

Screening Mammography: A Reassessment

Summary

AGENCE D'ÉVALUATION DES TECHNOLOGIES
ET DES MODES D'INTERVENTION EN SANTÉ

Screening Mammography : A Reassessment

Summary

Report prepared for AETMIS
by Wilber Deck
with the contribution of Ritzuko Kakuma

August 2005

The content of this publication was written and produced by the Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS). Both the original English report and its French version, titled *Mammographie de dépistage : une réévaluation*, are available in PDF format on the AETMIS Web site.

Scientific review

Alicia Framarin, MD, MSc, Deputy Scientific Director
Jean-Marie Lance, MSc, Senior Scientific Advisor

Editorial Supervision

Suzie Toutant

Page layout

Frédérique Stephan
Jocelyne Guillot

Proofreading

Suzie Toutant

Bibliography research

Micheline Paquin
Denis Santerre

Coordination

Lise-Ann Davignon

Communications and dissemination

Richard Lavoie

For information about this publication or any other AETMIS activity, please contact:

Agence d'évaluation des technologies et des modes d'intervention en santé
2021, Union avenue, Suite 1050
Montréal (Québec) H3A 2S9

Tel.: (514) 873-2563
Fax: (514) 873-1369
e-mail: aetmis@aetmis.gouv.qc.ca
<http://www.aetmis.gouv.qc.ca>

How to cite this publication:

Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS). Screening Mammography: A Reassessment. Report prepared by Wilber Deck with the contribution of Ritzuko Kakuma (AETMIS 05-03). Montréal: AETMIS, 2005.

Legal Deposit

Bibliothèque nationale du Québec, 2005

National Library of Canada, 2005

ISBN 2-550-44486-X (Printed version) (French edition ISBN 2-550-44487-6)

ISBN 2-5550-44486-8 (PDF) (French edition ISBN 2-550-44488-4)

© Gouvernement du Québec, 2005.

This report may be reproduced in whole or in part, provided that the source is cited.

MISSION

The mission of the *Agence d'évaluation des technologies et des modes d'intervention en santé* (AETMIS) is to contribute to improving the Québec health-care system and to participate in the implementation of the Québec government's scientific policy. To accomplish this, the Agency advises and supports the Minister of Health and Social Services as well as the decision-makers in the health care system, in matters concerning the assessment of health services and technologies. The Agency makes recommendations based on scientific reports assessing the introduction, diffusion and use of health technologies, including technical aids for disabled persons, as well as the modes of providing and organizing services. The assessments take into account many factors, such as efficacy, safety and efficiency, as well as ethical, social, organizational and economic implications.

EXECUTIVE

Dr. Luc Deschênes

Cancer Surgeon, President and Chief Executive Officer of AETMIS, Montréal, and Chairman, Conseil médical du Québec, Québec

Dr. Véronique Déry

Public Health Physician, Chief Executive Officer and Scientific Director

Dr. Reiner Banken

Physician, Deputy Chief Executive Officer, Development and Partnerships

Dr. Alicia Framarin

Physician, Deputy Scientific Director

Jean-Marie R. Lance

Economist, Senior Scientific Advisor

BOARD OF DIRECTORS

Dr. Jeffrey Barkun

Associate Professor, Department of Surgery, Faculty of Medicine, McGill University, and Surgeon, Royal Victoria Hospital (MUHC), Montréal

Dr. Marie-Dominique Beaulieu

Family Physician, Holder of the Dr. Sadok Besrouer Chair in Family Medicine, CHUM, and Researcher, Unité de recherche évaluative, Hôpital Notre-Dame (CHUM), Montréal

Dr. Suzanne Claveau

Specialist in microbiology and infectious diseases, Hôtel-Dieu de Québec (CHUQ), Québec

Roger Jacob

Biomedical Engineer, Coordinator, Capital Assets and Medical Equipment, Agence de développement de réseaux locaux de services de santé et de services sociaux de Montréal, Montréal

Denise Leclerc

Pharmacist, Board Member of the Institut universitaire de gériatrie de Montréal, Montréal

Louise Montreuil

Assistant Executive Director, Direction générale de la coordination ministérielle des relations avec le réseau, ministère de la Santé et des Services sociaux, Québec

Dr. Jean-Marie Moutquin

Obstetrician/Gynecologist, Executive Director, Centre de recherche, CHUS, Sherbrooke

Dr. Réginald Nadeau

Cardiologist, Hôpital du Sacré-Cœur, Montréal, Board Member of the Conseil du médicament du Québec

Guy Rocher

Sociologist, Professor, Département de sociologie, and Researcher, Centre de recherche en droit public, Université de Montréal, Montréal

Lee Soderström

Economist, Professor, Department of Economics, McGill University, Montréal

FOREWORD

SCREENING MAMMOGRAPHY: A REASSESSMENT

Screening mammography, a technique which is 50 years old, aims to advance the diagnosis of breast cancer in order to offer early treatment, thereby improving the chances of cure. The practice of mammography, in constant evolution, varies widely according to the equipment used, the interpretation of films, and program aspects such as the age of women when they are invited to their first screen, the interval between rounds of screening and participation rates.

In 1990, a report by Québec's *Conseil d'évaluation des technologies de la santé* (CETS) recommended that this practice be structured as part of a formal program which would include quality standards. A second CETS report published in 1993 underlined the absence of proof in favour of screening women younger than 50. Since 1998, the *Programme québécois de dépistage du cancer du sein* (PQDCS) offers systematic screening every 2 years to all women aged 50 to 69. Younger women can still obtain mammography with a prescription from their physicians. Many other countries have also started screening mammography programs, but there remains controversy regarding the age when screening should start, and recent studies have raised doubts about the value of screening mammography at any age.

In this context, the *ministère de la Santé et des Services sociaux* (MSSS) has asked the *Agence d'évaluation des technologies et des modes d'intervention en santé* (AETMIS) to re-examine the quality of the scientific evidence on which the PQDCS is based and on the pertinence of extending screening to women less than 50 years old. This report evaluates various aspects of the validity of screening trials and their pertinence with regard to the performance and quality assurance of a modern screening program such as the PQDCS.

The analysis indicates that most trials had serious problems with validity, making it difficult to use them to estimate the potential benefits of mammography. The best trials show a modest reduction in breast cancer mortality; this reduction is greater when the analysis is limited to women 50 to 69 years old. No trial was designed and conducted in such a way that the full potential of screening mammography could be realised. It is thus plausible that a modern program, conducted under conditions of superior quality, might obtain better results than the trials suggest.

In conclusion, a screening program targeting women 50 to 69 years old remains justified by the available data. This justification does not extend to younger women. However, it is possible that screening of individual women, based on a personalized risk assessment, might be of benefit to some younger women. This conclusion should be reviewed in several years, when results of the ongoing UK Age Trial become available. In the meantime, a modern mammography screening program like the PQDCS can benefit from measures which aim to maximise the quality of screening and to increase participation rates.

In submitting this report, AETMIS hopes to contribute to the optimal use of screening mammography for the benefit of all women.

Dr. Luc Deschênes
President and Chief Executive Officer

ACKNOWLEDGEMENTS

This report was prepared by Dr. Wilber Deck, MSc (epidemiology and biostatistics), with the collaboration of Ritzuko Kakuma, PhD candidate (epidemiology), at the request of the *Agence d'évaluation des technologies et des modes d'intervention en santé* (AETMIS).

The Agency would also like to thank the external reviewers, who have provided extremely helpful comments:

Dr. Jean-François Boivin

Full Professor, Department of Epidemiology and Biostatistics, McGill University, Montreal

Dr. Jacques Brisson

Professor, Faculty of Medicine, *Université Laval*, researcher, *Unité de recherche en santé des populations, Hôpital du Saint-Sacrement*, and associate expert, *Institut national de santé publique du Québec*

Dr. Heather Bryant

Research Scientist, Division of Population Health and Information, Tom Baker Cancer Centre, Calgary, Alberta

Julietta Patnick

Director, UK NHS Breast Cancer Programmes, United Kingdom

Dr. André Robidoux

Surgical Oncologist, Director, *Clinique des maladies du sein*, Hôtel-Dieu de Montréal (CHUM)

AETMIS would also like to thank a radiologist external reviewer (anonymous) for his comments.

DISCLOSURE OF CONFLICT OF INTEREST

None declared.

SUMMARY

INTRODUCTION

Eight trials examining the performance of screening mammography have been conducted in the USA, Sweden, the United Kingdom and Canada, beginning in 1963. A first report by the *Conseil d'évaluation des technologies de la santé* (CETS) published in 1990 concluded that screening mammography trials had shown reductions in mortality from breast cancer of 35%, with 45% in the subgroup of women aged 50 to 69. A second report in 1993 concluded that mammographic screening of younger women had not been shown to reduce mortality. By the year 1998, when Québec introduced the *Programme québécois de dépistage du cancer du sein* (PQDCS), all Canadian provinces and many other countries had organized screening programs in place. A recent Cochrane Collaboration Group review, challenging the belief that mammography screening is an effective tool for reducing breast cancer deaths, has raised concerns about the validity of the published randomized trials. This update addresses three questions:

- (1) What is the strength of the scientific evidence on which screening mammography programs are based?
- (2) What evidence is there in support of screening for women aged 40 to 49 years?
- (3) What are the implications of research studies for maximizing the effectiveness of modern programs such as the *Programme québécois de dépistage du cancer du sein* (PQDCS)?

METHODOLOGIC ANALYSIS

An evaluation of efficacy trials essentially aims to determine whether the conditions under which the trials were performed and the results that were obtained can guide decisions regarding the intervention in ques-

tion. Scientific evidence must satisfy three prerequisites in order for it to be the basis for such decision-making: relevance, validity, and precision. A study is **relevant** if it is designed to contrast two or more interventions that are options of interest to decision makers. A study is **valid** if it is designed, conducted and analysed in such a way as to ensure that no important biases affect the measured comparison of the effectiveness of the technologies that are compared. A study is **precise** if it allows for an estimation of efficacy that is not vulnerable to random effects.

Previous analyses of screening mammography trials have tended to emphasize trials' validity, and in particular factors that might bias trials' results in an unknown direction. These analyses have also calculated the precision of the estimates of breast cancer mortality reduction, and have narrowed confidence intervals around these estimates by combining trials' results in meta-analyses. However, since the issues of relevance and of bias in known direction have not been adequately addressed, they will be further developed in this reassessment, which also includes a meta-analysis.

Relevance

To better appreciate the concept of relevance, we introduce here the notion of *contrast*, which represents the opposition or temporary divergence between an experimental intervention, offered to the screening cohort, and a reference intervention offered to the control cohort. Since the question at issue here is the value of mammography screening, relevant trials are those which contrast screening mammography with no screening. We will thus not include in this analysis trials which compare in principal different screening strategies. In practice, the reference strategy (no screening) may include some uncontrollable screening activities,

which will weaken the contrast with the screening intervention.

Validity

A valid study must be a fair comparison between screening and no screening. Thus screening and control cohorts should have the same baseline risk of breast cancer mortality, should be treated equally in all regards except concerning the screening or control intervention, and should have the information on their outcome measured in a way that is independent of their assignment to the screening or control group. Validity can be compromised by bias of known direction and by bias of unknown direction.

In this evaluation, to further develop the notion of bias of unknown direction, we use the concept of *strength of contrast*. It corresponds to the degree to which a trial succeeds in realizing the divergence between the strategies compared and in measuring the effects that this divergence produces. Five elements are evaluated in this report which help assess the strength of contrast:

- the technical contrast, or the nature of the difference between screening and control interventions;
- the era in which these techniques are applied;
- the quality of the intervention, including quality control measures;
- rates of participation and contamination measured among screening and control cohorts; and
- the timing of the measurement of the effects of screening on mortality (or timing dilution).

For each trial, a score for the strength of contrast corresponds to the product of individual estimates for each of these elements, as assessed by two researchers in this analysis. For comparison's sake, we have applied this scale to a modern screening program, the PQDCS, in an analogous fashion.

As in other analyses, we also examine trials' biases of unknown direction, in particular

concerning randomization, the equivalence of the risk of breast cancer mortality between the screening and control cohorts, the equivalence of criteria for exclusion from the two cohorts, and the equivalence of the follow-up of the two groups. A score is attributed to each of these elements, and the sum of these scores constitutes a global validity score for each of the eight published trials.

Precision

Precision is evaluated by progressively combining the results of screening trials, weighted by their variance, adding in studies in order of their score on our validity scale.

ANALYSIS OF PUBLISHED TRIALS AND COMPARISON WITH THE QUÉBEC PROGRAM

The Québec Program involves screening mammography every two years offered to about 900,000 Québec women 50 to 69 years old. The program's structure, process and objectives are based on established breast cancer screening programs in place in Sweden, the United Kingdom and Australia. Based on pre-existing medical facilities, it includes approximately 80 screening centres, which must meet province-wide quality assurance standards. In 2003, participation rates were 46.7%. However, if all mammographic exams are included, including mammography outside the program and diagnostic mammography, mammography rates in women 50 to 69 years old reached 63%, compared to 50% before the onset of the program.

Eight screening mammography trials have published results; a ninth trial, involving younger women only (UK Age Trial), begun in 1991 in the United Kingdom, has not yet reported final mortality results. The eight trials are the following, in order of their initial years:

- Health Insurance Plan Trial (HIP), New York (1963);

- Malmö Mammographic Screening Trial (Malmö), Sweden (1976);
- Two-County Trial (TCS), Sweden (1977);
- Edinburgh Randomised Trial of Screening for Breast Cancer (Edinburgh), Scotland (1979);
- National Breast Screening Study #1: women 40 to 49 years old (NBSS-1), Canada (1980);
- National Breast Screening Study #2: women 50 to 59 years old (NBSS-2), Canada (1980);
- Stockholm Mammographic Screening Trial (Stockholm), Sweden (1981);
- Gothenburg Breast Screening Trial (Gothenburg), Sweden (1982).

Although the results of trials have been used to estimate how effective a screening program might be in reducing mortality, it is important to first compare the screening regimens in the trials with the screening regimens that modern programs such as the PQDCS have put in place. This comparison applies particularly to criteria of relevance and validity related to the notion of contrast. In several regards, trials have obtained elements of contrast which are stronger than those in the Québec program: some have included clinical breast exams along with mammography (HIP, NBSS-1 and -2), some have used annual mammography (HIP, NBSS-1 and -2) rather than every two years, and some have used expert readers and double reading (Swedish studies).

However, many of the conditions of screening trials resulted in significantly weaker contrasts than those of modern programs. In particular, all studies are from earlier eras when mammographic equipment and techniques were less refined. Many have used a single view of each breast, as opposed to modern standards using double views. Some studies have used intervals longer than two years (28 months in Stockholm, 33 months in Two-County). Participation rates have been as low as 53% (Edinburgh) and 54% (HIP) and as high as 87% (TCS) and 88% (NBSS), but in all studies the effective contrast has been reduced because women in the

control group also received mammography, ranging from about 5% (HIP) to as high as 15% (Gothenburg) and 20% (Stockholm). The timing of the relationship between the screening period and mortality results has caused significant dilution in all studies, compared to the steady-state reduction that a program could achieve after a suitable delay. In particular, the durations of all studies have been much shorter than the twenty years of screening that most programs propose (from age 50 to 69); some studies had only two rounds (Stockholm), three rounds (Two-County) or four rounds (HIP, Edinburgh, NBSS).

The Canadian NBSS-2 trial is not included in our meta-analysis, since the interventions contrasted in this trial are not relevant to answering the decisional question of the efficacy of screening mammography in Québec. Indeed, that trial did not compare screening with no screening, but rather the reference screening intervention (regular high-quality clinical breast examination and self-examination) with another screening regimen (the same exams, with added mammography).

SYNTHESIS OF RESULTS

Our review indicates marked differences in the quality of the design and execution of mammography trials, which have tended to fall short of modern quality standards, as indicated by their validity scores. Some trials were not randomized at all, and most studies have been poorly or inconsistently documented. In particular, most have not provided baseline characteristics of women in study and control groups, often because next to nothing was known about the control cohort. Exclusion of previously diagnosed cancers has been inconsistent in six of the eight studies. Blinding has not been attempted in any trial, either of patients or of care providers.

On the other hand, no study has been designed and conducted in such a way that the full potential of mammography screening could be determined, as indicated by their

strength of contrast scores. To the extent that published trials have not maximized the potential of mammography screening, there is thus a potential for modern programs to identify earlier lesions and, perhaps, to achieve greater reductions in breast cancer mortality than those that have been reported in the scientific literature.

Our evaluation indicates that three trials (HIP, Edinburgh and Two-County) contain flaws which preclude their use in estimating the effectiveness of screening. These exclusions are consistent with those of other reviewers that have judged the quality of individual trials. An examination of only traditional aspects of study validity that takes no account of relevance or contrast issues would cast strong doubt on the efficacy of mammography screening. The present meta-analysis has included successive study results by order of validity: good or medium quality (scores of 3 and over out of 4), poor quality (scores of 1.5 but less than 3), and flawed (scores less than 1.5).

For women of all ages, results show an inverse relation between the quality of the study and reduction in breast cancer mortality. If all studies are included, regardless of validity, mortality reduction is estimated at 23%, but this estimate diminishes to 15% when only studies of medium and poor quality are included, and 9% if only medium quality studies are included. Thus, the more valid studies tend to show lesser reductions, and confidence intervals sometimes include the null value¹.

In women under 50, many women were enrolled towards the end of their forties, and many of their cancers were not detected before they were in their fifties, so these results are pertinent to women who start screening in their late 40s, not at age 40. In the medium-quality studies of this age group, the cumulative risk reduction is 2%. A similar inverse relation is observed, since including data from studies with weaker validity increases mortality reduction to 8%. Mortal-

ity reduction is thus much smaller in younger women, and confidence intervals include the null value for all combinations of studies.

In women older than 50, on the other hand, results are more favourable. The only study of medium validity gives a risk reduction of 27%. Including data from studies of poor quality, the overall reduction would be 24%, and including all data irrespective of validity gives a reduction of 29%. These mortality reductions are substantially higher than in women of all ages combined. Confidence intervals in this sub-group are naturally wider, but no combination of studies includes the null value.

As for strength of contrast, no study came close to the standard of an ideal program with many years of regular screening using modern equipment, quality assurance, two-view mammography at intervals of two years or less, and with full participation. We estimate that, compared to this standard, the eight published mammography screening trials only achieved a strength of contrast of between 12 and 45% of what would be possible. Using this same standard, modern programs are likely to achieve considerably greater strength of contrast, with the Québec program estimated to obtain 63% of full potential.

DISCUSSION AND CONCLUSIONS

Question 1: What is the strength of the scientific evidence on which screening mammography programs are based?

There are serious concerns regarding the validity of most of the trials supporting mammography screening, based on methodologic weaknesses in the screening trials. Studies are highly heterogeneous with regard to the strength of the contrast that they studied, with numerous weaknesses identified in all the major studies, meaning that the potential of screening mammography has perhaps not been thoroughly explored. Using the best available data, one can conclude that there is fair evidence of moderate reduction of breast cancer mortality, of the

1. The null value indicates absence of efficacy. A confidence interval that includes the null value indicates that the observed results would be likely to be observed by chance even if no true efficacy is present.

order of 9 to 15%; data restricted to women over the age of about 50 show greater reductions, of the order of 24 to 29%. Furthermore, our analysis has demonstrated that modern mammography, carried out under quality conditions that maximize its performance, has the potential to identify cancerous lesions earlier in their progression, and this may allow for some further reduction in mortality.

Conclusion: *Existing scientific trials, despite their flaws, support mammography screening programs. In addition, there are good reasons to believe that modern, well-conducted screening programs may achieve earlier detection and diagnosis of breast cancer and, perhaps, greater reductions in breast cancer mortality than what has been found in screening trials.*

Question 2: What is the evidence in support of screening mammography for women aged 40 to 49 years?

There is much less data available to answer the question, since most study experience is in women over 50, even though some women in some of the studies started screening several years earlier than their fiftieth birthday. The best data available show no significant reduction in breast cancer mortality in women screened before the age of 50. In the absence of any convincing data that mammography is efficacious in this age group, harmful effects may outweigh any positive effects.

Conclusion: *Trial data published to date do not provide scientific justification to recommend screening for women younger than 50. However, this conclusion does not exclude the possibility that screening of individual women, based on a personalized risk assessment, could be of benefit. These conclu-*

sions should be reviewed when results from the UK Trial become available.

Question 3: What are the implications of research studies for maximizing the effectiveness of modern programs such as the Programme québécois de dépistage du cancer du sein (PQDCS)?

Although the PQDCS already includes rigorous control of the quality of films produced, certain aspects of the structure and process of trials examined under the rubric of strength of contrast can be transposed as additional quality norms. Notable among these are double reading of films and an annual reading volume sufficient to allow each radiologist to acquire and maintain the necessary expertise to detect breast cancer in its early stages. These aspects should also allow for a reduction in false positive rates and subsequent unnecessary diagnostic procedures. Moreover, high participation rates at each screening round will contribute to achieving and perhaps exceeding the mortality reductions obtained by screening trials.

Conclusion: *Modern screening programs such as the PQDCS may produce outcomes comparable or even superior to those observed in screening trials if they achieve a standard of quality equal to or better than the standard achieved by trials. Measures that should reduce false positive rates and assure high quality screening include making sure that high quality mammographic films are being produced, that readers have the necessary expertise to detect early cancer and avoid false positives, and double reading of a proportion of films. While participation rates should be as high as possible, efforts to increase participation should not overstate the benefits of mammography nor understate the risks and uncertainties which remain.*

*Agence d'évaluation
des technologies
et des modes
d'intervention en santé*

Québec 