Publikation distanziert hat. Die Antwort von Andy Oxman (Chairperson of Cochrane Collaboration Steering Group) zu diesem Thema ist deutlich: «Although Gøtzsche and Olsen have a Cochrane protocol in The Cochrane Library, they have not yet submitted a Cochrane review to the Cochrane Breast Cancer Group. What they published in the Lancet is not a Cochrane review and has no Cochrane status other than arising from two people who work in a Cochrane Centre. The Cochrane Collaboration had no control (editorial or otherwise) over this review» [43, 4]. Zurzeit wird vom Cochrane Steering Committee on breast cancer erwogen, einen offiziellen Bericht herauszugeben ([www.cochrane.org](http://www.cochrane.org)).

#### Schlussfolgerungen

Die Argumentation von Gøtzsche und Olsen kann man wie folgt zusammenfassen: Weil es Altersunterschiede zwischen der Interventionsgruppe und der Kontrollgruppe gibt, vermuten wir, dass die Randomisierung dieser Studien nicht «optimal» ist. Deshalb dürfen wir die Studienresultate nicht berücksichtigen. Wenn wir die Resultate dieser Studien ausschliessen, haben wir keinen Beweis mehr, dass das Screening wirksam ist, und folglich ist dieses nicht gerechtfertigt.

Aufgrund einer inadäquaten epidemiologischen Analyse ziehen Gøtzsche und Olsen Schlussfolgerungen, die sie als definitiv darstellen, die aber in Tat und Wahrheit unverhältnismässig sind. Aus wissenschaftlicher Sicht ist es ungerechtfertigt, unter Bezugnahme auf den Artikel von Gøtzsche und Olsen die Wirksamkeit des Mammographiescreenings in Frage zu stellen. Ihre allenfalls anregende statistische Auswertung ist weder stichhaltig noch objektiv. Sie genügt den Anforderungen nicht, die man zu Recht von Studien erwarten darf, die eine Massnahme von dieser praktischen Tragweite evaluieren. Sie wird von den internationalen Experten auf dem Gebiet der Brustkrebsvorsorge stark kritisiert und ist in den Kontext einer Randdebatte von Screeninggegnern auf der Suche nach Publizität einzuordnen.

#### Literatur

- 1 Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129-34.
- 2 Cates C, Senn S. Screening mammography re-evaluated (letter). *Lancet* 2000;355:750.
- 3 De Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000;355:80-1.
- 4 De Landtsheer JP, Delaloye JF, Hessler C, De Grandi P, Paccaud F, Levi F. Dépistage organisé du cancer du sein. L'expérience vaudoise. *Rev Med Suisse Romande* 2000; 120:501-10.
- 5 Duffy SW, Tabár L. Screening mammography re-evaluated (letter). *Lancet* 2000;355:747-8.
- 6 Hayes C, Fitzpatrick P, Daly L, Buttner J. Screening mammography re-evaluated (letter). *Lancet* 2000;355:749.
- 7 Law M, Hackshaw A, Wald N. Screening mammography re-evaluated (letter). *Lancet* 2000;355:750.
- 8 Moss S, Blanks R, Quinn MJ. Screening mammography re-evaluated (letter). *Lancet* 2000;355:748.
- 9 Nyström L. Screening mammography re-evaluated (letter). *Lancet* 2000;355:748-9.
- 10 Palmieri C, Fishpool S. Breast cancer screening. *Br Med J* 2000;321:567-8.
- 11 Wald N. Populist instead of professional. *J Med Screen* 2000;7:1.
- 12 Sjönell G, Ståhle L. Hälsokontroller med mammografi minskar inte dödlighet i bröstcancer. *Läkartidningen* 1999; 96:904-13.
- 13 Rosen M, Lundin A, Nyström L, Rutqvist LE, Stenbeck M, Talbäck M. Incidence and mortality of breast cancer during a 25-year period. International and regional comparisons. *Läkartidningen* 2000;97(4):294-9.
- 14 Rosen M, Nyström L, Stenbeck M. The National Board of Health and Welfare: stronger evidence basis for the benefits of mammography. *Läkartidningen* 2000;97(7):739-40.

- 15 Rosen M, Stenbeck M. The debate on mammography, the National Board of Health and Welfare is answering Sjönell and Ståhle. *Läkartidningen* 2000;97(8):859-60.
- 16 Rehnqvist N, Rosén M, Karlberg I. The National Board of Health and Welfare on mammographic screening: mortality analysis requires completely different methods. *Läkartidningen* 1999;96(9):1050-1.
- 17 Rosen M, Rehnqvist N. Breast screening. No need to reconsider breast screening programme on basis of results from defective study. *Br Med J* 1999;318(7186):809-10.
- 18 Rutqvist LE. Naturalförloppet, grova metoder ledde till felkalkyl om bröstcancer. *Läkartidningen* 1999;96:1210-11.
- 19 Tabår L. Assessment of the effect of mammographic screening can not be based on wrong data. *Läkartidningen* 1999;96(14):1763-4.
- 20 Tabår L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiol Clin North Am* 2000;38(4):625-51.
- 21 Tabår L, Fagerberg G, Duffy SW, Day NE. The Swedish two-county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* 1989;43:107-14.
- 22 Tabår L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Gröntoft O, et al. Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 1985;i:829-32.
- 23 Duffy SW, South MC, Day NE. Cluster randomization in large public health trials: the importance of antecedent data. *Stat Med* 1992;11(3):307-16.
- 24 Bjurstam N, Björnelid L, Duffy SW, Smith TC, Cahlin E, Eriksson O, et al. The Gothenburg breast screening trial. First results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. *Cancer* 1997;80(11):2091-9.
- 25 Bjurstam N, Björnelid L, Duffy SW, Prevost TC. Author reply. *Cancer* 1998;83:188-90.
- 26 Kopans DB. Canadian National Breast Screening Study. *Lancet* 1997;350(9080): 810.
- 27 Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. *Cancer* 1995;75(4):997-1003.
- 28 Baines CJ. A different view on what is known about breast screening and the Canadian National Breast Screening Study. *Cancer* 1994;74(4):1207-11.
- 29 Kopans DB, Halpern E, Hulka CA. Statistical power in breast cancer screening trials and mortality reduction among women 40-49 years of age with particular emphasis on the National Breast Screening Study of Canada. *Cancer* 1994;74(4):1196-203.
- 30 Kopans DB, Halpern E, Hulka CA. Mammography screening for breast cancer. Reply to the commentaries. *Cancer* 1994;74(4):1212-6.
- 31 Mettlin CJ, Smart CR. The Canadian National Breast Screening Study: an appraisal and implications for early detection policy. *Cancer* 1993;72:1461-5.
- 32 Miller AB, Baines CJ, To T, Wall C. Canadian national breast screening study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Can Med Assoc J* 1992;147(10):1459-76.
- 33 Miller AB, Baines CJ, To T, Wall C. Canadian national breast screening study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Can Med Assoc J* 1992;147(10):1477-88.
- 34 Miller AB, Baines CJ, To T, Wall C. Screening mammography re-evaluated (letter) *Lancet* 2000;355:747.
- 35 Gøtzsche PC. Openness about the Swedish breast cancer screening trials is needed. *Läkartidningen*. 2000;97(3):2-3.
- 36 Hurley SF, Kaldor JM. The benefits and risks of mammographic screening for breast cancer. *Epidemiol Rev* 1992;14:101-30.
- 37 Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993;341(8851):973-8.
- 38 De Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst* 1995;87(16):1217-23.
- 39 Tabår L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, Smith RA. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer* 1995;75(10):2507-17.
- 40 Tabår L, Fagerberg G, Duffy SW, Day NE, Gad A, Gröntoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992;30:187-210.
- 41 Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials* 1996;17:285-93.
- 42 Gøtzsche PC. Screening for colorectal cancer. *Lancet* 1997;349(9048):356. Discussion 358.
- 43 Dean PB. The Swedish mammography screening trials. Check up on your sources. *Läkartidningen* 2000;97(3):2.
- 44 Dean PB. The articles by Gøtzsche and Olsen are not official Cochrane review and lack scientific merit. *Läkartidningen* 2000;97(3):3.

La version française a paru dans le no 12/2001

Siehe auch Editorial in Nr. 12/2001