Expert review Panel

Does new information have a credible scientific impact on the reliability of the CNBSS' published?

A report from Prof. Kalager to the University of Toronto, March 1 2024

Table of Content

PART I Organization
1. Mandate
2. Organization of the report
3. Organization of the panel´s work

PART II Mandate and Methods
4. The review panel
5. Methods
    Interview schedule
    Interview guide
    Interviewees
    Conflict of interest

PART III Complaint material
6. Complaint material
7. Disclosure of conflicts of interest

PART IV Previous criticism and review
8. Canadian National Breast Screening Study (CNBSS)
9. Criticism of CNBSS 1993
10. Rebuttal C. Baines
11. Criticism of CNBSS 1995
12. Previous Review of CNBSS

PART V Detailed transcripts from interviews
13. Methods
14. Description of the trial
15. Recruitment
16. Coming forward
17. Randomization
18. Claims from witnesses
    19. How could randomization be subverted?
20. Women with lumps
21. Follow-up after mammography
22. Training and equipment
23. Review
24. Equipment compared to the standard at the time
25. Views
26. Mammography screening outside the trial

27. Comparability of other trials at the time
28. How can the quality be so poor?

PART VI Summary of the panel´s review and interview
29. Summary of interview transcripts
30. Summary of previous critique, rebuttal, and previous review
31. Summary of results from CNBSS that may or may not indicate violation of randomization or poor quality mammography
32. Data provided by [redacted]
33. Comparison with other studies
34. Was CNBSS an outlier?
35. Burden and standard of proof
Conclusion

Mandate

The Canadian National breast Study Expert Review Panel was "*established as a result of a formal request from the Canadian Cancer Society and a group of academic researchers to interview people who they say have come forward with information that may have an impact on randomization, recruitment of symptomatic women, and poor radiographic image quality in the Canadian National Breast Screening Study (CNBSS)*".

The panel had the following responsibilities:
1. Meet with both **Professor Martin Yaffe**, the corresponding author for the allegation, and **Professors Emeriti Anthony Miller** and **Cornelia Baines**, the study's principal investigators.

2. Interview as many as possible of the following five individuals identified in the allegation to ascertain whether information or other documentation they may provide would have a credible scientific impact on the reliability of the CNBSS' published recommendations:

- [redacted] **– radiologist**
- [redacted] **– radiological technologist**
- [redacted] **- radiologist**
- [redacted] **- radiologist**
- [redacted] **– radiologist**

Additional individuals may be interviewed at the panel's discretion, including any other research personnel identified by the CNBSS principal investigators.

3. Deliver a final report that details the panel's assessment of whether this **new information would have a credible scientific impact** on the reliability of the CNBSS' published.

Recommendations, and, if they wish, to place the CNBSS in the context of the broader literature.

4. Keep the Associate Vice-President, Research (Oversight and Compliance) updated about the process and outcome, as appropriate.

Organization of the report

The report is organized in six parts with altogether 35 chapters, and a conclusion.
Part 1-3 is a review of previous criticism, previous review and the ongoing debate and complaint material (the basis of the present review). Part 4-6 includes detailed transcripts from all interviews we conducted, and a summary of findings from the interviews, with some transcripts and our comments and interpretations.

The first part presents the mandate, the interview panel, the interviewees, reported potential conflicts of interest, and the method the panel used, including the interview guide.

The second part presents the complaint material. The complaint material consists of a Summary report, a publication, letters, and an open letter from Dr. Kopans. This part is not only a summary of the findings, but also includes our comments on the material. It also includes a chapter about disclosures of potential conflict of interest in medical guidelines, and how to recognize issues that may raise red flags of a guideline quality and trustworthiness.

The third part summarizes previous criticism of the Canadian National Breast Screening Study (CNBSS), rebuttals by the trialists, the previous review of CNBSS in 1995, and the discussion that followed that review.

The forth part contains detailed transcripts of all interviews, organized as follows: description of the trial, recruitment, coming forward, randomization, claims from witnesses of manipulation of the trial, how randomization could be subverted, women with lumps, follow-up after mammography, training, equipment compared to the standard at the time, views, mammography outside the trial, comparability to other trials, and a discussion of findings that may indirectly reflect quality of mammography in the CNBSS, reported in the following two publications: *Miller et al. 25 year follow-up for breast cancer incidence and mortality of the Canadian National Breast screening Study: randomized screening trial, BMJ 2014; Narod SA. On being the right size: a reappraisal of mammography trials in Canada and Sweden; The Lancet 1997*.

This part does not contain any comment from our side by merely reports the interviews. But we have organized the transcripts by headlines to improve readability.

The fifth part is a summary of findings from our interviews including our comments covering violation of randomization, mammography outside of the trial, quality, comparability to other trials, mammographic views, and reflections of CNBSS and other screening trials. In addition, we have included a summary of a previous review of the technical quality of the CNBSS, and a chapter on an offer to review the quality of mammograms we heard about in our interviews.

The sixth part is a summary of previous critique, rebuttal, the previous review and the published results from CNBSS that may indicate violation of randomization. We have included results from the CNBSS 25-year follow-up data by age-group provided by [redacted]. In addition,

we have included a brief comparison between the CNBSS with the two-county study, and with other randomized trials of mammography screening.

Finally, we present our conclusion

Organization of the panel´s work

The panel was asked to interview specified individuals *who have come forward with information that may have an impact on randomization, recruitment of symptomatic women, and poor radiographic image quality in the Canadian National Breast Screening Study (CNBSS).*

We, the panel, were asked to deliver a report that detailed our assessment of whether this new information would have a credible scientific impact on the reliability of the CNBSS' published.

We assessed all ***new*** information in two different ways: Assessing information by witnesses in documents brought to the panel (**complaint material**) and in interviews of individuals that may have information on the CNBSS.

In order to assess what was new information, we also reviewed **previous criticism** and a **previous review** including the discussion that followed the review in the Canadian Journal of the Medical Association.

As it is more than 40 years since the CNBSS was initiated, any new information may not be reliable because it is dependent on human memory that is malleable. Therefore, we also assessed the published results from the CNBSS.

Our work is illustrated in figure 3.1

Figure 3.1 The panel assessed and reviewed new information: Complaint material and Interviews; and previous information: Previous criticism and previous review, and findings from CNBSS.

PART II

Mandate and Method

The review panel

Isabelle Boutron

Isabelle Boutron is Professor of Epidemiology at the Centre of Epidemiology and Statistics, Université Paris Cité and director of Cochrane France. Her research activities particularly focus on bias in randomized controlled trials. She is co-convenor of the Cochrane Bias Methods group and member of the CONSORT steering committee.

Peter Jüni

Peter Jüni is Professor of Medicine and Clinical Trials in The Nuffield Department of Population Health at the University of Oxford, United Kingdom. He has run large scale clinical trials, meta-analyses and methodological work on bias in randomized trials.

In 2013, he served as a member of the Swiss Medical Board, an independent health technology assessment initiative that appraised the evidence on mammography screening. He co-authored the resulting report and two subsequent perspective articles in Annals of Internal Medicine and the New England Journal of Medicine (1, 2). In 2014/2015, he was in charge of the mammography screening program of the Canton of Bern, a Swiss canton with approximately 1 million inhabitants.

Mette Kalager

Mette Kalager is Professor of Medicine at the University of Oslo, Norway. She serves as the head of the medical curriculum for evidence based medicine, quality improvement and medical leadership in Oslo. She has been involved in large-scale population-based screening trials of colorectal cancer for more than 10 years. She has been working as a breast cancer surgeon and was the head of the mammography screening program in Norway 2004-2006.

She co-authored several papers on mammography screening, wrote an editorial regarding the Canadian National Breast Screening Study in the BMJ in 2014 (3) and co-authored a paper on breast screening guidelines with dr. Baines (4).

Methods

The panel interviewed 15 individuals between November 14 2022 and January 16 2023. All interviews were digital on Zoom (zoom.us) using camera and were transcribed directly by Zoom. At least two panel members were present in every interview. For most interviews, all panel members were present. Mette Kalager lead the interviews,. All panel members were involved in the questioning, asked questions, helped clarified potential misunderstandings, and made sure that interviewees had an opportunity to answer all questions in the interview guide.

The Zoom conference were set up by the University of Toronto. Each interview lasted between 40 minutes and 2 h and 23 minutes. The interview schedule is presented below.

All interviewees were asked to report their conflict of interest and all interviewees signed a consent form (template; supplement). All interviewees were asked the same questions, preferably in the same order, and all interviewees were given a chance to respond to every question, even if we knew some questions would not be relevant for the individual interviewee (interview guide, supplement).

All interviewees are de-identified in this report. The review panel has full knowledge who the interviewees are and whom we are quoting, i.e who said what.

The report the review panel has submitted to the University of Toronto (University) is de-identified for each interviewee but does not have redacted text. Each interviewee is given a specific pseudonym so it may be possible to follow the arguments and statement of interviewee throughout the interview. We believe this makes the report more transparent, facilitates reading and interpretation, and support conclusion.

We have marked text we believe should be redacted before the report is made publicly available.

Interview schedule (2022/2023):

November 14	*Committee members present: Peter, Mette*	
8:00 a.m. – 9:30 a.m.	*Allen*	49.58
9:30 a.m. – 11:00 a.m.	*Bell*	1.14.51
11:00 a.m. – 12:30 p.m.	*Fraser*	1.47.22
November 15	*Committee members present: Peter, Isabelle, Mette*	
8:30 a.m. – 10:00 a.m.	*Anderson*	1.58.45
10:30 a.m. – 12:00 p.m.	*Bennett*	1.32.56
12:00 p.m. – 1:00 p.m.	*Evans* (Cancelled)	
November 16	*Committee members present: Isabelle, Mette*	
8:30 a.m. – 10:00 a.m.	*Armstrong*	39.12
10:00 a.m.– 11:30 a.m.	*Gill*	1.12.19
12:00 p.m. – 1:00 p.m.	*Campbell*	39.21

November 17	*Committee members present: Peter, Mette*	
8:00 a.m. - 9:30 a.m.	*Adams*	2.23.02
11:00 a.m. – 12:30 p.m.	*Hill*	28.20
December 5	*Committee members present: Peter, Isabelle, Mette*	
8:00 a.m. – 10:00 a.m	*Barker*	1.57.26
10:00 a.m. – 11:30 a.m.	*Bell*	1.01.03
11:30 a.m. – 1:00 p.m	*Clark*	47.40
December 6	*Committee members present: Peter, Isabelle, Mette*	
10:30 a.m. – 12:00 p.m.	*Johnson*	1.22.08
January 16	*Committee members present: Peter, Isabelle, Mette*	
3:30 – 5:00 p.m.	*Davis*	49.40
5:00 – 6:30 p.m.	*Evans* (Cancelled)	

Interview guide

Before conducting the interviews, we developed an interview guide which we followed in all interviews. The interview guide included questions in the following three topics related to the Canadian National Breast Cancer Screening Study (CNBSS):

1. Randomization impairment
2. Poor mammography quality and equipment
3. Improper analyses and interpretation of study result (not part of our review)

We focused our interviews on the first two topics and requested any new concerns that interviewees may have.

We asked open questions, without a-priori assumptions on what had happened, and did not interpret the interviewee's or ask leading questions. We were interested in gathering information and get the respondents to answer truthfully. For full transparency, this report includes the transcribed interviews including all conversation we had with the interviewees and amongst the panel members during each interview.

Interviewees

We were asked to interview these individuals:

1. Interview; Study group:
    - Prof. Martin Yaffe
    - Prof Em Anthony Miller
    - Prof Em Cornelia Baines
2. Interview; Critics
    - [redacted] – radiologist
    - [redacted] – medical imaging technologist
    - [redacted] -radiologist

- [redacted] - radiologist
- [redacted] – radiologist

3. Interview hands-on staff

We interviewed all individuals in the study group (Drs. Yaffe, Miller, Baines). We interviewed three ([redacted]) of five critics that we were asked to interview while two of them ([redacted]) did not respond to multiple requests for an interview, by email and by phone.

In addition, the panel identified individuals who we believed had first-hand knowledge about the randomization process, either individuals that were members of guideline panels on breast cancer screening or who were involved in meta-analysis of randomized breast cancer screening trials (a total of 4 individuals).

[redacted] (5).

Prof. Martin Yaffe kindly provided us with a list of individuals ([redacted] he proposed we interview (supplement). For various reasons we were not able to interview three of the individuals suggested by Dr. Yaffe ([redacted] did not respond to our approach, and [redacted] (see below)). In total, we interviewed five individuals he proposed. Of these three were thought to have been working in the CNBSS; a radiologist, another physician (not a radiologist), and an epidemiologist.

[redacted]
We were not able to interview [redacted] (research coordinator). [redacted] had provided information in the Complaint material, but she cancelled several interviews we had scheduled with her, and ultimately she informed the panel that she is unable to meet with us at any time in the future (e-mail Jan 9 2023).

In his interview, Dr. Anthony Miller talked about [redacted] We got the impression that she oversaw the randomization at the different centers in the Canadian trials. We attempted to contact her, but did not find any information about her.

We also requested to get access to the allocation books of the trial. We were not able to find these.

Panel approached by [redacted]
Dr. [redacted] who is [redacted] reached out to [redacted] at the University of Toronto and requested to meet with the review panel.

She argued that it may be useful to interview her because "*there is a great deal of material within this document and it may help for me to highlight with the panel to identify the issues that were confirmed by multiple sources:*

1. *The quality of the mammography was extremely poor, and the training of the staff including technologists and radiologists was almost non-existent in certain centres.*
2. *The control plan was often not followed, and several centres found that although a mammographic finding was present, the surgeons would not believe it and would not operate on the abnormality until the lesion became palpable. The data of how often this occurred is published in the review paper which compared the rates of the recommended breast biopsies and breast surgeries to the proportions of when this was actually done. This is summarized in section 5, pages 4 and 5 of the attached study (wbab099 (6)).*
3. *Many women with symptoms of breast cancer were recruited and included in the study.*
4. *Several staff indicated that it would have been easy to allocate women into the mammography arm given that the clinical breast examination was done before the allocation to the screening arm or the usual care arm.*"

[redacted] also shared published documents (6, 7). All material she shared with [redacted] and we received from [redacted] is included in the supplement of this report.

We do not know how [redacted] knew about the panel's work, or who suggested that she should reach out. We decided not to interview her as she was not part of the Canadian National Breast Screening Study (CNBSS) and thus could not provide eyewitness information about the trial, she was not included in the mandate or proposed by anyone that we were asked to interview as part of our mandate, and because she was one of the co-signers of the complaint and we already had a sufficient number of complaint co-signers we interviewed.

[redacted]

[redacted]

References

1. Biller-Andorno H, Jüni P. Abolishing mammography screening programs? A view from the Swiss medical Board. N Engl j med 2014;370:1965-7. doi: 10.1056/NEJMp1401875.
2. Jüni P, Zwahlen M. Is it time to initiate another breast cancer screening trial. Ann Intern Med 2014;160:864-6. doi: 10.7326/M14-0569
3. Kalager M, Adami H-O, Bretthauer M. Too much mammography. BMJ 2014.doi:10.1136/bmj.g1403
4. Jørgensen KJ, Kalager M, Barratt A, Baines C, Zahl PH, Brodersen J, Harris RP. Overview of guidelines on breast screening: Why recommendations differ and what to do about it. Breast. 2016. doi: 10.1016/j.breast.2016.08.002.
5. Miller AB, Wall C, Baines C, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening study: randomized screening trial. BMJ 2014;248:g336
6. Seely JM, Eby PR, Yaffe MJ. The fundamental flaws of the CNBSS trials: A scientific review. Journal of Breast Imaging, 2022, 1-2, https://doi.org/10.1093/jbi/wbac099
7. Seely JM, Eby PR, Gordon PB, Appavoo S, Yaffe MJ. Errors in Conduct of the CNBSS Trials of Breast Cancer Screening Observed by Research Personnel, Journal of Breast Imaging, Volume 4, Issue 2, 2022, Pages 135–143, https://doi.org/10.1093/jbi/wbac009

PART III
Complaint material

Complaint material

In this chapter of our report, we provide both the factual information about the complaint, and provide the reader our comments and interpretations for the different topics related to the complaint.

Summary Report

The review panel was provided with the complaint material, including "*Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants*". The report summarizes finding of interviews with "*New evidence has come to light that brings absolute certainty that the Canadian National Breast Screening Studies (CNBSS) performed over 40 years ago were compromised and should not be used to inform any decisions on breast cancer screening policy.*" (1). We will hereafter call this report the Summary report.

It is not clear who authored and who funded the Summary report, but we found the following information on page 2 in the report (1):

> "*This review was compiled by* [redacted]*, a former federal government employee who worked on government reporting, provided policy analysis to senior executives, carried out internal issue reviews and conducted internal non-financial audits. The interviews were conducted by Dr. Jean Seely, Dr. Paula Gordon, and Dr. Sheila Appavoo between January 30, 2021 and April 22, 2021. All interviewed study staff have agreed to share the information and their names*".

Drs Seely, Gordon, and Appavoo who conducted the interviews are breast radiologist. This may influence the content and framing of the questions and replies, and the interpretation of the information (2).

Interviewees were identified as follows (3):

> "*All staff listed on CNBSS publications* [4, 5] *were invited to participate in telephone interviews. Attempts to locate and contact former staff took place via social media and included requests for contact information from biostatisticians, epidemiologists, physicists, radiologists, surgeons, and technologists who had been employed by CNBSS. Staff who responded to the invitation were verified to have worked on the trials and cross-validated with CNBSS publications and contacts with other former staff*".

It is not known how possible conflict of interest were assessed or disclosed, or how it was confirmed that all interviewees were indeed part of the CNBSS. We also do not know why the interviews were done at the time they were. The interview guide did not include any direct questions about violation of random allocation, whether staff had seen others or had violated allocation themselves, but included questions (Q3-Q4) such as these (3):

> Q-3: "*Who was responsible for placing women in the treatment group with mammograms versus the control group?*"
> Q-4: "*Did this person have knowledge of the results of the physical breast exam before placing the patient volunteer into one of the trial groups?*

As stated in the Summary report, the interviews were done from Jan 30 to April 22 2021. This means that most interviews presumably took place before the eye-witness claiming potential corruption of the study ( ) came forward, which was in , after a meeting at the , where gave a talk about concerns of nonrandom allocation in the CNBSS. (*9*):

> "*I know nothing about there being any questions about the study being corrupt or anything, until happens to make little comment in his presentation on , saying that the investigators had always wanted to talk to a person who was employed by the study, and they weren't allowed to. I made a little comment in the sidebar during his presentation, saying. Yes, the study was corrupt. I witnessed it, I mean, and that's where this is all coming from all these years later somebody is finally listening*."

What prompted the decision to do interviews and the specific timing (January 2021), is unknown, but , one of the informants in the Summary report, that turned out not to have worked on the CNBSS, said the following regarding how she was approached by Dr. Gordon, a professor of breast radiology, who participated in the interviews published in the Summary report (1):

> "*Several months ago, originally just out of nowhere. I got this email that suggested I could help the problems that they were looking at. Then I got another one, and she said, Oh, would you be interested? Would you be willing to do an interview? And I said, Well, you know I mean It's a long time ago. I didn't have a good time at that program, you know, and I mean I might be a little biased.*
>
> *Because when I agreed with what they* said [Paula Gordon and the summary report (1,3)]. *Well, they're trying there. It* [CNBSS] *might be used to deny mammograms to women forty to fifty, and I thought that's not a good thing, because I think, it save lives, you know*."

Published papers

Based on interview, Drs. Seely, Eby, Gordon, Appavoo and Yaffe published a report in Journal of Breast imaging in 2022, concluding they had new information that the "*CNBSS did not consistently and rigorously assess the true efficacy of screening mammography*" (3). In addition, three of the authors published a review article "*summarizing multiple weaknesses in the execution of the CNBSS trials*" in the same *Journal* (6).

The *Journal of Breast Imaging*, is the official journal of the Society of Breast Imaging, an association of breast imaging professionals, which aims to minimize breast cancer mortality and suffering for patients, their families, and society (7). Dr. Peter Eby, one of the co-authors of the first paper (3) is the Director-At-Large on the board of directors (7). Drs. Yaffe and Seely are on the editorial board of the journal (8).

Summary of claims

We have reviewed the Summary report and the publication in the Journal of Breast Imaging (1, 3) for evidence about violation of the randomization process. Below is a list of the claims (quotes). The text in square brackets are our comments.

- Women with palpable lumps or clinical findings were assigned deliberately to the mammogram arm, so the study was not randomized, and this skewed the results of the study (page 6 in (1)).
- [redacted] observed other issues with the CNBSS trials including: women with clinical signs and symptoms of breast disease being encouraged to enroll in the study; clinical breast exams being performed prior to enrolment in a study group; and suspected preferential channeling of women with late-stage breast cancer into the mammogram arm of the trials (page 8 in (1)).
- Dr. Yaffe [CNBSS study physicist; professor of medical biophysics. On the editorial board of Journal of Breast Imaging; see Conflict of Interests later] raised concerns that the nurses performed the clinical breast exam before the patient volunteer were placed in a trial group. This meant that they knew which patients had lumps and clinical findings. They could bypass the randomization process, and deliberately place these women into the mammography arm, compromising the trial results (page 9 in (1)). Further, he raised the point that that subversion may not have been with devious intent where it occurred, but because the staff did not understand the importance of randomization in the trials and acted in their desire to do what was in the patients' best interests (page 10 in (1)).
- [redacted] [medical imaging technologist working 2 days a week for one month at CNBSS] witnessed patient volunteers with palpable lumps or clinical findings being placed intentionally into the mammogram arm of the study after the nurse performed a physical breast exam (page 11 in (1)). She indicated that it was routine that if the nurse examiner found "*any sort of abnormal findings - they were put into the mammo arm of the study*." She further witnessed that women were removed from the mammogram list later in the day to accommodate women with these findings unless they also had abnormal findings.
- [redacted] [research coordinator for 4 years] stated with certainty that non blinded allocation was done for women with palpable lumps and women with masses were "*quietly*" placed into the mammogram group. Women with lumps were not turned away (page 12 (1)). She provided written information after her interview for the Summary report "*I don't have proof of this but will try to obtain the records which I think are still available. But I know that it was done. I was not witness to that happening but human behavior and the lack of knowledge around the rules of an RCT is likely.*"

[was scheduled for interview with the panel two times, but ended up declining the invitation. She informed the panel that she was unable to have a meeting with the panel at any time in the future].

- [medical imaging technologist 5 years] said if a woman did not want to go into the mammogram arm she was not put there but was assigned to the control group. She heard that if a patient volunteer had a lump, they would routinely be put into the mammogram arm (page 16 in (1)).
- [data collection, managing on of the two arms of the program. At the panel interview it became clear was not involved in the CNBSS]. She "*spoke up at least one meeting and with the liaison to the epidemiology department, indicating that women identified at the initial breast exam with lumps, nipple discharge, and pain should not proceed into the trial because the trial was about evaluating screening mammography, not diagnostic methods, and those women should instead be referred to their family doctor for diagnostic assessment for possible breast cancer*" (page 18 in (1)).
- [resident in radiology] remembers adding on patient volunteers who were seen by the study nurse because they had a lump. Stated she saw diversion of symptomatic women into the mammography arm as a resident when we were a site for the CNBSS (page 19 in (1)). [This is what is stated in the supplementary material:]
    - *"she remembers adding on patient volunteers who were seen by the study nurse because they had a lump.*
    - *she stated "I saw diversion of symptomatic women into the mammography arm as a*
    - *resident when we were a site for the CNBSS."*
    - *Could not answer the other questions".*

who is quoted in the Summary report (1) was the . We were not able to interview him as he did not respond to our approaches (by email and phone). He has expressed concerns about the study in 1993 (9), but we are not aware he was actually witnessing any manipulation of the randomization process.

Two of the staff ( ) were said to witness women with lumps being placed in the mammography arm of the study. Two others, ( ) had not witnessed this themselves but "*knew*" or "*had heard*" that women with lumps routinely were assigned to the screening arm.

The interviews were done in 2021, the NBCSS ended enrollment in 1986 (4, 5), 36 years before the interviews. One of the witnesses, only worked with the CNBSS for one month and it is unknown for how long worked at the NBCSS. In the Summary report she is recorded as a (1). In the paper, her title is (3).

The complaint material does not provide documentation of how the allocation could have happened, no details on what witnesses saw, how specifically any violation of allocation would have occurred (the randomization books were prespecified with allocation – arm), or how often this may have happened. The witnesses were not involved in the randomization process, but were a medical technologist and radiologist, respectively.

We have no further information that may confirm the claims of observation, many of them only made 30-35 years after they may have been observed.

Errors

Dr. Yaffe recommended we interview *Ms. Hill* ([redacted]) because "*she had worked for the CNBSS*". She was one of the informants being interviewed in the Summary report: New evidence from key informants" and in the paper in Journal of Breast imaging in 2022. This material was provided to the review committee prior to our interviews (1, 3).

About 20 minutes into her interview with us, it became clear that she had not been working in the CNBSS. The time period she claimed she worked in the trial did not match the time period of the trial. She worked at a center [redacted] that was part of the trial, but she worked there around 5-10 years after the trial ended.

We wanted to understand how she could have been an informant in the Summary report (1) and mistakenly been thought to be part of the CNBSS. This is what was said in the interview (text in brackets are our comments, made to clarify content):

Panel member: Who suggested that we would interview you, [redacted]?

Hill: Dr. Paula Gordon. She contacted me just by computer. [Paula Gordon, a breast radiologist and professor. She participated in the interviews published in the Summary report]

Panel member: So did you talk to her at all, or was it just like a written conversation that you had?

Hill: Yeah, it was just by email over the computer.

Panel member: How long ago was this?

Hill: Well, she [Paula Gordon], I think it was several months ago, originally just out of nowhere. I got this email that suggested I could help the problems that they were looking at. Then I got another one, and she said, Oh, would you be interested? Would you be willing to do an interview? And I said, Well, you know I mean It's a long time ago. I didn't have a good time at that program, you know, and I mean I might be a little biased, but our [redacted] thing was run by a radiologist rather than an epidemiologist, which you know kind of. Sometimes it doesn't work so well.

Because when I agreed with what they said [Paula B Gordon and the summary report]. Well, they're trying there. It [CNBSS] might be used to deny mammograms to women forty to fifty, and I thought that's not a good thing, because I think, it save lives, you know.

Panel member: I understand.

Hill: I did not even know anyone in that trial.

As evident from the transcription above, during the interview, we found out that that *Ms. Hill* was involved in the screening mammography program of [redacted] (10), not in the CNBSS.

It appears, *Ms. Hill* was contacted by dr. Paula Gordon (professor and breast cancer radiologist), one of people that interviewed former staff of the CNBSS and co-authors of "Errors in conduct of the CNBSS trials of breast cancer screening observed by research personnel" (3), and [redacted] wanted to do something good and hence accepted to be one of the informants.

Three days after we interviewed *Ms. Hill*, we received an email from Dr. Yaffe where he apologized for recommending *Ms. Hill*. Information about who and at what time we were interviewing was not publicly available. This information was only known to the review panel, one person at the University, and the interviewee.

We, the panel, was interested in knowing how Dr. Yaffe knew we had interviewed *Ms. Hill* and that we revealed she was not part of the CNBSS. Dr Yaffe explained that *Ms. Hill* has written to one of his colleagues and he was copied in on the correspondence. All information provided by her in the Journal of breast imaging was retracted and an error clarification was published in the Journal in February 2023 (11).

E-mail correspondence between the panel and Prof. Yaffe, Nov 21 and 22, 2022:

> "*I am writing on behalf of Prof. Kalager who is hoping you could share how you were aware that the Review Panel met with Ms. Hill, and how it was learned that she had not worked at the CNBSS*".

Dr. Yaffe responded:

> "*Pretty straightforward. Ms. Hill had been interviewed by my colleagues who had been informed that she had worked with CNBSS. Her transcribed comments have been documented in a publication. These comments led me to believe that she could provide useful information to the panel, so as you know, I suggested her name.*
>
> *After her interview (where I presume it was established that she had, in fact, not worked with CNBSS, she wrote to one of my colleagues and that correspondence was copied to me. I am concerned that only useful accurate information be provided to the panel and this prompted my last email to you. I have not met nor spoken with Ms. Hill at any time and other than obtaining her permission to be contacted by you (and the email copy to which I have just referred where I learned that I had been misinformed about her relationship to CNBSS), I have had no interactions with her*."

We do not know why *Ms. Hill* was wrongly included in the complaint material as informant as her role as staff of the CNBSS while she indeed was not and why this was not recognized by those who did the interviews for the Summary report. We are concerned that this may indicate

a lack of proper process and that other people included in the report who we did not interview may not have been part of the CNBSS.

Open letter

As part of the complaint material, the review panels received an open letter by [redacted] (supplement, complaint material). The letter is not dated, and we do not know if it is publicly available. The letter explains on how the "*new evidence*" was brought forward. [redacted] describes that he was contacted by a medical technician after he gave a talk in Toronto in [redacted].

[redacted] *confirms in his interview with us:*
[redacted] It wasn't until [redacted], when I gave a virtual talk in [redacted], actually to the [redacted]. My talk was: There's problems with the Canadian national breast screening study or facts about the Canadian breast screening study that Canadians need to know, and I went through all of the issues that we'll cover.

It's saying that you know we still didn't know if this non-random allocation took place because we weren't allowed, no one was allowed to interview the coordinators. And I had actually in my material that I sent to you. I gave you a copy of the letter that I had sent to, I think it was MacMahon saying that they really needed to interview these women [a copy of the letter was provided to the panel], and I also published. I think it was in the Canadian Medical Association Journal [18]. I think that I published that you need to interview the coordinators to find out what went on, and I'm told Tony [Miller] wouldn't allow that.

Panel member: Can I quickly ask you when you say `I'm told Tony wouldn't allow that´. Do you know who would know that for sure? What had happened? Who told you that?

[redacted] I don't know the answer to that. I know that one way or another. I learned that, Tony, wouldn't allow, which made complete sense to me. He was running the trial, and he knew that I and others have, you know, said it's very important to interview the coordinators. And quite frankly, if I was defending my trial, I would say absolutely, you got to interview the coordinators because they're going to tell you that they did it the right way. And the fact that they weren't, I mean. I was in writing, MacMahon, and Bailar knew it. The world knew it. If you read what I had published it wasn't a secret.

I don't know that Tony ever said I won't let it happen in public. But that was sort of, you know, because I was doggedly trying to get this done and it wasn't done. I have to assume, and I'm fairly sure, either Tony told me, or someone told me that he wouldn't allow it, but I can't. I can't give you absolute proof of that.

But it makes complete sense, and just if I forget to say later on. The quality of the mammography, if I had been in a trial, if I had been a radiologist in the trial, and I felt that I had done a great job, I would have been all over this guy [redacted] and other people who were complaining about the quality. What are you talking about? Our quality was terrific, and so on. I haven't heard from any of the radiologists, and I'm not aware of anyone who participated in the study as a radiologist defending the study. I mean, that's I think, astonishing. Now you can say Well, you know I didn't come out the way radiologist wanted, so they're not supporting it. I don't think so.

[redacted] and others, had made complains that the previous review of the CNBSS (12) did not include interviews with staff about the randomization process and the inclusion of symptomatic women in the screening arm. He did not convince Drs. Bailar and MacMahon who were the review panel at the time (13), to interview staff. He claims that Dr. Miller was responsible as he did not "*allow*" such interviews.

Contrary, Bailar and MacMahon did not confirm this and gave their reasons for not interviewing staff as follows: *"steering" of randomization would have been dependent on center and time and would not have happened in all centers, any potential witnesses that had not at the time for the review come forward would be unlikely to come forward to the review panel* (approximately 10-15 years after the study was enrolling women). Bailar and MacMahon tried to interview an eye-witness Dr. Kopans had suggested to the panel, but she did not respond although she was promised confidentiality (13).

[redacted]

Letters

We were also provided with letters from 83 "*breast cancer experts*" (supplement). All but one are radiologists, most of whom worked with mammography, 80% worked in the US and 16% in Canada.

The start of the letters reads:

> "*Enclosed please find 83 letters from breast cancer experts in Canada, the USA and Europe, who for decades, have decried the flaws in the Canadian National Breast Screening Studies. In spite of experts' wellfounded concerns about the design and execution of the trials, they were used by the Canadian Task Force on Preventive Health Care to craft their breast cancer screening guidelines, and they've had influence on screening policies globally. The trials are not only an embarrassment; they have undoubtedly contributed to tens of thousands of unnecessary deaths and suffering worldwide*".

Some of the letters were addressed to the Minister of Health in Canada and dated [redacted], a month after [redacted] gave his talk in [redacted] whereafter an eye-witness came forward. [redacted] described this in a paper (14):

> "*In [redacted], I presented a talk virtually to the [redacted], in which I outlined the concerns raised by the published data about the CNBSS and, in particular, the indications of nonrandom allocation. Soon after, I received an email from an attendee who had been an X-ray technologist in the CNBSS. She attested to the fact that she had witnessed nonrandom allocation of women with clinical evidence of breast cancer who were assigned out of random order to the mammography arms*" (14).

He also confirmed this in the interview with the panel.

Content

The wording and the references were similar in each letter and the content themes were:

1. Role of witness

The witness has information that may prove the NBCSS are "*flawed*", "*compromised*", and "*not a true RCT*" [Randomized Controlled Trial], have "*systematic errors*, and that "*women with known symptoms of cancer were deliberately placed in the screening cohort.*"

> "*I have recently become aware that persons involved with randomization during the CNBSS…has stepped forward to acknowledge….namely that patients with suspicious palpable masses were preferentially "randomized" to the screening mammography arm rather than the control arm*". The "*Mistakes" in the randomization, the only way that explains that the investigation arm included a higher number of advanced cancers*".

2. CNBSS an outlier

CNBSS is by several said to be the

> "*only trial that did not show mortality reduction from mammography screening, and also has marked overestimates of overdiagnosis*" and
>
> "*outlier among numerous RCTs*" and
>
> "*failing to demonstrate mortality reduction from breast cancer screening, contrary to all other large randomized controlled trials*".

3. Mammography screening is beneficial

> "*As a breast radiologist, I have firsthand knowledge of the benefits of screening women for breast cancer with mammography"*
>
> "*We know that the best way to save lives is to regularly screen all women*"
>
> "*The well-being of women around the world is at stake*".

4. Do not use data from CNBSS in guidelines and policy making

Most of the letters urged that the CNBSS should not be used in making guidelines and

> "*should not be used to determine breast cancer screening policy*".
>
> "*This study should be officially denounced and no longer included as sound evidence for the development of national guidelines*".

"*This study must be removed from the body of literature regarding breast screening randomized controlled trials and any recommendations based on this trial must be reviewed and revised*."

Papers that several of the letters referred to the following articles:

Article	
1	Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders DE, Sickles EA, To T, Wall C. Canadian National Breast Screening Study: assessment of technical quality by external review. AJR Am J Roentgenol. 1990 Oct;155(4):743-7; discussion 748-9. (15)
2	Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. AJR Am J Roentgenol. 1993;161:755-60. (9)
3	Burhenne LJ, Burhenne HJ. The Canadian National Breast Screening Study: a Canadian Critique. AJR Am J Roentgenol. 1993;161:761-63. (16)
4	Kopans DB. Breast Imaging – 3rd edition. Lippincott Williams & Wilkins: Philadelphia, 2007 (17)
5	Tarone RE. The excess of patients with advanced breast cancers in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995 75:997-1003. (18)

None of the letters contained new information, or information that may shed light on what happened in the CNBSS.

The individuals signing the letters(mostly breast imaging radiologists), supported the conclusion of the Summary report that the CNBSS "*should not be used to inform any decisions on breast cancer screening policy*" because *"[n]ew evidence has come to light that brings absolute certainty that the Canadian National Breast Screening Studies (CNBSS) performed over 40 years ago were compromised"* (1). This is similar to a statement Dr. Kopans made in 2017 (19).

Other aspects

When going through the complaint material and during our interviews, we noticed that there are some expert radiologists (academic and non-academic) and mammography technicians that have made some the claims at hand and authored critique of the CNBSS (*Chapter 9, 11*).

As experts in breast radiology, they have expressed concerns since the results of the CNBSS first was made available, about the quality of mammography, inclusion of women with breast symptoms or lumps and allocation of these women to the mammography arm of the trial, contamination of mammography screening outside the trial, and subversion of randomization by allocating women who wanted mammography to the mammography arm (possibly due to lack of mammography outside the trial).

In our interview with [redacted], we were made aware of correspondence between one of the key witnesses [redacted] and some of the radiologists.

In an email correspondence provided to the panel by [redacted] (Nov 16 2022), [redacted] thanked "[redacted]". Although no last names were mentioned, we

believe the email is sent to Drs. Paula Gordon, Daniel Kopans, [redacted], Jean Seely and [redacted], some (Drs. Gordon, Seely, and [redacted]) were part of the Summary report and authors of [redacted]:

Quote from the email:

> "*Hello Paula, Daniel, [redacted], Jennie &* [redacted]
> *I hope this email is finding you all well and in the same wonderful spirits that I find myself in tonight. Today, after 35.5yrs I find myself ready to cry with the joy of being heard after years of voicing my upset with the CNBSS & world wide.*
> *I have all of you & those with whom each of you has shared my original email sent to* [redacted] *on* [redacted] *after h Conference.*
>
> *A huge thank you as well to the unknown individual who took the time to forward my side note during* [redacted] *' lecture s had witnessed just how flawed the CNBSS was when I worked there in 1985.*
>
> *Thank you to my HEROS,*
> *Sincerely,*
> [redacted]

References

1. Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants.
2. Catlin M, Wilson DB, Redlich AD. et al. Protocol: Interview and interrogation methods and their effects on true and false confessions: An update and extension. Campbell systematic reviews 2023 https://doi.org/10.1002/cl2.1314
3. Seely JM, Eby PR, Gordon PB, Appavoo S, Yaffe MJ. Errors in Conduct of the CNBSS Trials of Breast Cancer Screening Observed by Research Personnel, Journal of Breast Imaging, Volume 4, Issue 2, 2022, Pages 135–143, https://doi.org/10.1093/jbi/wbac009
4. 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: 1459-76.
5. 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: 1477-88
6. Seely JM, Eby PR, Yaffe MJ. The fundamental flaws of the CNBSS trials: A scientific review. Journal of Breast Imaging, 2022, 1-2, https://doi.org/10.1093/jbi/wbac099
7. The Society of Breast Imaging, https://www.sbi-online.org Accessed Jan 25, 2024.
8. Editorial board of Journal of breast Imaging https://academic.oup.com/jbi/pages/Editorial_Board, Accessed Jan 6 2024.
9. Kopans DB, Feig SA. The Canadian National Breast Screening Study: A Critical Review. AJR 1993;161 :755-760 0361-803X193/1614-0755

10. Olivetti IA, Kan L, Mated D, King S. Screening mammography program of British Colombia: pattern of use and health care system costs. CMAJ 1999; 160:337-41.
11. Correction to: Errors in conduct of the CNBSS trials of breast cancer screening observed by research personnel. Journal of Breast Imaging, 2023, e1 doi:10.1093/jbi/wbac094 Correction Published Online: February 15, 2023, https://academic.oup.com/jbi/article/5/2/e1/7040330
12. Kopans DB. NBSS: opportunity to compromise the process. [letter to the editor] Can Med Assoc J 1997; 157:247
13. Bailar III, JC. MacMahon B. Randomization in the Canadian national breast Screening Study: a review for evidence of subversion. CMAJ 1997; 156:193-199
14. Kopans DB. Misinformation and facts about the breast cancer screening. Curr Oncol 2022;29:5644-5654. https://doi.org/10.3390/curroncol29080445
15. Miller AB, Howe CR, Wall C. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 1981; 4:233-246
16. Burhenne LJ, Burhenne HJ. The Canadian National Breast Screening Study: a Canadian Critique. AJR Am J Roentgenol. 1993;161:761-63.
17. Kopans DB. Breast Imaging – 3rd edition. Lippincott Williams & Wilkins: Philadelphia, 2007
18. Tarone RE. The excess of patients with advanced breast cancers in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995 75:997-1003.
19. Kopans DB. The Canadian national Breast Screening Studies are compromised and their results are unreliable. They should not factor into decisions about breast cancer screening. Breast Cancer res Treat 2017; 165: 9-15. Doi: 10.1007/s10549-017-4302-9

Disclosure of conflicts of interest

The complaint material (letters and Summary report) concluded that the CNBSS "*should not be used to inform any decisions on breast cancer screening policy*" (1). Guidelines can have a powerful effect on the behavior of clinicians, and clinicians may consider guidelines as "rules" for clinical practice and be reluctant to deviating from the "standard" (2).

> *Poor quality guidelines may also result when the available evidence is inadequate or conflicting, but guideline authors—who often document all the reasons the study results are unreliable—nonetheless combine the conflicted results to promote a single approach describing what clinicians should do, rather than acknowledging definitive recommendations would be inappropriate, and credible alternative approaches can be justified* (2).

We here discuss the topic of conflicts of interest and disclosures of such conflicts using the example of clinical guidelines because firstly this is what the Summary report refers to and secondly because adequate handling of conflicts of interest in guidelines provides a good example of handling of conflicts in science and academic in general.

Widespread financial conflicts of interest among the authors and sponsors of clinical practice guidelines have turned many guidelines into marketing tools of industry (2). Conflicts of interest may introduce bias in research and guideline making, because it may influence decisions and recommendations. Biased guidelines may cause harms to patients.

Content experts or topic specialists are especially likely to have a financial or professional conflict of interest or both, increasing the risk of bias (2, 3). For example, physicians´ specialty in one study was correlated with recommendations in breast cancer screening guidelines (3):

In 12 guidelines of mammography screening for women aged 40-49 years, 70% recommended routine (preform screening on regular basis) and 30% recommended non-routine (decision left to the discretion of the patient and physician) screening. Only half of the 12 guidelines reported conflict of interest disclosures for authors. Almost 90% of panel members were physicians and 22% had a financial conflict of interest (3).

The guideline panel recommending non-routine mammography, did not have radiologist in the guideline panel, whereas 5 of the 8 guidelines recommending routine guideline had radiologist in the panel. A guideline panel with a radiologist was 6 times more likely to recommend routine screening, but the association was not significant (odds ratio 6.05) (3).

Guideline panels recommended routine screening had 38% primary care physicians, and panels recommending non-routine screening had 90% primary physicians (p=0.02). With a primary care physician lead author, the odds of recommending non-routine screening was almost 4 times the odds when a primary care physician was not the lead author (3).

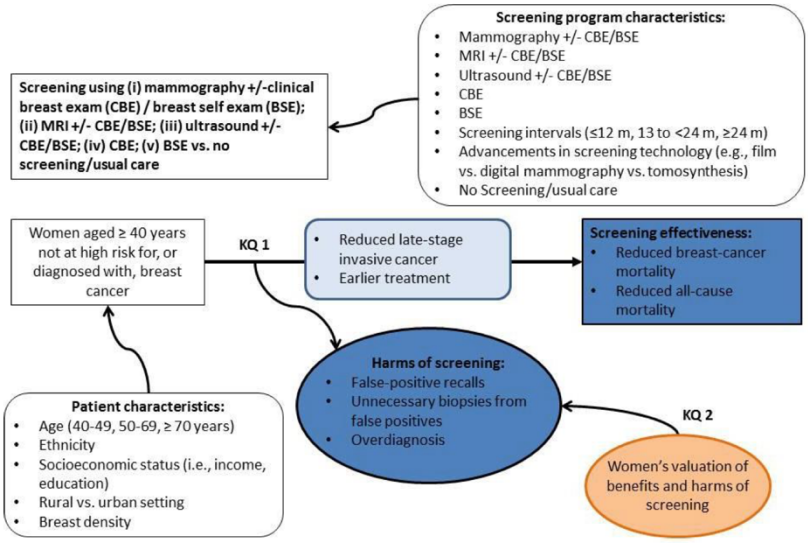
Publications on breast disease or cancer, by guideline lead authors, were associated with recommendations for routine screening (3). The authors discussed

> "*Specialists who derive their income from screening or treating breast cancer (e.g., radiologists and surgical or medical oncologists) might be more likely to recommend routine screening than physicians who do not derive most of their income from these activities (e.g., generalists).*
>
> *Further, specialist physicians might have values and preferences related to screening that are acquired from daily exposure to the burden and morbidity of breast cancer and that may differ from the values and preferences of generalist physicians*" (3).

Based on the same evidence, different guidelines arrive at different recommendations, dependent on the members of the panels and their professional background. Unfortunately, only half of the guidelines in this study disclosed the conflict of interest for their panel members (2), so we do not know whether conflict of interest, financial, professional and intellectual, played a role in or even biased the recommendations in the guidelines.

The bias may be introduced in all steps of making the guideline (figure 7.1) (4): *the key questions* (ex. For women not at high risk for breast cancer aged 40-49 years: does clinical breast exam reduce risk of advanced not treatable breast cancer, or does clinical breast exam increase risk of overdiagnosis, women's values of benefit and harms of clinical breast exam), *determination of study inclusion* (should we exclude the CNBSS-1 (age 40-49 years)?) *and exclusion criteria* (include only randomized trail s or include cohort studies) *as well as the selection of the body of evidence the methodology of the evidence review (e.g., systematic or not), criteria for assessment of the risk of bias, selective reporting of outcomes, methods for assessing the balance of benefits and harms, and transparency of the process for translating the body of evidence into recommendations* (3, 4).

Figure 7.1 analytic framework for breast cancer screening used in the breast cancer screening recommendations from the Canadian Task Force on Preventive Health Care (Fig 1 in (4))



Issues that may raise red flags and skepticism about a guideline are displayed in box 7.1 (2).

Box 7.1 Issues that may raise red flags of a guideline being biased (2)

Issue	Red flag
Sponsor of guideline	Professional society receiving substantial industry funding Proprietary company Undeclared or hidden
Committee chair	Financial COI*
Multiple panel members	Financial COI*
Suggestion of committee	Recruiting members that might preordain (decide prior to reviewing the evidence) a recommendation regarding a controversial topic such as mammography screening for women aged 40-49 years
Method expert	No method expert Regarding evaluation of evidence
External review	No external review
Patient or non-physician expert	No inclusion of non-physician expert or patient representative, community stakeholders

*Including either or a financial relationship with a proprietary healthcare company and/or who's clinical practice/specialty depends on test or interventions covered by the guideline

However, content experts are also necessary in developing guidelines, and the concerns are not simply about identifying and disclosing direct financial or indirect conflicts of interest (COIs), but also how to manage COIs in a fair, judicious, transparent manner (5).

The Canadian Task Force on Preventive Health Care has set up guidance on how to handle conflict of interest that adhere Guidelines International Network (GIN) principles for disclosure of interest and management of conflict of interest (6, 7) (*Chapter 35*).

The Guidelines International Network was founded in 2000 and are formally incorporated as a company and Scottish Charity (6). Their vision is to make trustworthy and accessible guidance for better health and is a network of organizations and individual s interested in evidence-based

guidelines. Among others GRADE, The Canadian Task Force on Preventive Health Care, McMaster University, Public Health Agency of Canada, Registered Nurses Association of Ontario, and WHO are all GIN member organizations (6).

The guiding principles are as follows (shortened (5)):

- Guideline developers should not include members with direct financial or relevant indirect COIs. Transparency about reasons for including panel members with COIs and how the COIs are managed. Conflicted members should be a minority on the panel.
- The definition of COI and its management applies to all members and should be determined before a panel is constituted.
- Use standardized forms for disclosure of interests, COI should be updated regularly, disclose interests publicly, easily accessible for users of guideline. Registries of disclosures could be used
- Chairs of guidelines should have no direct financial or relevant indirect COIs.
- Experts with relevant COIs may be permitted to participate in discussion of individual topics, but an appropriate balance of opinions should be sought.
- No member with a direct financial COI may decide about the direction or strength of a recommendation, not be present in the discussion.
- An oversight committee should be responsible for developing and implementing rules related to COIs.

Examples presented in box 7.2 (5).

Box 7.2 Examples of potential conflicts of interest (5)

Conflict	
Direct financial	*Direct payment for service*
	Consultant
	Payment for lectures, meetings
	Honoraria and gifts
	Stocks
Indirect	*Gain income from guideline*
	Academic advancement
	Standing
	Scientific interest
	Political, religious, ideological, other

In order to understand what may be perceived as a conflict of interest, one may use the red flags outlined above (box 7.1). A declaration of interest may not be a conflict of interest (5, 7). The Canadian Task Force on Preventive Health Care has set up guidance on financial and non-financial conflict of interest (7). The process starts with the individual declaring his or her interests related to the topic from the past 3 years in a standardized form. The assessment is judged by an oversight committee (*Chapter 35*).

The individuals we interviewed also had to state, at the beginning of the interview their conflicts of interest (*Chapter 5*). In addition to the COI the interviewees reported, we report

here on additional COI we believe should have been reported. There might be other undisclosed conflicts of interest (direct or indirect (Box 7.2):

[redacted]

For transparency we, the panel, have disclosed the following, that might be considered COI:
Six people that were part of the CNBSS:

Two medical imaging technologists: [redacted], and Ms. *Armstrong*
Three physicians: Drs. Cornelia Baines and Anthony Miller, *Clark*
One medical physicist: Dr. Martin Yaffe

In addition we have interviewed the following people:

Three breast imaging radiologists: Drs *Fraser*, [redacted], Daniel Kopans,
Three physicians and one PhD that were all scientist and have been involved in meta-analyses or guidelines for breast cancer screening: Drs. *Bell*, *Allen*, *Campbell*, *Gill*.

References

1. Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants.
2. Lenzer J, Hoffman J, Furberg C, Ioannidis J. Ensuring the integrity of clinical practice guidelines: a tool for protecting patients. BMJ 2013;347:f5535 doi: 10.1136/bmj.f5535
3. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, Schünemann H, Deyo R. Author´s specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiology 2012; 65:725-33
4. Canadian Task Force on Preventive Health Care. Breast cancer screening: protocol for an evidence report to inform an update of the Canadian Task Force on Preventive Health Care 2011 Guidelines, 2017. https://canadiantaskforce.ca/wp-content/uploads/2018/07/BCU-protocol-14Mar2017-FINAL_v3.pdf
5. Schünemann HJ, Al-Ansary LA, Forland F, Kersten S, Komulainen J, Kopp IB, et. al Principles for Disclosure of Interests and Management of Conflicts in Guidelines. Ann Intern Med. 2015;163:548-553. doi:10.7326/M14-1885
6. Guidelines International Network. Accessed Feb 22 at https://g-i-n.net/about-gin
7. Canadian Task Force on preventive Health care. Policy on disclosures of interests and management of conflict of interest, 2020 https://canadiantaskforce.ca/wp-content/uploads/2020/10/COI-Policy-202008Final-1.pdf
8.Mammography Education, Inc. accessed Feb 23 2024 at https://mammographyed.com/2019/05/03/dr-tabar-follow-up-to-cnbss-study/

[redacted]

PART IV
Previous criticism and review

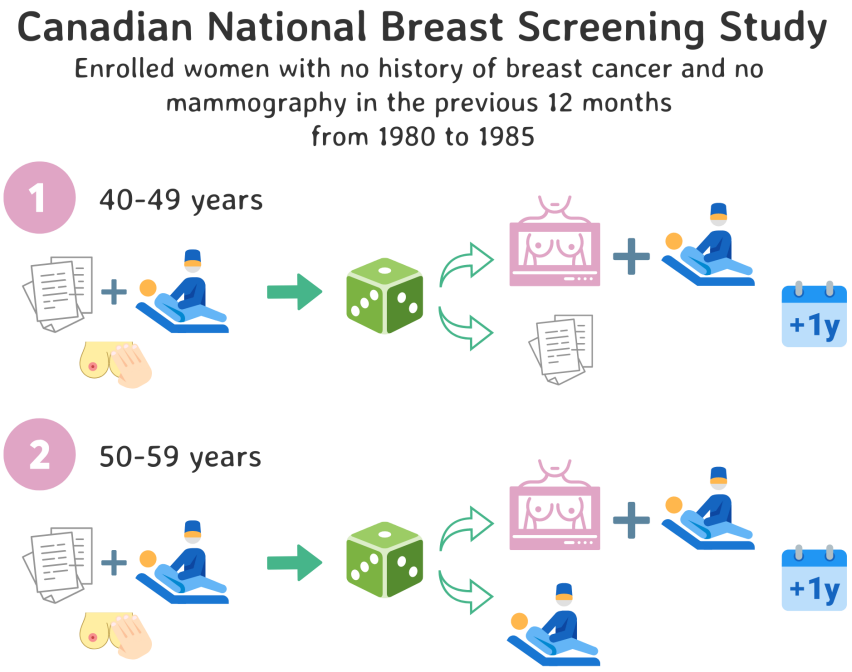
Canadian National Breast Screening Study

The Canadian National Breast Screening Study (CNBSS) recruited women with no history of breast cancer and no mammography in the previous 12 months from 1980 to 1985, figure 8.1. The CNBSS consisted of two trials, CNBSS one enrolled women age 40-49 years and randomized volunteers to undergo annual mammography screening and physical examination of the breasts, *screening arm* or to annual self-administered questionnaire mailed to each woman, *control arm*. The first 62% of the women who entered the study were eligible for a 4-year program; the remainder were offered a 3-year program (1).

CNBSS two enrolled women age 50-59 years and randomized volunteers to receive either annual mammography and physical examination of the breasts, *screening arm*, or to annual physical examination only, *control arm* (2).

Prior to randomization all women were taught self-breast exam, all women underwent physical examinations of the breasts, filled in a questionnaire and signed a consent form (1, 2).

Figure 8.1 Interventions prior to randomization and in the *screening arm* and *control arm* of CNBSS one and two (1, 2).



Dr. Baines clarified methods for recruitment, clinical breast exam, randomization, how women with symptoms and lumps were handled, clinical breast exam surgical- and pathological review, and follow-up procedures in the CNBSS (3). We have summarized her description in table 8.1 (same table as table 30.3, *Chapter 30*).

Table 8.1 Methods for recruitment, clinical breast exam, randomization including how women with lumps were handled, quality of mammograms, clinical examination, surgical- and pathological review, and follow-up procedures in the CNBSS (3)

Method	**Description from Baines**
Recruitment	Women aged 40-59, not pregnant, no prior breast cancer, no mammogram last 12 months, consented
Clinical breast examination	Was done because CBE was believed to reduce mortality Examiner asked about lumps or symptoms Physical examination and instruction in self-examination Clinical findings required referral to surgical review (within 1 week). Documented on examiners form Prior to randomization
Randomization	Women individually randomized Randomization lists in 4 separate books (age 40-44, 45-49, 50-54, 55-59) Each center Center coordinator entered the date and name on the first available line in appropriate book dependent on age, and assigned the woman her ID and randomization allocation ID and allocation was entered to all chart forms "*Skipping a line to achieve a desired allocation was not feasible because she could not predict when the next appropriately aged woman would arrive to fill the skipped slot."* All original randomization sheets were submitted to the central coordinating office where all sheets were examined for suspicious entries, inappropriate dates and lack of congruence with participant records Examiner told each woman her allocation
Women with lumps	All, irrespective of allocation arm, was referred to review clinics. No reason for examiner to allocate
Quality of mammograms	Documented in: - Baines CJ, McFarlane DV, Wall C. Audit procedures in the National Breast Screening Study: mammography interpretation. Can Assoc Radiol J. 1986;37:256-60. - Baines CJ, McFarlane DV, Miller AB. Sensitivity and specificity of first screen mammography in 15 NBSS centres. Can Assoc Radiol J. 1988;39:273-6. - Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the National Breast Screening Study. Invest Radiol. 1990;25:971-6.
Clinical examination	5 -10 min Documented in: - Baines CJ, Miller AB, Bassett AA. Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. Cancer. 1989;63:1816-22. - Baines CJ, To T. Changes in breast self-examination behavior achieved by 82,835 participants in the Canadian National Breast Screening Study. Cancer. 1990;66:570-6. - Miller AB, Baines CJ, Turnbull C. The role of the nurse-examiner in the National Breast Screening Study. Can J Public Health. 1991 ;82: 162-7.
Surgical review	Study surgeons appointed to each center If diagnostic follow-up was required, surgeons forwarded their recommendation to woman's physician Most recommendations were followed Women aged 40-49: 0.8% in screening arm and 1.5% in control arm had diagnostic mammogram
Pathology review	Reference pathologist at each center Reviewed all specimens from surgical procedures on participants

Follow-up procedures	3 or 4 years dependent on time of entering study; those enrolled in 1983-84, had only 3 years of routine follow-up High compliance with procedures Questionnaires Complete ascertainment of breast cancer and deaths After study interventions ended, passive follow-up, active follow-up of all women with breast cancer

Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334

We do not describe the CNBSS in more detail. More information can be found in the four main publications of the trial results, referred to in table 8.2.

Table 8.2 showing the publications presenting the main results from the CNBSS

1	Miller AB, Howe GR, Wall C. The National Study of Breast Cancer Screening Protocol for a Canadian Randomized Controlled trial of screening for breast cancer in women. Clin Invest Med. 1981;4(3-4):227-58. PMID: 6802546. (4)
2	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. CMAJ. 1992 Nov 15;147(10):1459-76. (1) Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543.
3	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. CMAJ. 1992 Nov 15;147(10):1477-88. (2) Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423088; PMCID: PMC1336544.
4	Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002:137; 305-312 (5)
5	Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 12 year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000;92:1490–9 (6)
6	Miller A B, Wall C, Baines C J, Sun P, To T, Narod S A 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 doi:10.1136/bmj.g366 (7)

References

1. 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: 1459-76.
2. 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: 1477-88
3. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334

4. Miller AB, Howe CR, Wall C. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 1981; 4:233-246
5. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002:137; 305-312
6. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 12 year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000;92:1490–9
7. Miller AB, Wall C, Baines C, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening study: randomized screening trial. BMJ 2014;248:g336.

Criticism of CNBSS - 1993

Before we summarize the findings of a previous review of the CNBSS (1), we summarize the main critique that was the basis for the review (2, 3, 4, 5), and the response to the criticism (6) by Dr. Cornelia Baines, one of the PIs of the CNBSS (6).

Criticism of the CNBSS, published in 1993 in Radiology and the American Journal of Radiology

The American Journal of Radiology

Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review (2)

In their critical review of CNBSS, Drs. Kopans and Feig focused on effect of mammography screening for women aged 40 to 49 years old (2). They claimed that

> "*major mistakes made in the design and implementation of the trial cast serious doubts on the applicability of its results*" (2).

They further cast doubt about (2):

- incomplete follow up of the control arm

> "*The investigators apparently linked the CNBSS data base with national registries; however, it is suggested in their publication that the national linkage was maintained only through the end of 1988. Thus, although subsequent provincial linkages were provided, it is unclear how complete the documentation is as to the number of breast cancers and breast cancer deaths among the control group for the additional months of follow-up reported*." (2).

- women with symptoms should not be included

> "*screening trials should exclude women with clinical symptoms of breast cancer, such as lumps, but these women were allowed to participate in the CNBSS. Inclusion of these women, who are unlikely to benefit from breast screening, increases the number of cancers in the trial (improving the statistical power) but dilutes any assessment of the effects of screening*." (2).

- late-stage disease disproportionately assigned to the screening arm

> "*Benefit may also be masked if women with late-stage cancer are disproportionately assigned to the study group*". (2).

Lack of statistical power

> "*None of the studies included sufficient numbers of women ages 40-49 to be able to provide proof of an expected benefit of 20-25%".* (2).

"*Not only was the death rate lower than the background mortality from breast cancer among Canadian women aged 40-49 but it did not even reach the level upon which their power calculation was based*". (2).

Contamination

"*This inadequate power was further diluted by the fact that 26% of the "unscreened" women in the control group had mammography outside the trial*" (2).

Kopans and Feig argue that even diagnostic mammograms are screening:

"*All mammograms "screen" the breasts*." (2).

Too short follow-up time

"*The preliminary results are further compromised by too short a follow-up period. The HIP trial, as well as the Swedish trials, suggests that mortality reduction for these women begins to appear approximately 8-10 years after screening begins*" (2).

Higher rate of women with advanced cancers (non-blinded randomization)

"*the randomization process allocated significantly more women with advanced cancers (four or more positive axillary nodes) to the screened group in the prevalence year (the first year of screening)*" (2).

Poor quality mammograms

" *Unfortunately, many of those reviewing the trial have looked only at the results and not at the flaws in the underlying design and statistical assumptions [12]. They have also ignored flaws in the execution of the trial, particularly deficiencies in mammographic technique and interpretation*" (2).

"*The screening centers were permitted to use whatever mammographic equipment they had available, since funding was not provided to purchase modern equipment. There was no special training for the technologists performing the mammograms, or for the radiologists interpreting the studies*" (2).

"*Thus, instead of evaluating the efficacy of high-quality mammography for screening, the designers of the trial decided to test the validity of mammographic screening as they assumed it was then being generally practiced at the time the study was forming (Miller AB, personal communication). In addition, despite concerns voiced by early advisers to the CNBSS, little effort was made to improve the quality of the mammography; indeed, two advisers (W. Wende Logan and Stephen A. Feig) resigned in protest over this issue*" (2).

Suboptimal mammographic interpretation

" *An internal review conducted by the CNBSS's own reference radiologist revealed that 42% of the cancers that were missed at screening and became palpable in the interval*

between screenings were visible on a previous mammogram and had been missed by the interpreting radiologist" (2).

Finally the complainers stated their believes:

"*We believe that the available data support the use of annual mammographic screenings for all women beginning at age 40*" (2).

Burhenne LJ, Burhenne HJ. The Canadian national breast screening study: A Canadian Critique. Am J Radiol 1993; 161: 761-63 (3)

In the same issue of the Journal, a paper by Drs. Burhenne and Burhenne (radiologists in Canada) was published (3). They echoed similar concerns, focusing on inadequate mammography technique, quality of mammography equipment (high interval cancer rates; interval cancer is a cancer appearing after a negative screening mammogram, but before next scheduled screening mammogram), inclusion of women with symptoms, and physical examination prior to randomization (which lead to more women with lumps in the mammography arm than in the control arm) (3). They also claim that "*Dr. Miller had difficulty accepting the benefits of screening*" and cites a paper of cervical cancer screening Dr. Miller had authored (3).

They also claim that "*Even though radiologic program consultants recommended pretrial training of radiologists, and* [redacted] *and* [redacted] *are on record as having offered to train CNBSS radiologists in mammographic screening*" (3). [redacted] we have not been able to find any record of this offer and Drs. Miller and Baines both deny such offer. (*Part VI*, *Chapter 29*).

Radiology

Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ. A critical appraisal of the National Breast Cancer Screening Study (4)

Boyd and coauthors express similar concerns as Drs. Kopans and Feig in their paper from 1993 (4).

CNBSS is not generalizable

As CNBSS is based on volunteers responding to recruitment and are not generalizable to the Canadian population. Women recruited in the CNBSS are more educated, had fewer children, had lower prevalence of smoking. Women with symptoms were included. These factors may influence their risk of breast cancer and breast cancer death (4).

Randomization

Randomization by local coordinator instead of central randomization. Allocation of randomization arm was known to the local coordinator before consent was given. Boyd and co-authors claim there was no evidence of bias of allocations (4):

"*Impossible, when inspecting a list of names to determine the order in which the names were entered*." (4).

"*Evidence of imbalance in allocation is, however, present for the first (prevalence) screening examination that took place immediately after allocation. In women aged 40-49 years, 24 had breast cancer with poor prognosis, with four or more lymph nodes involved when they entered the trial. Nineteen of the 24 were allocated to the screening arm of the trial, and in all but two of them, breast cancer was detected at physical examination*". (4).

"*This imbalance could have arisen from bias in the allocation process or from chance*. *The probability of an imbalance of at least this magnitude arising by chance can be calculated as a binomial and is 3.3 in 1,000. Chance is, therefore, a possible but not likely explanation*" (4).

Quality

"*The evidence, however, suggests that the quality of mammographic images obtained in the NBSS did not represent state of the art even for the time*" (4).

Compliance

"*High levels of compliance are described: 86% of women aged 40-49 years and 87% of those aged 50-59 years*"
"*No allowance for less than full compliance appears to have been made in the sample size calculations*" (4).

Contamination

"*Of women aged 40-49 years, 26% of the control group underwent mammography at least once during the study period; of those aged 50-59 years, 17% of those allocated to undergo physical examination only also underwent mammography at least once during the study*" (4). There was no reference to these numbers.

" *No allowance for contamination appears to have been made in the sample size calculations.*" (4).

Cointerventions

"*In this context, treatment of detected breast cancer is the principal cointervention to be considered, and the investigators describe a review of treatment (not yet published) that failed to show any treatment-related differences that were associated with mortality from breast cancer*" (4).

Follow-up

The authors cast doubt about the follow-up procedures of the CNBSS.

"*The extent to which these follow-up procedures were successful in establishing the health status of all NBSS participants at the time of analysis is not stated*" (4).

Statistical considerations and study power

In the CNBSS protocol, the trialists states that the HIP study does not provide guidance for effect of mammography screening among women aged 40 to 49, and:

> "*therefore, a sample size has been chosen on the basis of feasibility. The power of the test as a function of the relative death rate from breast cancer in the control and study groups is shown in Table 1 for the chosen sample size of 25,000 using an expected 5 year death rate from breast cancer in the control group of 212 per 100,000 and a one-sided test at the 5% level*" Figure 9.1 (7).

Figure 9.1 Sample size by relative breast cancer death (table 1 in 7)

TABLE 1. THE POWER OF THE TEST AS A FUNCTION OF THE RELATIVE DEATH RATE FROM BREAST CANCER IN THE CONTROL AND STUDY GROUPS

Relative death rate	Power with sample size of 25,000
1.4	0.50
1.5	0.59
1.6	0.69
1.7	0.77
1.8	0.83
1.9	0.87
2.0	0.91

Thus, when following 25,000 women in the age 40-49 for 5 years, the expected numbers of breast cancers death are 53 (25,000 x 212/100,000 = 53) versus the observed 28 deaths (follow-up period ranged from 5.4 to 12 (mean 8.5) years) (Box 12.1) (8). Boyd and co-workers claim:

> "*substantially fewer deaths occurred than were predicted. For example, it was predicted that in women aged 40-49 years, there would be approximately 74 deaths over 7 years in the control group*" (4).

It seems as the expected death rate in the control group at 5 years was too high, and almost double the observed rate (28 deaths/25,216 women = 111 per 100,000 women) (8).

The panel could not find information about the expected effect of the intervention (annual mammography and clinical breast exam) among women in the age 40 to 49 years (7). However, in the first publication of the results (8), the expected effect seems to have been a 40% reduction in breast cancer mortality:

"*The sample size was fixed to determine whether a reduction of 40% in the rate of death from breast cancer would be seen in the intervention group, as compared with the control group*." (8).

Boyd and co-authors conclude (4):

"*For women aged 40-49 years, the NBSS results appear to exclude the anticipated 40% in mortality reduction. The 95% confidence interval, however, is compatible with a reduction in mortality as large as 16%. This result is likely biased by the large imbalance in women with breast cancer with poor prognosis at baseline*."

"*The influence of this baseline imbalance will diminish as the number of women in the trial who die of breast cancer increases, and extensive further follow-up will be required to determine whether the present estimates of relative risk change.*" (4).

References

1. Bailar III, JC. MacMahon B. Randomization in the Canadian national breast Screening Study: a review for evidence of subversion. Can Med Assoc J 1997; 156:193-199
2. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. Am J Radiol 1993;161:755-60.
3. Burhenne LJ, Burhenne HJ. The Canadian national breast screening study: A Canadian Critique. Am J Radiol 1993; 161: 761-63.
4. Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ. A critical appraisal of the National Breast Cancer Screening Study. Radiology 1993;189:661-3.
5. Tarone RE. Breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995; 75:997-1003.
6. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334.
7. Miller AB, Howe CR, Wall C. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 1981; 4:233-246.
8. 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 Nov 15;147(10):1459-76. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543

Rebuttal, published in 1994

These claims have been repeated since these publications and are more or less similar to the claims the critics are claiming today and in all the documentations given to the panel in the complaint material. It appears the academic debate has had a standstill and that every rebuttal to their claims are being ignored.

Dr. Baines summarized the critique of CNBSS in Annals of Internal medicine in 1994 (1)

> "*what one wants to believe is easy to believe*." (1)

Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334 (1)

In the piece, she summarized the critique in a table, figure 10.1 (1).

Figure 10.1 Summary of aspects criticized in the CNBSS (table 1 in (1))

Table 1. Aspects of the National Breast Screening Study That Have Been Criticized

Design and execution
Entry and exclusion criteria
Power
Timing of randomization
Randomization
Subversion of the randomization procedure
Excess cases of advanced breast cancer in the mammography group
Excess number of deaths from breast cancer in the mammography group
Mammography
Delayed cancer detection
External review
Competence of radiologists and technologists
Lack of mortality reduction in women aged 50 to 59 years
Contamination of controls
Analysis
Lack of exclusion of screen-1 cancers
Premature publication
Follow-up and ascertainment

She then explained the methods used in CNBSS (6):

Design and execution

> *"Entry criteria included meeting the age criteria, not being pregnant, not having been diagnosed with breast cancer, not having had a mammogram in the 12 months before entry, and signing informed consent forms".* (1).

Randomization

Women were individually randomized

> *"Women who met the entry criteria and completed two questionnaires (yielding identifying and demographic data including risk factors for breast cancer) then signed informed consent forms in the examining room.*
>
> *The screen-examiner asked if the participant had breast symptoms (lump, pain, discharge) and recorded the responses. A physical examination of the breasts and instruction in breast self-examination followed, after which the examiner decided if the clinical findings required the participant's referral to the NBSS surgical review clinic (usually held within 1 week). This decision was documented on the examiner's form, and the woman was informed.*
>
> *The examiner then left the participant and approached the center coordinator or her deputy, who carried out the randomization procedure. Randomization lists, contained in four separate books that each included one quinquennium (40 to 44 years, 45 to 49 years, and so on), were provided to all centers. On learning the participant's age, the coordinator chose the appropriate book and entered the date and name on the first available line, thus assigning to the participant her identification number and allocation. The coordinator entered the number and allocation on all chart forms, and the examiner told the participant if she was to have mammography.*
>
> *From the examiner's perspective, it was not important to obtain a mammography allocation if a breast lump had been found because the participant would already be referred for surgical review. From the coordinator's perspective, skipping a line to achieve a desired allocation was not feasible because she could not predict when the next appropriately aged woman would arrive to fill the skipped slot."* (1).

Clinical examination of the breasts (CBE)

> *Breast examination and breast self-examination instruction and evaluation were done by locally trained nurse-examiners in all centers outside the province of Quebec and by physicians in Quebec. The duration of the clinical examination ranged from 5 to 15 minutes, depending on the size of the breasts and on the amount of verbal interaction required.* (1).

She referred to the following studies regarding the quality of CBE (box 10.1) (1)

Box 10.1 CNBSS publications regarding quality of clinical breast exam (CBE) (1)

Clinical examination	5 -10 min Documented in: - Miller AB, Baines CJ, Turnbull C. The role of the nurse-examiner in the National Breast Screening Study. Can J Public Health. 1991 ;82: 162-7. (2) - Baines CJ, Miller AB, Bassett AA. Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. Cancer. 1989;63:1816-22. (3)

- Baines CJ, To T. Changes in breast self-examination behavior achieved by 82,835 participants in the Canadian National Breast Screening Study. Cancer. 1990;66:570-6. (4)

Surgical review

Study surgeons at each center examined women with abnormalities and decided if diagnostic follow-up was required. They forwarded their recommendations to the woman's physician. Most recommended procedures were done. In women aged 40 to 49 years, diagnostic mammography was done as a consequence of the screening examination in 0.8% and 1.5% of the mammography and comparison groups, respectively. (1).

Pathology review

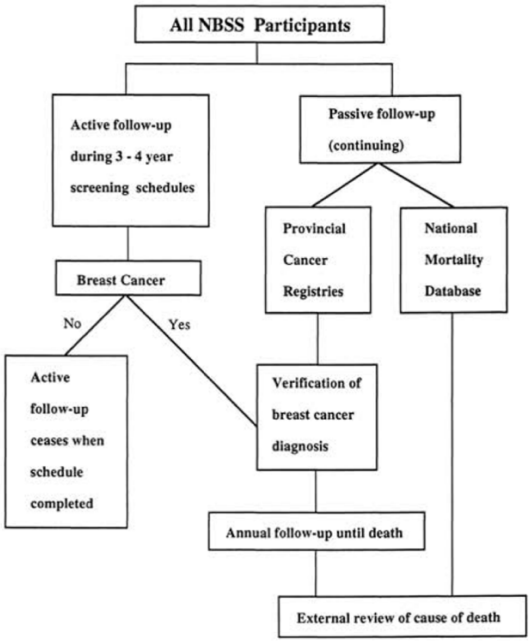
Reference pathologists appointed for each center reviewed slides from all surgical procedures done on participants from their center. (1).

Follow-up procedures

Breast cancer ascertainment was achieved by screening centers (in operation from 1980 to 1988) during each woman's screening schedule, linkage to provincial cancer registries yielding unknown cancer cases in program dropouts and in participants after schedule completion, and linkage to the National Mortality Database where breast cancer mentioned as a cause of death led to review of diagnostic and other records. Pathologic review verified breast cancer diagnoses. After verification, annual follow-up enabled ascertainment of death. The cause of death was reviewed externally. (1).

Follow up is described in figure 10.2.

Figure 10.2 shows flow-chart of follow-up for all CNBSS participants. (figure 1 in (1))



After describing the randomization process and methods used, Dr. Baines argued against some of the claims. She said some of the controversies was defused at an international workshop on breast cancer screening where it was observed that CNBSS results were consistent with findings from other screening studies (1).

Randomization

> *Crucial prerandomization variables are distributed equally across allocations, namely, the frequency of self-reported symptoms, a positive family history for breast cancer, and the referral rates to surgical review based on abnormalities found during physical examination (*figure 10.3) (1).
>
> *Speculation has arisen that abnormal physical findings at the initial visit would induce the examiners or the center coordinators to assign preferentially such women to the mammography group. Table 2 also rules out this concern* (figure 10.3) (1).

Figure 10.3 Distribution of variables documented before randomization (table 2 in (1))

Table 2. Distribution of Variables Documented before Randomization Showing Equivalence across Groups

Variable	40 to 49 Years		50 to 59 Years	
	Annual Mammography plus Physical Examination	Single Physical Examination	Annual Mammography plus Physical Examination	Annual Physical Examination Only
Self-reported symptoms at entry				
Total cohort, *n*	25 214	25 216	19 711	19 694
Lump, %	7.3	7.3	4.0	3.7
Pain, %	18.0	18.0	15.2	14.4
Discharge, %	2.6	2.7	1.4	1.4
Any symptom, %	23.5	23.5	18.2	17.7
Family history of breast cancer, *n*				
Total reporting	9493	9652	7606	7585
Mother	2051	2055	1489	1522
Sister	831	872	1258	1242
Daughter	2	4	20	18
Second degree	6609	6721	4849	4803
Referral to surgical review on the basis of abnormal physical findings, %				
Screen 1	14.1	14.6	11.0	11.2

> *In further response to claims that randomization was flawed, the original randomization sheets were re-examined to look for changes in script or pens used, crossing out of names, erasures, or problems with date sequences, with special attention given to the records of those who had died of breast cancer. No suspicious entries were found.* (1).

Excess advanced cancer

> *The phenomenon is not unique to the NBSS. Excess cases of advanced cancer in screened groups aged 40 to 49 years were reported previously by three Swedish trials (5-7).* (1).

Identical proportions of the two groups were referred to review clinics on the basis of abnormalities found on physical examination of the breasts.
Study surgeons consistently recommended more diagnostic interventions for the former [mammography arm] *than the latter* [control arm]. (1).

By year 5, cumulative numbers of cases with four or more positive nodes were 39 and 22 in the mammography and comparison groups, respectively, difference which, when expressed as a proportion of difference which, when expressed as a proportion of cases detected, 14.6% compared with 10.9%, becomes less dramatic. (1).

This imbalance triggered a review of all invasive breast cancers diagnosed in the first 5 years. When the mammography and control groups were compared the mean numbers of nodes dissected were 11 and 10, the proportions of cases in whom no nodes were dissected were 5% and 10%, and the proportions of cases in whom fewer than four node s were dissected were 10% and 14%, respectively. These comparisons indicate a potential for under-ascertainment of nodal involvement in the group not receiving mammography. (1).

Contamination

"Contamination" is said to have occurred because 26% of the control group aged 40 to 49 years reported receiving a mammogram. It was always expected that the control group would receive normal community care, meaning that symptomatic women would receive diagnostic mammography when clinically required. (1).

Indeed, it was the purpose of the study to determine whether intensive screening could improve outcomes compared with normal care. In fact, during their NBSS schedules, 14.5% (3651) of the control group had one examination, 7.8% (1968) had two examinations, and 4% (1036) had three or more. Single mammographic examinations are probably diagnostic. It is improbable that a single, or even a few, mammographic examinations over a 3- to 4-year period in 26% of the control group could obliterate any benefit achieved by annual mammography in almost 100% of the screened women. (1).

Quality of mammography and equipment, and clinical breast
She referred to the following studies (box 10.2) (1):

Box 10.2 CNBSS publications regarding quality of mammograms (1)

Quality of mammograms	Documented in: - Baines CJ, McFarlane DV, Wall C. Audit procedures in the National Breast Screening Study: mammography interpretation. Can Assoc Radiol J. 1986;37:256-60. (8) - Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the National Breast Screening Study. Invest Radiol. 1990;25:971-6. (9) - Baines CJ, McFarlane DV, Miller AB. Sensitivity and specificity of first screen mammography in 15 NBSS centres. Can Assoc Radiol J. 1988;39:273-6. (10)

Interpretation of an external review
Dr. Baines provides information about the external review (1):

> *The evening before the review, the two reviewers said their participation was conditional on being allowed to rate all mammograms by 1988 standards: The 1988 standards required a mediolateral oblique view. This was unfortunate because between 1980 and 1984 the NBSS protocol* [11] *required two-view mammography, including straight mediolateral and craniocaudal positioning, a decision determined in consultation with U.S. and Canadian expert radiologists before the initiation of the NBSS in 1980. Ironically, the director of the NBSS urged at that time that the mediolateral oblique view, already being used in the Swedish trials, be used. The radiologic consultants insisted on the straight mediolateral view because it conformed to contemporary North American practice. In 1985, when the screen-1 examinations were completed, the Policy Advisory Group formally approved a change in positioning to mediolateral oblique, although at least one center had implemented it in 1983.*
>
> *A fifth variable imposed by the two U.S. radiologists, namely a global rating, correlated so closely with the mediolateral scores that it was a proxy for the oblique view*.
>
> *To claim that in the first 2 years of the study, almost 50% of the mammograms were unacceptable is misleading* [12]. *Table 7 shows that only 4.9% of all NBSS mammograms performed between 1980 and 1988 were done in the study's first 2 years (1980 and 1981). The so-called unacceptable mammograms done in those 2 years make up only 2.1% of all NBSS mammography. A now familiar criticism is that even by year 4 more than 50% of mammograms were unacceptable. Table 7 clearly refutes this point* (figure 10.4) (1).

Figure 10.4 The proportion of mammograms scored poor of fair (Table 7 in (1))

Table 7. Proportions of Mammograms Scored as Poor or Fair at Technical Review Applied to Distribution of Mammogram Frequency by Calendar Year*

Calendar Year	Proportion of All Mammograms (%)	Proportion Rated as Poor or Fair (%)		
		In Calendar Year Using Four Criteria	As Proportion of All Mammograms	
			Using Four Criteria	Using Three Criteria
1980	1.3	39	0.5	0.4
1981	3.6	44	1.6	1.2
1982	7.3	39	2.8	2.1
1983	12.2	36	4.4	3.3
1984	19.1	32	6.1	4.4
1985	19.0	25	4.7	4.0
1986	16.7	11	1.8	1.5
1987	13.0	10	1.3	1.3
1988	7.7	NA†	NA	NA

* Scores were based on four criteria: craniocaudal view, mediolateral view, contrast and density, and image quality. The score for mediolateral views was censored for column 4.
† NA = not available. Sample for technical review did not include 1988 films.

> *Table 7* [figure 10.4] *also shows how these proportions are further reduced (to 0.4% to 4.4%) if scores for the mediolateral view are removed for reasons already described. Even then, the proportions of unsatisfactory mammograms are inflated because the reviewers' scoring for image quality was influenced by their disapproval of the straight mediolateral view. The persuasiveness of the reviewers-turned-critics is weakened when one considers their poor intraobserver and interobserver agreement combined with the small sample size on which the criticism is based* [13]*. Nor can technical quality be linked to the previously discussed excess of advanced cancers at screen 1. When centers are categorized into tertials by technical quality, no association with the distribution of advanced cancers is found* [14]*. Comparative data on technical quality are not available from any other screening trial.* (1)

Dr. Baines concluded

> *The uncritical acceptance of erroneous information has been widespread. One NBSS radiologist said he knew that some centers did not have dedicated mammography units and that others used Xeromammography. A U.S. radiologist asked study investigators, "When did the NBSS start using compression?" It always used breast compression, always used dedicated mammography units, and never used Xeromammography* [15]*. Myths can preempt evidence*" (1)

> "*It is revealing that two radiologists have dismissed as the "scientific fringe"* [16] *hypotheses that assume there may be a biological basis for breast screening's apparent lack of benefit in these women. In fact, there are many biological differences between premenopausal and postmenopausal women and in the natural history of breast cancer in these two groups. The breast screening controversy is an excellent example of socioscientific controversy. Such controversy is to be expected when established medical practice is challenged by results from randomized controlled trials. The failure of intracranial arterial bypass to reduce the risk for ischemic stroke when tested in a randomized multicenter trial* [15] *distressed surgeons who believed the procedure was useful* [17]. *In the case of breast screening, not only are radiologists distressed but also women who have been programmed to overestimate their risk for breast cancer* [18] *and who therefore need to be reassured.*

References

1. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334.
2. Miller AB, Baines CJ, Turnbull C. The role of the nurse-examiner in the National Breast Screening Study. Can J Public Health. 1991 ;82: 162-7.
3. Baines CJ, Miller AB, Bassett AA. Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. Cancer. 1989;63:1816-22.
4. Baines CJ, To T. Changes in breast self-examination behavior achieved by 82,835 participants in the Canadian National Breast Screening Study. Cancer. 1990;66:570-6.

5. Andersson I, Aspergren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ. 1988;297: 943-8.
6. Frisell J, Eklund G, Hellstrom L, Lidbrink E, Rutqvist LE, Somell A. Randomized study of mammographic screening—preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat. 1991;18:49-56.
7. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am. 1992;30:187-209.
8. Baines CJ, McFarlane DV, Wall C. Audit procedures in the National Breast Screening Study: mammography interpretation. Can Assoc Radiol J. 1986;37:256-60.
9 Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the National Breast Screening Study. Invest Radiol. 1990;25:971-6.
10 Baines CJ, McFarlane DV, Miller AB. Sensitivity and specificity of first screen mammography in 15 NBSS centres. Can Assoc Radiol J. 1988;39:273-6.
11. Miller AB, Howe CR, Wall C. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 1981; 4:233-246.
12. Kopans DB. The Canadian Screening Program: a different perspective. AJR Am J Roentgenol. 1990;155:748-9
13. Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders D, Sickles EA, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. Am J Roentgenol. 1990;155: 743-7.
14. 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 Nov 15;147(10):1459-76. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543
15 . The EC/IC Bypass Study Group. Failure of intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. N Engl J Med. 1985;313:1191-200.
16. 2. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. Am J Radiol 1993;161:755-60.
17. Sundt TM Jr. Was the international randomized trial of extracranial-intracranial arterial bypass representative of the population at risk? N Engl J Med. 1985;316:814-6.
18. Feuer EJ, Wun LM, Boring CC, Flanders WD, Timmel M, Tong T. The lifetime risk of developing breast cancer. J Natl Cancer Inst. 1993;85:892-7.

Criticism of the CNBSS in 1995

Tarone RE, Breast cancer in young women screened with mammography in the Canadian National Breast Screening Study, Cancer 1995; 75:997-1003 (1).

Dr. Tarone points to the imbalance of number of women with positive lymph nodes in the screening compared to the control arm among women aged 40-49 year (CNBSS 1) (1).

> "*The excess of patients with cancer with four or more positive lymph nodes in the 40-49-year mammography age group of the NBSS was statistically significant, even when expressed as a percentage of all invasive cancers diagnosed.*" (1)

Among women aged 40-49 years, 22% of women with breast cancer had four or more positive lymph nodes in the screening arm compared with only 8% in the control arm (Figure 11.1) (p = 0.043; assuming a Poisson distribution) (1). After 7 years of follow-up, 14% of women with breast cancer in the screening arm and 8% in the control arm had four or more positive lymph nodes (p = 0.037) (1).

Eight (47%) of the 17 women with breast cancer with four or more positive lymph nodes in the screening arm and one (20%) of 5 in the control arm died during the first 7 years of follow-up (1).

Figure 11.1 The proportion of patients with stage II invasive breast cancer in different randomized trials of mammography screening (table 1 in (1))

Table 1. Proportion* (%) of Patients With Invasive Breast Cancer Diagnosed With Four or More Positive Lymph Nodes in the Canadian National Breast Screening Study by Year and Method of Detection

	Age 40–49 yr		Age 50–59 yr	
	Mammogram and physical examination (%)	*Usual care (%)*	*Mammogram and physical examination (%)*	*Physical examination only (%)*
Screen 1:P†	17/65 (26)	5/60 (8)	9/71 (13)	11/64 (17)
Screen 1:MO‡	2/21 (10)	—	0/48 (0)	—
< 12 Months§	6/16 (38)	2/24 (8)	1/14 (7)	4/16 (25)
Screens 2–5:P	8/73 (11)	—	7/70 (10)	11/84 (13)
Screens 2–5:MO	1/46 (2)	—	5/79 (6)	—
Years 2–5§	10/65 (15)	15/148 (10)	8/59 (14)	7/110 (6)
≥ Year 6§	3/45 (7)	1/40 (2)	2/29 (7)	1/46 (2)
Total	47/331 (14)	23/272 (8)	32/370 (9)	34/320 (11)

* Denominator is the number of patients with invasive breast cancer diagnosed, and numerator is the number with four or more positive lymph nodes.

† P: detected by physical examination.

‡ MO: detected by mammography only.

§ Interval or incident cases.

Dr. Tarone also showed the proportion of stage II-IV breast cancer in women with breast cancer in several breast cancer screening trials (Figure 11.2 (1)). The Swedish two-county data are from patients diagnosed in the first 8 years of follow-up. The Stockholm data are from patients diagnosed in the first 5 and 6 years. The Malmö data are from patients diagnosed during 10 years of mammographic screening. The Edinburgh data are from patients diagnosed in the first 7 years. The HIP study are from patients diagnosed during the first 6 years (1).

How stage was defined in the different trials is not mentioned by dr. Tarone, and data on TNM stage distribution is not directly available for the CNBSS. Further, it is not known if clinical breast examination detection rate is higher for breast cancer with four or more lymph nodes compared to breast cancer with less than four positive lymph nodes.

The number of cancers that were detected by clinical breast exam was similar in both arms (age 40-49: 65 in the screening arm and 60 in the control arm; p = 0.72; age 50-59: 71 in the screening arm and 64 in the control arm; p = 0.61) (6), The number of women referred to a review clinic was similar in both arms for women aged 40 to 49 years (3 569 women (14,1%) in the screening arm, and 3674 women (14,6%) in the control arm) (2).

In most trials other than the CNBSS 1 (women aged 40-49 years), the proportion of advanced (stage II and more advanced) breast cancers were higher in the control arm. In contrast to the CNBSS, the control arm of the other trials were not offered any intervention.

Figure 11.2 The proportion of patients with stage II invasive breast cancer in different randomized trials of mammography screening (table 2 in (1))

Table 2. Proportion (%) of Patients With Invasive Breast Cancer Diagnosed at Stage II or Higher in Randomized Trials of Screening Protocols Including Mammography

Study	Screened group (%)	Control group
Two-county		
	460/1193 (39)	453/742 (61)
Stockholm*		
5 years	143/340 (42)	150/239 (63)
6 years†	173/397 (44)	210/393 (53)
Malmo		
All ages	190/486 (39)	231/393 (59)
Age < 55‡	— (42)	— (52)
Age ≥ 55‡	— (38)	— (63)
Edinburgh		
	228/354 (64)	221/256 (86)
HIP		
All ages	162/371 (44)	190/371 (51)
Age < 50	69/151 (46)	72/145 (50)
Age ≥ 50	93/220 (42)	118/226 (52)
NBSS		
All ages	200/701 (29)	156/592 (26)
Age < 50	102/331 (31)	66/272 (24)
Age ≥ 50	98/370 (26)	90/320 (28)

HIP: Health Insurance Plan of Greater New York; NBSS: Canadian National Breast Screening Study.

* To facilitate the direct comparison of screened and control groups, published control numbers were adjusted to reflect the larger size of the screening group.[22] Percentages are unaffected by this adjustment, but adjusted control numbers should not be used for formal significance testing.

† The control group was offered screening during 6th year of follow-up.

‡ Only percentages could be derived from relative frequency distribution.

§ Numerator is number of cases with any nodal involvement. International Union Against Cancer staging of NBSS cases has not been published; tumors with nodal involvement would be stage II or higher.

Removing women with advanced breast cancer form statistical analysis of CNBSS
Tarone suggested that "*a mortality analysis omitting all patients with advanced disease detected by physical examination at the initial screening visit is warranted to assess the impact of the initial imbalance of advanced disease*," because any effect of screening on breast cancer death, may be obscured by the excess mortality observed in the screening arm among women with breast cancer with four or more lymph nodes in the first years of follow-up (1).

Staging data have not been published for the CNBSS study, but used women diagnosed with a positive lymph node (figure 11.2 (1)). The denominator is number of cancers, not the rate (number of cancers divide by person-years), nor the risk (number of cancers divided by number of women enrolled).

References

1. Tarone RE. Breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995; 75:997-1003.
2. 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 Nov 15;147(10):1459-76. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543

Previous review of CNBSS

In March 1995, Professors John C. Bailar III and Brian MacMahon were asked by the Cancer Institute of Canada (NCI) to reviewed the randomization and allocation process in the CNBSS (1). The review was prompted by the result of CNBSS published in the Canadian Medical Association Journal in 1992 (2, 3), indicating that death rates were higher in the screening arm than in the control arm among women aged 40 to 49 years old at enrollment (Box 12.1).

Box 12.1 Breast cancer mortality after 5.4 to 12 years, in screening and control arm by age (2,3)

For women aged 40-49 years old, 38 women died in the screening arm and 28 in the control group, resulting in a 36% increase in death among women in the screening arm, (ratio proportion of death 1.36; 95% confidence interval 0.84 - 2.21). The follow-up period ranged from 5.4 to 12 (mean 8.5) years (2).

For women aged 50 to 59 years, 38 women died in the screening arm and 39 in the control group, resulting in a 3% decrease in death among women in the screening arm (ratio proportion of death 0.97; 95% confidence interval 0.62 to 1.52). The follow-up period ranged from 5.3 to 12 (mean 8.3) years (3).

It was suggested that nurses responsible for clinical breast exams, or coordinators at some centers may have subverted the randomization by allocating women to the screening arm (1).

According to Bailar and MacMahon (1), rumors of misallocation were fueled by repeated communications presentations at meetings, personal letters (to the director of NCI [Dr. J. David Beatty] and the reviewers [Drs. Bailar and MacMahon]), media reports, and publications in specialist journals, such as Radiology and American Journal of Radiology (4, 5).

Further, Dr. Robert E. Tarone (Biostatistics Branch, US National Cancer Institute) published a paper in 1995 pointing to the excess rate of advanced breast cancer among the CNBSS for women aged 40-49 years (2, 6).

Review by Drs. Bailar and MacMahon

Drs. Bailar and MacMahon focused on the CNBSS data for women 40 to 49 years old (CNBSS study 1) where the critique was most substantial and where the results showed more women with breast cancers with four or more lymph nodes in the screening arm compared to the control arm (1, 2, 6).

The review did not focus on quality of imaging, equipment, or any other technical aspects of mammography in the trial.

The review focused on centers where women in the screening arm had higher mortality than in the control arm (3 centers) and on centers where administrative problems were reported (2 centers).

The review also consisted of a document review and an assessment of the randomization. The review did not interview study staff. Bailar and MacMahon wanted to interview one radiology technician that was said to have personal knowledge about subversion of the randomization, but the technician did not want to participate in the review and declined the interview request (Figure 12.1) (1)

Document experts at a private investigation and security company (KPMG) were hired to assist in reviewing instances in which names of subjects were altered in the "allocation books" (the basic instrument used to assign, at random, participants to either the mammography or the usual-care arm) (1).

The KPMG investigators found no evidence of violation of randomization allocation or deliberate attempts to conceal any alterations. The investigators did find alterations in assignment books, but explained that these occurred for explainable reasons and do not indicate subversion. Among women where the name had been deleted or superimposed, one died of breast cancer. She was in the mammography arm (1).

The review concluded that

> "*The document experts found no evidence of a deliberate attempt to conceal the alterations. Even if there had been acts of subversion, they could only have been few in number and, given that there was only 1 death from breast cancer in the group reviewed, the alterations could have had only a trivial effect on the study findings as reported in 1992*" (1).

More details are provided in the next section.

From Bailar and MacMahon, CMAJ 1997; 156:193-199 (1)

We have quoted the text in the published review and made text excerpts from the publication (1). To improve readability of the published review, we have highlighted text that we believe is the most important findings in the review (1). Text in brackets are the panels comments.

1. Assessment of randomization

Randomization and allocation

The originally published protocol stated that after the woman completed the questionnaire and signed the consent form the form was to be "*passed to the center coordinator for randomization. The allocated regimen would not, however, be reported until after the physical examination was completed to avoid bias in decisions over possible abnormalities*."

… it is clear from other publications that, except in one centre (identified as centre 03), ***randomization occurred after the clinical examination****: the material completed at the registration desk accompanied the patient to the examination, and, after the information from the examination by the nurse or physician had been added, the folder was given to the coordinator for randomization*.

We are unclear (as seemingly were the study centre staff at the time of our visit) as to why this procedure became the normal practice in all but one of the centres. ***We believe that questions about whether clinical findings prompted circumvention of the randomization strategy would have been avoided if the strategy as originally specified had been followed.***

It was the task of the coordinator to enter each woman's name and identification number on the next available line in an "allocation book." These books were specific for centre and for 5-year age group. The lines were randomly allocated between mammography (MA) and usual care (UC). If the line on which the woman's name was placed indicated MA, the procedure was usually ***undertaken at the first visit****.*

If we understand the process correctly, except in centre 03 ***the nurses*** *(and probably also the coordinators)* ***were aware of the findings of the clinical examination when the allocation was made.***

Basis for charge

Herein lies the basis of the charge that examiners who thought that a woman should or should not have a mammogram, because of findings at clinical examination or personal information obtained during the examination (e.g., risk factors for breast cancer), may have compromised the randomization.

Abnormal findings

It should be noted that in each centre, ***a nurse or physician finding an abnormality****, regardless of the woman's allocation,* ***would so inform the coordinator****, and the case would usually be referred to a special review clinic for examination by the project surgeon.* ***The woman might then have a mammogram even if her allocation was to the UC arm****.*

However, because referral would not have ensured mammography, the charge has been made that there remained a motive for an examiner or a coordinator to subvert the randomization if for clinical or other reasons he or she believed that the subject should or should not have a mammogram.

To avoid subversion of randomization of this type, it is current practice to conceal the allocation from both the study subject and the person doing the randomization until or after the commitment of the subject to a particular arm of the study*. Randomization by telephone through a central study office is one method currently employed. This may not have been possible in a study the size of the NBSS, but a simple procedure (involving, for example, removal*

of labels identifying the study arm after the patient's name had been entered) would have strengthened the credibility of the process.

2. Review of allocation by KPMG

Three centers with excess deaths

The KPMG investigators examined the allocation books for evidence of alteration or substitution of names, one of the methods in which randomization could have been compromised. Because the goal of mammography is to prevent death (and morbidity) from breast cancer rather than to prevent the disease itself, and because time and expense were important issues, the investigators' ***review was limited to the three centres where there was an excess of deaths in the mammography arm*** *compared with the control arm among women 40–49.*

These were centre 02 (6 and 2 deaths respectively), centre 03 (7 and 4 deaths respectively) and centre 11 (4 and 2 deaths respectively). Although these centres were selected on the basis of data for the ***group aged 40–49****, the allocation books for women* ***50–59*** *at the same centres were also examined.*

*In addition, the books were examined for limited periods in two centres where the NBSS central office had suspected administrative problems: centre 01 (****from October 1980 to January 1982****) and centre 04 (****from May 1981 to August 1982****).*

When a name had to be changed (e.g., the original name had been entered in the wrong age book) the existing name was to be crossed out with a single line (so that the original name could still be read) and the correct name written above it.

Staff at the centres frequently used white-out or correcting tape to block out the original name. Fortunately, the KPMG investigators were able, in most instances, to ***decipher both the original name and the replacement****.*

Original name

The original name was then sought in the database, by both the KPMG investigators and the NBSS central staff. Information from the database sometimes provided an explanation for the change—for example, if the original name was later found in the book for a different age group in the same centre with the same date of recruitment, or there was confusion between a maiden name and a married name.

However, this process was incomplete: in some cases the original name was not found, and in others a difference in recruitment dates implied that the names referred to two different women with identical names, such occurrences being not uncommon in so large a database, especially, according to Wall [Claus Wall was the data manager of the CNBSS]*, in the francophone Quebec centres, where the range of names was smaller than in the anglophone centres.*

No cover-up

The KPMG investigators stated that the likelihood of successful cover-up of an alteration would be near zero under their examination. *Entries made in pencil could be erased well enough to obscure the original entry, although evidence would be left that a change had been made. The investigators added that they* ***found no evidence of any deliberate attempt to conceal alterations.***

Result of investigation

Clerical error

A total of ***30 182 records were inspected****, of which* ***467 (1.5%) required investigation****. Of the* ***467 records 219 (47%) indicated clerical errors*** *(e.g., given and family names in reverse order) and involved no change in the identity of the woman entered on the allocation line*.

Credible match

The remaining 248 records of women, whose names had been covered and substituted, revealed 147 with a "credible match" (i.e., the same name and date of entry, within specified ranges described by Wall), and 101 whose names were not found elsewhere. When an uncovered name differed from that superimposed, the two names were listed as "pairs" in Wall's report.

For 86 (59%) of the 147 women for whom a match for record 2 (uncovered by KPMG) was found elsewhere in the database, the matched woman was in a different age group from the woman in record 1 (name visible on the allocation list). It is likely that the original entry was made in the wrong age book and was subsequently corrected.

The remaining 61 subjects matched with another woman in the same age group, it is important ***to recognize that the items randomized in the study were not women but lines in the allocation books.*** *Which study arm a woman was assigned to depended on which line she occupied in the allocation book. If there was subversion of the randomization it had to have affected the woman's placement in the book, because the allocation assigned to the line could not be changed*.

Alterations

Thus, after we eliminated obvious clerical errors and instances in which a woman might have been first entered in the wrong age book, ***we discovered alterations on 97 lines allocated to mammography and on 65 lines allocated to usual care.***

Assuming an equal probability of allocation to either group, the split of 97/65 is highly unlikely (χ2 = 6.3, 1 degree of freedom; p ≅ 0.01, by McNemar's test).

Women allocated to mammography returned to the centre annually for repeat visits, and so there would be more opportunity to correct originally incorrect entries and to make changes, such as a name. Further, there was more interaction between the centre staff and the women assigned to mammography than between the staff and those assigned to usual care in the

period immediately after allocation because of the mammography procedure itself and other procedures associated with it. ***It is therefore not surprising that there would be more changes to the allocation pages of the mammography group.***

It is also possible that some of the alterations may have been made to free up a line allocated to mammography to make room for an improper allocation. However, the logistics of such a manoeuvre would have been challenging. *By the time a name would have been covered, the woman first entered would have probably been told of her allocation and her identification number would have been entered on several study forms.*

Seventeen (18%) of the 97 women whose names were entered onto the mammography lines were referred by the nurse (or physician) for surgical review, as compared with 8 (12%) of the 65 whose names were entered onto the usual care lines. *This difference is well within the bounds of chance.*

Clearly, whatever misallocation might have occurred by the overwriting of names could have had ***only a trivial effect*** *on the results as published in 1992* (1)*.*

The review quotes relevant literature (table 6.1). The review panel did not have access to the four last reports in tab 12.1.

Table 12.1 showing relevant literature for by Drs. Bailar and MacMahon paper (1)

1	Miller AB, Howe GR, Wall C. The National Study of Breast Cancer Screening Protocol for a Canadian Randomized Controlled trial of screening for breast cancer in women. Clin Invest Med. 1981;4(3-4):227-58. PMID: 6802546.(7)
2	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. CMAJ. 1992 Nov 15;147(10):1459-76. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543. (2)
3	Baines CJ. The Canadian National Breast Screening Study: a perspective on criticisms. Ann Intern Med 1994;120:326-34 (8)
4	KPMG Investigation and Securities Inc. Toronto, résumés of the investigators from this company who were assigned to review the NBSS documents are on record in the office of the NCIC executive director (9)
5	National Breast Cancer Screening Study (“NBSS”). Toronto: KPMG Investigation and Securities Inc, 1995 November 20.(10)
6	National Breast Cancer Screening Study (“NBSS”). Toronto: KPMG Investigation and Securities Inc, 1996 January 5. (11)
7	Wall C. Report in response to the KPMG reports of November 20, 1995, and preliminary report of January 5, 1996. Toronto: Department of Preventive Medicine and Biostatistics, University of Toronto, 1996 January 31. (12)

These three last reports contain personal confidential information and were said to be made available, but the panel were not able to retrieve these reports (1, 10-12).

Bailar and MacMahon also discussed the findings of CNBSS 1 (women aged 40-49 years old), where four and more lymph nodes were found in the screening arm (0.13%) than in the control arm (0.08%) (1, 6). (*Chapter 31*)

Removing women with advanced breast cancer form statistical analysis of CNBSS

"*Tarone* [6] *recommended that, because there was a higher proportion of patients with 4 or more positive lymph nodes among the subjects assigned to mammography than among those assigned to usual care, mortality analyses should have been undertaken after elimination of "advanced cases detected by physical examination at the initial screening visit.*

We agree with Tarone but note that, as Miller has pointed out, allocation to mammography may itself lead to surgery and the discovery of nodal involvement; therefore, elimination of, for example, patients with 4 or more positive nodes may introduce a bias favouring survival for the patients with no advanced disease in the mammography group.

The most appropriate way to identify such women would have been to note whether or not they had been referred by the examining nurse or physician for surgical evaluation. This referral would have occurred before mammography and, therefore, should be the least biased with respect to allocation.

We discussed this issue with Miller, who stated that such analyses had been done and the results did not substantially affect the initial study findings, but that the NBSS group had decided to delay publication of these results until the 10-year follow-up data were available"(1).

In the 11-16 years of follow-up of CNBSS study 1 (including women aged 40-49 years old) results for these suggested analyses, was done (13).

> "*Tarone suggested that women with cancer detected at screen 1 by breast physical examination should be excluded from both groups. Although the validity of excluding subgroups identified after the intervention as a result of mortality analyses is uncertain, Cox regression analysis performed after such exclusions results in an odds ratio of 0.93 (CI, 0.70 to 1.24). A similar analysis excluding women who reported a lump to the examiner at screen 1 yields an odds ratio of 0.88 (CI, 0.66 to 1.18)* (13)".

Without exclusion of any women the cumulative rate ratio comparing the screening arm with the control arm was 1.12 (CI, 0.82 to 1.53) (13).

A coordinator was dismissed

"*The central office became aware of rumours that the coordinator at one of the study centres was subverting the randomization to ensure mammography for some of her friends. When confronted, the coordinator firmly denied the allegations. However, after examining the allocation books the study director deemed it sufficiently likely that the rumours were true, and the coordinator was promptly removed from her position.*

Records from this centre were reviewed for the 14-month period when this coordinator was in charge. Of the 4945 records, 34 (0.69%) indicated insignificant alterations (i.e., a change that did not result in a different name appearing on the allocation line). The proportion of insignificant alterations is similar to that among the records reviewed from the other four centres (0.73%).

Of these 34 insignificant alterations 25 were on lines allocated to mammography and 9 on lines allocated to usual care. This ratio (25:9) is significantly different from the ratio of 17:17 that one might expect ($p \cong 0.01$). However, we have noted earlier in this report the reasons why such a discrepancy might exist. Only 1 of the patients with an insignificant alteration in her record died; her death occurred 7 years after entry into the study and was not attributed to breast cancer.

During the 14 months at this centre, there was only 1 significant alteration (i.e., a name substitution); it was on a line allocated to mammography. The woman was not found to have breast cancer. Overall, among the women aged 40–49 enrolled at this centre, 8 died of breast cancer: 4 in each study group (Dr. Anthony B. Miller: personal communication, 1995).

In addition, we explored the question of whether, in the centre where the coordinator was removed, the pattern of allocation itself was unusual, irrespective of whether any unusual allocation affected the results of the study.

During the total study period at this centre, 4111 women aged 40–49 years were allocated to mammography and 4120 to usual care. The corresponding figures for women aged 50–59 were 3208 and 3199. We requested data for the 4945 women for whom records were reviewed by KPMG (i.e., the women entered into the study during the 14 months when this coordinator was in office). The number of allocations to the mammography and control arms were virtually identical in both age groups: 1397 and 1394 in the 40–49 age group, and 1055 and 1060 in the 50–59 age group. There were 39 refusals: 22 among the women assigned to the mammography group and 17 among those assigned to the control group. ***It does not appear that the activities of the coordinator in question influenced either the pattern of allocation or the mortality results from this centre as included in the data reported in 1992****"* (1)

Why was staff from CNBSS not interviewed

Dr. Kopans wrote an open letter to Dr. MacMahon June 23 1995 urging Dr. MacMahon to interview staff from the CNBSS as part of the review.

Excerpts from the letter Kopans sent to MacMahon, June 23 1995 :

"*It has long been rumored in the technical community in Canada that this happened. A shift of tens to even a few hundred women from one group to the other would not alter the balance of demographic factors but could have major consequences for the trial. I would strongly urge that you interview those involved in examining the women, as well as those involved in assigning them to the study or control group. I suspect, however, that, if the process was compromised, individuals will be reluctant to admit to actions that may have such important consequences.*

Either they have to be assured of complete confidentiality (to avoid any retribution) or they must be interviewed under some form of oath."

Respond by Dr. Bailar and MacMahon

It is stated in the terms of reference that Bailar and MacMahon should contact and obtain information from one individual that had come forward with information that compromised the randomization process (figure 12.1) (1). Bailar and MacMahon explain why they did not interview the field staff (1):

"*First, we did not interview any of the field staff of the study, even though a few are still working in the participating centres. We felt that any "steering" of the randomization was likely to have been highly specific to location (centre) and possibly time, and that any examiner or coordinator who participated in or knew of active subversion of the randomization but did not come forward at the time would have been unlikely to admit it to us, more than 10 years later, even if he or she remembered the details.*

Second, we wrote to the individual (a radiology technician) quoted by Kopans as having "personal knowledge" of such details, but she did not respond to our letter, even though the letter assured her that her response would be kept confidential. Beatty [Executive director of the Canadian National Cancer Institute at the time, 1995 (1)] *reported to us that, before our review, he had, after several attempts, spoken by phone to this person, who told him that on one social occasion (Kopans was not present) she had made idle comments on this subject but was unaware of any substance to the charges. She declined, however, to put any of her statements in writing, despite Beatty's assurance of confidentiality. She had not been employed at the centre until about 2 years after the close of randomization*." (1).

Figure 12.1 The terms of reference for the review of the CNBSS by Drs. Bailar and MacMahon (1)

Appendix 1: Terms of reference for the review of the randomization procedures of the Canadian National Breast Screening Study

The following terms of reference are taken from a letter to the reviewers dated Mar. 30, 1995, from Dr. J. David Beatty, then executive director of the National Cancer Institute of Canada.

1. Review the randomization strategy, in the context of when the study was planned (1976), the size of the study, and the randomization strategies of other similar studies.
2. Review the original randomization sheets for evidence of possible randomization violations, such as overwritten names, erasures, unused slots, etc.
3. For each if the 24 subjects with breast cancer at the original screen who had four or more nodes positive, review the records for randomization violations (e.g., check the times of their original examination and when they were randomized).
4. At the centres in which substantial differences in breast cancer mortality are present between the mammography and usual-care groups, review the records for randomization violations of each of the women who had physical abnormalities at initial screen, who were referred to review and in whom cancer was detected by the screening process.
5. Contact an individual who was reported to have claimed to have first-hand information concerning compromise of the randomization process to obtain information about this matter.

We, the review panel, had similar experiences. One person who was suggested to be a witness [redacted] would not talk to us, and another person suggest being a witness was not part of the CNBSS [redacted]). Both of whom were part of the Summary report (14).

Accompanying editorial

In an accompanying editorial Boyd wrote (15) "*They present convincing evidence that there was no association between alterations, or administrative problems, and the rate of death from breast cancer*.

101 names are not accounted for

> *"These findings, however, are unlikely to quell completely the concern about the randomization process in the NBSS. The absence of name alterations had previously been cited by the NBSS investigators as evidence that randomization had not been subverted. We now know that names were altered and that there were more alterations in the mammography arm. Although 78% of these changes could be accounted for in*

some way (e.g., clerical errors), the remaining 22% (representing 101 names) could not. We know nothing of who these women were or why their names appeared once in the allocation books, were replaced by another name and never appeared again in the NBSS.

We also now know that a coordinator at one of the NBSS centres was suspected of assigning her friends to the mammography arm and that the suspicion was strong enough to remove her from her position. We do not know the method of subversion thought to have been used by this coordinator, nor do we know whether an examination of name alterations revealed what she was alleged to have done. No coordinators were interviewed. Although it is unlikely, as Bailar and MacMahon suggest, that any would admit wrongdoing, if such admissions had been made they would have provided powerful evidence." (15).

Boyd explains how this may have happened:

"*Apparently, great deviousness would not have been required to achieve a particular allocation. Simple, riskfree methods appeared to exist if a subject wished, with the cooperation of the coordinator at the centre, to be allocated to one of the two study arms. Suppose a subject wished to be allocated to the mammography arm? She would have had a 50/50 chance of being assigned to that group in any event. If the next allocation was to the control arm instead, the subject's name could have been entered onto the line with the next mammography allocation, leaving a gap in the allocation book, or she could have been advised to wait until the line for the desired arm was the next to be filled. In either case, it is unlikely that much time would have elapsed before a mammography allocation came up or a gap on the list was filled. Fifteen NBSS centres randomly assigned 90 000 women over 5 years or less: they must have been busy places*." (15).

Correspondence

Trial investigators response - changes were made, but none were suspicious

Dr. Bains wrote in a letter to the Editor (16):

"*NBSS investigators have never reported that name alterations did not occur among the entries for the 90 000 NBSS participants; alterations clearly did occur. We have reported that no suspicious changes in the random allocation sheets had been identified in the participants who died of breast cancer*.

The external review found that, of 97 unexplained alterations on lines allocating women to mammography, only 1 was associated with a woman who died of breast cancer, and breast cancer had not been diagnosed at the first screen in this case. This alteration was either overlooked by us or detectable only by forensic experts.

All NBSS randomization sheets were routinely and carefully examined each month at the national coordinating centre during the recruitment period. It is unrealistic to expect that

written entries could be made for 90 000 participants without errors requiring correction. The issue is not whether changes were made but whether suspicious changes were made. Suspicion, like beauty, is in the eye of the beholder. Those who are suspicious should demand equal scrutiny of random allocation procedures in all screening trials" (16).

Why names had to be changed

Dr. Boyd did not understand that names may have had to be erased and replied:

"*The need for any erasures in the randomization lists is far from clear. Although, as Baines states, the NBSS did enroll a large number of women, names were entered on randomization lists only after the completion of several procedures. Why, after a woman has completed 2 questionnaires, undergone a breast examination and signed a consent form, there should be any remaining doubt about her name, is something that I do not understand*" (17).

The review sparked a discussion between dr. Kopans and the trial investigators

In a letter to the editor regarding the review by Drs. Bailar and MacMahon (1), Dr. Kopans wrote (18):

"*there was opportunity to compromise the process, since the lists were open and multiple allocation numbers were frequently obtained ahead of time. As a result, lines could be skipped without any need for erasures or alterations"* (18).

Review adds little

"The most direct way to find out whether the process was compromised would be to ask those involved in the allocation and to provide them with anonymity and protection from retribution. This was not done. Consequently, the authors' review adds little to what is already known." (18)

Problems at other centers than those reviewed

"The reviewers confine themselves to evaluating 3 centres. Given that allocations were supposedly random and given the relatively small number of deaths due to cancer at each centre, the problems may not have occurred in the centres where the allocations appeared to be "imbalanced"; they may well have occurred in the centres where the allocations appeared "balanced." (18)

Independent review of follow-up and linkage in the CNBSS

"In the abstracts printed by the NIH for the conference, Miller wrote that "the number of breast cancer deaths are now 52 in each arm." At the meeting, he stated that this had been a "mistake" and that there were 82 deaths among the screened women and 67 among the controls. An independent review of the linkage and follow-up of deaths due to breast cancer in the NBSS should be undertaken to ascertain whether there are other "mistakes."" (18)

Absence of support

> *"Had I* [Dr. Kopans] *been a radiologist involved in the NBSS, and confident in what had transpired, I would have argued strenuously in support of the methods and results of the trial. I find the absence of such support surprising*." (18)

Baines, Miller and To replied (19):

> "*Dr. Kopans persists in raising concerns, most of which have previously been shown to be unwarranted."* (19).

Randomization was not open

> *"Randomization in the NBSS was not "open." Individualized randomization was achieved by a process in general use before distributed computing and electronic mail were available. Instead of telephone operators consulting prearranged lists, we had specially trained administrative staff handle our randomization process. Only they had access to the lists. The screen-examiners did not conduct the process, nor did they have access to the lists"* (19).

Balanced randomization

> *"The NBSS is the only screening study in the world that can completely document balanced randomization in the 2 allocation arms.*
> *Two external evaluations of randomization in the NBSS have failed to find evidence of falsification* [20]. *No other screening study has been subjected to equivalent scrutiny, although questions should have been raised not only by the Edinburgh trial but also by the recently published Gothenburg trial, in which screening did not detect a higher rate of breast cancer than in the control group."* (19).

Warning after hearsay

> *"The recent review of randomization in the NBSS was initiated after Kopans made a charge to the National Cancer Institute of Canada of scientific misconduct by one of us. This serious charge was based on hearsay from a radiographer previously employed at an NBSS centre; the radiographer had begun her employment after randomization had ceased, as Bailar and MacMahon discuss".* (19)
>
> *"It is not a "revelation" or an "imbalance," as Kopans claims, that women in a usual-care group, in whom breast cancer is mainly detected on clinical grounds, are treated at different institutions than those receiving screening mammography. What may have been a revelation to Kopans was that women with breast cancer in the usual-care group fared no worse than those who had been screened with mammography, although they had lesser degrees of axillary dissection and less extensive histologic examination of resected tissue".* (19)

No mistakes

"Kopans refers to "mistakes" in the data we submitted for the NIH consensus conference.[21] *At the conference, we reported 82 deaths due to breast cancer in the mammography arm and 72 in*

the usual-care group, not 82 and 67, as Kopans states. What Kopans fails to acknowledge is that at the conference other investigators presented revised figures that superseded the data in their abstract submitted months before. The purpose of all presentations at the conference was to give the most recent data." (19).

Enthusiasm

Radiologists such as Kopans, who rely on good survival from screen-detected case series to establish that a benefit exists, [22] *are unhappy because women 40 to 49 years of age with mammographically detected breast cancer in the NBSS achieved a 90% 10-year survival rate, and yet these good survival data do not translate into a reduced rate of death due to breast cancer. Kopans' zeal may be excessive* [23]. (19).

References

1. Bailar III, JC. MacMahon B. Randomization in the Canadian national breast Screening Study: a review for evidence of subversion. Can Med Assoc J 1997; 156:193-199
2. 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 Nov 15;147(10):1459-76. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543
3. 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 Nov 15;147(10):1477-88. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423088; PMCID: PMC1336544.
4. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. Am J Radiol 1993;161:755-60.
5. Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ. A critical appraisal of the National Breast Cancer Screening Study. Radiology 1993;189:661-3.
6. Tarone RE. Breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995; 75:997-1003.
7. Miller AB, Howe CR, Wall C. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 1981; 4:233-246.
8. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334.
9. KPMG Investigation and Securites Inc. Toronto, résumés of the investigators from this company who were assigned to review the NBSS documents are on record in the office of the NCIC executive director
10. National Breast Cancer Screening Study ("NBSS"). Toronto: KPMG Investigation and Securities Inc, 1995 November 20.
11. National Breast Cancer Screening Study ("NBSS"). Toronto: KPMG Investigation and Securities Inc, 1996 January 5.
12. Wall C. Report in response to the KPMG reports of November 20, 1995, and preliminary report of January 5, 1996. Toronto: Department of Preventive Medicine and Biostatistics, University of Toronto, 1996 January 31.

13. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002;137:305-312
14. Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants.
15. Boyd NF. The review of randomization in the Canadian National Breast Screening Study. Is the debate over? Can Med Assoc J 1997; 156:207-209
16. Baines CJ. NBSS: changes were made, suspicious changes were not. Can Med Assoc J, 1997;157 248-50
17. Boyd NF. The author responds. Can Med Assoc J, 1997;157 250
18. Kopans. DB. NBSS: opportunity to compromise the process. Can Med Assoc J, 1997;157 247
19. Drs. Miller, Baines and To respond. Can Med Assoc J, 1997;157 247-48
20. Cohen MA, Kaufert PA, MacWilliam L, Tate RB. Using an alternative data source to examine randomization in the Canadian National Breast Screening Study. J Clin Epidemiol 1996;49:1039-44
21. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 12 year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000;92:1490–9
22. Stacey-Clear A, McCarthy KA, Hall DA, Pile-Spellman E, White G, Hulka C, et al. Breast cancer survival among women under age 50: Is mammography detrimental? Lancet 1992;340:991-4
23. 16. Taubes G. How one radiologist turns up the heat. Science 1997;2275:1057.

PART V
Transcripts from interviews

While interview transcripts were included
in the report, they have been removed
to protect the privacy of the individuals interviewed
consistent with their consent form

PART VI

Summary of our review and interview

Summary of interviews

Violation of randomization

Eye-witnesses

We interviewed two medical radiation technologist who were working in the CNBSS.
One had been working 10 shifts at St. Michael's Hospital in Toronto in 1985. She worked shifts from noon to 8 pm. According to her statement, women with lumps were placed on the mammography list, whether or not that person initially was randomly allocated to the mammography arm.

> "*All the patients saw the nurse examiner first. Once they saw the nurse examiner. If the nurse examiner felt an anomaly, if that person wasn't on my list, it didn't matter, they became a person on my list, and instead somebody else was removed, which again did not go with what I understood about randomization, of what I had understood when I first started the program when I was hired for the interview.*"

She continued to explain that names of women without lumps were removed from the lists and replaced by another woman with palpable lumps.

> "*There were 20 to 24 patients every day that we're coming into the screening program. And the nurses were given a list of all the patients. We were given a list of 10 to 12 patients, depending on what the number was that day. But even the first 3 patients, even the very first patient that they brought us to do a mammogram on she wasn't on my list. And the technologist said to me, all this happens. The nurses see the patients first and when the nurses are seeing the patients, if they feel something is wrong, if there's an anomaly, they'll bring them to us to do a mammogram".*
>
> *So let's say Mrs. Jones was on my list. And then they brought in Mrs. Apple instead. Later in the day, Mrs. Jones, as long as Mrs. Jones didn't have a symptom, they may have taken Mrs. Jones off to replace Mrs. Apple that they did put on the list.*
>
> *There was a person above like a nurse navigator type individual. So they wrote the names of the actual individuals that did receive the mammogram. So the start of a day. I would have a proposed list. Right.*
> *So Mrs. Apple wasn't on the list. Mrs. Jones was. Mrs. Jones didn't have a symptom. Mrs. Jones's name would have been removed from that list at the end of the day. Her name didn't even exist on that list that I had, and Mrs. Jones name wasn't on the list that went in as being a person that had a mammogram. But Mrs. Apple's name did go on.*
>
> *However, if Mrs. Jones showed up and did have a symptom, then they would take somebody else off the list. I'm just giving an arbitrary at the end of the day. It still worked out to the same number."*

She stated that she saw names being written down on the lists and at the end of the day, the list would be given back to the coordinator. She never saw the list again.

> *"I could see the names they were writing in the book, but it wasn't me writing the names in the book. There was a list of patients that were supposed to be coming in, and if somebody didn't show, it would say no-show beside it. Many people who had the mammo only got filled in once that mammo was done. That list wasn't pre filled for them. They gave me a list, but then that list disappeared. I had to give the list back, so it would disappear."*
>
> "*They didn't write their names in until the end of the day. They may have had a proposed list beforehand. The proposed list in the start of the day disappeared."*

She was asked by the review panel if the lists were not completed as the women came in, but were rather completed at the end of the day.

> *"The nurses had their list. I had my list. The nurses were seeing a patient while I was doing a mammogram. And there was another person that was there overseeing. So at the end of the day she coordinated to make certain she may have been writing down names for the nurses as they were going in. But my list at the end of the day I had to give it to her to coordinate. She would write my names that I have seen, down at the end when I gave her the list"*.

She claims that she saw that this happened, not every day, but perhaps over half of the 10 days she worked in the program.

> "*I saw her writing names in the book, and I saw one section for the nurses one section for me, I mean. I didn't sit here, and hover over her head 24/7 she's writing. She was my boss.*
>
> *I didn't see it every day I worked, but I saw it over half the days that I worked. I'd say if I worked 10 shifts. I saw it 6 of the shifts.*"

The review panel asked her specifically if she actually did see anyone changing the book.

> *"So the ones that I did, that's when they would write it into the book. Afterwards they would write down the ones that I did in my book. Afterwards. The book that you're talking about like the page there, wasn't anybody written in there until I gave them my list".*

Backup randomizers

[redacted] were often backup randomizers when the center coordinators were sick or off duty. [redacted] did not report what the eye-witness reported to us.

> "[redacted] *we're often the backup randomizers when the center coordinators were sick or away, or off duty. I've never encountered a nurse who came in and told me what assignment her patients had to have. She asks for the information and I provided. That is the way it operated.*"

Another eye-witness said women had opinions on whether to have a mammogram or not, but did not observe that women were placed into the arm they were not randomized to. She only spoke to women who were assigned to mammography.

> "*It just seemed from talking to patients or patients talking to us, that the participants had a lot to say about whether they were going to have the mammography or not have a mammography… We met, of course, with the women who actually had a mammogram.*"

Not names, but IDs

Another eye-witness had worked on the trial from the beginning [1980] until 1988. She had been working as a medical radiation technologist at [redacted] since [redacted]. She did not mention that the list contained names, but numbers or IDs, which is also confirmed by the PI.

> "*So the clients would come over, and all they had was a paper envelope with a number on it, and then we would use that number to mark our films… We just basically did what we would do with a normal patient, except that we didn't take a history."*

[redacted] explained the randomization, were at each center there were four books with randomization schedules, each one for a quinquennium (age 40-44, 45-49, 50-54 and 55-59) and these were held secure by the center coordinator. The nurse would come to get the allocation from the coordinator after the clinical breast exam. And the assignment would be the next blank allocation in the appropriate age group.

> *"After the nurse had made her examination and given some instruction on self-examination. The nurse went to the Coordinator and asked for the randomization. There had to be a supervisory entity the week called the Center Coordinator. She did not examine. She did not do anything except make sure everything ran properly, and she was the only one who had the responsibility for doing the randomization.*
>
> *It was organized to be in four randomization books, one for each five year age-group. Sometimes, unfortunately, the Coordinator would put the woman in the wrong line, and then that would have to be corrected. It didn't happen very often.*
>
> *Each line had to be filled in with an ID number given to the participating women at enrollment and each line was pre-populated with the random allocation: for example mammogram, no-mammogram, no-mammogram etc*."

Possible to wipe out IDs and replace them?

The review panel asked [redacted] specifically whether it would be possible to wipe out a ID and switch it with another ID. They emphasized it could be possible, but would be very difficult , and that those who would do so had to be lucky not to be caught.

> "*That's what I am saying, she had to be lucky. Because if she skipped two lines and nobody else came in that day to fill in those 2 spaces. That would have not been acceptable. Because ultimately you would know the date for each ID person.*"
>
> "*No*.
>
> *I think it would be very difficult to do that without it being detected. Claus Wall was pretty good at these sorts of things. He'd be with me for a number of years he'd been the data manager for the clinical trials program when it was initiated several years before. Well, we didn't think it would be easy to manipulate it. As I said, the randomization sheets were carefully checked by the data manager, Claus Wall. And we did not find there were errors."*

[redacted] confirmed that fields in the randomization books were pre-populated for mammography and in order to re-assign a person, the coordinator had to leave gaps in the books. And would need to populate these gaps later to have an off-controlled group people.

One instance of subversion

In one instance a coordinator was subverting randomization. This would mean that she basically entered the name in a line that was indicating that the allocation would be to the experimental group to the mammography group.

> "*In one instance, in* [redacted] *we discovered that the Coordinator was, in fact, choosing who to put in what field, in in terms of the allocation. But other than that, and we discovered that very early on, and she was very quickly replaced. Other than that, I don't think we had any issues.*
>
> *We were very careful to check the randomization sheets, and we don't believe it happened elsewhere*".

[redacted] also referred to this event.

> "*It was simply a woman from a high socioeconomic level in* [redacted] *deciding that her friends needed to be included. She was fired.*
>
> *We would have no way of determining how many, but they were her personal friends, so would presume that not very many people. It was discovered very quickly. Could she corrupt a trial of almost 100,000 people? I doubt it very much. I also doubt that any of these women had breast cancers."*

The proportion of women for whom subversion may have occurred was likely to be small.

> "*90,000 women were entered in the book. Is it a surprise that entries were made in the wrong book? Or that women misreported their age? Then corrected themselves. So names were crossed out and re-entered in the correct book. A tiny minority.*
>
> *It was organized to be in four randomization books, one for each five year age-group. Sometimes, unfortunately, the Coordinator would put the woman in the wrong line, and then that would have to be correct. It didn't happen very often.*"

[redacted] continued:

> "*I think it's important to understand. that this was a trial across a whole continent with 15 centers and 15 independent institutions. We all had a very strong sense of their own relative importance compared to everybody else. From Halifax to Vancouver we hired people, we trained people and we visited them regularly. Tony visited them, I visited them, our National board visited them. We observed what was going on during the intake processes. And if, after all that, people wanted to deceive you, I guess it's always possible to be deceived.*
>
> *We had extraordinary trust in everybody that was operating as far as the protocol was concerned."*

Women with lumps

Women with lumps were referred to review clinics and part of the follow-up might have been mammography. We tried to understand whether the patients with lumps the eye-witnesses described had mammograms, were women that should have mammograms as part of their follow-up and even if they might have been assigned to the control arm. And that this was actually what she might have seen. She emphasized that this could not have happened.

> "*No, the reality is, they literally left the nurses room and came directly into my room. There was no going back and forth to see anybody else, no consultation with anybody. If the nurse did an exam, the minute they finish the exam they walk directly into my room. The nurse walked them into my room*.
>
> *So when the nurses brought the patients to me it was being emphasized. They were doing this because the patients needed it. The nurses said they were going to have to have a mammogram, anyway. So they were putting them in and just taking somebody else out".*

None of the interviewees who worked in the trial on a daily basis could explain the procedure to the reviewers. It seemed as some of the interviewees did not know that it was according to the trial protocol that all women with palpable lumps at the clinical exam prior to

randomization had been referred to a review clinic and probably receive a mammogram regardless of randomization arm.

> "*Now, if they were on the mammography arm, my understanding is that they would have had the physical examination. Then the mammogram, and then, if there was positive, I guess, on either of these, they would be referred to the review clinic.*
>
> *If it were control group, they would basically have the results of the clinical examination. They were still registered in the allocation book as part of the control group, and then they would go without the mammogram again to the review clinic*."

[redacted] explained that all women with lumps would be referred to the study surgeon.

> "*If there were an abnormal physical examination in either group, they went to the study surgeon*".

Typically, interviewees who had signed the complaint material believed women with lumps were assigned to the screening arm.

> "*You know, I mean that the nurses were palpating, and they, I understand them, felt sorry about those who had a palpable tumor. They were put in into the study group. This was an open secret*"

Further, some of the interviewees said that women with lumps should be excluded from the trial all together.

> "*It would be a diagnostic, not a screening mammogram. There's no point in screening a woman who is symptomatic. Screening is only for asymptomatic women. The purpose of screening is to find something that would not have been detected otherwise.*"

Witnesses coming forward in 2021

One of the witnesses who had tried to speak to colleagues and others about what she witnessed in 1985, was finally heard in 2021.

> *"I know nothing about there being any questions about the study being corrupt or anything, until* [redacted] *happens to make little comment in his presentation on* [redacted]*, saying that the investigators had always wanted to talk to a person who was employed by the study, and they weren't allowed to. I made a little comment in the sidebar during his presentation, saying. Yes, the study was corrupt. I witnessed it, I mean, and that's where this is all coming from all these years later somebody is finally listening*".

[redacted] confirmed this:

> "*When I gave the talk the virtual talk in [redacted], a technologist emailed me a couple of hours after the talk, and said that she has witnessed virtually everything that I said. It's my understanding that the Canadians who have tried to track down anyone who's still alive from the trial, found out that women were recruited from doctors' offices where they were because they had concerns about having breast cancer. So that provided more opportunity to put in women who had clinical evidence of breast cancer and more opportunity to, with the hope of taking good care of people, putting them out of random order into the screening arm.*"

The second witness the panel spoke to said:

> *"I really didn't understand until much later how these trials would work. When I started doing trials myself, for equipment and positioning and things like that, and it's like: You've got to have a standard. You've got to stick to your guidelines, and if you don't, then you have to disregard that. So I did not realize that I just thought that you know, if a woman came, and she felt that she had a problem, that I was doing something good to help her, and I wouldn't have known to tell anybody either, or who to tell*?"

Concern about unethical practices

The reason to come forward was explained by one of the witnesses:

> "*I feel that it was subverted from the get go. I don't think that they realize the outcome, that this was going to have for not just Ontario, not Canada, but worldwide. How this study was going to impact women*".
>
> *"I questioned it after I left, and after I started to realize what, I started to do more research and do more studies, and do more by continuing education. And when I started to realize what randomization and about studies and about everything else, and then, when I did adeem this study, I had to do an ethical study before I could even participate in adeeming this study. It really emphasized again how unethical the practices that I witnessed In 1985 were, and it's been bothering me since 1985.*
>
> *I mean, I've done mammography for as long as 1985 to now still. And I see how it can be such a benefit to catch it when it's so much smaller. So then it doesn't have a tiny chance to spread, to infiltrate, to grow to a stage 3 to stage 4. The whole idea is to try and catch it as small as we can, so trying to dissuade these women from having mammograms pre age 50.*
>
> *My career started on a bad note. I didn't like what I saw. I'm still working. I hope to work for another couple of years at least. I'd like to see in my career the wrongs of this study corrected. I realized that things happened in the past*."

Mass media

The witnesses also highlighted repeated media reports claiming women in their forties were denied mammography screening in the trial, causing unnecessarily premature death. According

to some reports, this was all due to the findings in the CNBSS, and that CNBSS was an outlier (Complaint material, supplement).

> "*If you go through the news, Yaffe, and Gordon put up these articles and they go, you know they go to the press, and they go directly to try to get to government, and they go to patient organizations that are closely held in some ways. But the articles will say some of that: "the task force guideline would kill a thousand women". But another article says, "four hundred". Another one says "the Canadian trial has led to four hundred deaths a year". I think in common parlance, this is known as `making things up´.*"

Anecdotes

Several of the interviewees mentioned that there was a lot of rumors and anecdotes about the CNBSS:

> "*I heard a lot of anecdotes*"

> "*And there's now some at least one person who was a technologist, or somebody who was not a technologist, somebody there who was working in the program, he [Kopans] said that she actually mis-allocated people. But that is all I know*".

> *"So much of what we know about what theoretically took place comes from Bailar and MacMahon. But rumors, you know we, those of us in the community. How's it? How is the NBSS going? And oh, yeah, and people would say, well, but you know they did this, and they did that. It was all rumor."*

How rumors may have had an impact on the witnesses coming forward in 2021-22, is unknown. The main witness that claimed she saw subversion only worked 10 shifts in the CNBSS She is a medical technologist and has been working with mammography her whole career.

Conflicts of interest

One of the interviewees pointed to conflicts of interest:

> "*Let me start by saying that I find this "trial" completely ridiculous. To go after a woman* [Cornelia Baines] *forty years after she published the trials, and she is now approaching* [redacted] *years of age, I find absolutely disgusting. And this has come forward because of people with a huge conflict of interest in relation to mammography screening, who have been haunting Cornelia Bains for forty years. Now it is more than enough. I'm very angry about this*".

> "*Why is this going on? I think this trial had a bad luck of running into a very dedicated group of people who are out to get one result, regardless of what there is. I don't think there's any concern about what their academic responsibilities are. Responsibilities, at least basically telling, you know, the truth*".

Result-driven allegation

Several of the interviewees said that there must have been subversion of the randomization and that women with symptoms and lumps were put in the screening arm, because of the results of CNBSS, especially study one (women aged 40-49 years).

> "*Did you hear about somebody who actually did this [subvert randomization]"?*
>
> *I can't say I did that. You know the reason that I assumed it happened was when you look at the data, for example, in the in CNBSS one"*.
>
> "*Mammography was not where it really needed to be to find small invasive cancers efficiently. And you know, with good sensitivity and specificity, that was a problem, and that certainly was one of the factors that affected the findings of the of the trial. I'm convinced of that. Because when you look at the statistics of the cancer finding, that speaks to it very clearly."*
>
> *"The experts in many ways, betrayed us. Because they didn't like our results. They didn't like our results.*
>
> *It was expected with this trial with individual randomization, was definitely going to show benefit for women in their forties. And since some of these collaborating external radiologists were on our policy advisory group, they knew in detail, our results. They knew long in advance of the publication, that we were not going to produce the results that they were expecting. That we had expected"*.

Breast cancer mortality

[redacted] commented on this in the interview.

> *"One of the things that I did was: I monitored the deaths, and where they were occurring. And I can tell you in the first 3 years of my being there, saying the excess mortality in women 40 to 49 was a real surprise until 1985 when Tabár published his first result. I saw he found the same thing. He never admitted it."*

Some of the findings in the CNBSS, were similar to what was found in other trials (1-3).
For example more women with advanced cancers were observed in the screening arm of CNBSS one (age 40-49 years old) (1). This was also observed in the two-county study (2), figure 33.2 (fig 1 in 2). Similarly, breast cancer mortality was higher the first years of follow-up (1, 2, 3).

Increased scrutiny of CNBSS

Others pointed to the fact that the CNBSS has been scrutinized more than any other trial, and that the amount of details published was higher than for any other breast cancer screening trial.

"*I've never seen a trial scrutinized like this, and even with that scrutiny which I believe was done through the University of Toronto, if I remember correctly, they didn't have any reason to be concerned. So we we're not in the position, when we have that kind of vetting from a highly competent review team and investigation, we had absolutely no reason to believe that it hadn´t been vetted properly. And we were comfortable that there wasn't any kind of excess risk here, beyond the normal risk that go along with any kind of trial with non-centralized randomization*".

"I can only tell you that the Canadian trial was the only one that, reported in such detail the attention that was paid to the actual quality. If you were to ask me the same question for all the other trials I would have to say, I have no idea. It's just not available in the literature. To my knowledge, I am aware that radiologists involved in some of those trials have declared that their mammography was the best in the world. But unlike the Canadian trial, I have no published evidence other than self-congratulatory information".

References

1. 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 Nov 15;147(10):1459-76. Erratum in: Can Med Assoc J 1993 Mar 1;148(5):718. PMID: 1423087; PMCID: PMC1336543
2. Tabár L, Gad A, Holmberg LH, et al. Reduction in mortality from breast cancer after mass screening with mammography. Lancet 1985, 829-832
3. Larsson LG, Andersson I, Bjurstam N, Fagerberg G, Frisell J, Tabár L, Nyström L. Updated overview of the Swedish randomized trials on breast cancer screening with mammography: Age group 40-49 at randomization. Monogr Natl Cancer Inst 1997;22:57–61

Mammography outside of the trial

The panel was presented with two claims:

One is that there was limited availability of mammography screening outside of the trial. This is used to explain why women with symptoms or lumps were placed in the screening arm of the CNBSS. The other claim was that there was contamination in the control arm, meaning women had mammograms outside of the trials.

[redacted] explained:

> *"If you look at the Winnipeg site, there was a review done by Cohen et al [1]. And they looked at women who actually had previous health insurance claims for breast cancer. But then, for whatever reason, later were put into the NBSS, and I think part of that is that mammography was very, and very limited availability, and the best mammography that was available was at the screening site. So they found of nine participants who actually had prior breast claims in their health records".*

Earlier in the interview, [redacted] had shown us a slide claiming that there was contamination in the control arm, with 15-16% screening in the control arm.

> "*[T]his slide here just shows some of the screening that was done in the control group outside of the trial. This is a study based on the Winnipeg screening site* [1] *and shows like fifteen, sixteen percent of a screening of women in the control group outside the trial. So that's going to reduce the effect size*".

We asked how there could have been limited availability and a lot of contamination at the same time.

> "*There was, but it could have been on the same machine just not in the trial. I mean, the machine was not used necessarily exclusively for the trial, so that one, you know some of the hours of the day it was used for the screening program, and some of the hours it might have been used for diagnostic. Or there may have been another mammography machine that was used for the diagnostic. I don't know the answer. I know that in Winnipeg the mammography that was done outside the screening was done using the zero radiography process. So it could have been either on the study machine, but outside of the study or it could be on another machine.*
>
> *They're possible to both. Both occur at the same time*".

Other of the interviewees confirmed that availability of mammography was limited in 1980-1985/6, outside of the trial, and that mammography was not popular among women.

> "*We didn't do a lot of mammography in the early eighties like screening demography or we didn't really do screening mammography. At that time we did mostly diagnostic."*

"*At the time it's important to recognize why mammography was much less generally available than it now is*"

"*[M]ammography was very and very limited availability, and the best mammography that was available was at the screening site."*

"*[I]t [mammography] wasn't all that popular*"

The Winnipeg experience

In 1998 a study about the extent of contamination in the CNBSS, was published (2). Two years prior another study used alternative data to examine the randomization in the CNBSS (1). [redacted] gave his interpretation of these studies.

"*The Winnipeg experience [2], of course, just applies to the one site where there was a study done by an epidemiologist who was interested in that. The same group actually, who looked at contamination, screening outside the study, they published on that as well. It was of somewhat limited ability, but one still could get a mammogram, and I guess if someone was determined, and ask their physician for a referral.*

But for nine women who had prior health claims for breast cancer, and then participate in the NBSS. Eight of those, and these are small numbers, but eight of those nine are found their way into the mammography arm, and in the younger group the women in their forties, four of four of those women ended up in the mammography arm" [1].

The Winnipeg experience was a cohort study including 10,107 women participating in the Winnipeg screening center in Manitoba. Data on data of birth, health care insurance number, postal code was used to link data to the Manitoba health insurance records. 96.8% of women were identified in the health insurance and data on diagnosis (ICD-9) and use of mammography in the private sector (hospital mammograms were not available) were recorded. Information on indication for mammography was not available, billings for bilateral mammograms were considered as "*screening*" mammograms after excluding women who had a visit to a physician for a breast problem or breast-related procedures in the previous 2 years or a referral for a mammogram from the CNBSS.

The study from 1998 compared the rates of screening mammography outside the trial with data from a questionnaire participants answered. The observed rates were in agreement with the self-reported numbers of mammograms (questionnaire).

For women aged 40 to 49 at enrollment, 5.3% in the screening arm and 21.8% in the control arm had a claim for at least one bilateral mammogram. After excluding non-screening mammograms these proportions fell to 2.2% and 14.1% ($p< 0.0001$). Data from the CNBSS (all

screening center, not only the Winnipeg site) showed that self-reported use of outside mammography was 7% in the screening arm and 26% in the control arm (3)

For women aged 50 to 59, 4.5% in the screening arm and 16.7% of the control arm had at least one claim for a bilateral mammogram. These proportions were 2.1% and 10.5% for screening ($P < 0.0001$). Data from the CNBSS (all screening center, not only the Winnipeg site) showed that self-reported use of outside mammography was 6% in the screening arm and 17% in the control arm (4). The rate of screening mammograms outside of the trial increased over time.

The study from 1996 used similar data to examine whether allocation could have been nonrandom. They examined health claims (ICD-9 diagnosis, figure 29.1, and use of mammography) in the 2 years prior to CNBSS participation (1).

Figure 29.1 International Classification of Disease- 9 diagnosis used as prior health claims (Table 2 in (1))

TABLE 2. Codes used in study[a]

Condition	Code
Breast cancer	174.0–174.9
Breast carcinoma *in situ*	233.0
Other malignancy	140.0–173.9
	175.0–208.9
Benign neoplasm breast	217.0–217.9
Benign mammary disease	610.0–610.9
Other disorders of breast	611.0–611.9
Pain in breast	611.71
Lump or mass in breast	611.72
Nipple discharge	611.79

[a]From the *International Classification of Diseases*, 9th rev.

Altogether 9477 women (96.6% of 9780) who had at least one physician or hospital claim during the 24 months prior to CNBSS enrollment. There was no difference in the screening and control arm (both age-groups) among women who did not have a claim.

There were no significant differences in the proportion of women with prior histories of benign breast disease, investigation, or mammograms ($p > 0.05$) in the screening and control arms of CNBSS (both age-groups) (Figure 29.2).

A detailed examination of all records from January 1, 1979 to December 31, 1992 was undertaken for women with breast cancer. Nine women had a claim with the diagnosis of breast cancer (ICD-9 174); 8 had been allocated to the screening arms of the trial ($p = 0.06$), 4 among women aged 40-49 all in the screening arm, and 5 among women aged 50-59 (4 in the screening arm and 1 in the control arm). Eight of these women had no subsequent claims for breast cancer. There were no differences across the study arms in the proportion of women who reported a prior history of breast disease or a prior xeromammogram for whom an insurance claim was found.

Figure 29.2 Number and percentage of women with one or more claims for different diagnosis and procedures in the 24 months prior to CNBSS enrollment (tab 3 and 5 in (1))

TABLE 3. Number and percentage of women with one or more claims for the following diagnosis in the 24 months prior to entry in the NBSS by whether randomized to mammography or no mammography arms of the NBSS

	Age 40–49 years					Age 50–59 years				
	Mammography (*N* = 2528)		No mammography (*n* = 2540)			Mammography (*N* = 2204)		No mammography (*n* = 2205)		
	N	Percent	*N*	Percent	*p* Value	*N*	Percent	*N*	Percent	*p* Value
Malignant neoplasm of the breast (ICD 174.0–174.9)	4	0.16	0	0.0	0.12	4	0.20	1	0.05	0.37
Benign neoplasm of the breast (ICD 217)	5	0.20	7	0.28	0.56	3	0.14	1	0.05	0.32
Benign mammary dysplasias only (ICD 610)	75	2.97	83	3.27	0.47	38	1.72	38	1.72	0.91
Other disorders of the breast only (ICD 611)	133	5.26	109	4.29	0.12	88	3.99	77	3.49	0.39
Multiple diagnoses (ICD 610 and ICD 611)	65	2.57	57	2.24	0.53	24	1.09	34	1.54	0.19
Any breast problem	282	11.16	256	10.08	0.28	159	7.21	151	6.85	0.60
Benign mammary dysplasia and any other diagnosis	143	5.66	143	5.63	0.90	67	3.04	72	3.27	0.74
Other disorders of the breast and any other diagnosis	205	8.11	172	6.77	0.09	117	5.31	112	5.08	0.73

TABLE 5. Number and percentage of women with one or more claims for the following procedures in the 24 months prior to entry in the NBSS by whether randomized to mammography or no mammography arms of the NBSS

	Age 40–49 years					Age 50–59 years				
	Mammography (*N* = 2528)		No mammography (*N* = 2540)			Mammography (*N* = 2204)		No mammography (*N* = 2205)		
	N	Percent	*N*	Percent	*p* Value	*N*	Percent	*N*	Percent	*p* Value
Cyst aspiration only	20	0.79	18	0.71	0.63	8	0.36	8	0.36	0.99
Needle biopsy only	6	0.24	2	0.08	0.25	2	0.09	2	0.09	0.69
Other biopsy only	14	0.55	10	0.39	0.41	10	0.45	10	0.45	0.82
Excision of breast lesion only	3	0.12	1	0.04	0.69	3	0.14	1	0.05	0.31
Xeromammography only	72	2.85	69	2.72	0.74	48	2.18	49	2.22	0.92
Multiple procedures	31	1.23	29	1.14	0.79	21	0.95	16	0.73	0.33
Any breast investigation	145	5.74	129	5.08	0.29	92	4.17	86	3.90	0.54
Cyst aspiration + any other procedure	36	1.43	27	1.06	0.26	12	0.55	13	0.59	0.84
Xeromammography + any other procedure	94	3.72	88	3.46	0.54	58	2.64	61	2.77	0.86
Other biopsies of the breast + any other procedure	30	1.19	32	1.26	0.99	31	1.41	21	0.95	0.09

The authors concluded:

> "*Using data external to the NBSS for Manitoba participants, the study found no definitive evidence to support a nonrandom allocation of women with prior breast disease to the mammography arms of the study. However, generalizability to the other NBSS centers cannot be assured*".

Clinical breast exam prior to randomization

Based on the interview with [redacted] clinical breast exam was done prior to randomization because one concerned that knowing whether a woman would or would not have a mammogram, would influence the result of the clinical breast exam. The nurse would for example, not report on an abnormality if the woman was randomized to the screening arm.

> "*Why did the randomization occur after the clinical exam? For a very, very obvious reason….That reason was, there was curiosity at the time of designing the trial, to know what conclusions the clinical breast exam would yield in the absence of knowledge that a mammogram was about to follow, because quite clearly, if I'm a nurse and I'm uncertain about the finding, and I know the one's going to get a mammogram, and then I think, I don't want to say anything about it…So that was bad news for the study*."

References

1. Cohen MM, Kaufert PA, MacWilliam L, Tate RB. Using an alternative data source to examine randomization in the Canadian National Breast Screening Study. J Clin Epidemiol 1996: 49:1039-44

2. Goel V, Cohen MM, Kaufert P, MacWilliam L. Assessing the extent of contamination in the Canadian National Breast Screening Study. Am J Prev Med 1998; 15:206-11. doi.10.1016/s0749-3797(98)00074-9.

3. 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;147:1459–76

4. 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;147:1477–88

Quality

The quality of the mammograms was claimed to be poor, even for the standard at the time of the CNBSS (1980-1985/6). The interviewees said:

> "*There's no question but they use poor equipment even for that time, and I think Professor Yaffe probably wrote that it multiple points."*
>
> "*It was a time when technology was changing and improving, and in Canada. And particularly at the sites in this study, there was a lot of old equipment that did not have the features that would have been considered state of the art as of 1980. That did not have all of the features that really would be required for high quality mammography.*"
>
> "*The facilities were told to do mammography the way they had been doing it, and no special effort was made to either upgrade their equipment, train the radiologists in interpretation or train the technologists and positioning. They just wanted to do it the way it was."*
>
> "*Dr. [redacted] [redacted], consultant radiologist in CNBCC] had criticized the quality of the mammograms.*"
>
> "*By the time I got there that they were they, they were not paying enough attention to the mammography part of it. If they had bought new equipment and sort of stopped the trial for six months until they got everybody up to speed, they probably would have come up with in retrospect, a different result."*

The panel were not able to get access to mammograms, nor to the equipment used in the trials.

British Colombia 1993

Dr. HJ Burhenne, a radiologist wrote a letter to the editor in the Canadian Medical association journal in 1993 (1):

> "*The NBSS was well designed, but there were problems in the quality of the mammography. Dr. Cornelia J. Baines was quoted in Toronto's Globe and Mail (Nov. 14, 1992) by Paul Taylor as saying that "the equipment used in the study was state of the art at the time." When the NBSS centre in British Columbia started up in 1983 it used an 11-year-old machine that was not replaced until late 1986! Dedicated film processing was not required. Poor mammographic technique does not permit scientific evaluation of the yield, and it results in late detection of cancers and greater node involvement at the time of diagnosis*".

We had eye-witnesses saying the quality was terrible and that the equipment was not up to the standard:

> "*Their [redacted]) quality of their equipment was like* ***horrendous****, and it actually been sitting in a garage, and they had been brought out of the garage before they installed it*."

> "*Professor [redacted] and I looked at the equipment at one of the centers, and found that it was operating poorly. It wasn't properly set up to do high quality mammography.*"

> "*And particularly at the sites in this study there was a lot of old equipment that did not have the features that would have been considered state of the art as of 1980."*

> "*Terrible positioning. One third of the breast is not on the film*."

> "*Don't you feel ashamed about this? What are you doing? This is terrible. This is crimey medicine*"

> "*I believe that there were, I think, fifteen different centers that were doing it, and I think the last two or three that got added to the centers probably started with better equipment than any of the other ones did."*

One of the mammography technicians did not have any comparison at the time, and it was not until she got educated she reflected that the quality might have been suboptimal in the CNBSS.

> "*But was that the state of mammography all over? I wouldn't have had a lot at that point, to compare it to, because it wasn't until after I left the [redacted], that I started doing a lot of education and conferences and meeting lots of different people."*

Pragmatic trial

Another interviewee pointed to the fact that the trail was a pragmatic trial and that the trial used the equipment available at the screening centers at the hospitals.

> "*Well, the equipment actually didn't belong to the trial in any sense. It belonged to the hospitals that we're participating, so they would have updated equipment as it became desirable*."

We asked if the hospitals would give poor equipment to the trial. The interviewee did not think so.

> "*No, I think that's a question that the external experts in radiology would have focused on. But I don't know what they were finding*".

Potentially biased views

Another interviewee had heard that quality was low, but was wondering if these claims were true.

> "*I know what's been written that it was lower quality. But I'm not sure that people saying that are not biased."*

We asked why the interviewee thought the people who claimed that the quality was poorer claimed that?

> "*Personally I think they say it because they don't like the results of the trial. I mean, you know they are all radiologists. This is their profession. This is what they do. It's hard for them not to be biased*."

Ridiculous accusation

Another interviewee thought it was a ridiculous accusation and that the quality of the mammography interpretation, the technique and the equipment should be reflected in tumor size of breast cancer detected.

> "*Yes, this is another ridiculous accusation. The standards of the mammograms being done in the Canadian trials were exactly what you expected at that point in time, and they found smaller tumors than in the two county study, so I can't see a problem*".

> "*The average tumor size was bigger in the two county trial, and yet they claimed that they had a huge effect on breast cancer screening which they did not have in Canada. And this is one reason why some people suspect that this trial is fraudulent*."

When we wondered if equipment might have different in the NBCSS and the two county study, the interviewee responded.

> "*It doesn't matter. It doesn't matter the slightest bit when the size of the tumors you find is actually small, then you are doing a good job. It's very easy*".

Tumor size

Tumor size of breast cancers is an indirect measure of quality, and size were comparable to other trials.

> "*Now, the indirect information: All you can expect of mammography, really, since it's a imaging technology focused on one organ, the breast, all you can expect is, how good it is to pick up a primary lesion that's in the breast. And so what you can go on, is the indirect parameter of quality by the size of the lesions that's picked up, and I do know from again published papers, that the size of the lesions detected by mammography in Canada were at least the match of some trials, many other trials that reported size in that era*."

> "*I don't believe that the failure of the NBSS study was because we didn't use up to date mammography. I believe the paradigm of the early detection is false. It is that the breast*

cancer will have spread by the time it's diagnosed. Because program detection, probability does not reduce women's mortality. "

Effect of poor quality

Some interviewees thought that the quality must have been poor because more advanced cancers were found in the screening arm and there was no or a very small effect of effect of mammography screening.

"*If the image quality was terrible that wouldn't have brought down the advance cancer rate compared to the advanced cancer rate in the control group. As if nobody had ever looked at the mammogram, you understand. But then there was even more problem. There were more advanced cancer. Now that smokes. You have to go to the bottom. Is it randomization? Is it contamination. Is it both? My humble opinion is both*".

"*No, I think the conclusion that can be drawn in retrospect. This is in retrospect. So yeah, understand that, you know we're always very smart in retrospect. The conclusion that can be drawn in retrospect is that the reason why the Canadian study at least the forty to forty-nine arm failed I mean, basically it showed no benefit. And we know that's a failure. Now, the reason that it failed primarily, at least from my point of view, is that the image quality made it fail."*

Reference

1. Burhenne HJ. Canadian national breast Screening Study. CMJA 1993;148:875-883

Comparison with other trials

To understand the quality of the mammograms and interpretation, we asked the interviewees how the trial quality was in comparison with the standard at the time and with other randomized controlled trials done in the 1970s to 1990s.

Swedish studies

A radiologist who had been training in the US, and had visited mammography screening sites in Sweden said this:

> "[The mammograms were] *I've seen mammograms from several of the Swedish trials. I've seen mammograms from the Canadian trial. I see mammograms from the Age trial. They're better, but they're later because the technology was better. I've seen mammograms from maybe not every single trial, but most of the trials. Aside from the HIP trial, which was even worse. But it was ten years earlier. Aside from that, the Canadian trial was far poorer than any of the other trials. By almost an order of magnitude."*

The panel was not aware of any publication of review of the Swedish trials and asked if there were any publications we should be aware of.

> "*It was my understanding that there was no image quality studies done of the Sweden trials"*.

[redacted] claimed the quality must have been much better in the Swedish studies then in the CNBSS.

> "*They certainly were not in the same league as what was being done in Sweden at the time. They were using the medial-lateral-oblique view, and my guess was I didn't visit those sites at the time, so I didn't actually put my hands on the equipment, but my expectation would be that this would be modern, at the time, equipment."*

Other trials

A conversation about the quality in the other trials followed:

> Panel member: "*So the mammograms that you saw from the Swedish trials, and since nothing is published to our knowledge, it's very hard for us to take that as a fact, so we need to know a little bit more.*
>
> Radiologist: "*I went to Sweden. And I saw them there. Because I wanted to get my mammography as good as theirs was, because theirs was the best in the world at that time."*

> "*The New York City trial. That's different. The New York City trial was done ten years before, and it was done at a time when there was no mammography. Basically, in clinical practice there might have been five people in the United States that were doing mammography. There was none, and they didn't know how to do it. They just did the best they could. I have no idea how they ever found the breast cancer in any of those patients, those images of virtually all white. Nonetheless, it was done at a time when, even there was no clinical breast exam. There was certainly no self-breast exam, and there was very little clinical breast exam*".

The quality of the HIP trial was also remarked by another interviewee.

> "*That that was not true of the study* [redacted] *was in. I didn't even know she was in the HIP trial* [New York] *where I found out, much to my huge alarm, that many of her mammograms were being taken by a standard x-ray machine. I can't speak with as much confidence about some of the other trials that we're going on in that era, because the procedures, at least in the literature, to my knowledge, are not quite as well described."*

> "*I would say, beyond a reasonable doubt, that the mammography in Canada was far better than in the HIP trial, which was the first randomized trial. I do know that there were reference radiologists, reviews of the parameters and the mammograms."*

[redacted] claimed the quality in the HIP trial was far better than the CNBSS

> "*I visited their place in New York. And I tell you their secret. Their mammograms were not at all as bad as in Toronto. But they were bad enough to be able to, well, miss, you know what I mean, one centimeter cancers*."

We wanted to clarify this, as we had the impression the HIP trial used regular X-ray equipment and not dedicated mammography devices.

> "*May I quickly intercept here? The HIP trial, obviously this was early birds, did not have dedicated mammography devices, so they use just regular X-ray devices. Correct?"*

However, the radiologist replied that if you manipulated regular X-ray equipment it may be fairly good.

> "*Yeah. But you see, you can manipulate the regular X-ray device by putting in filters. With the skillful engineer, we could have softened up somewhat the beam. In addition, we use the cone to cut out the scattered radiation. You can get very good image quality if you're smart enough how to rebuild the machine.*"

Comparison to other trials

One interviewee, a radiologist, did not think that comparison to other trials mattered.

> "*I have to say I'm not an expert in the other screening trials. And again, I would suggest you that that's irrelevant, because what that is asking is, did Professor Miller and Dr. Baines do anything differently than the other trials. And the answer is, that's not the question. The question is, what they did. Did that subvert the accuracy of the trial?*"

The panel tried to explain why we though these questions were relevant.

> "*No, your point of this being irrelevant. I think it's important just to agree to disagree on this aspect. The point here being, it may be that if we start to look into all of those trials, there might have been, obvious challenges, and it could be if the two of us, in a non-data driven way, started to pick out all the trials and look at it. What we would end up with is a complete lack of evidence that is credible for mammography*".

The discussion continued. The radiologist asked if we are nowhere with the evidence for mammography screening right now.

> "*So your argument is that the trials that showed a benefit all have problems, so we are nowhere at the moment. Is that it?"*

The interviewer argued that other trial may have had problems that we do not know about because the other trials were not scrutinized as much as the CNBSS.

> "*That the trials showing benefits have a lot of problems, too, and some of the problems have been discussed. Other problems we never really, might never know, because they have not been scrutinized as much as the Canadian trial*".

The interviewee argued that it is the CNBSS that are under investigation, not other trials.

> "*I apologize again for interrupting, but I would argue that's not the point. The point, if we're asking, did other trials make mistakes, and so on. That's fine. I'm happy to discuss that in another setting. But we're here, I think, to determine whether the Canadian national breast screening study data are reliable and should be used to advise women, or unreliable*".

Context

The interviewers were concerned that if the result of the CNBSS would have been different, there might not have been any claims of poor techniques, so context matters.

> "*We don't know if the same criticisms wouldn't have come up if the Canadian trials had been showing a significant result in favor of mammography. We don't know. But you see, that's the point there. So it is important to see the context, too. We all acknowledged that, you know, if we make a statement about this trial, we need to acknowledge the context*"

The interviewee did not agree and was concerned that women might die as a consequence of not being offered mammography screening.

> "*I would argue the context doesn't matter. I would argue the results and how they got there matter. Because the results are being used to advise women about whether or not they should participate in screening. And they're being told to wait until fifty. And there's all kinds of data that says you're gonna at least in my country, 100 000 women will likely die if they wait until starting screening at age fifty. So this, you know this is not just an intellectual exercise.*
>
> *The point is, should the Canadian National breast screening study data be withdrawn because they were corrupted by a non-random allocation. You know, I didn't think we were even going to talk about the quality of the mammograms, because you know it's like `Well, that was the mammography in Canada, in the eighties´, and that's the way it is.*
>
> *And most people don't even know what a good quality mammogram is, and that's been shown over and over again that the quality was terrible. No one has come out and so this was state of the art, mammography, and so on*".

Inclusion of "old", trials in guidelines

The panel discussed which trials should be used as evidence when making guidelines, and whether the quality of equipment and how different it is from today's standard it can be to be included in the evidence base.

> "*Yes, this is discussed quite fully* [in the guideline panel]. *To what degree, you know, changes in a lot of the equipment. We didn't have concern that this trial stood out comparing the others. That was one part of the discussion.*
>
> *There are two sets of discussions that we had quite lengthily, that was this trial different from what it should have been, or what others did in the time they're in? The answer was, we did not believe that that was It.*

Considering standards at the time

> *About standards and quality over time. It seems odd to me to be adjudicating what the standards are in the odds. You're asking me these questions. It's an odd query to be going over what kind of mammogram should have been used in the* 1980s".

Since the claims are that the mammograms had so low quality we need to understand what the standards were at the time and how it compares to today's standard, or what should then the comparison be?

"*I'm also, curious about that. If they were giving psychiatric drugs in the 1980 we don't use any more. We're not having integrity hearings about those trials.*"

[redacted] a radiologist did not support the BNBSS to be part of the evidence base for guidelines.

"*So I thought we were really just talking about the results. Is it legitimate to advise people based on the results and based on how randomized control trials should be run? They messed it up and the data are compromised.*"

Mammography views

The CNBSS used two mammographic views (mediolateral and craniocaudal) during the whole study period. They switched view from mediolateral to mediolateral oblique in around 1985 (1). In a review of the technical quality of CNBSS mammograms, mammograms were scored according to quality from 1980 to 1987. The largest increase in quality scores was when the view change from mediolateral to mediolateral oblique (1).

The CNBSS were criticized for not using the mediolateral oblique view.

> "*And the oblique was first described, I think, in 1976 by Lungren* [2] *So it was out there. Yeah, the two county trial* [redacted] *was out, you know, talking about it*".

The two county study of mammography screening used one view, the mediolateral oblique (MLO). [redacted] when asked what would be better one view mediolateral oblique or two views mediolateral and craniocaudal, in detecting small tumors, he replied two views.

> "*Oh, my God! It must be two views. As soon as we publish* [redacted]*, I really forced the Gothenburg trial designers of two things: Shortened the interval for God's sake, and definitely two views. Right?* "

This was confirmed by [redacted] (radiologist)

"*In the Swedish trials they were using the oblique view without the cranial-caudal. They only used one view, at least in the beginning. I don't know if they started doing the cranial caudal view during the trial or after the trial was over, and now they use both. That's because it's better to use both of them than to use just one, but they were using only one*." (ES)

[redacted] explained they used the one view because of concerns about radiation at the time. The two county study started in 1977.

> "*John Bailar III, genius, made a mistake in calculation in his Lancet at article which he published minutes before the National Board [Swedish National Board responsible for the two-county trial] decided to run the trial. And the conclusion was that radiation kills more women than that mammography ever can save. And of course, that made the National Board extremely scared."*

We asked why the mediolateral oblique view was not used in the CNBSS and the explanation was that it was not the standard at the time.

> "*In Canada? The standard was the true lateral and the caudo-cranial views. That's what everybody was doing, and that's why that was done in Canada because they wanted, their goal was to test current practice. In the United States there was an ongoing shift*

> *from the true lateral view to the ML view. It was gradual. It didn't happen overnight. In the more experience practices, it had already happened even before the trial began*".
>
> "*It was evolving much faster* [replacing ML with MLO view] *in the United States than in Canada. That's probably one of the reasons why the external reviewers were much more critical than the internal reviewer.*"

Dr. Baines confirms this in her paper in Annals of Internal Medicine in 1994 (3):

> *The 1988 standards required a mediolateral oblique view. This was unfortunate because between 1980 and 1984 the NBSS protocol required two-view mammography, including straight mediolateral and craniocaudal positioning, a decision determined in consultation with U.S. and Canadian expert radiologists before the initiation of the NBSS in 1980. Ironically, the director of the NBSS urged at that time that the mediolateral oblique view, already being used in the Swedish trials, be used. The radiologic consultants insisted on the straight mediolateral view because it conformed to contemporary North American practice.*
> *In 1985, when the screen-1 examinations were completed, the Policy Advisory Group formally approved a change in positioning to mediolateral oblique, although at least one center had implemented it in 1983.*

References

1. Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders D, Sickles EA, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. Am J Roentgenol. 1990;155: 743-7
2. Lundgren B. The oblique view at mammography. Br J Radiol 1977; 50:626–628
3. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334.

Review of the technical quality

The quality of CNBSS mammography had been questioned and it was decided to have a retrospective review of a representative sample of NBSS mammograms. In 1990 the review was published (1).

Three external radiologists, who had had no prior involvement with the CNBSS reviewed 830 mammograms. The sample included 10 randomly selected mammograms from each of 15 screening centers for each calendar year (1980 and 1987).

All mammograms were reviewed in randomized sequence by each reader and rated 0-3 for each of four criteria on positioning and image quality: craniocaudal positioning, straight mediolateral or mediolateral oblique positioning, contrast and density, and image quality. At the request of the reviewers, a fifth "global" score criterion was added. The total possible score was 0-12 (1).

1988 standard

The reviewers decided that all films would be judged by "1988 standards": all lateral views should be oblique, visualize substantial amounts of pectoral muscle and the axillary tail of the breast. The oblique projection, however, was not used routinely by most U.S. mammographers even in 1983 (2).

1980 when the NBSS was initiated, the protocol required straight mediolateral positioning. It was not until 1985, when all first screening examinations had completed, that a change to mediolateral oblique positioning was made, following a recommendation by the policy advisory group that monitored the conduct of the study.

This is how Dr. Baines summarized the review in 1994 (3):

> *The evening before the review, the two reviewers said their participation was conditional on being allowed to rate all mammograms by 1988 standards: The 1988 standards required a mediolateral oblique view. This was unfortunate because between 1980 and 1984 the NBSS protocol required two-view mammography, including straight mediolateral and craniocaudal positioning, a decision determined in consultation with U.S. and Canadian expert radiologists before the initiation of the NBSS in 1980.*
>
> *Ironically, the director of the NBSS urged at that time that the mediolateral oblique view, already being used in the Swedish trials, be used. The radiologic consultants insisted on the straight mediolateral view because it conformed to contemporary North American practice. In 1985, when the screen-1 examinations were completed, the Policy Advisory Group formally approved a change in positioning to mediolateral oblique, although at least one center had implemented it in 1983 (*3*).*

Drs. [redacted]

This is what Dr. [redacted], who was one of the radiologists who reviewed the mammograms, said:

"*Baines organized a review of their image quality that [redacted] was invited with Dr. [redacted] and [redacted], who was a [redacted] radiologist. [redacted], was, is at the University of [redacted]. We were invited up to review the quality of the mammograms, which at that time was a major concern. It still is.*

But that review showed that the quality was poor, unacceptable for much of the trial, but did get better over time.

I'm not aware of any expert in breast imaging and reading mammograms who has supported the quality of their mammograms.

When we reviewed what she had set up, and there was debate about what we were going to do. The medio lateral oblique had not been used, and she said, `Well, that's not really fair, because most people weren't using the media lateral oblique at the time´, and I think that's correct. Certainly in Canada it was correct. That didn't mean you were going to miss cancers by not using it, so we, I think we agreed.

At the time we would do two ratings. One would be the quality of the mammograms as they were done. And then, we could put a second rating based on what we thought the absence of the medio lateral oblique might have meant.

She insisted that we judge them, you know, based on what had been done as opposed to what should have been done.

And that [review] *showed that the quality was poor to unacceptable. For certainly the first two years of the trial, and it got better, but it was never, it certainly wasn't state of the art, the state of the view medio lateral oblique, with grids, and the proper processing*".

[redacted] a radiologist who was part of the Policy Advisory group and one of the radiologists who reviewed the mammograms. The Policy Advisory group was according to [redacted] "*put together by the agency that funded the trial, which was the National Cancer Institute of Canada. They were the ones who approached me, not the investigators*".

"*My experience throughout the several years that I was involved was that were very, very resistant to any suggestions for change". "I tried to give them the benefit of the doubt, even though they were very hard to work with. But I yeah, I've sat on these review panels many times, and usually when criticism is offered, it's accepted. And there's an attempt to improve, because when you know when you're on a review panel, what you're trying to do, if you're trying to salvage it, you're not trying to destroy it. These people just weren't interested."*

> "*Dr. Miller and Dr. Baines, who were the PI and the CoPI, they somehow convinced the NCI that they had to be the first authors, and that the people who did the study couldn't be the first authors and write the paper. This is politics" "The paper was written with a lot of spin. It made it look as good as it possibly could. In English, we have a term called sugar coating*."

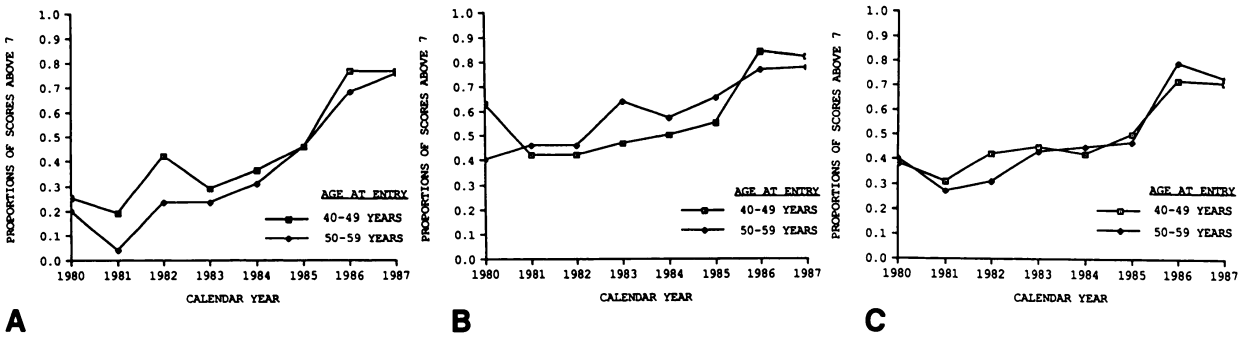
In an accompanying commentary in the same issue of the American Journal of Radiology, Dr. Kopans, who is also a co-author of the review (4), published his own perspective "The Canadian Screening program: A different perspective" (4).

> "[*A*]*lmost 50% of the mammograms obtained during the first 2 years of screening were judged to be unsatisfactory, and it is not until the final 2 years that satisfactory image quality was achieved in over 70% of the screenings. The paper suggests that the initial low scores were primarily due to the failure to use the then "new" mediolateral oblique projection. Two of the reviewers, in fact, felt that it was the sharpness, contrast, and overall quality of the mammograms that was judged to be poor*".
>
> "*What has not been revealed, however, is that at least two of these advisors, recruited during the early years of the trial, resigned because one was not even permitted to view the images, and the other's recommendation to improve image quality was not heeded earlier (W. Logan and S. Feig, personal communication)*".
>
> "*The decision to use the then "current" level of mammographic quality as a measure of efficacy was ill conceived*".

Results
Although Dr. Kopans published his own perspective, he did not discredit the findings of the review summarized in Figure 29.3 (Figure 2 in (1)).

The largest increase in scores was associated with a 1985 protocol change in which mediolateral oblique positioning replaced straight mediolateral positioning. The technical quality of NBSS mammograms improved over time; 49-75% of mammograms scored 2 or higher (satisfactory) between 1980 and 1985 for contrast and density, and image quality. The proportions of mammograms with satisfactory scores rose to 85-89% in 1986 and 1987 (1).

Figure 29.3 percentage of scores higher than 7 for radiologist A, B, C by age of inclusion (40-49 and 50-59) in CNBSS and calendar year (fig 2 in (1)). The maximum scores were 12



Interobserver agreement were obtained in three combinations (observer 1 with observer 2, 2 with 3, and 1 with 3) for the five criteria rated. Kappa statistics were used to determine the amount of agreement between the radiologists rating (kappa of 1 is full agreement). Kappa indicated not more than fair interobserver agreement. The highest kappas, 0.32-0.37 ($p < .0005$) occurred for the criterion of mediolateral oblique positioning. Agreement ranged from 41.7% to 54.1% for the ratings on the craniocaudal view; from 53.8% to 58.3% for the mediolateral view; from 45.6% to 47.6% for contrast and density; from 42.5% to 48.5% for image quality; and from 54.8% to 59.4% for global score (1).

However, the percentage of intraobserver agreement was higher than the interobserver agreement. The reviewers reviewed 44 cases two times, and kappa ranged from 0.50 to 0.79 for the global score and from 0.47 to 0.61 for the mediolateral oblique criterion (1).

References

1. Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders D, Sickles EA, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. Am J Roentgenol. 1990; 155: 743-7

2. Bassett LW, Gold RH. Breast radiography using the oblique position. Radiology 1983;149:585-587

3. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334.

4. Kopans DB. The Canadian Screening program: A different perspective AJR 1990; 155:748-49

Offer to review the quality of the mammograms

We interviewed [redacted] that told us he was invited by Dr Miller, and [redacted] he was asked to review 50 cases:

> "*I was so shocked. I mean you could hardly decide whether you were looking at a skull or a breast." "I'm going to be kind of very straightforward. Terrible positioning. One third of the breast is not on the film. Uh the processing. I tell you the diagnosis, this is ninety, second processing, which means: Be careful. This is a single emulsion film. Chest. X-ray is double emulsion. That that's all right for ninety seconds. But the single emulsion demands three minutes processing. So if you just dip the film into the processor and pull it out, that's the result. You see, when I mean. They see the bromide is still married, in the grains. So you understand.*
>
> *This is terrible. This is crimey medicine*.
>
> *I have pictures about it."*

We asked Dr. [redacted] if he could send us the pictures.

> "*It would be very interesting if you could send us those pictures*".

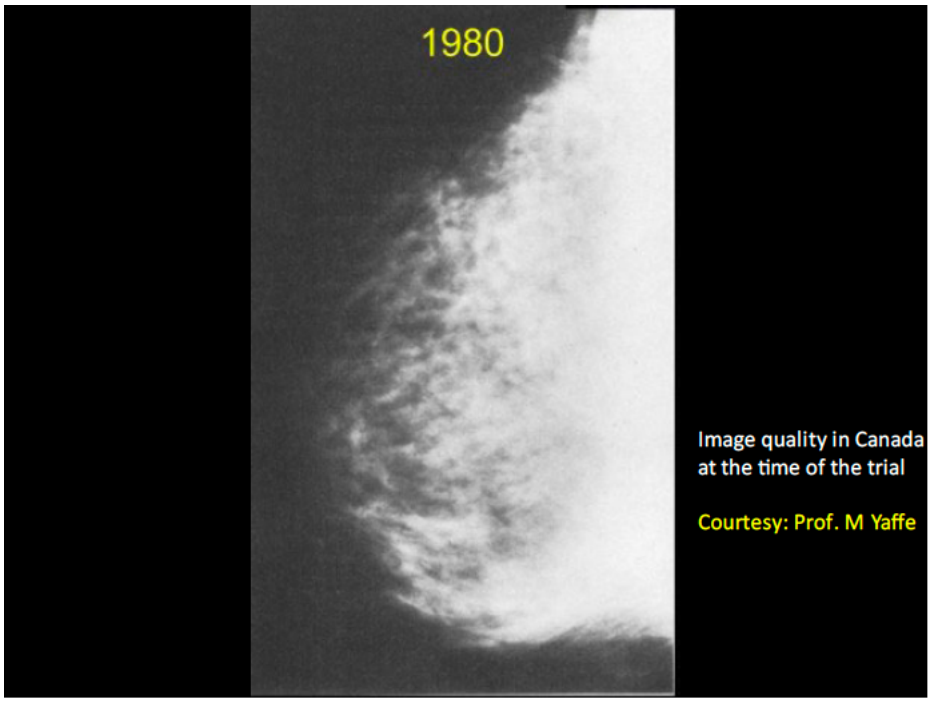
He responded he would send us pictures

> "*Very well. I will send you*".

We did not receive any pictures documenting the review, as [redacted] said he would provide. The panel were sent additional material that the panel asked for [1], including an email from [redacted], where a picture was included (Figure 29.4). However, this picture is not from the review [redacted] describes, but we have no other information about the origin of the image.

[redacted]

Fig 29.4 picture from an email sent form Dr. [redacted] (provided to the Panel by [redacted], Nov 16 2022)



[redacted] confirmed that there had been an offer to help

> "*Well, that's basically you just said it. They had offered to help. I think these were very motivated people who wanted to see the trial, do the best it could, and if there was an advantage, a mortality advantage. They wanted to maximize the ability to demonstrate it, and they realized that, having high quality, mammography was essential. So they were, they were, and remain, as you know, very committed to trying to do that well, and I think they offered their help. And in most cases. It was not accepted. I think it was perceived that it was interference with the trial.*

Dr. [redacted] had not heard of this:

> "*Do you mean somebody offered to train them for free, and they said, No, I'm not aware that that happened. You can't train people for free. It costs money, and if they had no budget to do it, because it was all committed to other aspects of the trial, they would have had to ask for more money.*"

[redacted] and [redacted] have no recollection of the review [redacted] describes. Both of them were specifically asked [redacted].

[redacted] said

> "*Oh, we do not know, and I did not think that happened. We do not think that it happened. No*".

Our conversation with [redacted] was as follows:

"*Can you give me the date of that meeting that he discusses?"*

Panel:

"*I don't know. I can't remember what date it was. I don't think he gave us a specific date. I think it was around 1986".*

Dr. [redacted]:

"[redacted] *was never invited to review. By 1986 I knew that he could not be trusted. I knew that he could not be trusted because he was pretending he had no excess mortality in the* [redacted] *paper.*

What we wanted to do, because his system of interpreting mammograms was brilliant, and, of course, much better than many of our regular radiologist did. We wanted him to show how to read a mammogram, and I wanted him to get that across to our study radiologists.

Now, probably what we did, the mammograms that we used for him to teach from. I don't remember this, but it may very well be, that they were NBSS mammograms. But we were not asking him to review mammograms. We were asking him to show our radiologists how he interpreted them.

What he is describing is totally non-existent in my head.

Reflections of CNBSS and other screening trials

CNBSS stood out because it had pre-randomization consent, use individual randomization, and had published more detail than any other trial

Some of the interviewees expressed that the reason for this review might have been because the CNBSS reported on many details and aspects

> "*It stood apart from some of the trials because it was pre-randomization consent. So, in other words, everyone who was randomized had consented to go into the trial. That's not always been true so, for example, in the first randomized mammography trial, the HIP Trial, as I recall, that was randomized from population registers and the control group didn't know that they were on a trial at the time. So this trial avoided the bias that's inherent in in that in that mode, and other trials like the two county trial, or I say, two county trials, were a population community based randomization. So there were cluster randomization. The same was true, I believe, for Edinburgh, so it was different in that regard. But it was the same as the age trial in the UK, and that there was individual randomization. The details of the CNBSS have been published, and more depth have been published more frequently and in actual more detail than some of the other trials. So, I couldn't comment on some of the other trials.*"

> "*Well, I just I think that, I don't really understand why this trial was so scrutinized when it to me, it was done better than all the other ones, as far as the randomization, doing the outcomes, you know, following procedures for consent. It just seems kind of mean to be honest."*

Transparency

One interviewee said the CNBSS were not transparent enough.

> "*I don't think there was enough transparency."*

But others contradicted this.

> "*I've been impressed particular with what Dr. Baines is written over the years and the time she's taken in this, and with just a level of transparency and conscientiousness, that she's gone through this, I mean, I think she's been much more willing, than most of us trialists, to actually openly go through.*
>
> *Going through and documenting the level of detail, she documents each of the aspects of the trial and the records that were kept*.

I find it difficult to believe we'd find any other trials not only having done it, but could do this. Could even to do this this kind of thing and, she's 30-40 years prior to what we're trying to get people to do with transparency today.

I mean this is this is not anywhere near the kind of things we look at when we're looking at maleficence in trials."

"*It's curious that we're not doing integrity hearings on every trial that uses out of date medical procedures. It's important to note that. I wonder if the University of Toronto is planning on going back through all these trials. Judging the utility of old trials or the applicability of old trials is not a research ethics question."*

Conflict of interest

"*And much more important the groups were highly comparable for a pretty large number of factors, so I can't see there is any hint of any problem here. I think it´s all about money and conflicts of interest. It's a despicable trial, I´m participating in right now, it should never have happened*".

We were wondering what the interviewee meant about "trial" and asked

"When you say "trial", you need to review that we are performing?"

The interviewee responded

"*Yes, it seems to me to be some kind of show trial".*

"*Well, what it means if you arrive at results that the mammography screening lovers don't like, then you are in for serious trouble, and this is what we see exactly here*".

Another interviewee claimed the criticism is mostly from radiologists.

"*And I also find it, the fact that radiologists are criticizing it. You don't have radiologists deciding what the data are and the guidelines. I believe that radiologists are good at talking about the quality of the mammography or the equipment, but do not have any methodologic training. They are not trained in epidemiology. They have no knowledge of methods or statistics, and so I find it quite odd that they are criticizing a trial when they don't have the credentials from my perspective to actually criticize it*."

Result driven issue

"*I do think that this is a results driven issue. In fact, you know, I could comment on this at the end, but you know we've had experience with this group. I think there's no boundaries here in terms of what goes on with, Dr Yaffe, Dr. Seely, Dr. Gordon, and a couple of others*.

I mean, this is a controversial topic [breast cancer screening]. And for whatever reason there's been a small group of people who've been quite dogged about this over the years, and have lashed on to it. I guarantee that if the CNBSS had found that there was a reduction in mortality due to screening, that no one would be raising any of these questions here."

"*You know mammography is a, it's just kind of an okay test. So to have a negative result is not a surprising thing. It's just kind of an okay test.*"

"*CNBSS was a high quality trial with valid results, not only for the era in which it was conducted, but for any, you know, for a long time. [N]one of the trials have today's adjuvant chemotherapy or treatment of cancer, and that is a deficiency of every single trial.*

The better therapy gets, the more it's going to mitigate the effect of mammography until therapy is perfect, and then mammography becomes completely irrelevant." ([redacted])

Misinformation

"And so we spent a lot of time managing misinformation. There'd be letters would go out to government officials. I spent a huge amount of my time managing this misinformation. It's interesting because there' be articles that go out with this group with sixty radiologists signing it. But the Canadian partnership against cancer has two hundred, some experts on it, nobody signed that letter. No epidemiologists signed that letter. But there is ongoing, you know misinformation."

Other reflections

Other reflections of the CNBSS were also made:

"*You are chairing a panel over one of the biggest shames in medicine in the 1980s period.*"

"*Everything we think we know about breast cancer is wrong, and you guys are just fiddling around, we are quibbling over the details, over the fundamentals. And this is why I don't like to get involved. You're quibbling over a bunch of details like a bunch of sixteenth century…"*

"*I do wonder. You know I again, I think you know, someone showing up in a room 35 or 40 years later, and saying, Oh, gosh, I'm a mammography technician who happens to attend ten academic talks 35 years later, and I think I saw this now, I'm the smoking gun*."

Summary of previous critique and previous review

The panel has reviewed the critique of the CNBSS, and has identified three issues: randomization impairment, poor quality of equipment, and mammography and improper analysis and interpretation of results.

A more detailed review is presented in *Chapter 9, 10, 11, 12*.

Previous criticism

We have summarized the criticism in tables 30.1 (1), 30.2 (2) and the rebuttal in tables 30.3 and 30.4 (3). Critique brought forward prior to the review by Drs. Bailar and MacMahon in 1995 (4) and perhaps prompted by the review is summarized in tables 30.1 and table 30.2 (1, 2). Dr. Cornelia Baines' responses to the review are summarized in table 30.3 (3).

Tab 30.1 Criticism raised by radiologists in the American journal of Radiology in 1993 (1)

Criticism by Kopans and Feig	**Arguments by Kopans and Fig**
Incomplete follow up in control arm	Unclear how complete follow-up was
Women with symptoms should be excluded	Dilutes effect of screening
Late stage disease disproportionately assign to screening arm	
Lack of statistical power	Lower breast cancer mortality in trial than in the Canadian population
Contamination	26% of unscreened women in the control group had mammography outside the trial
Diagnostic mammograms are screening	All mammograms screen the breasts
Too short follow-up time	Effect is expected beyond 8-10 years
Poor quality of mammograms	Deficiencies in mammographic technique and interpretations Use of any mammographic equipment available No specific training Poor quality system
Suboptimal mammographic intervention	42% of tumors were missed at screening

Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. Am J Radiol 1993;161:755-60.

Tab 30.2 Criticism raised in Radiology in 1993 (2).

Criticisms	**Arguments**
CNBSS not generalizable to the Canadian population	Volunteers (more educated, fewer children, lower prevalence of smoking)
Randomization	Local Allocation was not random More women aged 40-49 with advanced stage breast cancer in screening arm, chance an unlikely explanation
Quality of mammograms	Not state of art even for the time
Lack of statistical power	Lower breast cancer mortality in trial than in the Canadian population
Compliance	No allowance for less than full compliance in sample size calculations, compliance was 86-87%

Contamination	26% of women aged 40-49 years in the control group had mammography and 17% of women aged 50-59 allocated to physical examination, had mammography No allowance for contamination in sample size calculations
Co-interventions	Treatment, no treatment differences
Follow-up	Not stated how follow-up procedures captured health status of participants
Sample size estimates	Expected mortality rate was too high, No allowance for contamination or less than full compliance Expected too high mortality reduction (40%)

Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ. A critical appraisal of the National Breast Cancer Screening Study. Radiology 1993;189:661-3

Response Dr. Baines

Methods

Dr. Baines explained the methods regarding recruitment, clinical breast exam, randomization including how women with lumps were handled, quality of mammograms, clinical examination, surgical- and pathological review, and follow-up procedures (table 30.3) in a paper published in Annals of Internal Medicine in 1994 (3).

Table 30.3 Methods for recruitment, clinical breast exam, randomization including how women with lumps were handled, quality of mammograms, clinical examination, surgical- and pathological review, and follow-up procedures in the CNBSS (3)

Method	**Description from Baines**
Recruitment	Women aged 40-59, not pregnant, no prior breast cancer, no mammogram last 12 months, consented
Clinical breast examination	Was done because CBE was believed to reduce mortality Examiner asked about lumps or symptoms Physical examination and instruction in self-examination Clinical findings required referral to surgical review (within 1 week). Documented on examiners form Prior to randomization
Randomization	Women individually randomized Randomization lists in 4 separate books (age 40-44, 45-49, 50-54, 55-59) Each center Center coordinator entered the date and name on the first available line in appropriate book dependent on age, and assigned the woman her ID and randomization allocation ID and allocation was entered to all chart forms "*Skipping a line to achieve a desired allocation was not feasible because she could not predict when the next appropriately aged woman would arrive to fill the skipped slot.*" All original randomization sheets were submitted to the central coordinating office where all sheets were examined for suspicious entries, inappropriate dates and lack of congruence with participant records Examiner told each woman her allocation

Women with lumps	All, irrespective of allocation arm, was referred to review clinics. No reason for examiner to allocate
Quality of mammograms	Documented in: - Baines CJ, McFarlane DV, Wall C. Audit procedures in the National Breast Screening Study: mammography interpretation. Can Assoc Radiol J. 1986;37:256-60. - Baines CJ, McFarlane DV, Miller AB. Sensitivity and specificity of first screen mammography in 15 NBSS centres. Can Assoc Radiol J. 1988;39:273-6. - Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the National Breast Screening Study. Invest Radiol. 1990;25:971-6.
Clinical examination	5 -10 min Documented in: - Baines CJ, Miller AB, Bassett AA. Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. Cancer. 1989;63:1816-22. - Baines CJ, To T. Changes in breast self-examination behavior achieved by 82,835 participants in the Canadian National Breast Screening Study. Cancer. 1990;66:570-6. - Miller AB, Baines CJ, Turnbull C. The role of the nurse-examiner in the National Breast Screening Study. Can J Public Health. 1991 ;82: 162-7.
Surgical review	Study surgeons appointed to each center If diagnostic follow-up was required, surgeons forwarded their recommendation to woman's physician Most recommendations were followed Women aged 40-49: 0.8% in screening arm and 1.5% in control arm had diagnostic mammogram
Pathology review	Reference pathologist at each center Reviewed all specimens from surgical procedures on participants
Follow-up procedures	3 or 4 years dependent on time of entering study; those enrolled in 1983-84, had only 3 years of routine follow-up High compliance with procedures Questionnaires Complete ascertainment of breast cancer and deaths After study interventions ended, passive follow-up, active follow-up of all women with breast cancer

Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334

Rebuttal

Dr. Baines' rebuttal to the critique was also included in that paper and is presented in table 30.4.

Table 30.4 Criticism of CNBSS and Dr. Baines' counter arguments (3)

Criticism	**Arguments**
Design and execution	Design dependent on available knowledge and ethical considerations: Fear of radiation Individual informed consent Inconvenience of long-term commitment to scientific study offering interventions available outside of trial Medical professional autonomy Pragmatic trial Protocol approved of the Medical Research Council of Canada, National Cancer Institute of Canada, Canadian Cancer Society, Health and Welfare Canada, other institutions participating in the CNBSS Did not achieve statistical power Policy advisory group (international representation) monitoring

	Meticulously reporting "*unmatched*" any other screening study
Randomization	Demographic variables equally matched in screening and control arm (tables 31.1, 31.2) Allocation was not subverted Similar amount of women with self-reported symptoms at entry in both arms, both age groups (tables 31.3, 31.4) No suspicious entries in original randomization sheets
Excess advanced breast cancer in screening arm	More women with advanced cancer in screening arm women aged 40-49 y seen in other breast cancer screening RCTs: Malmö, Stockholm, Two-county (5, 6, 7) Similar amount of women with abnormal findings at CBE referred to review clinics, More women in screening arm were recommended diagnostic intervention by surgeon at review clinics compared to control arm No clustering of women with advanced stage cancer in the CNBSS centers, does not support subversion in some centers Difference in proportion of women with breast cancer with 4 or more lymph nodes cumulated at year 5 is less than at inclusion (screen 1): 14.6% screening arm vs. 10.9% control arm
Review of all invasive breast cancer detected in the first 5 years:	Mean numbers of removed nodes: 11 screening arm and 10 in control arm No-lymph nodes removed: 5% screening arm and 10% in control arm 4 or more lymph nodes: 90% screening arm and 86% in control arm Indicate under-ascertainment of nodes in control arm
Excess number of breast cancer deaths in screening arm	Excess number of breast cancer deaths in screening arm women aged 40-49 y seen in other breast cancer screening RCTs: Malmö, Stockholm, Two-county (5, 6, 7) and meta-analysis of Swedish trials with 12-year of follow-up shows a 12%, not statistically significant difference (8)
Quality of mammography	Poor quality and lack of training may reduce sensitivity of mammography and lower breast cancer detection rates
Sensitivity and detection rates	Sensitivity CNBSS comparable to the Stockholm, Two-county studies for women aged 40-49 years: 81%, 53%, 62% respectively Breast cancer detection rates and interval cancer rates for women aged 40-49 years comparable to the Two-county studies Among women aged 40-49, incidence ratio comparing incidence in prevalent round with incidence in control arm was 3.25 for CBE and mammography and 2.0 in mammography alone in the CNBSS and 1.99 in Two-county study (7)
Interpretation of an external review	"*A letter documenting factual inaccuracies in the commentary is rarely cited*" (8) Review random sample of 853 of 100,000 mammograms between 1980-1987, not weighted by center recruitment Did technical quality improve over time? Reviewers blinded for center, woman age, calendar year of mammogram, 2 of 3 invited experts were aware that there were more deaths from breast cancer in screened than control women age 40 to 49 years – may have influenced their rating Technical assessment: craniocaudal position, mediolateral position, contrast and density, and image quality

	Scoring scheme: 0 (poor), 1 (fair), 2 (satisfactory) 3 (good) Quality improved over time The proportions of unsatisfactory mammograms are inflated because scoring was influenced by disapproval of the mediolateral view (not using the oblique view)
Competence of radiologist	Radiologist specialists met requirement of the Royal College of Radiologist, some were US board certified Pre-CNBSS experience in diagnosing breast cancer Audits on centers radiologist receiving frequent memoranda regarding technical quality and interpretation of the mammograms Annual meeting radiologists Invited experts attended some meetings Study radiologists received regular updates on performance indicators and cancer detection Technologists received feedback from the reference radiologist as well as from the reference physicist 40% of women were screened in centers with new mammography units when they opened 18% were screened with new units purchased after centers had opened
Women age 50- 59 y	Comparing mammography screening and clinical breast exam with clinical breast exam
Contamination	26% of women aged 40-49 years in the control group is expected due to symptoms and follow-up of symptoms and lumps found at enrollment
Analysis	Remove women with symptoms or screen detected cancers from screening arm would bias comparison with control arm
Follow-up	For 80% of the study population at the time of the analysis, 19 to 44 months had elapsed between center closure and record linkage with the relevant provincial cancer registries Any post-study under-ascertainment of breast cancer in women 40 to 49 years at entry is minimal
Use of survival	Invalid endpoint, affected by lead time

Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334

Previous review

The previous review of the CNBSS by Bailar and MacMahon concluded (4):

> "*The document experts found no evidence of a deliberate attempt to conceal the alterations. Even if there had been acts of subversion, they could only have been few in number and, given that there was only 1 death from breast cancer in the group reviewed, the alterations could have had only a trivial effect on the study findings as reported in 1992*" (4).

Tables 30.5 and 30.6 show a summary of the findings from the review of the allocation books and thinkable methods for subversion with reasons why they were unlikely.

Table 30.5 In a previous review, the allocation books revealed (4)

Allocation books revealed	
Clerical error	30 182 records were inspected, of which 467 (1.5%) required investigation. Of the 467 records 219 (47%) indicated clerical error, no change in the identity of women entered on the allocation line
Credible match	147 of the 248 women revealed a credible match.
Alterations	Discovered 162 alterations, 97 lines allocated to mammography (18% referred to review clinic; 4% dead) and 65 controls (12% referred to review clinics; 3% dead). 97/65 alterations highly unlikely (p=0.01) Explained by women allocated to mammography returned to the centre annually, more opportunity to correct and to make changes to incorrect names
Conclusion	Whatever misallocation occurred, had only a trivial effect on the results.

Bailar III, JC. MacMahon B. Randomization in the Canadian national breast Screening Study: a review for evidence of subversion. Can Med Assoc J 1997; 156:193-199

Table 30.5 In a previous review, these methods for possible violation of random allocation were identified (4)

Possible methods for violation	
Overwriting names	
Allocate a name to the next desirable line	Considerable risk: another eligible woman of the same age group would have to appear the same day to fill the gap Every month allocation books were sent to central office checking gaps and errors, found indication of subversion, coordinator fired*
*Indication that a coordinator subverted randomization	It does not appear that the activities of the coordinator in question influenced either the pattern of allocation or the mortality
Ask women to wait or come back another day	When the coordinator would know a line allocated to the desired intervention would be available

*The central office became aware of rumors that a coordinator was subverting the randomization to ensure mammography for some of her friends. The coordinator was fired (see more *Chapter 12*)

Women with advanced breast cancer

Bailar and MacMahon also agreed with Tarone that women with advanced breast cancer should be excluded from statistical analyses of breast cancer mortality (4, 9) and that the most appropriate way to identify such women would be to identify these women prior to randomization, as this would be least biased with respect to allocation (4, 9).

In the 11-16 years of follow-up of the CNBSS study 1 (including women aged 40-49 years old), these suggested analyses were indeed performed, with the following results (10):

> "*women with cancer detected at screen 1 by breast physical examination should be excluded from both groups. Although the validity of excluding subgroups identified after the intervention as a result of mortality analyses is uncertain, Cox regression analysis performed after such exclusions results in an odds ratio of 0.93 (CI, 0.70 to 1.24). A*

similar analysis excluding women who reported a lump to the examiner at screen 1 yields an odds ratio of 0.88 (CI, 0.66 to 1.18)" (10).

CI: 95% confidence interval

In comparison, without exclusion of any women, the cumulative rate ratio comparing the screening arm with the control arm was 1.12 (95% CI, 0.82 to 1.53) (10). In the later 25-years follow-up, when the age-group 40-49 and 50-59 were analyzed together, in a mortality analysis where women with prevalent breast cancers were excluded, the hazard ratio comparing the screening arm with the control arm, was 0.90, (95% CI 0.69 to 1.16) (11).

After the screening period ended, however, breast cancer was diagnosed in 5.8% of women in the mammography arm and in 5.9% of women in the control arm, showing that the risk of breast cancer was similar in the two arms (P=0.80) (11).

Bailar and MacMahon thought that once these results were reported, "*this criticism of the study would end*" (4). But as we know now, it did not.

However, criticism and disagreement are important parts of the academic endeavor and should be encouraged.

References

1. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. Am J Radiol 1993;161:755-60.
2. Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ. A critical appraisal of the National Breast Cancer Screening Study. Radiology 1993;189:661-3
3. Baines C. The Canadian national Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334
4. Bailar III, JC. MacMahon B. Randomization in the Canadian national breast Screening Study: a review for evidence of subversion. Can Med Assoc J 1997; 156:193-199
5. Andersson I, Aspergren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ. 1988;297: 943-8.
6. Frisell J, Eklund G, Hellstrom L, Lidbrink E, Rutqvist LE, Somell A. Randomized study of mammographic screening—preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat. 1991;18:49-56.
7. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am. 1992;30:187-209.
8. Nystrom L, Rutqvist LE, Wall S, Lindgren A, Linqvist M, Ryden S, et al. Breast cancer screening with mammography: overview of Swedish randomised controlled trials. Lancet. 1993;341:973-
8. Miller AB, Baines CJ, Sickles EA. Canadian National Breast Screening Study [Letterl. Am J Roentgenol. 1990;155:1133-4.

9. Tarone RE. Breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995; 75:997-1003.
10. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002;137:305-312
11. Miller A B, Wall C, Baines C J, Sun P, To T, Narod S A 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 doi:10.1136/bmj.g366

Current debate (2021-2024)

We identified different claims in the complaint material regarding potential flaws in the CNBSS. We summarize these into five claims:

- women with symptoms and lumps were included as participants in the study,
- women with symptoms or palpable lumps, or with late stage breast cancer, were placed in the mammography arm,
- clinical breast exams were done prior to randomization,
- women without abnormal finding were removed from the mammography list.

The current debate and the criticism identified in the complaint material is summarized in tables 30.6, 30.7.

Table 30.6 Summarized claims to prove that CNBSS were "*compromised*" (1,2)

Reference	Claims
1, 2	Women with symptoms or lumps were included in the study
1, 2	Women with symptoms or palpable lumps placed in mammography arm
1, 2	Women with late stage breast cancer in mammography arm
1, 2	Clinical breast exam prior to randomization, bypass randomization
1, 2	Women without abnormal finding removed from the mammography list

References:

1. Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants.
2. Seely JM, Eby PR, Gordon PB, Appavoo S, Yaffe MJ. Errors in Conduct of the CNBSS Trials of Breast Cancer Screening Observed by Research Personnel, Journal of Breast Imaging, Volume 4, Issue 2, 2022, Pages 135–143, https://doi.org/10.1093/jbi/wbac009

Included in the complaint material were 85 letters signed mostly by radiologists. The claims and suggestions in the letters can be summarized into four topics (table 30.7).

Tab 30.7 Summarized arguments from the - Letters:

	Claim/wish	Quotes
1	Witness come forward	CNBSS flawed, compromised, not a true RCT, systematic error
2	CNBSS is an outlier	Only study that did not show mortality reduction, marked overestimates of overdiagnosis
3	Mammography screening is beneficial	Best way to save lives Well-being of women around the world at stake
4	No use data from CNBSS in guidelines and policy making	Study should be officially denounced Study must be removed from body of literature on screening

Based on the criticism and the claims, we have reviewed the main findings from the CNBSS and its various publications: Balance of demographic variables, reported self-detected lumps, number of women referred to "review clinics", numbers of cancers, interval cancers, palpable lumps, tumor size less than 10 mm/10 mm and larger), numbers of positive lymph nodes, four or more positive lymph nodes, death from breast cancer, overdiagnosis, at prevalent screening by age group 40-49 and 50-59 at randomization.

References

1. Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants.
2. Seely JM, Eby PR, Gordon PB, Appavoo S, Yaffe MJ. Errors in Conduct of the CNBSS Trials of Breast Cancer Screening Observed by Research Personnel, Journal of Breast Imaging, Volume 4, Issue 2, 2022, Pages 135–143, https://doi.org/10.1093/jbi/wbac009

Results from the CNBSS that may or may not indicate violation of randomization or poor quality of mammograms

Women in the screening arm were offered mammography and clinical breast exam. The number of women who were referred to review clinic based on findings are presented for both modalities. In the control arm, women were offered clinical breast exam and were referred to review clinic if they had lumps or reported symptoms. Clinical breast exam were offered to all women at inclusion (first screening round or prevalent screening round), and women in the screening arm were offered clinical breast exam and mammography annually for both age-groups. Women in the control arm among women aged 50-59 years were offered clinical breast exam annually (Figure 8.1) (1, 2).

Breast cancer screening with mammography or clinical breast exam aims at detecting breast tumors at a stage when they can be cured and prevent women from dying from breast cancer.

The aim of screening is to detect small, curable cancers and prevent late-stage cancers and thus death from breast cancer.

One of the claims is that women with palpable lumps were intentionally placed in the screening arm, violating their randomized allocation. If this was true, one would expect that the amount of women with palpable lumps were higher in the screening arm than in the control arm of the study in the first screening round (prevalent round).

It has been undisputed that once a woman were allocated to the screening or control arm, they were not switched later during the trial. We have therefore focused on the first (the prevalent) screening round.

We tested whether differences between the screening and control arm in any findings from the CNBSS may indicate whether women with symptoms or lumps were placed in the screening arm of the trial violating randomized allocation. We also looked for indicators of poor mammography quality.

We used two-sided 95% confidence intervals and 5% as two-sided our significance level, assuming random sampling, and normal distribution (3). We did not correct for multiple comparisons, but have also present 99% confidence intervals for differences, if one prefers that.

Balance in screening and control arm

If women were placed in the screening arm and randomization was subverted, we would expect that demographic variables differed in the two arms. As seen in the tables below, this was not the case (1, 2)

Women aged 40-49 years

Figure 31.1 Balance of demographic characteristics among women 40-49 years in the CNBSS (Table 1 in (1))

Table 1: Demographic characteristics of women aged 40 to 49 years upon entry into the Canadian National Breast Screening Study (NBSS) and women in the general population

Characteristic	Study group;* no. (and %) of women: MP group (n = 25 214)	UC group (n = 25 216)	% of women in Canada
Marital status	(n = 25 170)	(n = 25 171)	
Never married	1 639 (6.5)	1 636 (6.5)	6.1
Married	20 296 (80.6)	20 321 (80.7)	81.2
Separated or divorced	2 679 (10.6)	2 706 (10.8)	9.6
Widowed	556 (2.2)	508 (2.0)	3.1
No. of live births†	(n = 23 472)	(n = 23 459)	
0	2 331 (9.9)	2 369 (10.1)	7.6
1	2 353 (10.0)	2 447 (10.4)	9.7
2	7 943 (33.8)	7 774 (33.1)	27.2
3	6 281 (26.8)	6 186 (26.4)	24.2
4	2 848 (12.1)	3 008 (12.8)	15.4
5	1 032 (4.4)	1 072 (4.6)	7.6
≥ 6	684 (2.9)	603 (2.6)	8.1
Reproductive status	(n = 25 214)	(n = 25 216)	
Premenopausal	16 739 (66.4)	16 922 (67.1)	–
Perimenopausal	298 (1.2)	298 (1.2)	–
Postmenopausal	1 239 (4.9)	1 204 (4.8)	–
Underwent hysterectomy and oophorectomy	1 586 (6.3)	1 492 (5.9)	–
Underwent hysterectomy	4 873 (19.3)	4 834 (19.2)	–
Underwent oophorectomy	171 (0.7)	165 (0.7)	–
Unknown	308 (1.2)	301 (1.2)	–
Level of education	(n = 23 001)	(n = 22 926)	
Grade 8	1 892 (8.2)	1 956 (8.5)	25.1
Grade 9–13	7 011 (30.5)	7 026 (30.6)	40.4
Trade or business school	8 933 (38.8)	8 735 (38.1)	22.3
University	5 165 (22.5)	5 209 (22.7)	12.2
Family history of breast cancer, family member	(n = 9 493)	(n = 9 652)	
Mother	2 051 (8.1)	2 055 (8.1)	–
Sister	831 (3.3)	872 (3.5)	–
Daughter	2 (0.0)	4 (0.0)	–
Second-degree relative‡	6 609 (26.2)	6 721 (26.7)	–
Place of birth	(n = 25 214)	(n = 25 216)	
North America	21 246 (84.3)	21 266 (84.3)	76.1
Europe	3 325 (13.2)	3 284 (13.0)	18.1
Elsewhere	601 (2.4)	645 (2.6)	5.8
Not available	42 (0.2)	21 (0.1)	–
Cigarette smoking status	(n = 25 214)	(n = 25 216)	
Never smoked	12 074 (47.9)	12 034 (47.7)	53.9
Smoked, no. of cigarettes			
1–10	1 968 (7.8)	1 931 (7.7)	7.8
11–20	2 355 (9.3)	2 351 (9.3)	13.0
> 20	2 306 (9.1)	2 249 (8.9)	14.0
Used to smoke	6 511 (25.8)	6 651 (26.4)	11.4
Occupation	(n = 23 905)	(n = 23 922)	
Not in workforce§	7 912 (33.1)	7 874 (32.9)	32.0
Clerical	5 200 (21.8)	5 289 (22.1)	22.4
Medical or health related	2 615 (10.9)	2 574 (10.8)	5.4
Teaching	2 148 (9.0)	2 156 (9.0)	4.7
Managerial or administrative	1 824 (7.6)	1 807 (7.6)	4.3
Science or technology related	768 (3.2)	781 (3.3)	1.6
Sales, service	2 162 (9.0)	2 210 (9.2)	17.6
Other	1 276 (5.3)	1 231 (5.1)	12.0

*MP = mammography and physical examination (PE) of the breasts, UC = usual care (single PE of the breasts and annual follow-up by mailed, self-administered questionnaire).
†Single women not included for comparability with the Canadian population.
‡Includes aunts, cousins and other relatives.
§Includes women who were housewives, retired or unemployed.

Women aged 50-59 years

Figure 31.2 Balance of demographic characteristics among women 40-49 years in the CNBSS (Table 1 in (2))

Table 1: Demographic characteristics of women aged 50 to 59 years upon entry into the Canadian National Breast Screening Study (NBSS) and women in the general population

Characteristic	Study group;* no. (and %) of women: MP group (n = 19 711)	PO group (n = 19 694)	% of women in Canada
Marital status	(n = 19 684)	(n = 19 655)	
Never married	1 184 (6.0)	1 243 (6.3)	6.6
Married	15 554 (79.0)	15 438 (78.5)	75.5
Separated or divorced	1 653 (8.4)	1 593 (8.1)	7.7
Widowed	1 293 (6.6)	1 381 (7.0)	10.3
No. of live births†	(n = 18 429)	(n = 18 338)	
0	1 696 (9.2)	1 662 (9.1)	9.2
1	1 598 (8.7)	1 557 (8.5)	10.5
2	4 231 (23.0)	4 349 (23.7)	21.6
3	4 714 (25.6)	4 638 (25.3)	20.5
4	3 138 (17.0)	3 070 (16.7)	14.8
5	1 561 (8.5)	1 543 (8.4)	8.8
≥ 6	1 491 (8.1)	1 519 (8.3)	14.6
Reproductive status	(n = 19 711)	(n = 19 694)	
Premenopausal	2 653 (13.5)	2 736 (13.9)	–
Perimenopausal	440 (2.2)	447 (2.3)	–
Postmenopausal	8 634 (43.8)	8 671 (44.0)	–
Underwent hysterectomy and oophorectomy	2 754 (14.0)	2 713 (13.8)	–
Underwent hysterectomy	4 735 (24.0)	4 635 (23.5)	–
Underwent oophorectomy	132 (0.7)	121 (0.6)	–
Unknown	363 (1.8)	371 (1.9)	–
Level of education	(n = 18 935)	(n = 18 032)	
Grade 8	2 374 (12.5)	2 303 (12.8)	33.6
Grade 9–13	7 011 (37.0)	6 036 (33.5)	38.1
Trade or business school	6 662 (35.2)	6 650 (36.9)	18.9
University	2 888 (15.3)	3 043 (16.9)	9.4
Family history of breast cancer, family member	(n = 19 711)	(n = 19 694)	
Mother	1 489 (7.6)	1 522 (7.7)	–
Sister	1 258 (6.4)	1 242 (6.3)	–
Daughter	20 (0.1)	18 (0.1)	–
Second-degree relative‡	4 849 (24.6)	4 803 (24.4)	–
Place of birth	(n = 19 711)	(n = 19 694)	
North America	16 695 (84.7)	16 861 (85.6)	76.1
Europe	2 702 (13.7)	2 518 (12.8)	20.6
Elsewhere	284 (1.4)	275 (1.4)	3.3
Not available	30 (0.2)	40 (0.2)	–
Cigarette smoking status	(n = 19 711)	(n = 19 694)	
Never smoked	10 198 (51.7)	10 261 (52.1)	55.6
Smoked, no. of cigarettes			
1–10	1 470 (7.5)	1 411 (7.2)	8.6
11–20	1 689 (8.6)	1 734 (8.8)	12.8
> 20	1 290 (6.5)	1 260 (6.4)	10.6
Used to smoke	5 064 (25.7)	5 028 (25.5)	12.4
Occupation	(n = 18 323)	(n = 18 330)	
Not in workforce§	8 423 (46.0)	8 387 (45.8)	46.9
Clerical	3 672 (20.0)	3 660 (20.0)	16.6
Medical or health related	1 366 (7.5)	1 357 (7.4)	3.8
Teaching	918 (5.0)	925 (5.0)	3.0
Managerial or administrative	1 019 (5.6)	1 026 (5.6)	3.2
Science or technology related	411 (2.2)	443 (2.4)	1.2
Sales, service	1 723 (9.4)	1 736 (9.5)	15.9
Other	791 (4.3)	796 (4.3)	9.3

*MP = mammography and physical examination (PE) of the breasts, PO = physical examination of the breasts only.
†Single women not included for comparability with the Canadian population.
‡Includes aunts, cousins and other relatives.
§Includes women who were housewives, retired or unemployed.

Reported self-detected lumps at prevalent screen

The number of women aged 40-49 years who reported that they had detected a lump in their breast was 1847 out of 25,214 randomized (7.3%) in the screening arms and 1835 out of 25,216 (7.3%) in the control arm (appendix tab 1. In (4))

The point estimates are identical, and there is no statistical significant difference in the two arms. To reach a statistical difference at the 95% CI level, the difference between the two groups had to be 0.45% (0.60% at the 99% CI level). The difference observed in CNBSS was 0.0%.

Women aged 50-59 years

The amount of women who had reported that they had detected a lump in their breast was 4.0% in the screening arms and 3.7% in the control arm (table 2 in (4))

Given a 95% confidence level, there is no statistical significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.38% (0.50% at the 99% CI level). The difference observed in CNBSS was 0.3%.

We expect the number of women with self-detected lumps to be similar in the screening and the control arm

Women aged 40-49 years: Screening arm 7.3%; Control arm 7.3% (n*ot statistically different*)
Women aged 50-59 years: Screening arm 4.0%; Control arm 3.7% (n*ot statistically different*)

This does not provide evidence support that women with symptoms and lumps were placed in the screening arm

Number of women referred to "review clinics"

The number of women with palpable lumps is not directly reported in the papers from the study, but according to protocol and confirmed by Dr. Baines, all women with detected lumps were referred to a "*review clinic*". These numbers are reported (fig 31.3-31.4; tables 2 in (1,2)), and there was no difference between the two arms (1, 2):

Women aged 40-49 years

In the first screening round among women aged 40-49 years in the screening arm, 3,569 women (14,1%) were referred to review clinics after a clinical breast exam, either alone or in combination with mammography, as compared to 3,674 women (14,6%) in the control arm 3 referred to review clinics after a clinical breast exam (figure 31.3).

In the consecutive rounds, the numbers of women referred to review clinics were 1,640 (7.3%), 1,342 (6.1%), 1,201 (5.5%), and 705 (5.3%) in the screening arm. As women in the control arm were not screened or had clinical breast exam after the prevalent round, no women were referred to review clinics after the prevalent round (figure 31.3).

Given a 95% confidence level, there is not a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.61% (0.80% at the 99% CI level). The difference observed in CNBSS was 0.5%.

Figure 31.3 Number of women aged 40-49 in the two arms of CNBSS that were referred to review clinic by screening round (Table 2 in (1)).

Table 2: Frequency of referral to review clinic, by screen

Group; method by which abnormality detected	Screen; no. (and %) of women				
	1	2	3	4	5
MP group	(n = 25 214)	(n = 22 424)	(n = 22 066)	(n = 21 839)	(n = 14 146‡)
Mammography (Ma) only	1 118 (4.4)	557 (2.5)	531 (2.4)	489 (2.2)	306 (2.2)
PE only	3 137 (12.4)	1 505 (6.7)	1 216 (5.5)	1 115 (5.1)	681 (4.8)
Ma and PE	432 (1.7)	135 (0.6)	126 (0.6)	86 (0.4)	69 (0.5)
All PE*	3 569 (14.1)	1 640 (7.3)	1 342 (6.1)	1 201 (5.5)	750 (5.3)
UC group†	(n = 25 216)				
PE	3 674 (14.6)	–	–	–	–

*Number of women whose abnormality was detected by PE alone or in combination with Ma.
†No figures are given for screens 2 through 5 because women in the UC group were not eligible for rescreening.
‡Only 62% of the women were eligible for the fifth screen.

MP group: mammography screening arm; PE: clinical breast exam; UC: control arm

Women aged 50-59 years

A total of 39,405 women in the 50-59 year age group were enrolled in the study, 19,711 women in the screening arm and 19,694 in the control arm (2).

In the first screening round among women in the screening arm 2,164 women (11.0%) were referred to review clinics after a clinical breast exam, either alone or in combination with mammography. In the control arm 2,207 women (11.2%) were referred to review clinics after a clinical breast exam (figure 31.4).

In the consecutive rounds, the numbers of women referred to review clinics were 1,001 (5.7%), 676 (3.9%), 522 (3.2%), and 305 (3.2%) in the screening arm, and 1,032 (5.9%), 710 (4.1%), 642 (3.8%), and 366 (3.8%) in the physical exam control arm. Slightly more women in the control arm than the screening arm were referred to review clinics in the subsequent screening round (figure 31.4).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.62% (0.81.% at the 99% CI level). The difference observed in CNBSS was 0.2%.

Figure 31.4 Number of women aged 50-59 in the two arms of CNBSS that were referred to review clinic by screening round (Table 2 in (2)).

Table 2: Frequency of referral to review clinic, by screen

Group; method by which abnormality detected	Screen; no. (and %) of women				
	1	2	3	4	5
MP group	(n = 19 711)	(n = 17 669)	(n = 17 347)	(n = 17 193)	(n = 9 876†)
Mammography (Ma) only	1 208 (6.1)	452 (2.6)	368 (2.1)	322 (1.9)	170 (1.7)
PE only	1 855 (9.4)	930 (5.3)	619 (3.6)	518 (3.0)	278 (2.8)
Ma and PE	309 (1.6)	71 (0.4)	57 (0.3)	34 (0.2)	27 (0.3)
All PE*	2 164 (11.0)	1 001 (5.7)	676 (3.9)	552 (3.2)	305 (3.1)
PO group	(n = 19 694)	(n = 17 453)	(n = 17 143)	(n = 16 918)	(n = 9 755†)
PE	2 207 (11.2)	1 032 (5.9)	710 (4.1)	642 (3.8)	366 (3.8)

*Number of women whose abnormality was detected by PE alone or in combination with Ma.
†Only 62% of the women were eligible for the fifth screen.

MP group: mammography screening arm; PE: clinical breast exam; PO: control arm.

We expect the number of women referred to review clinics to be higher in the screening arm than in the control arm, as sensitivity for detecting breast cancer is higher with mammography compared to clinical breast exam (5).

Slightly more women in the control arm than in the screening arm were referred to a review clinic (1, 2).
Women aged 40-49 years: Screening arm 14,1%; Control arm 14,6% (n*ot statistically different*)
Women aged 50-59 years: Screening arm 11,0%; Control arm 11,2% (n*ot statistically different*)

This does not provide evidence to support that women with symptoms and lumps were placed in the screening arm

Number of cancers

Women aged 40-49 years

The number of women aged 40-49 years in the screening arm was 25,214 and in the control arm 25,216.

In the prevalent screening round, the number of cancers varied slightly in two publications (1, 6). In the 1996 publication a total of 86 cancers (0.34% of all women) were found in the screening arm, 21 cancers by mammography alone and 65 by clinical breast exam and mammography. In the control arm 60 (0.24% of all women) were found by clinical exam (Tab 7 in (1)).

In the 2002 publication a total of 87 cancers were detected in the screening arm (21 by mammography alone and 66 by clinical breast exam and mammograph). In the control arm 58 caners were detected. (supplement 2 in (6)).

Given a 95% confidence level, there is a statistical significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.09% (0.12% at the 99% CI level) The difference observed in CNBSS was 0.10%.

Women aged 50-59 years

A total of 39,405 women aged 50-59 years were enrolled in the study, 19,711 women in the screening arm and 19,694 in the control arm (2).

In the prevalent screening round, in the screening arm 119 (0.60% of all women) breast cancers were detected, 48 cancers by mammography alone and 71 by clinical breast exam and mammography. In the control arm 64 (0.32% of all women) breast cancers were detected.

Given a 95% confidence level, there is a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.13% (0.18% at the 99% CI level) The difference observed in CNBSS was 0.28%.

The number of cancers in the intervention arm was 42% higher (0.34%) than in the control arm (0.24%) among women aged 40 to 49 years old. The number of cancers were twice as high in in the intervention arm (0.60%) compared to the control arm (0.32%) among women aged 50 to 59 years old.

Expectation

We expect detection of breast cancer to be higher in the screening arm than in the control arm. With modern mammography techniques and equipment, the expected rates in the screening arm compared to no-screening, would be 3 times higher (7). However, since women in the control arm in the CNBSS were offered clinical breast exam, and detection rates with clinical breast exam may be higher than no examination, this may influence the ratio of breast cancer detected in the screening arm versus breast cancer detected in the control arm (8). We would expect a lower ratio comparing mammography screening and clinical breast exam with clinical breast exam versus mammography with no mammograph.

We expect detection of breast cancer to be higher in the screening arm than in the control arm (8).

Women aged 40-49 years: Screening arm 0.34%; Control arm 0.24% (1.4 times higher)
Women aged 50-59 years: Screening arm 0.60%; Control arm 0.32% (1.9 times higher)
(Statistically difference in screening and control arm, for both age groups)

This does not provide evidence to support that women with symptoms and lumps were placed in the screening arm

Interval cancer, after first (prevalent) screening round

Women 40-49 years

In the screening arm, 16 (0.063%) interval cancers and in the control arm 24 (0.095%) interval cancers were detected after the prevalent screening round (1).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.05% (0.06% at the 99% CI level). The difference observed in CNBSS was 0.03%.

Women 50-59 years

In the screening- and control arm, 14 (0.07%) and 16 (0.08%) interval cancers were detected after the prevalent screen (2).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.05% (0.07% at the 99% CI level). The difference observed in CNBSS was 0.01%.

We expect the interval cancer rate to be high if the mammography equipment and interpretation were of low quality

Women aged 40-49 years: Screening arm 0.063%; Control arm 0.095% (n*ot statistically different*)
Women aged 50-59 years: Screening arm 0.07%; Control arm 0.08% (n*ot statistically different*)

This does not support that quality of mammography was poor

Palpable lumps

Women 40-49 years

The number of detected cancer at screening was higher in the screening arm (87) than in the control arm (58) in the CNBSS (1).

The number of women in the screening arm was 25,214 and 25,216 in the control arm.

A total of 66 cancers (0,26% of all women) were detected by clinical breast exam in the screening arm and 58 (0,23% of all women) were detected in the control arm (6).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.09% (0.11% at the 99% CI level) The difference observed in CNBSS was 0.03%.

Women 50-59 years

The number of cancers detected at screening (prevalent screening) was twice as high in the screening arm than the control arm among women aged 50-59 years (2).

A total of 39,405 women were enrolled in the study, 19,711 women in the screening arm and 19,694 in the control arm (2).

The number of cancers detected at prevalent screening by clinical breast exam was 70 (0.36%) in the screening arm as compared to 64 (0.32%) in the control arm (9).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.11% (0.15% at the 99% CI level). The difference observed in CNBSS was 0.04%.

Expectation

If the randomization was manipulated and women with lumps were placed in the screening arm, we would expect the number of women with palpable lumps to be higher in the mammography arm compared to the control arm of the study. In the CNBSS, the number of women who had cancers detected by clinical breast exam was reported.

We expect detection of breast cancer by clinical breast exam to be similar in the screening and the control arm.

Women aged 40-49 years: Screening arm 0.26%; Control arm 0.23% (n*ot statistically different*)
Women aged 50-59 years: Screening arm 0.36%; Control arm 0.32% (n*ot statistically different*)

This does not support that women with symptoms and lumps were placed in the screening arm

The amount of women with lumps were similar in both arms for the age 40-49 and for 50-59 years. Even if the amount of women with lumps were similar in both groups, more women were diagnosed with cancer in the screening arm compared to the control arm. The number of cancers in the intervention arm was 42% higher (0.34%) than in the control arm (0.24%) among women aged 40 to 49 years old. The number of cancers were twice as high in in the intervention arm (0.60%) compared to the control arm (0.32%) among women aged 50 to 59 years old.

Tumor size

Women with large tumors are more likely to be detected by clinical breast exam. Tumors smaller than 10 mm are difficult to detect without mammography.

Women 40-49 years

Less than 10 mm

The number of women in the screening arm was 25,214 and in the control arm 25,216 (1).

In the prevalent, first screening round, 7 (0.028%) cancers in the screening arm and 7 (0.028%) cancers in the control arm smaller than 10 mm were detected (6).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.03% (0.04% at the 99% CI level). The difference observed in CNBSS was 0.00%.

10 mm and larger

In the prevalent, first screening round, 70 (0.28%) cancers were detected in the screening arm, and 47 (0.19%) cancer in the control arm were detected (6).

Given a 95% confidence level, there is a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.08% (0.11% at the 99% CI level). The difference observed in CNBSS was 0.09%.

Unknown size

In our calculation of percentage above, we excluded breast cancers with unknown size. The number of cancers with unknown size was 10 (0.040%) in the screening arm and 4 (0.016%) in the control arm (not statistically different at the 95% nor the 99% confidence level) (6).

Of all cancers detected in the prevalent screening round, 8% (7/87) were less than 10 mm, 80.4% (70/87) were 10 mm and larger, and 11.5% (10/87) were of unknown size in the screening arm. In the control arm 12.1% (7/58) were less than 10 mm, 81.0% (47/58) were 10 mm and larger, and 6.9% (4/58) were of unknown size.

Figure 31.5 Number and size of breast cancer among women aged 40-49 in the two arms of CNBSS by year of follow-up (Appendix Table 2 in (6)). Mammography group: Screening arm; Exam: clinical breast exam; Usual care group: control arm

Appendix Table 2. **Cases of Cancer according to Year, Size of Invasive Breast Tumors, and Mode of Detection**

Year and Tumor Size	Cases of Cancer, *n*							
	Screening Detected Cancer				Interval Cancer		Incident Cancer	
	Mammography Group			Usual Care Group	Mammography Group	Usual Care Group	Mammography Group	Usual Care Group
	All	Mammography*	Mammography + Exam†					
Year 1								
<9 mm	7	4	3	7	3	2	–	–
10–14 mm	17	4	13	6	1	4	–	–
15–19 mm	21	5	16	12	1	5	–	–
20–39 mm	28	4	24	26	4	8	–	–
≥40 mm	4	0	4	3	6	3	–	–
Unknown	10	4	6	4	2	2	–	–
Total	87	21	66	58	17	24	–	–
Years 2–5								
<9 mm	27	17	10	–	2	–	3	10
10–14 mm	18	10	8	–	7	–	3	24
15–19 mm	21	5	16	–	10	–	3	28
20–39 mm	32	7	25	–	13	–	8	59
≥40 mm	10	2	8	–	3	–	3	16
Unknown	13	7	6	–	5	–	5	18
Total	121	48	73	–	40	–	25	155
Years 6–9								
<9 mm	–	–	–	–	–	–	17	14
10–14 mm	–	–	–	–	–	–	22	32
15–19 mm	–	–	–	–	–	–	29	22
20–39 mm	–	–	–	–	–	–	63	49
≥40 mm	–	–	–	–	–	–	15	17
Unknown	–	–	–	–	–	–	58	69
Total	–	–	–	–	–	–	204	203
Total to year 9								
<9 mm	34	23	13	7	5	2	20	24
10–14 mm	35	14	21	6	8	4	25	56
15–19 mm	42	10	32	12	11	5	32	50
20–39 mm	60	11	49	26	17	8	71	108
≥40 mm	14	2	12	3	9	3	18	33
Unknown	23	11	12	4	7	2	63	87
Total	208	69	139	58	57	24	229	358

* Detected by mammography alone.
† Detected by physical examination alone or with mammography.

Women 50-59 years

Less than 10 mm

The number of women was 19,711 in the screening arm and 19,694 in the control arm (2).

In the prevalent, first screening round, 17 (0.086%) cancers in the screening arm and 3 (0.015%) cancers in the control arm smaller than 10 mm were detected (9).

Given a 95% confidence level, there is a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.04% (0.06% at the 99% CI level). The difference observed in CNBSS was 0.071%.

10 mm and larger

In the prevalent, first screening round, 87 (0.44%) cancers were detected in the screening arm, and 61 (0.31%) cancer in the control arm were detected (9).

Given a 95% confidence level, there is a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.12% (0.16% at the 99% CI level). The difference observed in CNBSS was 0.13%.

Unknown size

In our calculation of percentage above, we excluded breast cancers with unknown size. The number of cancers with unknown size was 14 (0.071%) in the screening arm and 3 (0.015%) in the control arm (statistically different at 95% confidence level).

Of all cancers detected in the prevalent screening round, 14.4% (17/118) were less than 10 mm, 73.7% (87/118) were 10 mm and larger, and 11.9% (14/118) were of unknown size in the screening arm. In the control arm 4.7% (3/64) were less than 10 mm, 90.6% (58/64) were 10 mm and larger, and 4.7% (3/64) were of unknown size.

Figure 31.6 Number and size of breast cancer among women aged 50-59 in the two arms of CNBSS by year of follow-up (Table 1 in (9)). MP: Screening arm; MA: mammography; PE: clinical breast exam; PO: Control arm

Table 1. Size of invasive breast cancers, ascertained in the first 9 years of follow-up, by year and mode of detection*

	Screen-detected cancer, No.				Interval cancer, No.		Incident cancer, No.		All cancers, No.	
Tumor size, mm	All MP	MP/MA	MP/PE	PO	MP	PO	MP	PO	MP	PO
Year 1										
<9	17	13	4	3	0	1	—	—	17	4
10–14	18	5	13	9	1	3	—	—	19	12
15–19	27	14	13	13	7	4	—	—	34	17
20–39	40	10	30	32	6	4	—	—	46	36
≥40	2	0	2	4	0	1	—	—	2	5
Unknown	14	6	8	3	0	3	—	—	14	6
Total	118	48	70	64	14	16	—	—	132	80
Years 2–5										
<9	31	24	7	3	2	8	3	4	36	15
10–14	36	24	12	20	6	12	2	5	44	37
15–19	36	16	20	16	6	9	8	7	50	32
20–39	32	7	25	35	14	25	7	20	53	80
≥40	6	1	5	7	3	5	4	3	13	15
Unknown	8	6	2	3	5	13	8	8	21	24
Total	149	78	71	84	36	72	32	47	217	203
Years 6–9										
<9	—	—	—	—	—	—	22	31	22	31
10–14	—	—	—	—	—	—	30	23	30	23
15–19	—	—	—	—	—	—	28	27	28	27
20–39	—	—	—	—	—	—	42	50	42	50
≥40	—	—	—	—	—	—	6	11	6	11
Unknown	—	—	—	—	—	—	47	75	47	75
Total	—	—	—	—	—	—	175	217	175	217
Total to year 9										
<9	48	37	11	6	2	9	25	35	75	50
10–14	54	29	25	29	7	15	32	28	93	72
15–19	63	30	33	29	13	13	36	34	112	76
20–39	72	17	55	67	20	29	49	70	141	166
40–49	8	1	7	11	3	6	10	14	21	31
Unknown	22	12	10	6	5	16	55	83	82	105
Total	267	126	141	148	50	88	207	264	524	500

*MP = mammography plus physical examination arm, MA = detected by mammography alone, PE = detected by physical examination with or without mammographic findings, PO = physical examination only arm.

We expect detection of breast cancer by size to be different in the screening and the control arm.

Women aged 40-49 years
Tumors less than 10 mm: Screening arm 0.028%; Control arm 0.028% (n*ot statistically different*)
Tumors 10 mm or larger: Screening arm 0.28%; Control arm 0.19% (*statistically different*)

Women aged 50-59 years:
Tumors less than 10 mm: Screening arm 0.086%; Control arm 0.015% (*statistically different*)
Tumors 10 mm or larger: Screening arm 0.44%; Control arm 0.31% (*statistically different*)

This does not support that women with symptoms and lumps were placed in the screening arm

Number of positive lymph nodes

Women 40-49 years

The number of women in the screening arm was 25,214 and in the control arm 25,216.

The number of women with positive lymph nodes detected at prevalent screening was 33 (0.13%) in the screening arm, and 21 (0.08%) in the control arm (6).

Of all detected tumors, in the screening arm 60.5% (52/86) had node negative breast cancer, 38.4% (33/86) had positive lymph nodes, and 1.1% (1/86) had unknown lymph node. In the control arm 56.7% (34/60) had node negative breast cancer, 35.0% (21/60) had positive lymph nodes, and 8.3% (5/60) had unknown lymph node status (6).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.06% (0.07% at the 99% CI level). The difference observed in CNBSS was 0.05%.

Women 50-59 years

The number of women in the screening arm was 19,711 in the screening arm and 19,694 in the control arm (2).

The number of women with positive lymph nodes detected at prevalent screening was 31 (0.16%) in the screening arm, and 22 (0.11%) in the control arm (9).

Of all detected tumors, in the screening arm 63.0% (75/119) had node negative breast cancer, 26.1% (31/119) had positive lymph nodes, and 10.9% (13/119) had unknown lymph node. In the control arm 57.8% (37/64) had node negative breast cancer, 34.4% (22/64) had positive lymph nodes, and 7.8% (5/64) had unknown lymph node status.

Given a 95% confidence level, there is no statistically significant difference in the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.07% (0.10% at the 99% CI level). The difference observed in CNBSS was 0.05%.

We expected the number of women with lymph nodes positive breast cancer to be higher in the screening arm

Women aged 40-49 years: Screening arm 0.13%; Control arm 0.08% (1.2 times higher) (not *statistically different*)
Women aged 50-59 years: Screening arm 0.16%; Control arm 0.11% (1.5 times higher) (*not statistically different*)

This does not support that women with symptoms and lumps were placed in the screening arm

Four or more positive lymph nodes

Women 40-49 years

Yaffe et al. have pointed to the fact that more cancers with four or more lymph nodes were found in the screening arm as compared to the control arm in the first publication of CNBSS 1 (women in their 40 s) in 1992. This is used to suggest that women with palpable lumps were placed in the screening arm of the study for women aged 40 to 49 years (10, 11).

The number of women in the screening arm was 25,214 and in the control arm 25,216.

A total of 24 cancers with four or more positive axillary lymph nodes were found at the prevalence screen, 19 in the screening arm (0.075%) and 5 in the control arm (0.020%). Of these 19, 17 were palpable (Table 7 in (1)).

More women with four or more positive lymph nodes were found in the screening arm.

Of 25,214 women in the screening group 19 had four or more positive lymph nodes (0.075%) whereas in the control arm 5 of 25,216 women (0.020%) had four or more positive lymph nodes.

Figure 31.7 Number of women aged 40-49 in the two arms of CNBSS that were referred to review clinic by screening round (Table 2 in (1)). MP group: mammography screening arm; PE: clinical breast exam; UC: control arm

Table 7: Nodal status of all invasive tumours, by year*

Year; no. of nodes involved	Screen-detected cancer				Interval cancer		Incident cancer	
	MP							
	All	Ma alone	PE†	UC	MP	UC	MP	UC
Year 1								
None	52	17	35	34	7	13	–	–
1–3	14	1	13	16	2	5	–	–
≥ 4	19	2	17	5	6	2	–	–
Unknown	1	1	0	5	1	4	–	–
Total	86	21	65	60	16	24	–	–
Years 2–5								
None	82	36	46	–	20	–	10	92
1–3	24	7	17	–	8	–	4	20
≥ 4	9	1	8	–	5	–	5	15
Unknown	4	2	2	–	4	–	9	21
Total	119	46	73	–	37	–	28	148
Year 6 or more								
None	–	–	–	–	–	–	20	18
1–3	–	–	–	–	–	–	3	2
≥ 4	–	–	–	–	–	–	3	1
Unknown	–	–	–	–	–	–	19	19
Total	–	–	–	–	–	–	45	40
All years								
None	134	53	81	34	27	13	30	110
1–3	38	8	30	16	10	5	7	22
≥ 4	28	3	25	5	11	2	8	16
Unknown	5	3	2	5	5	4	28	40
Total	205	67	138	60	53	24	73	188

*For all dashes there was no cancer in this category by study design.
†Tumours detected at physical examination, alone or in combination with Ma.

Given a 95% confidence level, there is a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.04% (0.05% at the 99% CI level). The difference observed in CNBSS was 0.055%.
The probability of such imbalance occurring randomly is estimated to be 0.0033 (11).

Expectation
One might argue that in the screening arm, where more tumors are detected, more women will have an axillary dissection and hence more lymph nodes might be found. Since statistically more tumors were detected (at the 95% significance level) in the screening arm, it is likely that more women with lymph nodes should be detected as well. However, it is not clear how different, if any, the number of women with breast cancer with node involvement should be.

We do not know, why the CNBSS presented number of lymph nodes in two categories, three and less versus four and more lymph nodes.

Women 50-59 years

A total of 39,405 women were enrolled in the study, 19,711 women in the screening arm and 19,694 in the control arm (2).

The number of women with four or more lymph nodes detected at prevalent screening, was 9 (0.046%) in the screening arm, and 11 (0.056%) in the control arm (2).

Given a 95% confidence level, there is not a statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.04% (0.06% at the 99% CI level). The difference observed in CNBSS was 0.01%.

Figure 31.8 Number of women aged 50-59 with breast cancer and node status, in the two arms of CNBSS by year after enrollment (Table 2 in (2)). MP group: mammography screening arm; PE: clinical breast exam; PO: control arm.

Table 7: Nodal status of all invasive tumours, by year*

Year; no. of nodes involved	Screen-detected cancer				Interval cancer		Incident cancer	
	MP			PO	MP	PO	MP	PO
	All	Ma alone	PE†					
Year 1								
None	75	32	43	37	6	6	–	–
1–3	22	8	14	11	4	4	–	–
≥ 4	9	0	9	11	1	4	–	–
Unknown	13	8	5	5	3	2	–	–
Total	119	48	71	64	14	16	–	–
Years 2–5								
None	98	53	45	49	17	42	10	26
1–3	28	12	16	21	6	13	2	4
≥ 4	12	5	7	11	4	5	4	2
Unknown	11	9	2	3	5	6	11	12
Total	149	79	70	84	32	66	27	44
Year 6 or more								
None	–	–	–	–	–	–	11	24
1–3	–	–	–	–	–	–	4	3
≥ 4	–	–	–	–	–	–	2	1
Unknown	–	–	–	–	–	–	12	18
Total	–	–	–	–	–	–	29	46
All years								
None	173	85	88	86	23	48	21	50
1–3	50	20	30	32	10	17	6	7
≥ 4	21	5	16	22	5	9	6	3
Unknown	24	17	7	8	8	8	23	30
Total	268	127	141	148	46	82	56	90

*For all dashes there was no cancer in this category by study design.
†Tumours detected at physical examination, alone or in combination with Ma.

We expected the number of women with four or more lymph nodes to be higher in the screening arm. We are unsure how much higher

Women aged 40-49 years: Screening arm 0.075%; Control arm 0.020% (3.75 times higher) (*statistically different*)
Women aged 50-59 years: Screening arm 0.046%; Control arm 0.056% (control 1.22 times higher) (*not statistically different*)

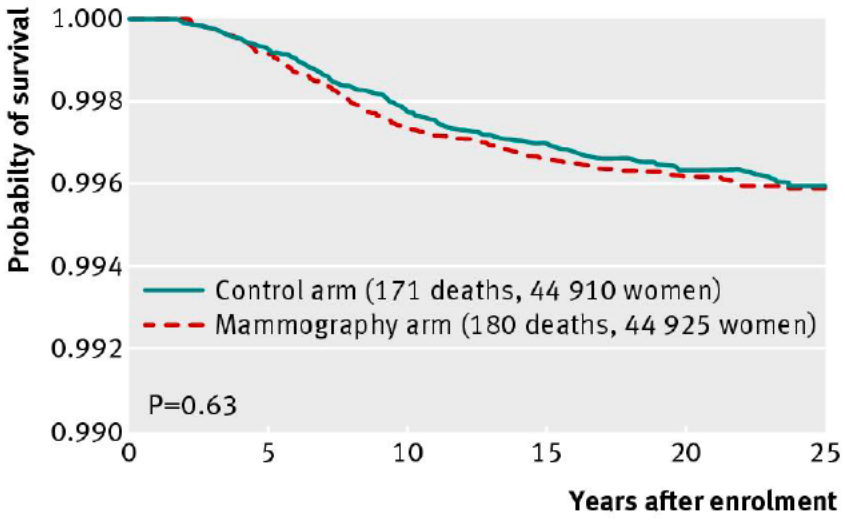
This does neither support nor rule out that women with symptoms and lumps were placed in the screening arm for women aged 40-49 years

Death

In 25-year follow-up of CNBSS, 180 women in the screening arm and 171 women in the control arm died of breast cancer. The overall hazard ratio for death from breast cancer diagnosed during the screening period associated with mammography was 1.05 (95% confidence interval 0.85 to 1.30). The findings for women aged 40-49 and 50-59 were similar (12).

During the entire study period, 500 women in the screening arm and 505 in the control arm died of breast cancer. Thus, the cumulative mortality from breast cancer was similar between women in the mammography arm and in the control arm (hazard ratio 0.99, 95% confidence interval 0.88 to 1.12) (12).

Figure 31.9 Breast cancer mortality from women diagnosed with breast cancers during in the screening period in CNBSS (Figure 4 in (12))



Results stratified by age-group were not published in the BMJ paper, but we were able to retrieve age-stratified data from Dr. [redacted] (*Chapter 32*).

Women 40-49 years

The number of women in the screening arm was 25,214 and in the control arm 25,216.

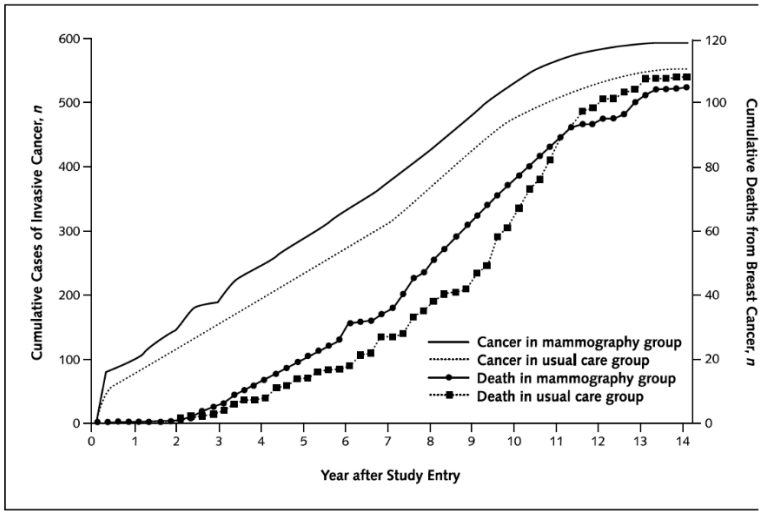
The number of women who died after 11 to 16 years of follow-up, was 105 (0.42%) in the screening arm, and 108 (0.43%) in the control arm (6).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.11% (0.15% at the 99% CI level). The difference observed in CNBSS was 0.01%.

The mortality rate in the screening arm was higher than in the control arm until around 11 years of follow-up. At a mean follow-up of 8.5 years after enrolment, 38 (0.15%) women died of breast cancer in the screening arm and 28 (0.11%) died in the control arm (REF Miller 1992)

Given a 95% confidence level, there is not a statistically significant difference between the two arms, at a mean follow-up of 8.5 years. To reach a statistical difference at the 95% CI level, the difference had to be 0.06% (0.08% at the 99% CI level). The difference observed in CNBSS was 0.04%.

Figure 31.10 Cumulative breast cancer incidence and mortality among women aged 40-49 years in the screening (mammography group) and control arm (usual care group) of CNBSS after 11-16 years of follow-up (Fig 2 in (6))



Women 50-59 years

The number of women in the screening arm was 19,711 and in the control arm 19,694 (2).

The number of women who died at 13 years of follow-up, was 107 (0.54%) in the screening arm, and 105 (0.53%) in the control arm (9).

Given a 95% confidence level, there is no statistically significant difference between the two arms. To reach a statistical difference at the 95% CI level, the difference had to be 0.14% (0.19% at the 99% CI level). The difference observed in CNBSS was 0.01%.

The cumulative rate ratio was 1.02 (95% confidence interval = 0.78–1.33), and the mortality rate ratio was higher in the screening arm than the control arm at each year of follow-up (Figure 31.11) (9).

Figure 31.11 Cumulative breast cancer mortality among women 50-59 years in the screening (MP) and control arm (PO) of CNBSS at 13 years of follow-up (Table 4 in (9))

Table 4. Cumulative number of deaths from breast cancer to June 30, 1996, by study arm and time of breast cancer detection*

Time of detection	Study arm, No. of deaths	
	MP	PO
Including only breast cancers identified to 5 years from entry		
Screen 1	25	12
Screens 2–5	28	22
Interval 1	7	7
Intervals 2–5	10	20
Incidents 2–5	4	7
Total	74	68
Cumulative breast cancer death rates per 10 000†	3.42	3.15
Mortality rate ratio (95% CI)	1.09 (0.78–1.51)	
Including breast cancers identified to 6 years from entry		
Among breast cancers detected to year 5	74	68
Among breast cancers detected during year 6	10	8
Total	84	76
Cumulative breast cancer death rates per 10 000†	3.89	3.52
Mortality rate ratio (95% CI)	1.10 (0.81–1.51)	
Including breast cancers identified to 7 years from entry		
Among breast cancers detected to year 6	84	76
Among breast cancers detected during year 8	9	7
Total	93	83
Cumulative breast cancer death rates per 10 000†	4.30	3.84
Mortality rate ratio (95% CI)	1.12 (0.83–1.50)	
Including breast cancers identified to 8 years from entry		
Among breast cancers detected to year 7	93	83
Among breast cancers detected during year 8	6	6
Total	99	89
Cumulative breast cancer death rates per 10 000†	4.58	4.12
Mortality rate ratio (95% CI)	1.11 (0.84–1.48)	
Including breast cancers identified to 9 years from entry		
Among breast cancers detected to year 8	99	89
Among breast cancers detected during year 9	5	8
Total	104	97
Cumulative breast cancer death rates per 10 000†	4.81	4.49
Mortality rate ratio (95% CI)	1.07 (0.81–1.41)	
Including breast cancers identified 9 or more years from entry		
Among breast cancers detected to year 9	104	97
Among breast cancers detected beyond year 9	3	8
Total	107	105
Cumulative breast cancer death rates per 10 000†	4.95	4.86
Mortality rate ratio (95% CI)	1.02 (0.78–1.33)	

*MP = mammography plus physical examination arm, PO = physical examination-only arm, and CI = confidence interval.
†Based on 216 133 person-years of observation in the mammography plus physical examination arm and 216 042 in the physical examination-only arm.

If mammography screening worked, we expect the number of women to die of breast cancer to be lower in the screening arm than the control arm.

Women aged 40-49 years: Screening arm 0.42%; Control arm 0.43% (*not statistically different*)
Women aged 50-59 years: Screening arm 0.54%; Control arm 0.53% (*not statistically different*)

This does neither support nor rule out that women with symptoms and lumps were placed in the screening arm

Overdiagnosis

Women 40-49 years

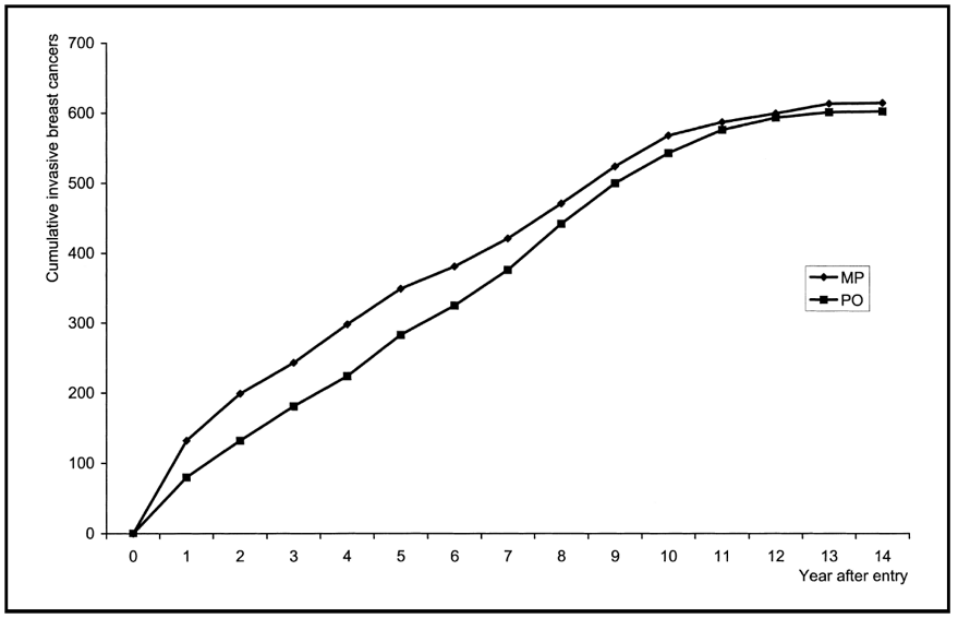
After 11-16 years of follow-up, an excess of breast cancers in the screening arm was observed Fig 31.10 (6). "*Unless the lead time gained by mammography exceeds 10 years, an excess 40*

cases of invasive breast cancer detected by mammography persist. This represents 58% of the 69 cases of nonpalpable invasive breast cancer in the mammography group" (6).

Women 50-59 years

The number of cancers after approximately 11 years are similar in the screening and control arm (figure 31.12). There seem to be no indication of overdiagnosis after 11 years for women aged 50 to 59 years old in the CNBSS.

Figure 31.12 Cumulative breast cancer among women 50-59 years in the screening (MP) and control arm (PO) of CNBSS at 13 years of follow-up (Figure 2 in (9))



At the end of the screening period, an excess of 142 breast cancer cases was observed in the mammography arm compared with the control arm (666 versus 524) . Fifteen years after enrolment, the excess became constant at 106 cancers (12). A total of 484 cancers were screen detected, the 106 excess cancers represent 22% (106/484) of screen detected invasive breast cancers that were overdiagnosed (12).

If mammography screening worked, we expect the number of women with breast cancer to be higher in the screening arm than the control arm, especially in the first years after enrollment.

Women aged 40-49 y: Higher breast cancer incidence (fig. 31.10) 14 y after enrollment
Women aged 50-59 y: Higher breast cancer incidence (fig. 31.11) 11 y after enrollment

This does not support that women with symptoms and lumps were placed in the screening arm
y: years

References

1. 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: 1459-76.
2. 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: 1477-88.
3. SAPIO research, Accessed February 15 2024 at https://sapioresearch.com/significant-difference-calculator/
4. Baines C. The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334.
5. Ngan TT, Nguyễn NT, Minh HV, Donnelly M, O´Neill C. Effectiveness of clinical breast examination as a stand-alone screening modality: an overview of systematic reviews. BMC Cancer 2020; 20,1070. https://doi.org/10.1186/s12885-020-07521-w
6. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002;137:305-312
7. European guidelines for quality assurance in breast cancer screening and diagnosis. EUREF Accessed Feb 24 2024 at https://euref.org/european-guidelines/5th-edition/
8. Kosters JP, Gotzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. Cochrane Database Syst Rev. 2003;2:CD003373.
9. Miller AB, To T, Baines CJ, Wall C. Canadian national Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000;92:1490–9
10. Yaffe MJ. Seely JM, Kopans DB. The randomized trial of mammography screening that was not – A cautionary tale. J Med Screen 2022 https://doi.org/10.1177/09691413211059461
11. Boyd NF, Jong RA, Yaffe MJ, et al. A critical appraisal of the Canadian national breast cancer screening study. Radiology 1993 Dec; 189: 661–663.
12. Miller A B, Wall C, Baines C J, Sun P, To T, Narod S A 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 doi:10.1136/bmj.g366

Data provided by [redacted] December 2022:

We interviewed Dr. [redacted] in November of 2022. We asked if he could provide us with results from the 25-year follow up stratified by age-groups, 40-49 and 50-59. The 25 year follow up publication in the BMJ showed results for the two age-groups together (1). As the control arm differed between the age-groups and because the claims that randomizations was subverted and women with symptoms or lumps were placed in the mammography arm of the trial, was primarily for the age group 40-49, we would like to see the results by age.

Dr. [redacted] provided us with the data a couple of weeks after the interview (supplement). The data are not peer-reviewed, but we thought the analysis by age was informative and have chosen to present it in this Report as is. We believe the data is similar to the data used in the 25-year follow up paper in the BMJ in 2013 (1). The result presented here is to be used as evidence, evidence with similar value as what interviewees are saying in the interview, and should not be treated similarly as results published in a scientific paper.

Dr. [redacted] provided us with an analysis by age group.

Dr. [redacted] sent us a letter with the results. He explains (text edited, shortened):

> "*The specific allegation is that we identified women with breast masses (lumps) prior to randomisation and some of these women by-passed the randomisation process and were directed by nurses to the mammography arm.*
> - *Mammography arm was enriched for patients with a palpable mass.*
> - *The number of deaths from breast cancer in the mammography arm was inflated. Had these deaths not been counted, a benefit of mammography would probably have been seen.*
>
> *If this were the case we would*
> - *expect there to be an excess of women with palpable breast cancers at the first round of screening in the mammography arm compared to the no mammography arm (but not in subsequent rounds).*
> - *expect the excess of deaths from cancer due the assignment of these women to be deaths from prevalent cancer and the majority of these would accrue in the first five years after the first mammogram.*
>
> *If mammography were in fact effective, but the benefit was obscured by the faulty randomisation of women with palpable cancers, then we expect that, after excluding all women with prevalent palpable cancers from the study, we would see a benefit from mammography*."

Women aged 40-49 years

Palpable cancers

Total number of palpable cancers detected at the first round of screening:

Screening arm:	84; 25 died of breast cancer.
Control arm:	85; 24 died of breast cancer.

Based on these results, there was no significant excess of palpable cancers in the mammography arm at the first round.

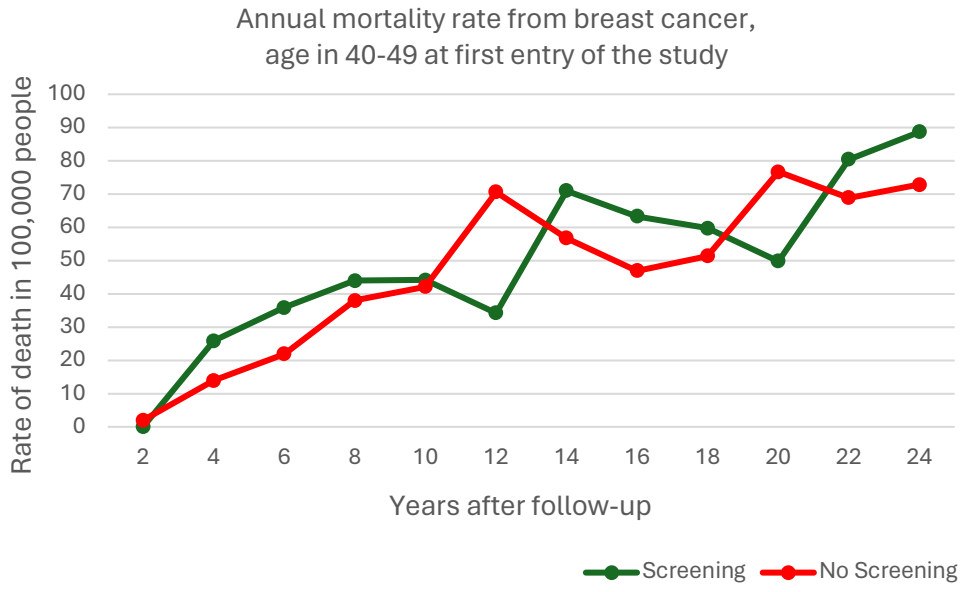
Non-palpable cancers
Screening arm: 27 (all mammogram-detected); 4 died of breast cancer
Control arm: 0 (non-screened with mammography)

Total number of cancer detected in round 1
Screening arm: 111
Control (no-screening) arm: 85

The overall hazard ratio for deaths from breast cancer detected in all five rounds of screening associated with mammography was 1.09
The hazard ratio for cancers detected in the first round of screening was 1.21
The hazard ratio for cancers detected in years 2-5 was 1.04.

No-short term increase in annual breast cancer mortality in the screening arm followed by a long term decline, was observed (see figure 32.1)

Figure 31.1 Annual mortality rate from breast cancer women aged 40-49 years in the screening and control (no screening) arm by years of follow-up



Women aged 50-59 years
Palpable cancers
At the first round of screening: (1 to 1 randomization)
Screening arm: 90; 24 died of breast cancer.
Control arm: 85; 23 died of breast cancer.

Non-palpable cancers
Screening arm: 52 (all mammogram-detected); 11 died of breast cancer
Control arm: 0 (non-screened with mammography)

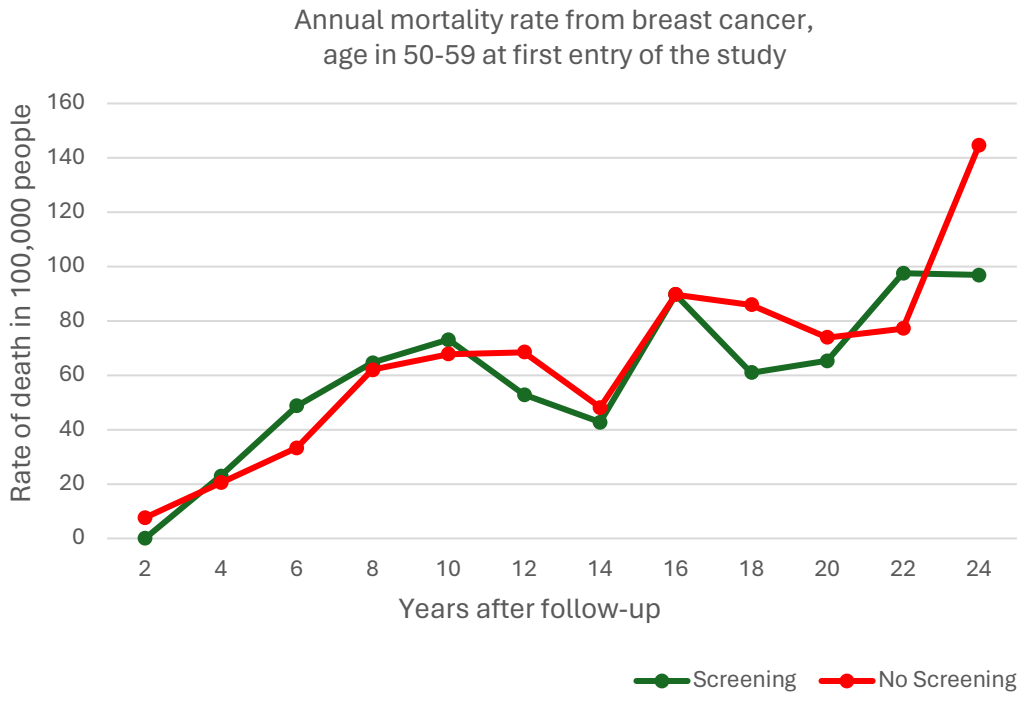
Total cancer detected round 1
Screening arm: 142
Control arm: 85

The overall hazard ratio for deaths from breast cancer detected in all five rounds of screening associated with mammography was 1.02
The hazard ratio for cancers detected in the first round of screening was 1.74
The hazard ratio for cancers detected in years 2-5 was 0.79.

The decline in the hazard ratio from 1.74 to 0.79 is expected because the date of diagnosis of the 52 non-palpable cancers was advanced to the first screening round and the 11 deaths from these were attributed to year one rather than to later years.

No-short term increase in the annual breast cancer mortality in the mammography group followed by a long term decline, was observed (see figure 32.2)

Figure 32.2 Annual mortality rate from breast cancer women aged 50-59 years in the screening and control (no-screening) arm by years of follow-up



Dr. [redacted] explained:

> "*It has also been stated by our critics that a screening trial should be limited to asymptomatic women and that women with palpable lesions should be ineligible for 'screening' from the outset. That is they should be identified and removed from the trial at outset. This is a valid study design provided that the women with palpable masses are removed from both arms.*
>
> *In the Swedish two county trial, the control group was women in the general population. They were never invited for screening. They were never examined. There was no mechanism to identify controls with a palpable mass in the Swedish trial let alone remove them. This is true of most other trials as well. If one were to exclude women with a palpable mass from the screened arm and not from the unscreened arm – as our critics suggest - then a screening benefit would be observed.*
>
> *In the CNBSS study, the hazard ratio for deaths from breast cancer associated with mammography after removing all palpable cancers detected in year 1 was 1.01*
>
> *Finally our detractors claim that our conclusions have led to the deaths of hundreds of women diagnosed under age 40 by denying them screening. Yet, they fail to cite the results of the UK AGE trial published in Lancet Oncology in 2020* [4] *which was a randomised trial of women assigned to screening at age 40 versus age 50. After 22 years of follow up there was no benefit to beginning screening at age 40 compare to age 50 (HR = 0.91)."*

References

1. Miller A B, Wall C, Baines C J, Sun P, To T, Narod S A 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 doi:10.1136/bmj.g366
2. 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: 1459-76.
3. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002;137:305-312
4. Duffy SW, Vulkan D, Cuckle H, Parmar D, Sheikh S, Smith RA et al. Effect of mammography screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomized, controlled trial. Lancet Oncol 2020; 21: 1165–72

Comparison with other studies

Breast cancer mortality and advanced stage disease for women aged 40-49 years

In the interview [redacted] said:

> "*We were interested in what experts would tell us. And the experts in many ways, betrayed us. Because they didn't like our results. They didn't like our results.*
>
> *It was expected after the …..* [not possible to hear] *and his publication of his results. This trial* [CNBSS] *with individual randomization, was definitely going to show benefit for women in their forties. And since some of these collaborating external radiologists were on our policy advisory group, they knew in detail, our results. They knew long in advance of the publication, that we were not going to produce the results that they were expecting. That we had expected.*
>
> *You know you asked me before a question that I didn't answer, which is what I did in this study. But in any case, one of the things that I did was: I monitored the deaths, and where they were occurring. And I can tell you in the first 3 years of my being there, saying the excess mortality in women 40 to 49 was a real surprise until 1985 when Tabár published his first result. I saw he found the same thing. He never admitted it*".

Breast cancer mortality for women aged 40-49 years

Baines is referring to table 7 in the Two county study (1)).

Altogether 13% of the control group had a mammographic examination (1), approximately 89% of women in the screening group had a single view mammogram every 2 or 3 years. After an average follow-up of six years, and two rounds of screening, there was a reduction in mortality (Figure 33.1)and the number of advanced stage breast cancer (stage II) (Figure 33.3)(1).

For women aged 40-49 years old, the study observed 16 (8 in Kopparberg and 8 in Östergötland) deaths among 19937 individuals (population) in the study group and 10 deaths among 15 678 individuals in the control group, given a relative risk of 1.26 (95% confidence interval 0.56-2.84) (1).

Fig 33.1 Number of deaths in the screening and control arm of the Kopparberg and Östergötland (the two county study) and the relative risk of death comparing the screening arm (study) and control arm (control) by age-group (Table 7 in (1))

TABLE VII—DEATHS FROM BREAST CANCER BY AGE AT ENTRY

—	Kopparberg county		Östergötland county		—
	Deaths	Population	Deaths	Population	
Age-group 40–49:					
Study	8	9625	8	10 312	Combined χ_1^2 = 0·31
Control	3	5053	7	10 625	Relative risk = 1·26, 95% confidence interval (0·56, 2·84)
Age-group 50–74:					
Study	43	29 426	28	28 722	Combined χ_1^2 = 9·14
Control	36	13 793	40	27 311	p = 0·003 (2-sided) Relative risk = 0·61 95% confidence interval (0·44, 0·84)

The higher number of deaths in the first 6-8 years of follow is also observed in the other Swedish breast cancer screening trials: Malmö, Kopparberg/Östergötland (the Two-county trial), Stockholm, and Gothenburg (2). Breast cancer mortality is higher in the screening arm than the control arm the first 6-7 years of follow-up (Figure 33.2) (2).

After median trial time of 7 years, a median follow-up time of 12.8 years, and a screening interval of 18–24 months, breast cancer mortality was reduced by a 23% (relative risk 0.77; 95% CI: 0.59–1.01). The reduction was not statistically significant for women aged 40-49 years old (2).

Figure 33.2, Cumulative breast cancer mortality in the screening arm (IG) and control arm (CG) in the Swedish randomized breast cancer screening trials (Figure 1 in (2)).

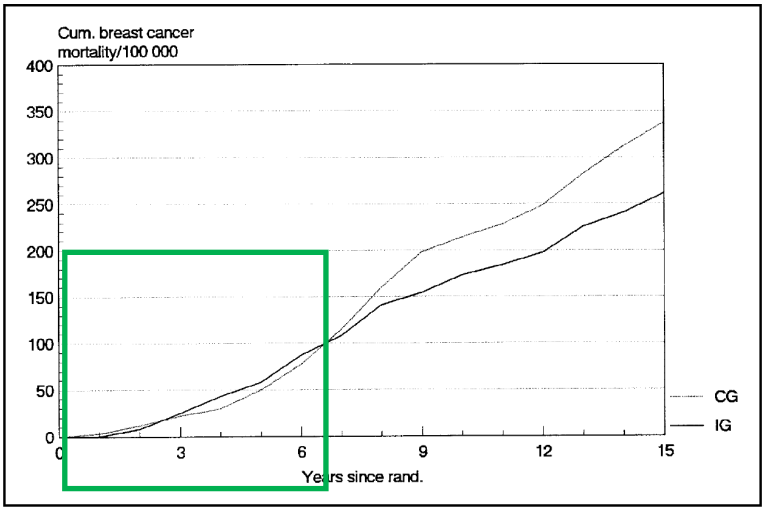


Fig. 1. Cumulative number of breast cancer deaths (BCDs)/ 100,000 women 40–49 years at randomization in the invited (IG) and control group (CG) by year since randomization.

In CNBSS, among women aged 40-49, the mortality rate in the screening group was higher in approximately 11 years after enrollment (figure 31.10) (3), and seemed to have levelled out thereafter (figure 32.1) (4).

Advanced stage breast cancer for women aged 40-49 years

In the Two-county study, the number of women with advanced cancers was higher in the screening arm than in the control arm at inclusion (Figure 33.3) (1). This was also found in the CNBSS where 0.13% of women aged 40-49 years in the screening arm had positive lymph nodes (usually defined as stage II, or advanced cancers), and 0.08% of the women in the control arm. The difference was not statistically different (Chapter 31) (5).

Figure 33.3 Number of advanced (stage II) breast cancer of the Two county study (Kopparberg and Östergötland) in the screening arm (study) and control arm (control)* (Figure 1 in (1)).

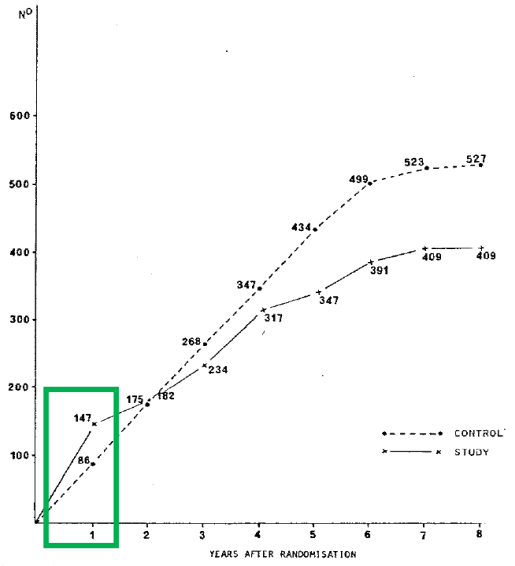


Fig 1— Cumulative number of women with stage II or more advanced breast cancer by time since randomisation, women aged 40–74 at entry.

> * "*The figures for the control group are adjusted for the different size of the control group in Kopparberg country"* (1).

In the CNBSS, the number of women aged 40-49, diagnosed with breast cancer with four or more lymph nodes in the prevalent round, was statistically significantly higher in the screening arm (0.075%) as compared to the control arm (0.020%), Chapter 31 (5).

We are not aware of any other publication of a randomized breast cancer screening study where number of women aged 40-49 years diagnosed with four or more lymph nodes are presented, so we are not able to compare this to other trials.

In a study comparing advanced disease in different randomized trials on breast cancer screening (6). The study showed incidence rate of advanced stage for all age-groups in the trials, for the whole intervention period, and not only for the prevalent round, but it may still serve as a valid comparison of advanced breast cancer in the trials (Figure 33.4) (6).

Advanced stage was defined as tumors 20 mm and larger, because this threshold distinguishes tumor classes T1 and T2 in both the TNM and American Joint Committee on Cancer (AJCC) evaluation of breast cancer stage at diagnosis (7, 8). If size was not available, stage at diagnosis, which is based on cancer size, node status, and existence of metastasis in distant organs were used (6).

As shown, the incidence rate of advanced breast cancers for CNBSS 1 (women aged 40-49 years) is similar to that in the Gothenburg trial (women aged 39-59), Stockholm trial (women aged 39-65)

Figure 33.4 randomized trials on breast cancer screening showing number of women enrolled, age at inclusion, intervention time, attendance rate, definition of advanced breast cancer, number of women with advanced breast cancer and cumulative incidence rate of breast cancer (Table 2 in (6))

Table 2. Summary Results of Randomized Trials on Mammography Breast Cancer Screening

Trial	Age at Entry (years)	No. of Patients in Trial for Analysis: Intervention	No. of Patients in Trial for Analysis: Control	Median Duration of Trial (years)*	Attendance Rate, First Round (%)	Definition of Advanced Breast Cancer	No. of Patients With Advanced Breast Cancer: Intervention	No. of Patients With Advanced Breast Cancer: Control	Cumulative Incidence of Advanced Breast Cancer per 1,000†: Intervention	Cumulative Incidence of Advanced Breast Cancer per 1,000†: Control	Relative Risk in Intervention Group v Control Group: For Advanced Breast Cancer†	Relative Risk in Intervention Group v Control Group: 95% CI	Relative Risk in Intervention Group v Control Group: For Breast Cancer Mortality‡	Relative Risk in Intervention Group v Control Group: 95% CI
Greater New York Health Insurance Plan (HIP), United States	40-64	30,239	30,256	5	67	Stage II+	160	188	5.29	6.21	0.85	0.69 to 1.05	0.83	0.70 to 1.00
Two-County Trial (TCT), Sweden	40-74	77,080	55,985	8	85	Stage II+	524	555	6.80	9.91	0.69	0.61 to 0.78	0.68	0.59 to 0.80
Malmo Mammographic Screening Trial (MMST), Sweden	45-70	21,088	21,195	14	74	Stage II+	190	231	9.01	10.90	0.83	0.68 to 1.00	0.82	0.67 to 1.00
Stockholm trial, Sweden	39-65	40,318	19,943	5	82	Stage II+	172	97	4.27	4.86	0.88	0.68 to 1.12	0.91	0.65 to 1.27
Goteborg trial, Sweden	39-59	21,650	29,961	4.8 (for age 40-49 years); 7.0 (for age 50-59 years)	84	≥ one node involved	85	144	3.93	4.81	0.80	0.61 to 1.05	0.77	0.60 to 1.00
Trial on women 40 years old at entry, United Kingdom	39-41	53,884	106,956	7	68§	Size ≥ 20 mm	171	386	3.17	3.61	0.88	0.73 to 1.05	0.83	0.66 to 1.04
National Breast Screening Study-1 (NBSS-1), Canada	40-49	25,214	25,216	5	86	Size ≥ 20 mm	111	115	4.40	4.56	0.97	0.74 to 1.25	0.97	0.74 to 1.27
National Breast Screening Study-2 (NBSS-2), Canada	50-59	19,711	19,694	5	87	Size ≥ 20 mm	114	136	5.78	6.91	0.84	0.65 to 1.07	1.02	0.78 to 1.33

*Period during which screening was offered to women in intervention group and not to women in control group; follow-up may have lasted for several more years.
†Relative risk and 95% CI calculated by us except for the Goteborg trial, for which we used the relative risk in Bjurstam et al.[15]
‡See Patients and Methods for selection of relative risks. Number of breast cancer deaths after 13 to 18 years of follow-up for all trials.
§Overall, 81% of women attended at least one mammography round.

Two tales in the debate

The following statements illustrate the debate:

On the complaint side:

> *"You just need to look at the end. Result and you say they couldn't have been well-trained. They couldn't have had good equipment, and you don't increase mortality, increase mortality. If you have good randomization, and no contamination. should have been absolutely zero difference. That´s what I am saying. This is crimy medicine. Don't increase mortality, for god sake."*

On the trialists side:

> *"We were interested in what experts would tell us. And the experts in many ways, betrayed us. Because they didn't like our results. They didn't like our results.*
>
> *Anybody will believe anything if they don´t like your results"*

Tumor size and interval cancer

Tumor size and rate of interval cancer may be an indirect indicator of quality, and are used as quality indicator in the European quality guidelines (9).

Tumor size

Tumor size may indicate the quality of the mammography. We have presented the tumor size as published in the Two-county and CNBSS by age group (3, 10-12). As shown, there is no strikingly difference in the screening arms (using mammography only) in size distribution. Tumors less than 10 mm is 11.0% for women aged 40-49 years (21.7% for women age 40-74) in the two county study, and 19.0% for women aged 40-49 in CNBSS (27.1% for women aged 50-59 years) (3, 10-12; table 33.1).

Table 33.1 Tumor size distribution in the screening and control arm in the Swedish Two-county study and the CNBSS by age and screening method (3, 10-12)

	Sweden			Canada					
	First screen[1] n (%)	Control group[1] n (%)	40-49[**,2] Ma n (%)	50-59[4] Ma n (%)	50-59[4] Ma+ n (%)	50-59[*,4] Control n (%)	40-49[3] Ma n (%)	40-49[3] Ma + n (%)	40-49[3] Control n (%)
Total n	284	590	17	48	70	64	21	66	58
1-9 mm	74 (21.1)	42 (7.1)	2 (11.0)	13 (27.1)	4 (5.7)	3 (4.7)	4 (19.0)	3 (4.5)	7 (12.1)
10-14	84 (29.6)	91 (15.4)	6 (37.0)	5 (10.4)	13 (18.6)	9 (14.1)	4 (19.0)	13 (19.7)	6 (10.3)
15-19	56 (19.7)	116 (19.7)	3 (16.0)	14 (29.1)	13 (18.6)	13 (20.3)	5 (23.8)	16 (24.2)	12 (20.7)
20 and larger	70 (24.6)	287 (48.6)	5 (29.4)	10 (20.8)	32 (45.7)	36 (56.3)	4 (19.0)	28 (42.4)	29 (50.0)
Unknown	1			6 (12.5)	8 (11.4)	3 (4.7)	4 (19.0)	6 (9.1)	4 (6.9)

Ma: mammography only, MA+: mammography and clinical breast exam

[**]No information on the control group for women aged 40-49 years

[*]Control group were offered yearly physical exam

1. Duffy SW, L, Fagerberg G, Gad A, Grontoft O, South MC, Day NE. Breast screening, prognostic factors and survival – results from the Swedish two-county study. Br.J.Cancer 1991; 64:1133-1138

2. Tabar L, Duffy SW, Burhenne LW. New Swedish breast cancer detection results for women aged 40-49. Cancer 1993;72 (4) 1438-48.

3. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002;137:305-312

4. Miller A, To T, Baines CJ, Wall C. Canadian national breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000; 92 (18)1490-1499

Interval cancer rate

Interval cancer rates may be associated with quality of mammography and interpretation of mammograms, but there may be factors related to risk of cancer among the women screened (included breast density and use of hormone therapy), and length of interval between two screens.

The interval cancer rates in the CNBSS-1 screening arm (invasive plus in situ) were highest in the first and second years, 0.75 per 1000 (19 cancers) and 0.71 per 1000 (16 cancers) for women aged 40-49, respectively (Figure 33.5) (13). The interval cancer rate in CNBSS are higher than in the Two-county study (13).

Figure 33.5 Comparison of interval cancer rates in CNBSS and Two-County study by age-group, tab 5 in (13)

Table 5. Rates per 1000 Women for Interval Cancers Ascertained in Two Screening Programs

Screen	Age at Entry			
	40 to 49 Years		50 to 59 Years	
	NBSS*	Swedish Two-County Study†	NBSS*	Swedish Two-County Study†
Round 1	0.75	0.48	0.76	0.19
Round 2	0.71	0.69	0.57	0.35
Round 3	0.36	0.23	0.46	0.43
Round 4	0.46	NA‡	0.52	NA
Round 5	0.64	NA	0.51	NA

* Mammography allocation only. NBSS = National Breast Screening Study.
† Only interval cancers occurring in the first 12 months after a screening examination are included.
‡ NA = not available.

Some claim the interval cancer rate in the CNBSS is high (14):

> "*The overall false negative rate for 108 interval cancers in the CNBSS was 2.5 per 1000 women, which contrasts with 1 .93 per 1000 women for the Breast Cancer Detection Demonstration Project, 1 .45 per 1000 for the Health Insurance Plan of Greater New York, and 0.92 per 1000 for the Screening Mammography Program of British Columbia (SMPBC). The high interval cancer rate of the CNBSS is again apparently related to the poor mammographic technique and inadequate interpretation*" (14).

We are not sure how the authors came to their interval cancer rates, as we cannot identify any published data that gives this interval cancer rate (14).

There are guidelines for what is an acceptable level of interval cancer in screening programs in Europe (9) the interval cancer ratio. The interval cancer ratio is the ratio of interval cancers divided by the rate of non-screening cancers (in a randomized trial the control arm), and is used as a marker of quality (9).

In the CNBSS, the rate of interval cancer in the first screening round was 0.75 per 1000 women screened aged 40-49 years and 0.76 per 1000 women screened aged 50-59 years, and breast cancer rates in the control arm was 2.46 per 1000 women aged 40-49 years (13, 5), and 3.45 per 1000 women aged 50-59, in the control arm (15). This gives an interval cancer ratio of 30%.

In the two-county study, interval cancer ratio is 92% for women aged 40-49 years and 20% for women aged 50-59 (16), and the interval cancer ratio is 21% for women aged 39-49 years in the Gothenburg trial (17).

The interval cancer ratio is acceptable if it is 30% within the first year and 50% in the second year, and the desired level is below 30% and 50% (9).

Even if the rate of interval cancer were higher in the CNBSS than in the Two-county study (Figure 33.5) (13), to interpret what is a high rate, the interval cancer rate should be compared to the cancer rate in the control group. The result from CNBSS is within what the European screening programs have achieved in recent years, with interval cancer ratio varying from 22% to 51% (18, 19).

References

1. Tabár L, Gad A, Holmberg LH, et al. Reduction in mortality from breast cancer after mass screening with mammography. Lancet 1985, 829-832.
2. Larsson LG, Andersson I, Bjurstam N, Fagerberg G, Frisell J, L, Nyström L. Updated overview of the Swedish randomized trials on breast cancer screening with mammography: Age group 40-49 at randomization. Monogr Natl Cancer Inst 1997;22:57–61.
3. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up. Ann Intern Med 2002;137:305-312.
4. Miller A B, Wall C, Baines C J, Sun P, To T, Narod S A 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 doi:10.1136/bmj.g366.
5. 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: 1459-76.
6. Autier P, Hery C, Haukka J, Boniol M, Byrnes G. Advanced Breast Cancer and Breast Cancer Mortality in Randomized Controlled Trials on Mammography Screening. J Clin Oncol 2009, 27:5919-5923.
7. American Joint Committee on Cancer: AJCC Cancer Staging Manual (ed 6). New York, NY, Springer, 2002, pp 171-180.
8. International Union Against Cancer (UICC): TNM Classification of Malignant Tumours (ed 6). New York, NY, Wiley-Liss, 2002.
9. European guidelines for quality assurance in breast cancer screening and diagnosis. EUREF Accessed Feb 24 2024 at https://euref.org/european-guidelines/5th-edition/
10. Duffy SW, L, Fagerberg G, Gad A, Grontoft O, South MC, Day NE. Breast screening, prognostic factors and survival – results from the Swedish two-county study. Br J Cancer 1991; 64:1133-1138.
11. Tabar L, Duffy SW, Burhenne LW. New Swedish breast cancer detection results for women aged 40-49. Cancer 1993;72 (4) 1438-48.
12. Miller A, To T, Baines CJ, Wall C. Canadian national breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000; 92 (18)1490-1499

13. Baines C. The Canadian national Breast Screening Study: A Perspective on criticism. Ann Intern Med 1994; 120: 326-33.
14. Burhenne LJ, Burhenne HJ. The Canadian national breast screening study: A Canadian Critque. Am J Radiol 1993; 161: 761-63.
15. 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: 1477-88
16. Tabar L, Duffy SW, Yen MF, Warwick J, Vitak B, Chen HH, Smith RA. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. J Med Screen 2002;9:159–162.
17. Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L. et al. The Gothenburg Breast Screening Trial. Cancer. 2003;97(10):2387-96. doi: 10.1002/cncr.1136
18. Andersen SB, Tornberg S, Lynge E, Von Euler-Chelpin M, Njor SH. A simple way to measure the burden of interval cancer in breast cancer screening. 2014; BMC Cancer 14:782 (2014).
19. Steponaviciene L, Vincerzevskiene I, Briediene R, Urbonas V, Vanseviciute-Petkeviciene R, Smailyte G. Breast Cancer Screening Program in Lithuania: Interval Cancers and Program Sensitivity After 7 Years of Mammography Screening. Cancer Control. 2019 Jan-Dec; 26(1): 1073274819874122.

Was CNBSS an outlier?

We did not perform a systematic review of all randomized trials of mammography screening, but relied on previously published meta-analyses and publications from individual trials.

Overview of randomized breast cancer screening trials

We have summarized randomization methods, calendar year, age-group included, screening intervals, number of screening rounds, number of women in the screening and control arms, the intervention in the screening and control arms, the views used, and the attendance rate in the randomized trials of mammography screening (Table 34.1) (1-3).

Table 34.1 Summarizing randomization methods, calendar year, age-group included, screening intervals, number of screening rounds, number of women in the screening and control arms, the intervention in the screening and control arms, the views used, and the attendance rate in the randomized trials of mammography screening. Data abstracted from different publications (1-3)

Study	Random-ization method	Calendar year	Age group (y)	Screen interval (m)	Number of rounds	Number of women in the screening arm	Number of women in the control arm	Intervention	Control	View	Att. rate (%)
HIP	Individual	1963	40-64	12	4	31,092	31,000	Mx + CBE	UC	2: ML + CC	65.0
Malmö	Individual	1976	45-70	18-24	8	21,088	21,195	Mx	UC	2,1: MLO + CC, MLO	74.0
Two county	Cluster	1977	40-74	24-33	4	78,085	56,782	Mx	UC	1: MLO	89.2
Edinburgh[1]	Cluster	1979	45-64	12	4	23,226	21,904	Mx + CBE (2y) PE (2 y)	UC	2,1: MLO + CC, MLO	61.3
CNBSS-1	Individual	1980	40-49	12	4-5	25,214	25,216	Mx + CBE	UC	2: ML (MLO) + CC	89.4
CNBSS-2	Individual	1980	50-59	12	4-5	19,711	19,694	Mx + CBE	CBE	2: ML (MLO) + CC	90.4
Stockholm	Individual	1981	40-64	24-28	2	40,318	20,000	Mx	UC	1: MLO	80.7
Gothenburg[2]	Cluster; Individual	1982	39-59	18	5	21,638	29,961	Mx	UC	2,1	84.1
UK Age	Individual	1991	40-49	12		53,890	106,971	1-Mx / 2-MM[a]	UC	2,1: MLO + CC, MLO	68.0
Estimate of overall attendance rate											78.7

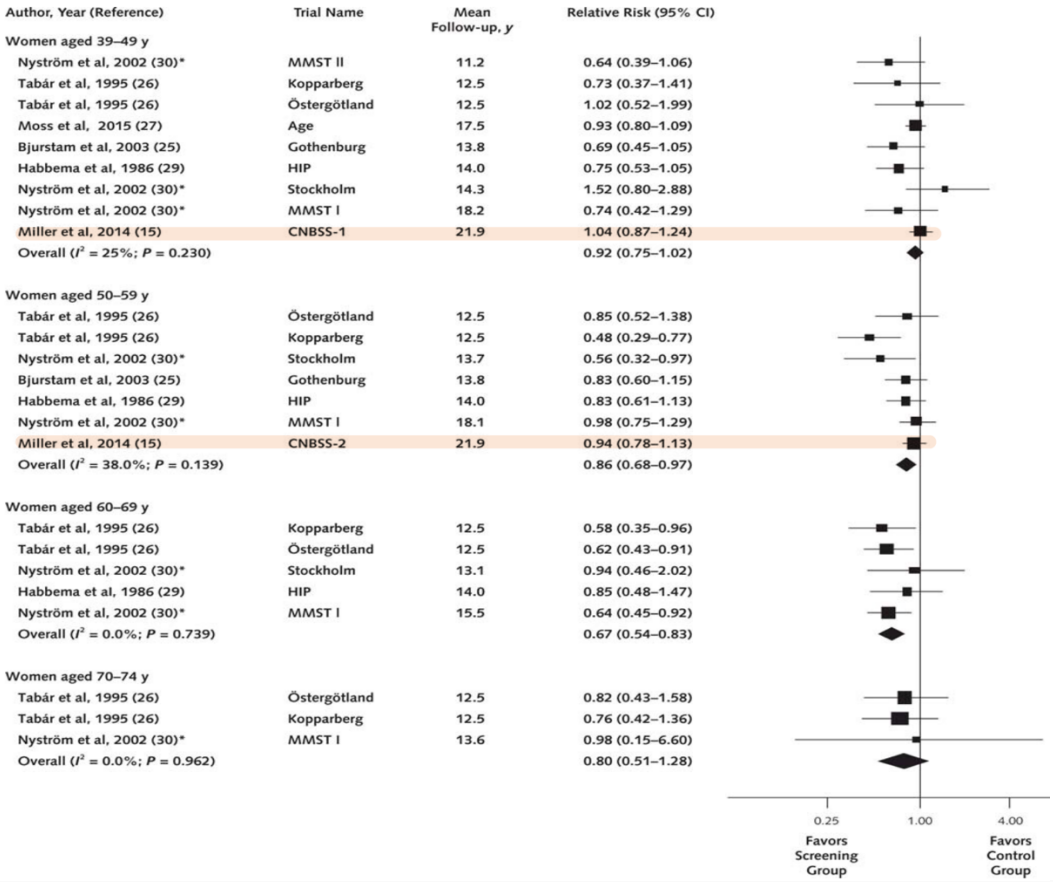
Abbreviations: y: year; M: months; Mx: mammogram; CBE: clinical breast examination; UC: usual care; ML: mediolateral view; MLO: mediolateral oblique; CC: craniocaudal

2: two-view mammogram; 2,1: two-view mammogram in the first screening, single-view thereafter; 1: single-view mammogram
[1]The Edinburgh trial: offered mammography and CBE year 1, 3, 5, 7; and CBE alone year 2, 4, 6
[2]We were not able to find the views used in the Gothenburg trial

Breast cancer mortality

Fair-quality evidence from a meta-analysis of mammography trials indicated relative risks (RRs) comparing breast cancer mortality in the screening- and control arms, of 0.92 (95% CI, 0.75-1.02) for women aged 39 to 49 years (9 trials; 3 deaths prevented per 10 000 women over 10 years); 0.86 (95% CI, 0.68-0.97) for those aged 50 to 59 years (7 trials; 8 deaths prevented per 10 000 women over 10 years); 0.67 (95% CI, 0.54-0.83) for those aged 60 to 69 years (5 trials; 21 deaths prevented per 10 000 women over 10 years); and 0.80 for those aged 70 to 74 years (95% CI, 0.51-1.28) (3 trials; 13 deaths prevented per 10 000 women over 10 years). All-cause mortality was not reduced with screening. (4)

Figure 34.1 Forest plot of relative risk of breast cancer mortality comparing screening arm with control arm in different screening trials by age-group at inclusion using the longest follow-up times available (figure 1 in (4)). CNBSS is marked.

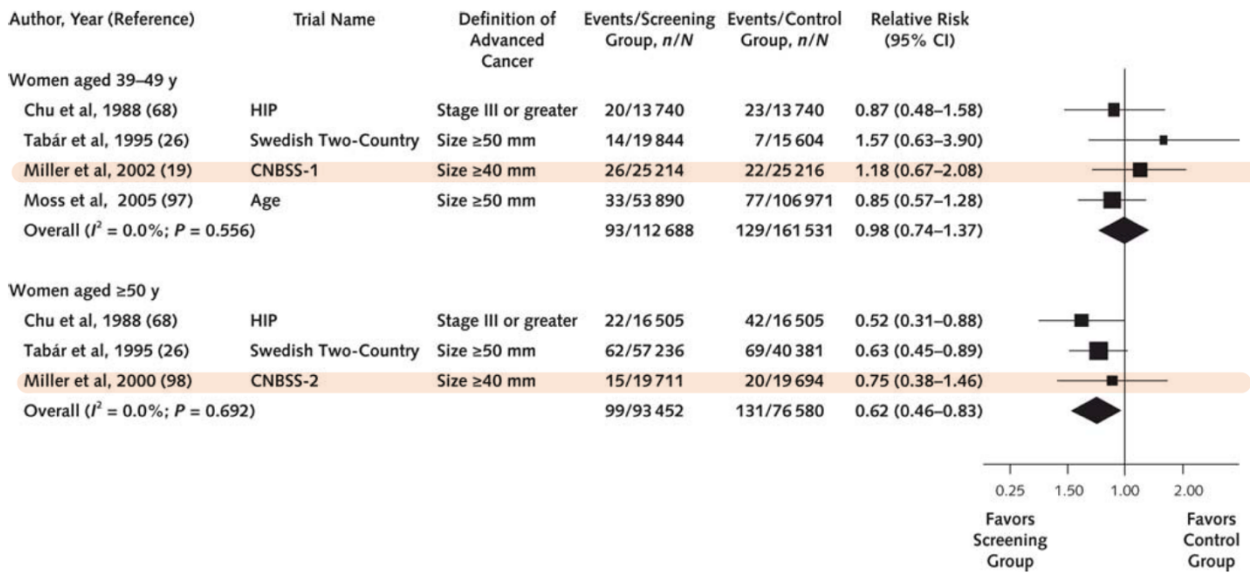


CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of New York; MMST = Malmö Mammographic Screening Trial.

Advanced stage breast cancer

The RR of advanced stage breast cancer indicated no difference with screening for women aged 39 to 49 years (RR, 0.98; 95% CI, 0.74-1.37) but reduced risk for those aged 50 years or older (RR, 0.62; 95% CI, 0.46-0.83) (Figure 34.2) (4).

Figure 34.2 Forest plot of relative risk of advanced cancer comparing screening arm with control arm in different screening trials by age-group at inclusion using the longest follow-up times available (figure 1 in (4)). CNBSS is marked.



All trials were classified as fair quality in the meta-analysis (4).

The meta-analysis included the most severe disease categories available from the trials, included stage III and IV disease (i.e., regional and metastatic, respectively), size 40 to 50 mm or greater, or 4 or more positive lymph nodes. (4).

Overdiagnosis

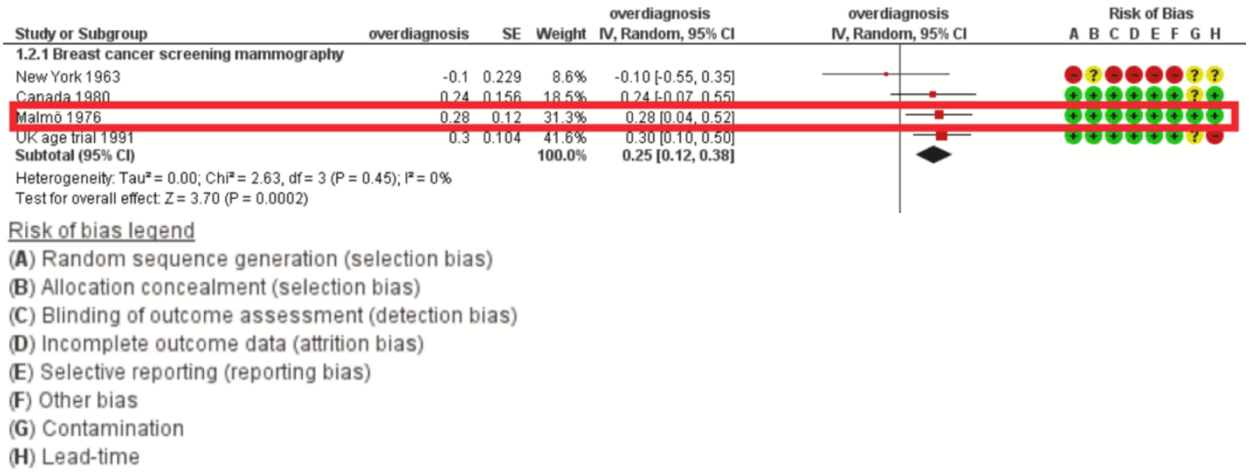
A meta-analysis of overdiagnosis (detection of breast cancers that would never progress to cause symptoms and/or death during an individuaľ s lifetime) included four randomized screening trials that reported cancer incidence for screened and control arm participants (, and the number of screen-detected cancers (figure 34.3). The meta-analysis reported risk of bias (figure 34.4) and found estimates of overdiagnosis ranging from −10 to 30% associated with breast cancer screening with mammography (figure 34.4). Overdiagnosis is the most serious harm of cancer screening and was estimated to be 25% (95% CI, 12%-38%) (5).

Figure 34.3 Incidence of cancer in screening trials that reported cancer incidence for screened and non-screened participants, and the number of screen-detected cancers (table 3 in (5)).

Table 3. Incidence of cancer.

Trial name or location, and year	Group size (Control/ intervention)	Control group cumulative incidence	Intervention group cumulative incidence (at end of follow-up)	N screen-detected	Estimation of over-diagnosis (Overdiagnosis/ 1000 screen-detected cancers)
Breast cancer screening with mammography					
Canada 1980[25], [26], [27]	44,910/ 44,925	3133	3250	484	240/1000
Malmö 1976 cohort 1 and 2 * [28], [29]	13,133/ 13,107	698	780	298	280/1000
New York 1983[30], [31]	31,092/ 31,092	439	426	132	- 98/1000
UK age trial 1991#[32],	53,883/ 106,953	821	482	229	299/1000

Figure 34.4 Forest plot of overdiagnosis in different screening trials and risk of bias associated with each included trial. Risk of bias were assessed for six domains, and the trial with no risk of bias was highlighted (red box) by the authors (figure 2 in (5)).



As shown, CNBSS is not an outlier in breast cancer incidence, overdiagnosis, mortality, tumor size, nor advanced cancer (Chapter 33 and 34).

References

1. Paci E, Alexander FE. Study design of randomized controlled clinical trials of breast cancer screening. JNCI Monographs 1997; 1997 (22):21-25
2. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database of Systematic Reviews 2013, Issue 6. Art. No.: CD001877. DOI: 10.1002/14651858.CD001877.pub5. Accessed 18 February 2024.
3. Chen THH, Yen AMF, Fann JCY, Gordon P, Chen SLS, Chiu SYH, et al. Clarifying the debate on population-based screening for breast cancer with mammography: A systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. Medicine 2017 Jan; 96(3): e5684.
Published online 2017 Jan 20. doi: 10.1097/MD.0000000000005684
4. Nelson HD, Fu R, Cantor A, Pappas MA, Dages M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 US preventive Service Task Force recommendation. Ann Intern Med 2016; 164:244-55. doi:10.7326/M15-0969.
5. Voss T, Krag M, Martiny F, Heleno B, Jørgensen KJ, Brodersen JB. Quantification of over diagnosis in randomized trials of cancer screening: an overview and re-analysis of systematic reviews. Cancer Epidemiology. 2023:84 102352. doi: 10.1016/j.canep.2023.102352

Burden and standard of proof

Judiciary system

The burden of proof is a standard of evidence or facts needed to satisfy a legal decision. It is associated with the principle "*semper necessitas probandi incumbit ei qui agit*", translated to "the necessity of proof always lies with the person who lays charges" (1-3).

The standard of proof is the degree to which a party must prove its case to succeed. The burden of proof is the requirement or threshold to satisfy that standard (4).

Inquisitory system vs. adversarial or accusatorial system

There are two legal systems: the inquisitorial, and the adversarial or accusatorial system (5, 6). In the adversarial system, two or more opposing parties gather and present evidence and arguments to a judge or a jury. In the inquisitory system, the judge is responsible for gathering evidence which is necessary to resolve a case.

In the adjudication process the opposing parties compete against each other, and the judge serves as a referee to ensure fairness to the accused. The adversarial system assumes that the best way to get to the truth of a matter is through a competitive process to determine the facts (7).

The inquisitorial system is associated with the civil legal system. It is characterized by extensive investigation and interrogation (7). In an inquisitorial hearing, the judiciary must be highly proactive, and is explicitly tasked with positively ascertaining the truth, rather than enabling the parties to do so (8).

The inquisitorial process can be described as an official inquiry to ascertain the truth, whereas the adversarial system uses a competitive process to determine the facts (7). The judge questions the witnesses and/or the defendant or plaintiff in an inquisitory system, while he/she does not in an adversarial system (7). The judge has more power in the inquisitorial system than in the adversarial system.

The concept of burden of proof has limited meaning in an inquisitorial system, as it is the court or judge that has the responsibility to establish the facts and determine the outcome (8).

Beyond a reasonable doubt or preponderance of the evidence

In a criminal case of law, the burden of proof is that the proof given is beyond a reasonable doubt. There is no strict definition on the limit, amount, or credibility of proof, but the burden of proof is lower in a civil case than in a criminal case.

More likely than not

In a civil case, the burden of proof is on a `preponderance of the evidence´, meaning the balance of probabilities where the evidence laid forward is such that the chance of it being true is greater than the chance of it not being true, and more than chance, which is above 50%.

The probability is that the evidence has a higher chance of being true than not being true. Or simply that the evidence is "*more likely than not*" to be true. A more precise statement is that "the weight [of the evidence, including in calculating such a percentage] is determined not by the amount of evidence, but by its quality."(9, 10).

"Proof" as in the phrase "standard of proof" and "probabilities" in the phrase "balance of probabilities" are words which go naturally with the concept of evidence relating to fact but are less perfect with evaluative assessments (8).

Clear and convincing evidence

In the US court system, there is a burden of proof that is in the middle of preponderance of the evidence and evidence beyond a reasonable doubt; called clear and convincing evidence. The level of evidence needed in cases to deem for example patients to involuntary hospitalization requires clear and convincing proof (evidence). Clear and convincing proof means that the evidence must be more probable to be true than not, and in addition the one who tries the fact must have a firm belief or conviction in its factuality. The concept of clear and convincing evidence has not been officially adopted in Canada (9).

Circumstantial evidence

Circumstantial evidence is not drawn from direct observation of a fact (11). Different pieces of circumstantial evidence that may enlighten different sides of a matter or a chain of circumstances pointing to existence of fact may be required to draw a conclusion (12). Circumstantial evidence may not be enough to prove a truth but may be strengthened if the evidence is tied to an explanation or inference supporting a finding.

In many cases, circumstantial evidence is all which is available of evidence. The judge or the jury must piece the evidence together and determine whether it leads to a reasonable conclusion about the fact, or not.

Direct evidence

Direct evidence is direct proof of a fact such as video surveillance of a crime or an eyewitness. Direct evidence is more reliable than circumstantial evidence and does not need inference about the fact. However, eyewitnesses may not always be reliable, especially when relying on the memory of an incident several years ago (13, 14).

Eyewitness

Memory is malleable, for example eyewitnesses may be led to believe that they saw a stop sign when they actually saw a yield sign (15). Eyewitness misidentifications are known to have

played a role in 70% of the 353 convictions that have been overturned on the basis of DNA evidence since 1989 (13). Memory may be contaminated by emotions, beliefs or knowledge which appeared after the fact, and accuracy and completeness of a person recall decreases over time. Factors, like the passage of time and exposure to misleading information after the event, can distort stored memories, thereby potentially reducing the reliability of eyewitness testimonies (16). Memory is reconstructive by nature, prone to alteration and influence, and may indeed, trick us.

Hearsay evidence

In Canada, in most situations, the judge will not allow testimony based on what a witness has heard from another person – this is called **hearsay** evidence (17, 18).

Our approach to evidence

Although the panel members are not lawyers, nor is this a civil or criminal court, or a court of law, we have been guided by the described evidence systems and applied the following principles when we have decided if ***new evidence has a credible scientific impact*** *on the reliability of the CNBSS* (8):

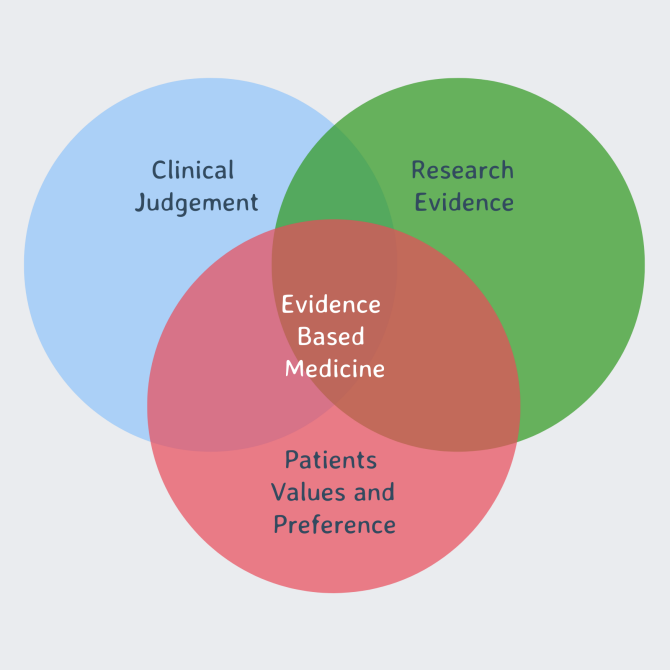
> *1. Where the matters in issue are facts, the standard of proof required in non-criminal proceedings is the preponderance of probability.*
>
> *2. The balance of probability standard means that the panel must be satisfied that the event in question is more likely than not to have occurred.*
>
> *3. The balance of probability standard is a flexible standard. This means that when assessing this probability the panel will assume that some things are inherently more likely than others. This concept was memorably encapsulated by Lord Hoffmann, when he observed:*
>
> > *'It would need more cogent evidence to satisfy one that the creature seen walking in Regent's Park was more likely than not to have been a lioness than to be satisfied to the same standard of probability that it was an alsatian.'*
>
> *4. The more serious the allegation the less likely it is that the event occurred, and thus the stronger and more cogent should be the evidence before a court determines that on the balance of probabilities, the event did occur. (8)*

These are the principles laid down to apply in non-criminal proceedings in the general civil courts. The balance of probabilities is the appropriate standard (8).

Evidence based medicine

Evidence is also at the core of medicine, and evidence-based medicine, which is integration of best research evidence with clinical expertise and patient values, figure 35.1 and is the acknowledged theory of how to use evidence to make decisions in medicine.

Figure 35.1 Evidence based medicine (EBM) (19)



The panel has considered the following evidence, that touches clinical judgement, and research evidence in EBM:

- complaint material
- previous criticism
- rebuttal from the trialists
- previous review
- results from CNBSS that may or may not indicate violation of randomization and poor quality
- close to 19 hours of interviews.

We have also considered possible conflicts of interest when making our decision. Our evaluation of conflicts of interest was guided by conflict-of-interest policies for clinical guidelines. Evidence used in guidelines is valued and graded, and individuals included on guideline panels should preferably not have any conflicts of interest.

Disclosure of conflict of interest

The Canadian Task Force on Preventive Health Care

The Canadian Task Force has set up guidance on how to handle conflict of interest that adhere to the Guidelines International Network (GIN) principles for disclosure of interest and management of conflict of interest (COI) (20, 21).

"*Another underpinning principle is that the Task Force distinguishes between declaration of interests and conflict of interest. The declaration of one or more interests does not necessarily indicate that a real or potential conflict of interest is present. The Task Force carries out an evaluation of declared interests to determine whether they represent a conflict of interest*" (21).

Financial and non-financial interests

Financial interest is when an individual or organization (individual representing the organization) receives or has received income or other form of monetary support or financial benefit that is related or relevant to, or could reasonably be perceived to be impacted by, the topic (21).

Non-financial interest is when an individual or organization (if an individual is participating on behalf of an organization) has academic, professional, or other personal interests or relationships that are related or relevant to the topic. The process is shown in figure 35.2 and starts with the individual declaring his or her interests related to the topic from the past 3 years in a standardized form. The assessment is judged by an oversight committee and declared interests do not automatically imply a conflict of interest. The Task Force avoids inclusion of individuals with a conflict of interest (21).

Figure 35.2 General overview of the Task Force procedures for the disclosure of interests, and the assessment, identification, and management of conflict of interest (COI) (fig 1 in (21)

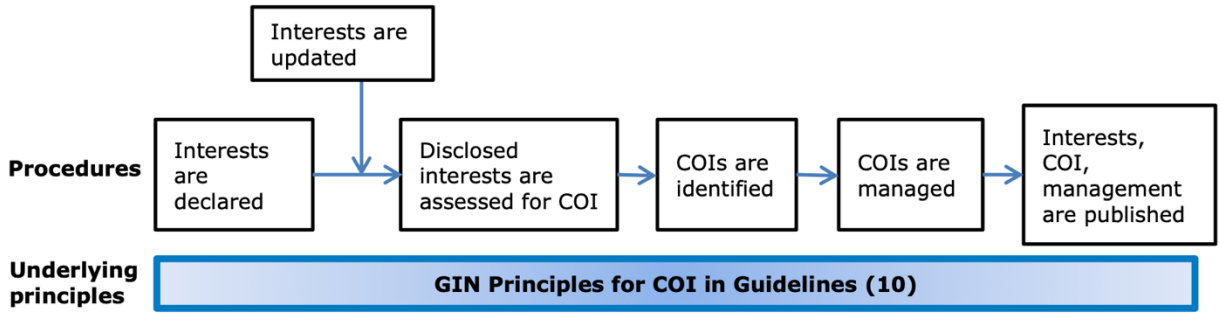


Figure 1. General overview of Task Force procedures for the disclosure of interests, and the assessment, identification, and management of COI.

Content experts

Clinical and content experts are invited to serve as external advisors to guideline topic working groups. They attend meetings, review documents and the final guideline, but are not members of the working group and do not have impact or vote on the direction and strength of the recommendations in the guideline. Content experts are allowed to have conflict of interest if the interests are not too extensive to potentially impair their credibility of input. However, if

reviewers are highly knowledgeable and reviewers without such conflicts are unavailable, they may still serve as content experts. In these cases, an appropriate balance of opinion from those who provide feedback is mandatory (21).

Stakeholders

As for content experts, conflict of interest for stakeholders are assessed and managed, and also anticipated, particularly where they represent advocacy organizations. Stakeholders are asked for their perspective and are sought to ensure Task Force has considered a range of views and implications. Conflict of interest among stakeholders do not exclude them, but they do not vote on recommendations and their declared interests and COIs are taken into account when interpreting their input (21).

The panel's evaluation of evidence

The panel was asked to assess ***new*** information given by witnesses in documents brought to the panel, and to interview such individuals that may have information on the CNBSS (Mandate, *Chapter 1*). The panel was specifically asked to interview the following individuals:

> Professor Martin Yaffe (corresponding author for the "allegation")
> Professors Emeriti Anthony Miller and Cornelia Baines (CNBSS principal investigators)
> [redacted] (*Part V; Chapter 13-28*).

The panel was asked to interview specific individuals *who have come forward with information that may have an impact on randomization, recruitment of symptomatic women, and poor radiographic image quality in the Canadian National Breast Screening Study (CNBSS).*

We, the panel, were asked to deliver a report that detailed our assessment of whether this new information would have a credible scientific impact on the reliability of the CNBSS' published.

New information

We assessed all new information in two different ways: Firstly, through documents brought to the panel (**complaint material**) and secondly by conducting **interviews** with individuals that may have information on the CNBSS (figure 35.3).

The complaint material is presented in *Chapter 6* of the report, and a summary of the complaint material is presented in *Chapter 30*.
A detailed transcript of the dialog with the 15 interviewees is presented in *Part V, Chapter 13-28.* A summary of the transcribed interviews including comments to give perspective to the dialog and quotes from the interviewees is presented in *Chapter 29*.

Previous criticism and review

In order to assess what was new information, we also reviewed previous criticism and related rebuttals by the trialists, the previous review, and the discussion that followed.
This is presented in *Chapters 9, 10, 11, and 12*. Summary of previous criticism and review is presented in *Chapter 30*.

Published findings from CNBSS

As it is more than 40 years since the CNBSS was initiated, we consider that that new information may not be reliable because it is dependent on human memory that is malleable. Therefore, we also assessed the published results from the CNBSS to find evidence for explanation or inference supporting a witness claim.

Our work is illustrated in figure 35.3

Figure 35.3 The panel assessed and reviewed new information: Complaint material and Interviews; and previous information: Previous criticism and previous review, and results from CNBSS publications.

New evidence

Hearsay

Most of the new information presented in the complaint material is hearsay, and the panel has treated this evidence as in the Canadian judiciary system (17, 18). We did not consider hearsay as credible evidence. When we heard hearsay during our investigation, we checked whether the claims had any explanation or inference supporting the claims in the published documents, including assessment of the published results from the CNBSS and other trials.

As an example, it turned out during our investigation that one of the informants in the Summary review actually did not work at the CNBSS and was thus not informative for our mandate. This illustrates how memory may trick people to remember things they have not experienced.

Further, our interview with one of the new witnesses (a medical imaging technologist) revealed that she had not directly witnessed any violation of the allocation but only had heard rumors of it. She "*knew*" or "*had heard*" that women with lumps routinely were assigned to the screening arm but had never observed it.

Two eyewitnesses were identified in the Summary report (22). One research coordinator, Ms. [redacted] stated with certainty according to the Summary report that non blinded allocation was done for women with palpable lumps. However, she also provided written information saying, "*I was not witness to that happening"* (22), (*Chapter 6*), so it is unclear if she actually was an witness or not. We were not able to interview her, because she cancelled several of our scheduled interviews, and ultimately informed the panel that she was unable to meet with us "*any time in the future*" (email Jan 9th 2023).

A similar instance happened in the previous review of the CNBSS, where an eye-witness, a medical imaging technologist, was said to have personal knowledge that could have compromised the randomization process but declined to be interviewed (*Chapter 6*) (23).

Eye-witness

We interviewed another eye-witness ([redacted]) who provided information in the Summary report. She was also mentioned by Dr. Yaffe in his interview. She is a medical imaging technician. What she said was consistent throughout her interview. She said (*Chapter 14-28, Chapter 29*):

> "*All the patients saw the nurse examiner first. Once they saw the nurse examiner. If the nurse examiner felt an anomaly, if that person wasn't on my list, it didn't matter, they became a person on my list, and instead somebody else was removed"*.

She explained the "book", presumably the randomization book, was not filled out at the start of the day. She had a separate list, that she had to give back to the coordinator:

> "*The book wasn't filled out at the start of the day, for who was completed from my component. There was a list of patients that were supposed to be coming in, so let's say 24, and if somebody didn't show, it would say no-show beside it. But the ones that were for the mammo component, for the days that I worked, many people who had the mammo only got filled in once that mammo was done. That list wasn't pre filled for them. They gave me a list, but then that list disappeared. I had to give the list back, so it would disappear*."

> "*Afterwards they would write down the ones that I did in my book. Afterwards. The book that you're talking about like the page there, wasn't anybody written in there until I gave them my list.*"

> "*They* [women] *literally left the nurses room and came directly into my room. The nurses said they were going to have to have a mammogram, anyway. So they were putting them in and just taking somebody else out*".

[redacted]´s explanation does not fit with how Dr. Baines and others explained the randomization books (Box 35.1). Her shift usually started at noon, and one may assume that routines which are in line with Dr. Baines´ description happened before her shifts, which all started at noon:

> "*I was a brand new graduate in* [redacted] *I was hired for a 3 month contract. I actually only stayed for about 10 shifts because it was 2 days a week. It was Tuesday and Thursdays working 12 to 8, as the mammography technologist for the screening program. I didn't stay for the full 3 months*"

Box 35.1 How Dr. Baines explained the randomization and randomization lists (24)

Randomization	Women individually randomized Randomization lists in 4 separate books (age 40-44, 45-49, 50-54, 55-59) Each center Center coordinator entered the date and name on the first available line in appropriate book dependent on age, and assigned the woman her ID and randomization allocation ID and allocation was entered to all chart forms "*Skipping a line to achieve a desired allocation was not feasible because she could not predict when the next appropriately aged woman would arrive to fill the skipped slot."* All original randomization sheets were submitted to the central coordinating office where all sheets were examined for suspicious entries, inappropriate dates and lack of congruence with participant records Examiner told each woman her

[redacted] came forward a long time after the study, at a meeting at the [redacted] meeting in [redacted], where [redacted] gave a talk about concerns of nonrandom allocation in the CNBSS, more than 35 year after she worked at the CNBSS.

[redacted] has expressed clear opinions about mammography screening, and has been working with mammography her whole career, which may result in a conflict of interest as judged by the standards of the Canadian Task Force (21):

> "*I mean, I've done mammography for as long as* [redacted] *to now still. And I see how it can be such a benefit to catch it when it's so much smaller. So then it doesn't have a tiny chance to spread, to infiltrate, to grow to a stage 3 to stage 4. The whole idea is to try and catch it as small as we can, so trying to dissuade these women from having mammograms pre age 50.*
>
> *My career started on a bad note. I didn't like what I saw. I'm still working. I hope to work for another couple of years at least. I'd like to see in my career the wrongs of this study corrected*".

[redacted] only worked 10 shifts at once center in the CNBSS, in 1985, which is a long time ago. How factors, like the passage of time and exposure to information after her participation in the CNBSS have affected her memory is uncertain. It is well-known, that passage of time and information may distort stored memories and reduce the reliability of eyewitness testimonies (16).

Discussions and speculations over the years may have influenced the witnesses and what they describe they saw more than 35 years ago. They may also have been exposed to claims that there was violation of randomization and poor quality mammography in the CNBSS, such as trough the news or scientific papers, attending talks or lectures, over time.

Certainty

The Summary report claims that the CNBSS "*should not be used to inform any decisions on breast cancer screening policy*" because *"[n]ew evidence has come to light that brings absolute certainty that the Canadian National Breast Screening Studies (CNBSS) performed over 40 years ago were compromised"* (22).

We heard hearsay and interviewed one eyewitness who gave a consistent testimony. The eyewitness confirmed what was previously known, that the randomization process in the CNBSS was vulnerable to subversion, but *"even if there had been acts of subversion, they could only have been few in number and…could have had only a trivial effect on the study findings"* (23).

Conclusion

We have considered the complaint material, the previous review, the published results from the CNBSS that may indicate violation of randomization and poor-quality mammography, and all interviews we have performed, summarized in *Chapter 29, 30, 31, 33 and 34* in this report.

After reviewing the complaint material, the previous review, the published findings from the CNBSS that may indicate violation of randomization and poor-quality mammography, and 19 hours interviewing individuals identified in the allegations several of them with first-hand knowledge about the CNBSS, we do not believe that there is new credible evidence that the CNBSS is compromised with regard to its scientific impact or reliability of findings.

We used the principles in the preponderance of probability (*more likely than not*) in the judiciary system (8).

Based on the evidence published in this report, we do not think the new evidence brought forward is beyond a reasonable doubt, nor beyond our predefined burden of proof that the evidence is "*more likely than not*". We therefore conclude:

> **The new information does not have a credible scientific impact on the reliability of the CNBSS.**

References

1. US Legal. Burden of proof. Accessed Feb 19 2024 at https://civilprocedure.uslegal.com/trial/burden-of-proof/
2. Cornell Law School. Beyond a reasonable doubt Accessed Feb 19 2024 at https://www.law.cornell.edu/wex/beyond_a_reasonable_doubt
3. Stanford Encyclopedia of Philosophy. The legal concept of evidence. Accessed Feb 19 2024 at https://plato.stanford.edu/entries/evidence-legal/
4. ICLR (UK) Incorporated Council of Law Reporting for England and Wales. Knowledge. Accessed Feb 19 2024 at https://www.iclr.co.uk/knowledge/glossary/standard-and-burden-of-proof/
5. Britannica, inquisitorial procedure. Accessed Feb 19 2024 at https://www.britannica.com/topic/inquisitorial-procedure
6. Encyclopedia.com Inquisitorial system Accessed Feb 19 2024 at https://www.encyclopedia.com/law/encyclopedias-almanacs-transcripts-and-maps/inquisitorial-system
7. UNODC United nations Office on Drugs and Crime. E4J University module series: Adversarial versus inquisitorial legal systems. Accessed Feb 19 2024 at https://www.unodc.org/e4j/en/organized-crime/module-9/key-issues/adversarial-vs-inquisitorial-legal-systems.html
8. Principles in practice. Cooper J. The burden and the standard of proof. https://www.judiciary.uk/wp-content/uploads/2016/01/cooper-burden-standard-of-proof-spring2008.pdf
9. Canadian Criminal Evidence/standard of proof. Accessed Feb 19 2024 at https://en.wikibooks.org/wiki/Canadian_Criminal_Evidence/Standard_of_Proof
10. Wikipedia, Burden of proof. Accessed Feb 19 2024 at https://en.wikipedia.org/wiki/Burden_of_proof_(law)
11. Britannica, inquisitorial procedure. Accessed Feb 19 2024 at https://www.britannica.com/topic/circumstantial-evidence
12. Wickipedia Circumstantial evidence Accessed Feb 19 2024 at https://en.wikipedia.org/wiki/Circumstantial_evidence
13. Wixted JT, Mickes L, Fisher RP et al. Rethinking the reliability of eyewitness memory. Perspectives on Psychological Science 2018; https://doi.org/10.1177/1745691617734878
14. The innocence project. Eyewitness misidentification; https://innocenceproject.org/eyewitness-misidentification
15. Loftus EF, Miller DG, Burns HJ. Semantic integration of verbal inflation into a visual memory. J Exp Phsychol Hum Learn 1978;4:19-31
16. StudySmarter. Forensic Psychology Accessed Feb 19 2024 at https://www.studysmarter.co.uk/explanations/psychology/forensic-psychology/reliability-of-eyewitness-testimony/
17. Government of Canada Civil and criminal cases. Accessed Feb 19 2024 at https://www.justice.gc.ca/eng/csj-sjc/just/08.html

18. Immigration and refugee board of Canada, weighing evidence - chapter 4: standard of proof and burden of proof. Accessed Feb 19 2024 at https://www.irb-cisr.gc.ca/en/legal-policy/legal-concepts/Pages/EvidPreu04.aspx
19. Guyatt G. et al Evidence-Based Medicine working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268: 2420-5
20. Guidelines International Network. Accessed Feb 22 at https://g-i-n.net/about-gin
21. Canadian Task Force on preventive Health care. Policy on disclosures of interests and management of conflict of interest, 2020 https://canadiantaskforce.ca/wp-content/uploads/2020/10/COI-Policy-202008Final-1.pdf
22. Canadian National Breast Screening Studies (CNBSS) Summary report: New evidence from key informants.
23. Bailar III, JC. MacMahon B. Randomization in the Canadian national breast Screening Study: a review for evidence of subversion. CMAJ 1997; 156:193-199
24. Baines C. The Canadian national Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326-334