ETMIS 2005; VOL. 1 : N° 3

Mammographie de dépistage : une réévaluation

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

Québec

ETMIS 2005; Vol. 1 : N° 3

Mammographie de dépistage : une réévaluation

Rapport préparé pour l'AETMIS
par Wilber Deck
avec la collaboration de Ritzuko Kakuma

Août 2005

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

Le contenu de cette publication a été rédigé et édité par l'Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS). Ce document est également offert en format PDF sur le site Web de l'Agence.

Révision scientifique
Alicia Framarin, m.d., M. Sc., directrice scientifique adjointe
Jean-Marie R. Lance, M. Sc., conseiller scientifique principal

Coordination
Lise-Ann Davignon

Traduction
Carmen Léger

Révision linguistique
Suzie Toutant

Montage
Frédérique Stephan
Jocelyne Guillot

Correction d'épreuves
Suzanne Archambault
Suzie Toutant
Brigitte Turmel

Vérification bibliographique
Denis Santerre

Communications et diffusion
Andrée Gravel
Richard Lavoie

Pour se renseigner sur cette publication ou toute autre activité de l'AETMIS, s'adresser à :

Agence d'évaluation des technologies et des modes d'intervention en santé
2021, avenue Union, bureau 10.083
Montréal (Québec) H3A 2S9

Téléphone : (514) 873-2563
Télécopieur : (514) 873-1369
Courriel : aetmis@aetmis.gouv.qc.ca
www.aetmis.gouv.qc.ca

Comment citer ce document :

Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS). Mammographie de dépistage : une réévaluation. Rapport préparé par Wilber Deck avec la collaboration de Ritzuko Kakuma. ETMIS 2005;1(3)1-80.

Dépôt légal
Bibliothèque nationale du Québec, 2005
Bibliothèque nationale du Canada, 2005
ISSN 1915-3082 ETMIS (Imprimé), ISSN 1915-3104 ETMIS (PDF)
ISBN 2-550-44487-6 (Imprimé), ISBN 2-550-44488-4 (PDF)

© Gouvernement du Québec, 2005.

La reproduction totale ou partielle de ce document est autorisée, à condition que la source soit mentionnée.

LA MISSION

L'Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS) a pour mission de contribuer à améliorer le système de santé québécois et de participer à la mise en œuvre de la politique scientifique du gouvernement du Québec. Pour ce faire, l'Agence conseille et appuie le ministre de la Santé et des Services sociaux ainsi que les décideurs du système de santé en matière d'évaluation des services et des technologies de la santé. L'Agence émet des avis basés sur des rapports scientifiques évaluant l'introduction, la diffusion et l'utilisation des technologies de la santé, incluant les aides techniques pour personnes handicapées, ainsi que les modalités de prestation et d'organisation des services. Les évaluations tiennent compte de multiples facteurs, dont l'efficacité, la sécurité et l'efficience ainsi que les enjeux éthiques, sociaux, organisationnels et économiques.

LA DIRECTION

Dr Luc Deschênes,
chirurgien oncologue, président-directeur général de l'AETMIS, Montréal, et président du Conseil médical du Québec, Québec

Dr Véronique Déry,
médecin spécialiste en santé publique, directrice générale et scientifique

Dr Reiner Banken,
médecin, directeur général adjoint au développement et aux partenariats

Dr Alicia Framarin,
médecin, directrice scientifique adjointe

M. Jean-Marie R. Lance,
économiste, conseiller scientifique principal

LE CONSEIL

Dr Jeffrey Barkun,
professeur agrégé, département de chirurgie, Faculté de médecine, Université McGill, et chirurgien, Hôpital Royal Victoria, CUSM, Montréal

Dr Marie-Dominique Beaulieu,
médecin en médecine familiale, titulaire de la Chaire Docteur Sadok Besrour en médecine familiale, CHUM, et chercheur, Unité de recherche évaluative, Pavillon Notre-Dame, CHUM, Montréal

Dr Suzanne Claveau,
médecin en microbiologie-infectiologie, Pavillon L'Hôtel-Dieu de Québec, CHUQ, Québec

M. Roger Jacob,
ingénieur biomédical, coordonnateur, Immobilisations et équipements médicaux, Agence de développement de réseaux locaux de services de santé et de services sociaux de Montréal, Montréal

Mme Denise Leclerc,
pharmacienne, membre du Conseil d'administration de l'Institut universitaire de gériatrie de Montréal, Montréal

Mme Louise Montreuil,
directrice générale adjointe aux ententes de gestion, Direction générale de la coordination ministérielle des relations avec le réseau, ministère de la Santé et des Services sociaux, Québec

Dr Jean-Marie Moutquin,
médecin spécialiste en gynéco-obstétrique, directeur scientifique, Centre de recherche clinique, CHUS, Sherbrooke

Dr Réginald Nadeau,
médecin spécialiste en cardiologie, Hôpital du Sacré-Cœur, Montréal, et membre du Conseil d'administration du Conseil du médicament du Québec, Québec

M. Guy Rocher,
sociologue, professeur titulaire, département de sociologie, et chercheur, Centre de recherche en droit public, Université de Montréal, Montréal

M. Lee Soderstrom,
économiste, professeur, département des sciences économiques, Université McGill, Montréal

AVANT-PROPOS

MAMMOGRAPHIE DE DÉPISTAGE : UNE RÉÉVALUATION

La mammographie de dépistage, pratiquée depuis environ 50 ans, vise à obtenir un diagnostic précoce du cancer du sein afin d'améliorer les chances de guérison. La technique, en constante évolution, comporte des variations importantes en ce qui concerne le matériel utilisé, l'interprétation des clichés et les aspects spécifiques d'un programme de dépistage tels que l'âge auquel les femmes y sont invitées, l'intervalle entre les cycles de dépistage, le nombre de cycles et les taux de participation.

Au Québec, un rapport du Conseil d'évaluation des technologies de la santé (CETS) recommandait en 1990 que cette pratique soit encadrée par un programme doté de normes de qualité. Un deuxième rapport du CETS publié en 1993 soulignait l'absence de preuves en faveur du dépistage pour les femmes de moins de 50 ans. Depuis 1998, le Programme québécois de dépistage du cancer du sein (PQDCS) offre un dépistage systématique tous les deux ans aux femmes de 50 à 69 ans. Les femmes plus jeunes peuvent toujours obtenir une mammographie sur ordonnance de leur médecin. Plusieurs pays ont instauré des programmes de dépistage mammographique, mais des controverses persistent sur l'âge à partir duquel le dépistage devrait être offert, et des publications récentes ont soulevé des doutes sur la valeur de la mammographie, quel que soit l'âge.

Dans ce contexte, le ministère de la Santé et des Services sociaux (MSSS) a demandé à l'Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS) de réexaminer la qualité des preuves scientifiques sur lesquelles se fonde le PQDCS et la pertinence d'offrir le dépistage aux femmes de moins de 50 ans. Le présent rapport évalue différents aspects de la validité des essais publiés sur le sujet et leur pertinence par rapport à la performance et à la qualité d'un programme de dépistage moderne comme le PQDCS.

L'analyse indique que la plupart des essais ont de sérieux problèmes de validité, ce qui rend difficile de s'en servir pour estimer les bienfaits potentiels de la mammographie. Les meilleurs essais montrent une réduction modeste de la mortalité par cancer du sein. Cette réduction est plus importante quand l'analyse se limite aux femmes de 50 à 69 ans. Aucun essai n'a été conçu ni mené de façon à réaliser tout le potentiel du dépistage mammographique. Il est donc plausible qu'un programme moderne, mené dans des conditions de qualité supérieure, obtienne de meilleurs résultats que les essais.

En conclusion, un programme de dépistage ciblant les femmes de 50 à 69 ans reste justifié selon les données scientifiques disponibles. Cette justification ne s'étend toutefois pas aux femmes plus jeunes. Cependant, il n'est pas exclu que le dépistage individuel, fondé sur une évaluation personnalisée du risque de ces femmes, puisse être bénéfique. Cette conclusion devra être revue dans quelques années, lorsque les résultats d'un essai en cours au Royaume-Uni seront publiés. Enfin, un programme moderne de dépistage mammographique comme le PQDCS pourra tirer avantage de la mise en place de mesures visant à assurer une qualité de détection encore supérieure et à accroître le taux de participation.

En soumettant ce rapport, l'AETMIS souhaite contribuer à l'utilisation optimale du dépistage mammographique, et ce, dans l'intérêt de toutes les femmes.

D^r^ Luc Deschênes
Président-directeur général

REMERCIEMENTS

Ce rapport a été préparé à la demande de l'AETMIS par le **D^r^ Wilber Deck,** M. Sc. (épidémiologie et biostatistique), avec la collaboration de **M^me^ Ritzuko Kakuma**, candidate au doctorat en épidémiologie.

L'AETMIS aimerait remercier les lecteurs externes suivants pour leurs précieux commentaires sur ce rapport :

D^r^ Jean-François Boivin
Professeur titulaire, département d'épidémiologie et biostatistique, Université McGill, Montréal

D^r^ Jacques Brisson
Professeur à la faculté de médecine de l'Université Laval, chercheur à l'Unité de recherche en santé des populations de l'Hôpital du Saint-Sacrement (CHA universitaire de Québec) et, expert associé à l'Institut national de santé publique du Québec

D^re^ Heather Bryant
Chercheure, *Division of Population Health and Information, Tom Baker Cancer Centre*, Calgary, Alberta

M^me^ Julietta Patnick
Directrice, *NHS Breast Cancer Programmes*, Royaume-Uni

D^r^ André Robidoux
Chirurgien-oncologue, directeur de la clinique des maladies du sein, Hôtel-Dieu de Montréal (CHUM)

L'AETMIS aimerait également remercier un lecteur externe radiologue (anonyme) pour l'ensemble de ses commentaires.

DIVULGATION DE CONFLITS D'INTÉRÊTS

Aucun conflit à signaler.

RÉSUMÉ

Depuis 1963, huit études portant sur la mammographie de dépistage ont été menées aux États-Unis, en Suède, au Royaume-Uni et au Canada. Selon un premier rapport du Conseil d'évaluation des technologies de la santé (CETS) publié en 1990, les essais randomisés sur la mammographie de dépistage ont montré une réduction de la mortalité par cancer du sein de 35 %, et une réduction de 45 % chez les femmes âgées de 50 à 69 ans. Un deuxième rapport datant de 1993 concluait qu'il n'était pas prouvé que le dépistage mammographique pouvait réduire la mortalité chez les femmes de moins de 50 ans. Dès 1998, année de l'instauration du Programme québécois de dépistage du cancer du sein (PQDCS), toutes les provinces canadiennes et plusieurs pays avaient mis en place des programmes de dépistage.

Une étude du groupe de collaboration Cochrane, qui contestait l'efficacité du dépistage mammographique pour réduire les décès par cancer du sein, a soulevé des doutes quant à la validité des essais randomisés publiés. La présente mise à jour se penche particulièrement sur trois questions :

1) Quelle est la solidité des arguments scientifiques à l'appui des programmes de dépistage mammographique ?
2) Quelles sont les preuves à l'appui de la mammographie de dépistage pour les femmes de 40 à 49 ans ?
3) Quelles sont les implications des données scientifiques des essais sur les possibilités de maximiser l'efficacité de programmes modernes comme le Programme québécois de dépistage du cancer du sein (PQDCS) ?

CADRE MÉTHODOLOGIQUE

Toute évaluation d'essais sur l'efficacité d'une intervention en santé vise essentiellement à déteminer si les conditions de réalisation de ces études et les effets obtenus peuvent guider la prise de décisions sur son implantation. Trois conditions essentielles sont habituellement recherchées, soit la pertinence, la validité et la précision. De façon générale, la **pertinence** permet de bien circonscrire l'intervention étudiée par rapport à d'autres options visant les mêmes effets. La **validité** implique qu'une étude est conçue, menée et analysée de sorte qu'aucun biais important n'affecte la mesure de la différence dans l'efficacité des stratégies comparées. Enfin, la **précision** exige que cette étude fournisse une estimation de l'efficacité qui ne soit pas vulnérable aux effets aléatoires.

Les analyses des essais sur la mammographie de dépistage se sont surtout intéressées à la validité des études, et tout particulièrement aux facteurs qui expliquent les biais à direction inconnue. Ces analyses ont aussi abordé directement la précision de l'estimation du taux de mortalité par cancer du sein et ont pu réduire les intervalles de confiance des résultats en les combinant à l'aide de la technique de méta-analyse. Toutefois, puisque la pertinence et les biais à direction connue de ces différents essais n'ont pas été adéquatement traités dans ces analyses, ils seront examinés plus à fond dans la présente réévaluation, qui comporte également une méta-analyse.

Pertinence

Pour mieux cerner le concept de pertinence, nous introduisons la notion de contraste. Le *contraste* représente l'opposition ou la divergence temporaire entre l'intervention expérimentale offerte à la cohorte de dépistage et la stratégie de référence appliquée à une cohorte témoin. Comme la question décisionnelle principale est l'utilité de la mammographie de dépistage, les études pertinentes seront celles qui mettront en contraste le dépistage mammographique et, idéalement, le non-dépistage. Ne seront donc pas retenues dans l'analyse les études comparant formellement différentes stratégies de dépistage, incluant ou non la mammographie. Dans la pratique, la stratégie de référence (non-dépistage) d'un essai pourra

contenir certaines interventions non contrôlables de dépistage, ce qui affaiblira le contraste avec la stratégie expérimentale.

Validité

En règle générale, pour être valide, une étude doit mettre en contraste de manière adéquate une intervention de dépistage et une intervention témoin. Ainsi, le groupe de dépistage et le groupe témoin doivent avoir au départ le même risque initial de mortalité par cancer du sein, être traités de la même façon à tous égards sauf pour l'intervention de dépistage ou témoin, et les résultats doivent être mesurés de la même manière. La validité peut être compromise par deux types de biais : les biais à direction connue et les biais à direction inconnue.

Dans la présente évaluation, pour approfondir les biais à direction connue, nous utilisons la notion de *force du contraste*. Elle correspond au degré auquel l'essai réussit à faire pleinement ressortir la divergence entre les deux stratégies comparées et les effets qui en découlent. Cinq éléments ont été évalués dans ce rapport pour marquer la force du contraste :

- le contraste technique, ou la nature de la différence entre les interventions de dépistage et témoin;
- l'époque à laquelle les techniques ont été utilisées;
- la qualité technique des interventions et les mesures de contrôle de la qualité;
- les taux de participation et de contamination notés dans les groupes de dépistage et témoin; et
- le choix du moment pour mesurer les effets du dépistage sur la mortalité (ou la dilution temporelle).

Pour chaque essai, le score de la force de contraste correspond à la multiplication des valeurs individuelles estimées pour chaque élément (selon l'appréciation des deux chercheurs de la présente analyse). À titre de comparaison, un programme de dépistage moderne, soit le PQDCS, sera évalué de manière analogue.

Comme l'ont fait d'autres recensions, nous examinons aussi les biais à direction inconnue. Ils concernent notamment la randomisation, les équivalences entre les cohortes de dépistage et témoin à l'entrée dans l'étude, les critères d'exclusion et l'équivalence des suivis pour la mesure des résultats. Un score a été attribué à chacun de ces éléments et leur somme a permis d'établir un score global de validité pour évaluer chacune des huit études publiées.

Précision

La précision est évaluée par la combinaison progressive des résultats des huit études, pondérés par leur variance, selon l'ordre de leur score sur l'échelle d'évaluation de la validité.

ANALYSE DES ESSAIS PUBLIÉS ET COMPARAISON AVEC LE PROGRAMME QUÉBÉCOIS

Le Programme québécois offre un dépistage par mammographie tous les deux ans à environ 900 000 femmes du Québec âgées de 50 à 69 ans. Sa structure, ses processus et ses objectifs s'inspirent de programmes reconnus de dépistage du cancer du sein établis en Suède, au Royaume-Uni et en Australie. Implanté dans des établissements médicaux existants, le programme compte environ 80 centres de dépistage qui doivent respecter des normes d'assurance de la qualité en vigueur à l'échelle de la province. En 2003, le taux de participation s'élevait à 46,7 %. Toutefois, si l'on tient compte de toutes les mammographies, incluant les mammographies hors programme ou de diagnostic, le taux de mammographies chez les femmes de 50 à 69 ans atteignait 63 % en 2001-2002, alors qu'il n'était que de 50 % avant l'instauration du programme.

Les résultats de huit essais sur la mammographie de dépistage ont été publiés. Un neuvième essai ciblant des femmes plus jeunes (*UK Age Trial*), amorcé en 1991 au Royaume-Uni, n'a pas encore communiqué de résultats définitifs sur les taux de mortalité.

Les huit essais étaient les suivants, dans l'ordre :

- *Health Insurance Plan Trial* (HIP), New York (1963);
- *Malmö Mammographic Screening Trial* (Malmö), Suède (1976);
- *Two-County Trial* (TCS), Suède (1977);
- *Edinburgh Randomised Trial of Screening for Breast Cancer* (Édimbourg), Écosse (1979);
- *National Breast Screening Study*, volet 1 : femmes de 40 à 49 ans (NBSS-1), Canada (1980);
- *National Breast Screening Study*, volet 2 : femmes de 50 à 59 ans (NBSS-2), Canada (1980);
- *Stockholm Mammographic Screening Trial* (Stockholm), Suède (1981);
- *Gothenburg Breast Screening Trial* (Göteborg), Suède (1982).

Même si les résultats des essais ont servi à évaluer l'efficacité d'un programme de dépistage pour réduire la mortalité, il est important de comparer d'abord les protocoles de dépistage de ces essais avec les modalités de dépistage de programmes modernes comme le PQDCS. Cet examen applique particulièrement les critères de pertinence et de validité élaborés autour du concept de contraste. À plusieurs égards, les essais ont réalisé des conditions de contraste plus marquées que le Programme québécois : quelques-uns ont inclus l'examen clinique des seins associé à la mammographie (HIP, NBSS), certains ont offert une mammographie par année (HIP, NBSS 1 et 2) plutôt que tous les deux ans, et d'autres ont fait appel à des lecteurs experts et à des deuxièmes lectures (études suédoises).

Par contre, plusieurs conditions de réalisation de ces essais ont contribué à l'obtention de contrastes considérablement plus faibles qu'avec les programmes modernes. Ainsi, toutes les études datent d'une époque où les appareils et les techniques mammographiques étaient moins perfectionnés. Plusieurs n'ont réalisé qu'une seule incidence par sein, alors que les normes modernes en exigent deux. Certains essais ont offert des mammographies à des intervalles de plus de deux ans (28 mois à Stockholm, 33 mois dans l'étude suédoise *Two-County*). Les taux de participation ont varié de 53 % (Édimbourg) et 54 % (HIP) jusqu'à 87 % (TCS) et 88 % (NBSS), mais dans tous les essais, un certain pourcentage des femmes du groupe témoin ont eu des mammographies, allant de 5 % environ (HIP) jusqu'à 15 % (Göteborg) et 20 % (Stockholm), ce qui affaiblit le contraste réel. Dans toutes les études, la relation temporelle entre la période de dépistage et la période où la mortalité a été mesurée a causé une dilution importante de la réduction des taux de mortalité que pourrait atteindre un programme en régime établi après un délai suffisant. Plus particulièrement, la durée de toutes les études a été bien en deçà des 20 ans de dépistage que proposent la plupart des programmes (pour les femmes de 50 à 69 ans) : avec leurs longs intervalles, les essais de Stockholm et *Two-County* ne comportaient que deux et trois cycles de dépistage respectivement, alors que les essais HIP, Édimbourg et NBSS proposaient quatre cycles à intervalles de 12 mois seulement.

L'essai canadien NBSS-2 n'est pas inclus dans la présente méta-analyse. Bien qu'il ait un score de validité élevé et un score de force de contraste moyen, les interventions mises en contraste dans cet essai ne sont pas pertinentes pour répondre à la question décisionnelle sur l'efficacité de la mammographie de dépistage. En effet, il compare deux stratégies de dépistage : la stratégie témoin comprenait un examen physique systématique ainsi qu'un enseignement de l'auto-examen des seins et l'incitation à le pratiquer, alors que la stratégie expérimentale comprenait les mêmes interventions, avec en plus une mammographie annuelle.

SYNTHÈSE DES RÉSULTATS

La présente analyse a relevé des différences marquées dans la qualité des plans d'étude et de l'exécution des essais sur la mammographie, souvent inférieure aux normes modernes, ce qui a compromis leur validité. Certains essais n'étaient pas randomisés du tout, et la plupart des comptes rendus étaient incomplets ou présentaient des contradictions. Tout particulièrement, la plupart des essais n'ont pas fourni les caractéristiques de départ des femmes des groupes expérimental ou témoin, souvent parce qu'on disposait de peu d'informations sur la cohorte témoin. Les cas de cancer préalablement diagnostiqués n'ont pas été systématiquement exclus dans six des huit essais, et aucun essai n'a tenté la mise à l'insu, ni des patientes, ni du personnel médical.

Il appert qu'aucune étude n'a été conçue ni menée de manière à permettre de déterminer la pleine mesure du potentiel du dépistage par mammographie, phénomène que nous avons examiné avec le critère de force du contraste. Si l'on admet que les essais publiés n'ont pas maximisé le potentiel du dépistage mammographique, on peut croire que les programmes modernes permettraient de déceler les lésions à un stade moins avancé et, peut-être, d'obtenir ainsi des réductions plus importantes de la mortalité par cancer du sein que celles dont fait état la littérature scientifique.

La présente évaluation permet de constater que trois essais (HIP, Édimbourg et *Two-County*) comportent des lacunes telles qu'ils ne peuvent servir à évaluer l'efficacité du dépistage. D'autres chercheurs ayant évalué individuellement la qualité de chaque essai les ont exclus pour la même raison. Si l'on examinait seulement les aspects traditionnels de la validité, sans égard à la notion de pertinence ou de contraste, on pourrait douter sérieusement de l'efficacité du dépistage par mammographie. La présente méta-analyse a inclus successivement les résultats des essais par ordre croissant de qualité : qualité bonne ou moyenne (score de validité de 3 ou 4 sur un total de 4), qualité faible (score de 1,5 ou 2, avec force de contraste adéquate), mauvaise qualité (score de 0,5).

Pour l'ensemble des femmes (tous âges confondus), les résultats témoignent d'une relation inverse entre la qualité des essais et la réduction de la mortalité par cancer du sein. Si toutes les études sont incluses, quelle que soit leur qualité, la réduction estimée s'élève à 23 %, mais cette estimation diminue à 15 % lorsque seules les études de qualité moyenne ou faible sont retenues, et enfin à 9 % dans les études de qualité moyenne. Ainsi, les études les plus valides tendent à montrer des réductions moins importantes, et leurs intervalles de confiance comprennent parfois la valeur nulle[1]. Bien que la précision augmente au fur et à mesure que les études sont incluses dans la méta-analyse, elle ne peut compenser une baisse globale de la validité.

Dans les groupes de moins de 50 ans, de nombreuses femmes ont intégré les études vers la fin de la quarantaine. Par conséquent, plusieurs cas de cancer chez ces femmes n'ont pas été détectés avant qu'elles atteignent la cinquantaine : les résultats des études s'appliquent donc à ces femmes, et non à celles qui commenceraient le dépistage à 40 ans. Dans les études de qualité moyenne portant sur ce groupe d'âge, la réduction cumulative du risque est de 2 %. On observe une relation inverse similaire puisque, si l'on inclut les données des études de moindre validité, la réduction globale se situe à 8 %. La réduction de la mortalité est donc beaucoup plus faible chez les femmes plus jeunes, et les intervalles de confiance comprennent la valeur nulle pour toutes les combinaisons d'études.

Dans les groupes de 50 ans et plus, cependant, les résultats sont meilleurs. La seule étude de validité moyenne montre une réduction du risque de 27 %. Si l'on inclut les données des études de faible qualité, la réduction globale est de 24 %, et si l'on tient compte de toutes les données sans égard à

1. La valeur nulle indique une absence d'efficacité. Un intervalle de confiance qui comprend la valeur nulle ne permet donc pas de démontrer l'efficacité de l'intervention.

leur validité, la réduction est de 29 %. Ces réductions de la mortalité sont de beaucoup supérieures à celles que l'on retrouve chez les femmes de tous âges confondus. Les intervalles de confiance dans ce sous-groupe sont naturellement plus grands, mais aucune combinaison d'études ne comprend la valeur nulle.

Quant à la force du contraste, aucune des huit études publiées ne s'approche de la norme d'un programme idéal offrant un dépistage régulier sur de nombreuses années avec du matériel moderne, des mesures d'assurance de la qualité, des mammographies à deux incidences tous les deux ans ou moins et un taux de participation optimal. En effet, selon le score obtenu pour la force du contraste, ces essais n'atteindraient que 12 à 45 % de leur plein potentiel. À titre comparatif, le PQDCS, un programme moderne, obtiendrait un score plus élevé, soit 63 %.

DISCUSSION ET CONCLUSIONS

Question 1 : Quelle est la solidité des arguments scientifiques à l'appui des programmes de dépistage mammographique ?

La validité de la plupart des essais qui appuient le dépistage mammographique suscite de sérieuses réserves en raison des nombreuses faiblesses méthodologiques qui y ont été relevées. En outre, les contrastes étudiés sont hautement hétérogènes et indiquent que le plein potentiel de la mammographie de dépistage n'a pas été réalisé. À la lumière des meilleures données disponibles, on peut conclure que le dépistage entraîne une réduction modérée de la mortalité par cancer du sein de l'ordre de 9 à 15 %. Les données restreintes aux femmes de plus de 50 ans montrent des réductions plus importantes, de l'ordre de 24 à 29 %. Par contre, la présente analyse a montré que la mammographie moderne, effectuée dans des conditions de qualité qui maximisent sa performance, a le potentiel de déceler les lésions cancéreuses à un stade plus précoce, ce qui pourrait réduire davantage la mortalité.

Conclusion : *Les études scientifiques existantes, malgré leurs lacunes, appuient les programmes de dépistage mammographique. De plus, il y a de bonnes raisons de croire que des programmes de dépistage modernes et bien menés pourraient permettre de détecter et de diagnostiquer le cancer du sein à un stade plus précoce et peut-être d'obtenir des réductions de la mortalité par cancer du sein plus importantes que celles qu'ont enregistrées les essais sur le dépistage.*

Question 2 : Quelles sont les preuves à l'appui de la mammographie de dépistage pour les femmes de 40 à 49 ans ?

Pour ce groupe d'âge, les données se font plus rares. En effet, la plupart des études portent sur des femmes de plus de 50 ans, bien qu'un certain nombre de femmes dans certaines études aient commencé le dépistage quelques années avant leur cinquantième anniversaire. Les meilleures données disponibles n'indiquent aucune réduction significative de la mortalité par cancer du sein chez les femmes soumises à un dépistage avant l'âge de 50 ans. En l'absence de données convaincantes sur l'efficacité de la mammographie pour ce groupe d'âge, il est possible que les effets néfastes dépassent les effets bénéfiques.

Conclusion : *Les résultats des essais publiés jusqu'à ce jour ne fournissent pas de justification scientifique permettant de recommander le dépistage pour les femmes de moins de 50 ans. Toutefois, ces résultats n'excluent pas la possibilité que le dépistage individuel, fondé sur une évaluation personnalisée du risque de ces femmes, puisse être bénéfique. Il sera opportun de revoir ces conclusions lorsque les résultats de l'essai du Royaume-Uni seront publiés.*

Question 3 : Quelles sont les implications des données scientifiques sur les possibilités de maximiser l'efficacité de programmes modernes comme le Programme québécois de dépistage du cancer du sein (PQDCS) ?

Même s'il persiste des incertitudes quant à l'efficacité de la mammographie de dépistage,

les programmes modernes ont le potentiel d'obtenir de meilleurs résultats que les essais scientifiques. Bien que le PQDCS exerce déjà un contrôle rigoureux de la qualité des clichés produits, certains aspects de la structure et des processus des essais qui ont été examinés sous la rubrique de la force du contraste peuvent être transposés en normes de qualité additionnelles. Parmi ces aspects, citons la double lecture des clichés et un volume de lecture annuel suffisant pour permettre à chaque radiologiste d'acquérir et de maintenir l'expertise nécessaire pour détecter les cancers du sein à un stade précoce. Ces deux aspects contribuent à réduire les taux de faux positifs et les taux subséquents d'interventions diagnostiques inutiles. Par ailleurs, un taux de participation élevé aux différents cycles de dépistage permettrait aussi d'atteindre et même de dépasser la réduction de la mortalité obtenue dans les essais.

Conclusion : *Les programmes modernes de dépistage du cancer du sein comme le PQDCS pourront atteindre des résultats comparables, voire supérieurs à ceux des essais si leur fonctionnement assure une qualité de mammographie analogue à celle des essais ou meilleure. La qualité des clichés mammographiques, l'expertise des lecteurs et l'instauration de mesures comme la deuxième lecture d'une certaine proportion des clichés constituent des moyens susceptibles de réduire les taux de faux positifs et d'assurer la qualité du dépistage. Bien qu'il faille viser des taux de participation aussi élevés que possible, les efforts déployés en ce sens ne devraient pas surestimer les avantages de la mammographie ni minimiser les incertitudes et les risques éventuels.*

TABLE DES MATIÈRES

LA MISSION i
AVANT-PROPOS iii
REMERCIEMENTS iv
RÉSUMÉ v
1 INTRODUCTION 1
1.1 Demande d'évaluation 1
1.2 Programmes de dépistage 2
1.3 Controverse entourant la qualité des essais sur la mammographie 2
1.4 Les nouvelles questions à l'étude 2
2 MÉTHODOLOGIE 4
2.1 Concepts généraux : pertinence, validité et précision 4
2.2 Contraintes méthodologiques des études sur la mammographie de dépistage 5
2.3 Mise au point du contraste 6
2.4 Force du contraste 6
2.4.1 Contraste technique 6
2.4.2 Époque 7
2.4.3 Qualité du dépistage 7
2.4.4 Participation et contamination 8
2.4.5 Facteur temps : choix du moment pour mesurer les effets du dépistage sur la mortalité (dilution temporelle) 9
2.4.6 Contraste thérapeutique 10
2.4.7 Échelle de la force du contraste 10
2.5 Biais à direction inconnue 10
2.5.1 Randomisation 11
2.5.2 Comparabilité de base 11
2.5.3 Traitement des exclusions 11
2.5.4 Uniformité de l'information relative aux résultats 12
2.5.5 Échelle de validité 12
2.6 Précision : minimiser les effets du hasard 12
2.7 Indicateurs de résultats pertinents à la question de recherche 13
3 ANALYSE DES ESSAIS PUBLIÉS ET COMPARAISON AVEC LE CONTEXTE QUÉBÉCOIS 15
3.1 Le dépistage au Québec 15
3.2 Essais publiés sur la mammographie de dépistage 16
3.2.1 New York (HIP) 16
3.2.2 Malmö (MMST) 16
3.2.3 *Two-County* (TCS) 17
3.2.4 Édimbourg 17
3.2.5 Canada-1 (NBSS-1) 18
3.2.6 Canada-2 (NBSS-2) 18
3.2.7 Stockholm 19
3.2.8 Göteborg 19
3.2.9 Royaume-Uni (*UK Age Trial*) 19
3.3 Comparaison de la structure et des processus des essais avec le PQDCS 23
3.3.1 Structure et processus 23
3.3.2 Participation 24
3.3.3 Dilution temporelle 25
3.3.4 Force du contraste 25

4 SYNTHÈSE DES RÉSULTATS 27
4.1 Évaluations réalisées par d'autres chercheurs 27
4.2 Approche d'analyse de la preuve 29
4.3 Résultats sur la mortalité par cancer du sein dans tous les groupes d'âge 32
4.4 Résultats sur la mortalité par cancer du sein chez les femmes plus jeunes 33
4.5 Résultats sur la mortalité par cancer du sein chez les femmes plus âgées 35
5 DISCUSSION ET CONCLUSIONS 37
ANNEXE A LA MISE EN CONTRASTE DU DÉPISTAGE ET SA DURÉE 41
ANNEXE B FACTEUR TEMPS : CHOIX DU MOMENT POUR ÉVALUER LES EFFETS DU DÉPISTAGE SUR LA MORTALITÉ PAR CANCER DU SEIN 45
ANNEXE C ÉCHELLE DE LA FORCE DU CONTRASTE 52
ANNEXE D ÉCHELLE DE VALIDITÉ 57
ANNEXE E ANALYSE DES NEUF ESSAIS 60
RÉFÉRENCES 73

LISTE DES TABLEAUX ET DES FIGURES

Tableau 1 Description des neuf essais 21
Tableau 2 Scores de validité 22
Tableau 3 Scores de la force du contraste 22
Tableau 4 Scores pour la structure et les processus des essais et du PQDCS 25
Tableau 5 Recensions précédentes : essais inclus et exclus 27
Tableau 6 1re étape : essais pertinents pour l'évaluation d'un programme moderne de dépistage 30
Tableau 7 2e étape : essais par ordre de validité 30
Tableau 8 3e étape : force du contraste des essais 31
Tableau B-1 Paramètres particuliers des essais 50
Tableau B-2 Calcul de la dilution temporelle (exemple de l'essai de Göteborg) 51
Tableau C-1 Échelle de qualité et score attribué 53
Tableau C-2 Évaluation de la force du contraste 55
Tableau C-3 Calcul du score de chaque essai : qualité 56
Tableau C-4 Différence entre les cohortes : participation - contamination 56
Tableau C-5 Calcul du score de chaque essai : force du contraste 56
Tableau D-1 Scores de validité attribués aux essais publiés sur la mammographie 59

Figure 1 Synthèse des résultats sur la mortalité par cancer du sein (tous âges confondus) 33
Figure 2 Synthèse des résultats sur la mortalité par cancer du sein (femmes plus jeunes) 35
Figure 3 Synthèse des résultats sur la mortalité par cancer du sein (femmes plus âgées) 36
Figure A-1 Modèle de base de la mise en contraste du dépistage 42
Figure B-1 Modèle de base de la mise en contraste du dépistage 45
Figure B-2 Modèle d'un décalage de 5 ans 47
Figure B-3 Modèle d'un décalage de 10 ans 47
Figure B-4 Profil de mortalité, décalage variant continuellement entre 5 et 10 ans 48

1 INTRODUCTION

1.1 DEMANDE D'ÉVALUATION

Le ministère de la Santé et des Services sociaux du Québec a demandé à l'Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS) de réexaminer la pertinence d'inclure les femmes de 40 à 49 ans dans son programme de dépistage mammographique.

Le Conseil d'évaluation des technologies de la santé (CETS, prédécesseur de l'AETMIS) avait déjà réalisé deux rapports sur la mammographie de dépistage. Le premier, publié en novembre 1990, avait pour objectif d'évaluer les données scientifiques sur les effets sur la santé du dépistage du cancer du sein afin d'estimer les répercussions d'un programme éventuel au Québec. Le rapport était basé sur les résultats de quatre essais contrôlés randomisés : le *Health Insurance Plan Trial* (HIP) de New York [Shapiro *et al.*, 1985], le *Malmö Mammographic Screening Trial* [Andersson *et al.*, 1988], le *Two-County Trial* de Suède [Tabár *et al.*, 1985a] et *l'Edinburgh Randomised Trial of Screening for Breast Cancer* [Roberts *et al.*, 1984]. À l'époque, quatre autres essais étaient en cours : deux au Canada, le NBSS-1 et le NBSS-2 [Miller *et al.*, 1992a; 1992b], et deux en Suède, soit à Stockholm [Frisell *et al.*, 1986] et à Göteborg [Andersson *et al.*, 1983]. Les résultats de ces derniers essais n'avaient cependant pas encore été publiés, et l'analyse de 1990 n'avait pu en tenir compte. Le premier rapport du Conseil se fondait sur les données sur la mortalité par cancer du sein comptabilisées à partir de la cinquième année de chaque essai. Selon ce rapport, le dépistage réduirait la mortalité par cancer du sein de 35 % (intervalle de confiance [IC] à 95 % : de 24 à 47 %). Lorsque l'analyse était restreinte aux femmes âgées de 50 à 69 ans, la réduction était de 45 % (IC 95 % : de 33 à 58 %) [CETS, 1990]. Étant donné le nombre élevé de mammographies déjà effectuées au Québec en 1990, les recommandations avaient été faites dans le but d'optimiser les activités de dépistage et de diagnostic existantes, particulièrement en ciblant le groupe d'âge pour lequel elles offraient les plus grands bienfaits, soit les femmes de 50 à 69 ans, et d'améliorer la qualité du dépistage.

Le Conseil a publié une deuxième étude sur le dépistage du cancer du sein en mars 1993. Elle avait pour objectif d'évaluer les bienfaits possibles du dépistage mammographique pour les femmes âgées de 40 à 49 ans. Depuis 1990, de nouvelles données provenant des quatre essais déjà révisés dans le premier rapport avaient été publiées, et les résultats préliminaires de trois autres études, soit les deux essais du Canada (NBSS-1 et NBSS-2) [Miller *et al.*, 1992a; 1992b] et celui de Stockholm [Frisell *et al.*, 1991], avaient fait l'objet de publications. Bien que seul l'essai NBSS-1 ait spécifiquement examiné le groupe d'âge de 40 à 49 ans, tous les essais ont fourni des données sur les femmes de moins de 50 ans. Encore une fois, l'analyse était limitée aux décès par cancer du sein apparus après cinq ans de suivi, sauf pour l'essai canadien, qui ne donnait que les taux de mortalité cumulatifs, ce qui ne permet pas d'analyser les taux de mortalité pour chaque année. En combinant les résultats des cinq essais les plus récents disponibles à l'époque, le CETS a obtenu une augmentation de la mortalité par cancer du sein de 1 % chez les femmes soumises au dépistage (IC 95 % : de - 26 à + 28 %). Il a conclu que le dépistage mammographique chez les femmes de moins de 50 ans ne réduisait pas la mortalité, mais que la qualité de la preuve n'était pas suffisante pour exclure la possibilité d'un avantage [CETS, 1993].

Depuis 1993, les chercheurs des sept essais cités dans le rapport de 1993 ont publié de nouvelles données basées sur des suivis plus poussés, et les premiers résultats de l'essai de Göteborg, en Suède, ont paru [Bjurstam *et al.*, 2003]. En règle générale, ces données plus récentes confirmaient l'efficacité théorique du dépistage mammographique, surtout

chez les femmes âgées de 50 à 69 ans. Un autre essai britannique entrepris en 1991 n'avait pas encore publié ses résultats [Moss, 1999].

1.2 PROGRAMMES DE DÉPISTAGE

En 1998, le Québec et les neuf autres provinces canadiennes avaient instauré des programmes de dépistage, justifiés en grande partie par la preuve de leur efficacité fournie par les essais sur le sujet. Dans tous ces programmes, les femmes âgées de 50 à 69 ans sont invitées, soit par leur médecin, soit par lettre, à subir une mammographie de dépistage tous les deux ans. Les femmes plus jeunes ont accès à la mammographie de dépistage dans les mêmes centres si leur médecin la leur prescrit, mais la plupart des provinces, dont le Québec, ne recrutent pas activement les femmes plus jeunes [Santé Canada, 2003].

1.3 CONTROVERSE ENTOURANT LA QUALITÉ DES ESSAIS SUR LA MAMMOGRAPHIE

Un rapport relativement récent a semé le doute sur l'efficacité du dépistage mammographique pour réduire les décès par cancer du sein. Gøtzsche et Olsen, membres du groupe de collaboration nordique Cochrane, ont publié les résultats de leur revue systématique de la littérature sur le dépistage mammographique dans le *Lancet* en 2000 [Gøtzsche et Olsen, 2000], et dans un rapport complet du groupe de collaboration Cochrane en 2001 [Olsen et Gøtzsche, 2001]. Dans ces rapports, ils ont mis en doute la validité des essais randomisés publiés et le choix de la mortalité par cancer du sein comme indicateur d'efficacité. Selon eux, deux des sept[2] essais étaient de mauvaise qualité, trois étaient de faible qualité, et deux étaient de qualité moyenne. Lorsqu'ils ne tiennent compte que des essais de qualité moyenne, ils concluent que le dépistage mammographique n'a pas d'effet significatif sur la mortalité par cancer du sein (risque relatif [RR] de 0,97; IC 95 % : de 0,82 à 1,14) ni sur la mortalité globale (toutes causes confondues) (RR = 1,05; IC 95 % : de 0,83 à 1,33). Lorsqu'ils tiennent compte des essais de faible qualité, ils constatent que le dépistage a un effet significatif sur la mortalité par cancer du sein (RR = 0,68; IC 95 % : de 0,58 à 0,78), mais pas sur la mortalité globale (toutes causes confondues) (RR = 1,01; IC 95 % : de 0,98 à 1,04). Ils n'ont pas jugé opportun de combiner les résultats des essais de moyenne et de faible qualité en raison de leur hétérogénéité et des divergences méthodologiques entre les deux groupes d'essais.

Les doutes soulevés par Gøtzsche et Olsen ont semé la controverse, et plusieurs chercheurs ont répliqué par des corrections, des réfutations et de nouvelles données. Les organisations nationales et internationales responsables des guides de pratique clinique et des programmes de dépistage ont été forcées de réexaminer les preuves sur lesquelles reposaient leurs recommandations et leurs activités.

1.4 LES NOUVELLES QUESTIONS À L'ÉTUDE

Depuis l'évaluation du CETS de 1993, des programmes de dépistage ont été instaurés dans les provinces qui n'en avaient pas déjà. De plus, plusieurs chercheurs ont publié des mises à jour des résultats de leurs études, et des débats scientifiques sur l'âge auquel les femmes devraient commencer le dépistage, la validité des nombreux essais sur le dépistage ainsi que la pertinence de la mammographie de dépistage en général ont eu lieu. La situation ayant évolué, les questions de recherche examinées dans ce rapport sont :

1) Quelle est la solidité des arguments scientifiques à l'appui des programmes de dépistage mammographique ?

2. Ces chercheurs considèrent les essais canadiens comme faisant partie d'un seul essai. Dans le présent rapport, le NBSS-1 et le NBSS-2 sont considérés comme deux essais distincts, puisque les interventions qui y sont comparées sont de nature différente.

2) Quelles sont les preuves à l'appui de la mammographie de dépistage pour les femmes de 40 à 49 ans ?
3) Quelles sont les implications des données scientifiques sur les possibilités de maximiser l'efficacité de programmes modernes de dépistage comme le Programme québécois de dépistage du cancer du sein (PQDCS) ?

Tout d'abord, les aspects méthodologiques des essais sur le dépistage mammographique seront examinés, et les éléments pertinents pour les prises de décisions sur les programmes de dépistage seront ensuite dégagés. Pour répondre aux deux premières questions, les essais publiés seront passés en revue en fonction du contexte québécois. Bien que les programmes de dépistage établis ne soient pas structurés pour être comparés et ne puissent être utilisés pour évaluer l'efficacité du dépistage, la comparaison de leurs caractéristiques avec celles des essais permettra d'évaluer dans quelle mesure les résultats de ces derniers pourraient être reproduits dans chaque programme de dépistage.

Dans les sections qui suivent, tous les essais publiés sur le dépistage mammographique seront examinés à la lumière de ces questions, notamment leurs forces et leurs faiblesses méthodologiques. Enfin, les résultats de recherche les plus pertinents pour les prises de décisions seront dégagés et appliqués au contexte québécois.

2 MÉTHODOLOGIE

2.1 CONCEPTS GÉNÉRAUX : PERTINENCE, VALIDITÉ ET PRÉCISION

Pour qu'une étude sur l'efficacité d'une intervention en santé puisse étayer la prise de décisions, elle doit remplir trois conditions essentielles : la pertinence, la validité et la précision. De façon générale, la **pertinence** permet de bien circonscrire l'intervention étudiée par rapport à d'autres options visant les mêmes effets. La **validité** implique qu'une étude est conçue, menée et analysée de sorte qu'aucun biais important n'affecte la mesure de la différence dans l'efficacité des stratégies comparées. La validité peut être compromise par deux types de biais : les biais à direction connue et les biais à direction inconnue. Enfin, la **précision** exige que l'étude fournisse une estimation de l'efficacité qui ne soit pas vulnérable aux effets aléatoires. Des résultats précis présentent des erreurs types négligeables, d'où leur association à de petits intervalles de confiance.

Les recensions des essais sur la mammographie de dépistage se sont surtout intéressées à la validité, et tout particulièrement aux facteurs associés aux biais à direction inconnue. Les questions de validité telles que la randomisation, les équivalences entre les cohortes de dépistage et témoin à l'entrée dans l'étude, les critères d'exclusion et l'équivalence des suivis pour la mesure des résultats ont été examinées de près. Ces recensions se sont aussi penchées sur la précision de l'estimation du taux de mortalité par cancer du sein, car chaque étude prise individuellement présentait de grands intervalles de confiance. En combinant les résultats de ces études, technique connue sous le nom de méta-analyse [DerSimonian et Laird, 1986], ces auteurs ont obtenu des intervalles de confiance plus petits. Toutefois, les aspects de la pertinence et de la validité, notamment les biais à direction connue, n'ont pas été traités à fond dans les différents essais.

C'est pourquoi la présente analyse mettra davantage l'accent sur les conditions nécessaires pour assurer la pertinence et la validité dans l'examen de chaque essai randomisé. L'approfondissement de ces aspects permet de jeter un éclairage nouveau sur l'interprétation de la littérature scientifique sur le dépistage mammographique, et peut aussi contribuer à mieux structurer un programme de dépistage. Notre démarche analytique sera complétée par le recours à la méta-analyse afin de bien aborder l'enjeu de la précision.

Contraste

Afin de mieux cerner les conditions nécessaires pour assurer la pertinence et la validité d'une étude, nous introduisons la notion particulière de contraste.

> Définition : Le ***contraste*** d'une étude expérimentale est l'opposition ou la divergence temporaire entre l'intervention expérimentale offerte à la cohorte de dépistage et la stratégie de référence appliquée à une cohorte témoin.

Comme la question décisionnelle principale est l'utilité de la mammographie de dépistage, les études pertinentes seront celles qui mettront en contraste le dépistage mammographique et, idéalement, le non-dépistage. Ne seront donc pas retenues dans l'analyse les études comparant formellement différentes stratégies de dépistage, incluant ou non la mammographie. Dans la pratique, la stratégie de référence (non-dépistage) d'un essai pourra contenir certaines activités non contrôlables de dépistage, ce qui affaiblira le contraste avec la stratégie expérimentale.

Force du contraste

Bien que la validité soit toujours un enjeu majeur dans la conduite d'essais sur le dépistage mammographique, de façon générale, les recensions se sont davantage concentrées sur le contrôle des biais à direction inconnue. Ces biais sont susceptibles de se produire lorsque les groupes de dépistage et témoin ne pré-

sentent pas au départ le même risque de mortalité par cancer du sein ou ne sont pas soumis aux mêmes interventions (à l'exception de la présence ou de l'absence de dépistage). Ces biais peuvent également se manifester lorsque la mesure des résultats dans les deux groupes n'est pas indépendante de l'affectation des sujets au groupe de dépistage ou témoin.

Tout en étudiant les biais à direction inconnue, la présente évaluation traite de façon plus approfondie des biais à direction connue, et le fait à l'aide de la notion de force du contraste.

> Définition : La ***force du contraste*** d'un essai correspond au degré auquel la divergence entre l'intervention de dépistage et la stratégie témoin est atteinte et maintenue, et auquel les effets de cette divergence sont adéquatement captés.

Les caractéristiques principales d'un contraste fort sont la qualité technique de l'intervention de dépistage, le degré de participation au dépistage du groupe expérimental et l'absence de dépistage dans le groupe témoin, ainsi que le choix et l'analyse de résultats (indicateurs et mesures) qui reflètent pleinement les différents effets des deux interventions.

Un contraste faible aura tendance à fausser les résultats d'une étude vers la direction de l'hypothèse nulle, selon laquelle l'effet de l'intervention de dépistage et celui de l'intervention témoin sont de même valeur. Ce concept sera développé davantage à la section 2.4, à l'annexe A (La mise en contraste du dépistage et sa durée), à l'annexe B (Facteur temps : choix du moment pour évaluer les effets du dépistage sur la mortalité par cancer du sein) et à l'annexe C (Échelle de la force du contraste).

2.2 CONTRAINTES MÉTHODOLOGIQUES DES ÉTUDES SUR LA MAMMOGRAPHIE DE DÉPISTAGE

L'étude du dépistage par mammographie est plus difficile que celle d'autres types d'interventions, car la mammographie est soumise à trois contraintes méthodologiques. Premièrement, elle vise à déceler une maladie (cancer du sein) qui, sans traitement, connaît habituellement une évolution clinique s'étendant sur plusieurs années. Alors que plusieurs cancers sont fatals dans les premières années qui suivent le diagnostic (d'où l'emploi de la statistique de survie sur cinq ans), le cancer du sein a un début particulièrement insidieux, et les décès surviennent souvent plus de cinq ou même dix ans après le diagnostic. Cela est encore plus vrai lorsque le diagnostic est précoce, ce qui arrive couramment lors d'un dépistage. À cause de ce long délai entre le diagnostic et le décès, un suivi de plusieurs années est nécessaire si l'on veut établir avec précision le résultat définitif d'une intervention qui vise à réduire la mortalité.

Deuxièmement, le dépistage mammographique n'est pas une intervention ponctuelle, mais comporte plutôt des examens de dépistage périodiques qui s'étendent sur plusieurs années. Apparaissent alors toutes sortes de facteurs de variabilité dans les plans de l'étude ou du programme, comme les intervalles entre les mammographies, sans compter le défi que représente la définition de concepts comme le taux de participation, l'intervalle moyen entre les mammographies et le nombre total de séances de dépistage compte tenu du fait que la participation varie largement d'une femme à l'autre.

Troisièmement, la durée de mise en contraste des essais sur le dépistage diffère grandement de la durée des programmes de dépistage

mammographique proposés ou en place dans divers pays. Dans les huit essais publiés sur la mammographie, la durée de mise en contraste du dépistage variait de 3,5 ans (New York) à 8,8 ans (Malmö). Par contre, un programme de dépistage offre habituellement des mammographies pendant 20 ans s'il commence lorsque les femmes atteignent l'âge de 50 ans, et même plus longtemps si elles sont invitées à l'intégrer avant l'âge de 50 ans ou si elles continuent après l'âge de 70 ans. Bien entendu, un essai ayant une période de mise en contraste aussi longue exigerait d'énormes ressources, et les résultats ne seraient disponibles que 30, voire 50 ans après le début de l'étude. Plutôt que de comparer une cohorte de jeunes femmes invitées à participer à un programme de dépistage sur 20 ou 30 ans, presque tous les essais ont choisi d'étudier une période de mise en contraste plus courte[3], et visaient à vérifier l'hypothèse selon laquelle la mammographie de dépistage peut réduire la mortalité par cancer du sein, et non le degré de réduction de la mortalité que pourrait donner un programme complet sur 20 ou 30 ans. Ce choix est fait pour des raisons pratiques, mais l'extrapolation des résultats des essais pour orienter les décisions sur des programmes de dépistage, qui sont beaucoup plus longs, a des implications qui seront examinées dans la section suivante.

2.3 MISE AU POINT DU CONTRASTE

Pour évaluer la pertinence d'un essai sur le dépistage mammographique, il faut d'abord vérifier quelle technique de mammographie est à l'étude et à quoi elle est comparée. Différentes modalités de dépistage ont été proposées, évaluées et intégrées dans divers programmes de dépistage. Tout particulièrement, les techniques de mammographie, dont le type de cliché (incidence médiolatérale, médiolatérale oblique [MLO], craniocaudale [CC], etc.), ne sont pas normalisées. D'autres aspects techniques comme la compression des seins, le positionnement des seins, le moment du dépistage à l'intérieur du cycle menstruel ainsi que la préparation diététique et cosmétique n'ont généralement pas été décrits dans la littérature publiée, mais ils varient probablement d'une étude à l'autre.

Quel que soit l'essai, le dépistage mammographique doit être comparé à une autre stratégie. Pour connaître le degré de contraste atteint, il est particulièrement important de savoir quel est le but de la mammographie (dépistage ou diagnostic) et s'il y a eu ou non examen clinique des seins et auto-examen des seins dans la cohorte de dépistage et dans la cohorte témoin. La plupart des essais n'ont pas donné de détails sur la stratégie de comparaison, se contentant de la qualifier de « soins usuels » (ou une variante). Étant donné que les « soins usuels » varient selon les lieux et les époques, on ne peut que supposer en quoi ils consistaient dans chaque étude.

2.4 FORCE DU CONTRASTE

Pour qu'un essai donne une mesure valide du potentiel de la mammographie de dépistage, il doit réaliser un contraste fort entre l'intervention de dépistage et l'intervention témoin. Évidemment, si un essai ne réussit pas à obtenir un tel contraste, ses résultats seront biaisés vers l'hypothèse nulle[4]. Les éléments qui contribuent à marquer le contraste entre l'intervention de dépistage et la situation ou stratégie témoin seront examinés ici. Ces éléments serviront de base à l'élaboration d'une échelle permettant de mesurer la force du contraste de chaque étude.

2.4.1 Contraste technique

Le contraste technique réfère à la différence entre les interventions offertes au groupe de dépistage et au groupe témoin. En règle générale, dans les essais analysés, on retrouve

3. L'*UK Age Trial* constitue une exception intéressante et fera l'objet d'une discussion ultérieure.

4. Toute recherche est fondée sur un test d'hypothèse, qui amène les chercheurs à confronter deux hypothèses statistiques : l'hypothèse nulle (il n'y a aucune différence ou association) et l'hypothèse alternative (il y a une différence ou association). Ici, l'hypothèse nulle signifie qu'il n'y a pas de différence dans l'effet observé (taux de mortalité) entre les femmes qui ont reçu l'intervention de dépistage (groupe expérimental) et celles qui ne l'ont pas eue (groupe témoin).

trois principaux types de contrastes techniques : 1) invitation à la mammographie uniquement *versus* aucune invitation (toutes les études suédoises); 2) invitation à la mammographie et à l'examen clinique des seins *versus* aucune invitation (HIP, Édimbourg, NBSS-1); 3) invitation à la mammographie, à l'examen clinique des seins et à l'auto-examen des seins *versus* invitation à l'examen clinique et à l'auto-examen des seins. Si certains contrastes techniques ne semblent pas pertinents pour les questions à l'étude, d'autres, jugés pertinents, présentent néanmoins des forces de contraste différentes qu'on peut apprécier. Par exemple, une mammographie associée à un examen clinique des seins *versus* aucun dépistage offre un contraste plus marqué qu'une mammographie seule *versus* aucun dépistage.

2.4.2 Époque

Le dépistage mammographique a été instauré dans les années 1950 et réalisé à partir de clichés produits par des appareils de radiographie standard. Le premier appareil conçu expressément pour la mammographie date de 1966 [GE Medical Systems, 2004], et la technologie n'a cessé d'évoluer depuis. C'est en 1963 que la première étude sur le dépistage mammographique a été mise sur pied (HIP), et depuis, chaque décennie comporte son lot de nouvelles études. Certains experts estiment que les progrès qualitatifs les plus importants ont été réalisés au début des années 1980. Sickles [1997], par exemple, a démontré que la précision des images obtenues par la mammographie moderne en pratique clinique aux États-Unis était supérieure à celle des appareils utilisés dans les huit essais sur le dépistage.

2.4.3 Qualité du dépistage

Bien que la technique mammographique de base n'ait pas beaucoup changé depuis 40 ou 50 ans, elle s'est progressivement raffinée de sorte que les techniques de compression, la distance foyer-cliché, la résolution et le contraste des clichés, l'exposition à la radiation, le développement des clichés, la lecture en négatoscope et l'archivage se sont généralement améliorés avec le temps. La mammographie est une technique dont la qualité dépend d'une longue chaîne d'activités humaines, chacune devant être exécutée avec expertise pour donner la pleine mesure de son potentiel. La production de clichés techniquement supérieurs représente un défi particulier, les clichés mammographiques exigeant une plus haute résolution que les clichés radiologiques standard. Distinguer les subtilités entre des anomalies radiographiques bénignes et malignes pour obtenir des taux élevés de détection du cancer tout en gardant le nombre de faux positifs aussi bas que possible exige une grande compétence. Entre autres aspects d'un dépistage de qualité, on peut citer des mesures de contrôle de la qualité telles la double lecture de clichés et l'expertise des professionnels qui lisent les clichés. De plus en plus, la norme en mammographie évolue vers la prise de deux incidences plutôt qu'une seule, pour chaque sein. Dans certaines études (*Two-County*, Stockholm), on n'a utilisé qu'une seule incidence, ou encore des incidences de type autre que médiolatéral oblique (NBSS 1 et 2), du moins pour certains examens. On a aussi parfois recours à une seule incidence comme cliché initial, avec une seconde incidence lorsque les seins sont denses ou que les résultats de la mammographie sont positifs. Dans les protocoles les plus récents, on tend à utiliser les incidences médiolatérale oblique et craniocaudale pour chaque sein. Enfin, l'intervalle[5] entre les mammographies devrait être assez court pour permettre de détecter la plupart des nouvelles tumeurs alors qu'elles sont encore petites. Dans les essais publiés sur la mammographie, l'intervalle suggéré entre les examens mammographiques variait grandement, allant de 12 mois (HIP, NBSS 1 et 2), à 18 mois (Malmö, Göteborg), à 24 mois (*Two-County*, Édimbourg), et à 28 mois

5. Bien que les études proposent toujours un intervalle bien défini ou variable, les intervalles réels entre les mammographies peuvent être plus courts ou, le plus souvent, plus longs, car les sujets sont libres de participer au moment qui leur convient le mieux. Toutefois, cette variation individuelle sera examinée ici sous le thème « participation », et le terme « intervalle » conservera le sens proposé par les concepteurs de l'étude.

(Stockholm). Dans certains cas, il y avait des intervalles différents à l'intérieur de la même étude (Malmö, *Two-County*).

2.4.4 Participation et contamination

Outre ces différences qualitatives, les études ont divergé dans la capacité de maintenir le même taux de participation pendant toute la durée de la mise en contraste. On obtient un plus grand contraste en atteignant un taux élevé d'observance (participation au programme) dans le groupe de dépistage et en limitant le recours à la technique expérimentale dans le groupe témoin (contamination).

Par souci de validité, les chercheurs de tous les essais ont analysé les résultats obtenus dans le groupe de dépistage (expérimental) et dans le groupe témoin, que les femmes du groupe de dépistage aient subi ou non une mammographie, ou que les femmes du groupe témoin en aient eu une ou non. Cette approche est connue sous le nom d'« analyse en intention de traiter ». Du point de vue du maintien de la validité, cette approche est adéquate, étant donné qu'elle conserve l'égalité initiale entre les deux groupes formés par la randomisation. Cependant, cette analyse se fait au prix d'une réduction du contraste mesuré entre les femmes soumises au dépistage et celles qui ne l'étaient pas. Comme les femmes de la cohorte expérimentale n'ont pas toutes eu une mammographie et que plusieurs femmes de la cohorte « témoin » en ont subi une, les résultats globaux des essais peuvent, au mieux, montrer l'effet « moyen » auquel on peut s'attendre dans une population à qui le dépistage est offert. Si le but de l'essai est d'anticiper l'effet de l'instauration d'un programme de dépistage dans une population qui a peut-être déjà eu une mammographie et qui participera au programme de manière sous-optimale, l'analyse en intention de traiter sera non seulement valide, mais aussi pertinente.

Cependant, pour la femme qui envisage de subir une mammographie de dépistage, l'analyse en intention de traiter sous-estime habituellement l'effet du dépistage. De toute évidence, une femme qui ne participe pas au programme de dépistage qui lui est offert n'en tirera aucun bienfait. Il ne fait aucun doute non plus qu'une femme qui participe au dépistage en tirera un bienfait supérieur à la moyenne de la cohorte expérimentale, qui inclut des non-participantes[6]. Les résultats de l'efficacité théorique à l'échelle populationnelle représentent une moyenne pondérée de l'efficacité théorique pour les participantes (inconnue) et les non-participantes (zéro).

Pour que l'analyse comparative soit juste, il faut mesurer les résultats de toutes les femmes à qui le dépistage était offert (paradigme de l'intention de traiter). Cependant, un essai qui ne réussit pas à maintenir un haut degré de contraste ne pourra montrer le plein effet du dépistage, et ses résultats doivent être interprétés en conséquence. Bien que, d'une perspective sociétale, il soit justifié de projeter les résultats d'une étude pour prédire l'effet du dépistage dans une population donnée, il ne faut pas oublier que ces résultats ne s'appliquent ni à la personne qui ne participe pas au programme (qui ne tire aucun bienfait de l'offre de dépistage), ni à celle qui y participe de façon optimale (qui devrait en tirer plus de bienfaits que la moyenne de la population invitée au dépistage).

Glasziou [1992] a traité de ce problème par rapport au dépistage du cancer du sein et a proposé une formule d'ajustement des résultats[7]. Bien qu'un tel ajustement quantitatif n'ait pas été fait ici, un procédé équivalent a

6. La contraception illustre bien ce phénomène, où l'on distingue l'efficacité d'une « méthode » contraceptive de l'efficacité de l'« usage » d'un contraceptif. La première est la protection moyenne dont jouit un groupe de personnes qui utilise la méthode contraceptive; ce groupe comprend des personnes qui oublient de prendre le contraceptif ou utilisent mal la méthode. Le deuxième type d'efficacité représente la protection dont jouissent les individus qui utilisent le contraceptif, et elle est, bien sûr, beaucoup plus élevée. L'efficacité de la mammographie de dépistage a toujours été considérée comme l'efficacité d'une « méthode », mais il est sûr que l'efficacité de l'« usage » d'une femme individuelle est plus élevée. Cela illustre également la différence entre l'usage idéal (efficacité théorique) et l'usage réel, sous-optimal (efficacité pratique) pouvant s'appliquer à un contexte donné.

7. Glasziou propose essentiellement une formule (attribuée à Newcombe) par laquelle l'estimation obtenue avec l'analyse en intention de traiter peut être ajustée par un facteur $1/\Delta$, où $\Delta = p_1 + p_2 - 1$; p_1 et p_2 étant les taux d'observance dans les groupes de dépistage et témoin. Comme p_2 est le taux d'observance dans le groupe témoin, $1 - p_2$ est le taux de contamination, et $p_1 + p_2 - 1 = p_1 - (1 - p_2)$, de sorte que le facteur d'ajustement peut être exprimé comme $1/(p - c)$, p étant le taux de participation dans la cohorte de dépistage et c, le taux de contamination dans la cohorte témoin. Un tel ajustement pour la non-participation et la contamination a aussi été nommé estimation « causale » par Baker et ses collaborateurs [2002].

été utilisé, qui consiste à considérer la participation sous-optimale pour interpréter la force du contraste obtenue par une étude donnée.

2.4.5 Facteur temps : choix du moment pour mesurer les effets du dépistage sur la mortalité (dilution temporelle)

La durée du suivi pendant lequel seront évalués les effets d'un protocole de dépistage constitue un élément important de l'analyse des essais sur le dépistage. Pour évaluer le potentiel de réduction de la mortalité d'un programme, il est essentiel, rappelons-le, d'examiner les données relatives aux décès suffisamment longtemps après le début du dépistage. Cette question a été traitée assez longuement dans le rapport du CETS de 1990, où l'on a tenté de déterminer la réduction potentielle de la mortalité obtenue par un programme de dépistage en régime établi après une période initiale sans effets. Miettinen et ses collaborateurs [2002] résument ainsi la situation : « […] la réduction du taux de létalité probablement attribuable au dépistage se solde par une diminution des décès par cancer chez les femmes soumises au dépistage *seulement après un délai suffisant*, et non au moment de leur entrée dans l'essai; il faut donc se concentrer sur les décès qui se sont produits dans le segment approprié du suivi – pas trop tôt après l'entrée dans l'étude, ni trop tard après la fin du dépistage » [trad.] (les italiques sont de nous).

Il est important de préciser que la plupart des études n'ont pas donné les réductions du taux de mortalité par cancer du sein pour la période d'effet maximal, mais ont plutôt indiqué la mortalité cumulative à partir du premier jour suivant le début du dépistage. Les résultats des essais sur le dépistage sont habituellement publiés en termes de ratio des taux cumulatifs de mortalité par cancer du sein entre les femmes soumises au dépistage et les femmes qui ne l'ont pas été. Ces taux cumulatifs sont calculés à partir du nombre total de décès par cancer du sein apparus dans chaque cohorte durant toute la période de suivi, soit du début du dépistage à la fin du suivi, divisé par le nombre de personnes-années d'observation. Cette pratique est largement acceptée : « On compare la mortalité attribuable au cancer du sein dans les deux branches de l'essai, de la date de la randomisation à la fin du suivi » [trad.] [Paci et Alexander, 1997].

Les effets d'un programme de dépistage sur la mortalité par cancer du sein se manifestent seulement cinq ans ou plus après le début du dépistage. Le fait de cumuler les décès survenus dans les premières années dilue les effets du programme de dépistage (dilution précoce). Plus le suivi est prolongé, plus les effets des interventions pratiquées pendant la période de mise en contraste (ratios des taux « en régime établi ») comptent dans les taux de mortalité cumulatifs. Toutefois, si le suivi se prolongeait au-delà de la zone de pleine réduction de la mortalité, on commencerait à comptabiliser les décès des cas de cancer du sein apparus après la période de mise en contraste, diluant encore une fois l'effet optimal du dépistage (dilution tardive).

Les effets de la dilution tant précoce que tardive peuvent être corrigés, mais pas éliminés, si l'on considère seulement les décès survenus chez les femmes dont le cancer a été diagnostiqué pendant la mise en contraste du dépistage. Cette stratégie, appelée « modèle d'évaluation » [Larsson *et al.*, 1997], a été utilisée dans les récentes revues des études suédoises. Cependant, une telle analyse ne peut s'appliquer qu'aux essais dans lesquels un cycle de dépistage est offert au groupe témoin à la fin de la période d'étude. Le cycle témoin de dépistage est nécessaire pour détecter les cas « latents », c'est-à-dire les cas qui ne sont pas encore symptomatiques. L'absence d'un tel cycle témoin créerait un biais lié au devancement du diagnostic *(lead-time bias)* qui ferait tendre les résultats vers l'hypothèse nulle, puisque les cas sont plus susceptibles d'être détectés précocement dans la population soumise au dépistage [Nyström *et al.*, 1993]. Il en résulte que cette analyse ne peut être appliquée qu'aux études suédoises *Two-County*, Stockholm et Göteborg, où un

cycle témoin a été offert aux deux cohortes à la fin de la mise en contraste du dépistage.

Dans ses rapports précédents, le CETS [1993; 1990] a tenté de corriger le problème de dilution précoce en examinant seulement les décès par cancer du sein apparus au moins cinq ans après le début du dépistage, ce qui lui a permis d'exclure les données des premières années, qui contribuent le plus à diluer l'effet du dépistage sur la mortalité[8]. Le « modèle d'évaluation » suédois contrecarre plus efficacement la dilution, précoce et tardive, en ne tenant pas compte des cas détectés avant et après la mise en contraste du dépistage. Dans la présente analyse, cette ligne de pensée a été poussée plus loin par l'évaluation du degré de dilution des réductions de la mortalité qu'ont engendré les techniques analytiques utilisées par les investigateurs, et ce, pour chaque étude. Pour plus de détails sur cette approche, voir l'annexe A (La mise en contraste du dépistage et sa durée) et l'annexe B (Facteur temps : choix du moment pour évaluer les effets du dépistage sur la mortalité par cancer du sein).

2.4.6 Contraste thérapeutique

Le dépistage est efficace s'il réussit à déceler les lésions avant qu'elles soient détectées autrement ou plus tard, permettant ainsi un traitement plus efficace parce qu'administré à un stade précoce. Jusqu'ici, ce sont surtout les aspects de la performance diagnostique favorisant la détection du cancer du sein à un stade précoce qui ont été examinés. Cependant, les thérapies instaurées depuis 40 ans, tout particulièrement la chimiothérapie adjuvante, ont amélioré considérablement les résultats des traitements du cancer du sein, et le taux de survie est plus élevé à tous les stades de la maladie. Par exemple, aux États-Unis, la survie à cinq ans dans le cas d'un cancer du sein localisé est passée de 72 % dans les années 1940, à 97 % aujourd'hui [ACS, 2004]. Ces progrès peuvent soit augmenter l'utilité thérapeutique des diagnostics précoces, soit, plus probablement, la réduire [Miller *et al.*, 2000a]. Il est donc possible que les avantages associés aux diagnostics radiologiques précoces aient été atténués par une meilleure prise en charge thérapeutique, tant aux stades précoces que plus avancés du cancer. Comme le présent rapport n'a pas examiné cette question, les évaluations à venir pourraient s'appuyer sur les modèles qui étudieront, d'une part, la contribution du dépistage à l'obtention d'un diagnostic plus précoce, et d'autre part, les taux relatifs de survie attribuables aux traitements actuels à différents stades du cancer.

2.4.7 Échelle de la force du contraste

Afin de comparer le degré auquel le plan, l'exécution et l'analyse de chaque étude ont permis de produire un contraste marqué, une échelle reproductible permettant d'évaluer les facteurs jugés pertinents à l'évaluation de la mise en contraste de l'étude a été élaborée aux fins de la présente analyse. Ces facteurs sont : le contraste technique, l'époque à laquelle la mammographie a été réalisée, la qualité technique de la mammographie, les taux de participation et la dilution temporelle découlant de la méthode d'analyse choisie. Cette échelle a ensuite été appliquée à chacun des huit essais sur la mammographie, de pair avec une échelle de validité décrite à la section 2.5.5 et à l'annexe D. L'annexe C (Échelle de la force du contraste) la présente en détail et explique plus à fond son application aux différentes études.

2.5 BIAIS À DIRECTION INCONNUE

L'essai contrôlé randomisé est un outil de recherche conçu pour permettre une comparaison valide entre deux interventions (ou

8. Cette ligne de pensée n'est pas nouvelle. À propos de l'étude HIP, Shapiro [1977] explique que « la série sur la mortalité dépassant cinq ans après l'entrée dans l'étude se limite alors aux décès par cancer du sein chez les femmes ayant eu un diagnostic au cours des cinq premières années. Cette approche réduit l'effet d'atténuation qu'engendrerait l'inclusion des taux de mortalité chez les femmes dont le cancer du sein a été détecté longtemps après la fin du cycle d'examens de dépistage, car le groupe expérimental, avec le temps, se retrouve dans la même situation que les femmes du groupe témoin […] Ce problème ne peut être résolu de façon complètement satisfaisante, mais on peut minimiser le facteur d'atténuation en excluant les cas diagnostiqués après la cinquième année » [trad.].

plus) en y soumettant des cohortes de personnes comparables pour toutes les caractéristiques susceptibles d'influer sur le résultat de l'étude, sauf l'intervention elle-même. Toute différence significative observée dans les résultats de ces cohortes sera ainsi attribuable à l'intervention, et non aux autres facteurs. Tous s'entendent pour dire que les caractéristiques du plan des essais contrôlés randomisés les rendent moins vulnérables aux biais. Lorsque l'objet d'étude est l'effet visé d'une intervention, le risque de biais est grand si les cliniciens affectent les patients à une intervention en se fondant sur les bénéfices possibles que ces derniers pourraient en tirer (l'indication) [Miettinen, 1983]. Les recensions systématiques de la littérature sur la mammographie de dépistage ont donc généralement exclu les études non expérimentales.

Les normes régissant le plan expérimental, le déroulement et l'analyse des essais contrôlés randomisés ont évolué depuis l'essai novateur réalisé en 1948 par Bradford Hill et ses collaborateurs [MRC, 1948], et ces normes constituent la base logique de l'évaluation de la validité des essais contrôlés randomisés sur la mammographie de dépistage. Dans la section précédente, les aspects de la mise en contraste du dépistage qui contribuent à la force du contraste ont été discutés. Il est clair que toute lacune dans la mise en contraste aura tendance à fausser la mesure de l'efficacité différentielle vers le zéro. Les caractéristiques des essais contrôlés randomisés − comme la randomisation − conçues pour assurer une comparaison valide seront examinées ici. L'absence de ces caractéristiques explique l'introduction de biais. Cependant, il est difficile de savoir si ces biais rapprocheront ou éloigneront les résultats de l'hypothèse nulle. Ces caractéristiques, expliquées ci-dessous, peuvent donc être considérées comme des facteurs de validité à direction inconnue.

2.5.1 Randomisation

La répartition aléatoire des patientes dans le groupe de dépistage ou le groupe témoin devrait, en principe, assurer que tous les autres facteurs de risque de mortalité par cancer du sein, tant connus qu'inconnus, sont répartis également dans les deux groupes. La randomisation est devenue la norme parce que l'imprévisibilité d'une répartition au hasard ne permet pas à un patient ou à un médecin de choisir l'un ou l'autre groupe [Chalmers, 2001]. D'autres pratiques, comme l'affectation des patients selon le jour de leurs visites à la clinique ou leur date de naissance, sont plus vulnérables à ce genre de sélection. Les essais qui pratiquent ce genre de répartition ne peuvent être appelés essais contrôlés randomisés, bien qu'on y réfère souvent par ce nom [Frisell *et al.*, 1986; Tabár *et al.*, 1985a]. La randomisation peut aussi être appliquée à des groupes de personnes [Roberts *et al.*, 1984], mais cette formule a des implications sur l'analyse des résultats.

2.5.2 Comparabilité de base

Alors que la répartition aléatoire des femmes dans les groupes de dépistage et témoin est faite pour former des cohortes dans lesquelles les facteurs de risque de mortalité par cancer du sein sont répartis également, la description de la répartition de ces facteurs de risque dans les groupes permet de vérifier si la randomisation est réussie, du moins pour les facteurs de risque connus. Le compte rendu de l'analyse descriptive de la comparabilité de base des groupes est l'un des critères de qualité proposés par Chalmers et ses collaborateurs [1981]. Les paramètres qui revêtent une importance particulière dans un essai sur la mammographie sont les facteurs de risque connus et présumés de décès par cancer du sein, notamment l'âge lors de la randomisation, le nombre de naissances vivantes, le statut ménopausique et les antécédents familiaux de cancer du sein.

2.5.3 Traitement des exclusions

Pour maintenir l'égalité entre les cohortes, toute exclusion subséquente à la randomisation doit être appliquée également aux groupes de dépistage et témoin. Un diagnostic de cancer du sein antérieur à la randomisation est un

motif d'exclusion fréquent et important dans les essais sur le dépistage du cancer du sein. Comme environ le tiers des femmes atteintes d'un cancer du sein en meurent, une application inégale de ce critère d'exclusion pourrait créer un énorme biais, habituellement en faveur du dépistage. Par exemple, si les femmes du groupe témoin ne sont pas interviewées selon le même protocole que les femmes affectées à la mammographie, des femmes ayant un cancer insoupçonné pourraient être incluses dans le groupe témoin. Une exclusion subséquente, fondée sur les dossiers médicaux, par exemple, repose trop sur la foi des données qui y sont inscrites, celles-ci pouvant ne pas contenir les informations obtenues par entrevue personnelle.

2.5.4 Uniformité de l'information relative aux résultats

Afin de maintenir une comparaison valide, il est important que les informations sur les résultats obtenus soient aussi exactes dans les deux groupes de femmes. Par exemple, l'exactitude peut être compromise si les diagnostics sont établis selon les protocoles du programme de dépistage chez les femmes du groupe expérimental et que des diagnostics similaires passent inaperçus dans le groupe témoin. Ce biais, qui pourrait atténuer la mesure de l'efficacité du dépistage, peut être partiellement corrigé par la validation des indicateurs de résultats grâce à l'évaluation à l'insu de la cause du décès et la validation des causes de décès par autopsie.

Une mauvaise classification des causes de décès compromet sérieusement la validité des essais sur le dépistage, et ce, pour plusieurs raisons. Premièrement, les femmes qui meurent d'un cancer du sein sont souvent à un âge où d'autres maladies, dont d'autres cancers, sont présentes et pourraient aussi être des causes de mortalité. Les autopsies peuvent aider à distinguer les différentes causes mais, depuis 50 ans, le taux d'autopsies diminue, se situant entre 30 et 40 % dans les essais publiés. De plus, les complications consécutives au traitement des cancers diagnostiqués ne sont pas nécessairement attribuées au cancer du sein, et cette sous-estimation du véritable taux de mortalité par cancer du sein qui, selon une estimation de Brown et ses collaborateurs [1993], se situerait aux environs de 9 %, pourrait jouer en faveur du dépistage mammographique.

2.5.5 Échelle de validité

Les revues systématiques de la littérature scientifique doivent établir les critères d'inclusion et d'exclusion des études, et ce, quelle que soit l'intervention étudiée. L'élaboration d'un indice permet d'évaluer de façon reproductible le degré auquel chaque étude respecte les critères de validité. Les facteurs qui font habituellement partie de l'évaluation de la validité des essais sur le dépistage ont déjà été traités. D'autres chercheurs ont mis au point des échelles de mesure de la validité, mais elles sont difficiles à appliquer à la mammographie de dépistage. Par exemple, l'échelle de Jadad et ses collaborateurs [1996] attribue deux points pour le procédé à l'insu sur un total de cinq points, mais aucun essai sur la mammographie de dépistage n'a été réalisé à l'insu, ni des patientes, ni du personnel soignant. L'échelle d'évaluation élaborée ici, comme d'autres, implique une estimation subjective des caractéristiques qualitatives d'un essai rigoureux. Une analyse plus détaillée de cette échelle et de son application à chaque essai est présentée à l'annexe D (Échelle de validité).

2.6 PRÉCISION : MINIMISER LES EFFETS DU HASARD

La technique de la méta-analyse a été tout d'abord élaborée à partir des travaux de Gauss et Laplace afin de réduire les « erreurs d'observation » par la combinaison des observations, d'abord en astronomie, et subséquemment dans des domaines comme la médecine [Chalmers, 2001]. Ainsi, on peut combiner des essais dont la taille est trop petite pour donner des résultats statistiquement significatifs afin d'obtenir une plus grande précision. Même si les essais sur la mammographie de dépistage ont tous fait

appel à des milliers de femmes dans chaque cohorte, seul un petit pourcentage d'entre elles sont atteintes d'un cancer du sein, et un nombre encore plus petit en meurt, soit moins de 100 par cohorte. Le problème est encore plus sérieux dans les sous-groupes, par exemple pour estimer l'efficacité de la mammographie chez les femmes âgées de moins de 50 ans.

Cependant, des estimations précises avec de petits intervalles de confiance sont moins importantes que la pertinence des mises en contraste et la validité de l'estimation des effets de l'intervention étudiée. Autrement dit, un résultat imprécis mais valide est de beaucoup préférable à un résultat précis mais non valide. Nul besoin, donc, d'élaborer une échelle d'évaluation de la précision des études, l'erreur type servant de mesure d'imprécision naturelle et étant disponible pour chaque étude. Les techniques de synthèse courantes qui utilisent l'inverse de l'erreur type comme pondérateur permettent de combiner les résultats de différentes études pour obtenir une estimation de la précision. Contrairement aux études aux contrastes faibles ou aux résultats non valides, il n'y a aucun avantage à exclure les résultats imprécis, car la précision se trouve augmentée par la seule force du nombre lorsque les résultats de ces études sont combinés.

2.7 INDICATEURS DE RÉSULTATS PERTINENTS À LA QUESTION DE RECHERCHE

Le but du dépistage du cancer du sein est de réduire la morbidité et la mortalité par cancer du sein sans augmenter l'incidence de la morbidité et de la mortalité dues à d'autres causes. Cet objectif implique bien sûr une réduction de la morbidité et de la mortalité globales. Les huit essais sur le dépistage mammographique dont les résultats ont été publiés ont tous comparé les taux de mortalité par cancer du sein entre les femmes invitées au dépistage et celles qui ne l'étaient pas. Plusieurs études ont aussi fourni des données sur d'autres causes de décès et la mortalité globale.

Bien que la mammographie de dépistage vise à réduire la mortalité par cancer du sein, il est important de vérifier l'étendue des effets indésirables du dépistage sur la santé. Le rapport du CETS publié en 1993 en mentionne deux : les faux positifs, qui entraînent des biopsies chirurgicales et par aspiration, et les risques liés à la dose de radiation absorbée par les femmes qui subissent des mammographies. Cette liste est loin d'être exhaustive : le dépistage peut causer d'autres préjudices, dont l'anxiété que provoque un résultat positif [Lampic *et al.*, 2001], les interventions diagnostiques qui découlent du dépistage (complications chirurgicales ou anesthésiques), ou les interventions thérapeutiques résultant d'un diagnostic confirmé (effets secondaires de la chimiothérapie ou de la radiothérapie pour les lésions décelées par dépistage).

Dans leur méta-analyse, Gøtzche et Olsen [2000] ont ravivé le vieux débat sur la pertinence d'utiliser la mortalité par cancer du sein comme indicateur de performance de la mammographie de dépistage. Craignant que la mammographie puisse causer autant de décès de diverses causes qu'elle sauve de vies menacées par le cancer du sein, ils estiment que la mortalité globale est un indicateur plus approprié. L'utilisation du taux de mortalité attribuable au cancer du sein est source d'ambiguïté, car les décès qui résultent du processus de dépistage proprement dit ou des traitements subséquents pour cibler la maladie ne sont généralement pas comptabilisés avec les décès causés par le cancer du sein. Mais comme le soulignent Black et ses collaborateurs [2002], « les règles qui ont servi à déterminer quels décès sont comptabilisés dans le taux de mortalité attribuable à la maladie sont rarement publiées avec les résultats des essais » [trad.].

De toute évidence, on ne devrait pas essayer de réduire la mortalité par cancer du sein sans se préoccuper des autres causes de mortalité. D'un point de vue pratique, lorsqu'il faut décider si l'on veut se prévaloir d'un programme de dépistage mammographique, il faut considérer tant les effets intentionnels du dépistage (principalement la réduction de la

mortalité par cancer du sein) que ses effets involontaires (principalement les effets secondaires d'interventions diagnostiques et thérapeutiques agressives, mais aussi d'autres effets non intentionnels comme les risques liés à la radiation [Jung, 2001]). Cependant, on ne ferait que brouiller les cartes si l'on mêlait ces deux catégories en utilisant le taux de mortalité globale comme principal indicateur de résultat. Afin d'avoir la puissance requise pour obtenir cet indicateur, les études devraient avoir une taille prohibitive. Il serait donc préférable de considérer deux questions distinctes : 1) est-il possible de réduire la mortalité par cancer du sein à l'aide d'un dépistage régulier de longue durée ? et 2) y a-t-il d'autres causes de décès ayant des liens plausibles avec le dépistage (complications chirurgicales, morts cardiaques postradiothérapie, suicides, etc.) dont les taux sont plus élevés chez les femmes soumises au dépistage ou chez les femmes qui ne l'ont pas été ?

Dans ce cas, il pourrait s'avérer approprié d'étudier les effets du dépistage à partir du début de l'étude.

L'objectif des études publiées étant de répondre à la première question, elles ne s'intéressent qu'à la mortalité par cancer du sein. Il ne fait aucun doute que l'effet spécifique du dépistage mammographique sur la mortalité par cancer du sein constitue une information utile pour les femmes qui envisagent de subir une mammographie de dépistage. Cependant, les réponses quantitatives à la deuxième question sur les effets secondaires ne sont pas encore disponibles. Les sections suivantes examineront de plus près les études sur l'efficacité théorique[9] de la mammographie de dépistage en prenant comme indicateur de résultat la mortalité par cancer du sein. Il ne faut toutefois pas oublier que le manque de preuves sur les préjudices possibles de la mammographie de dépistage n'est pas synonyme de preuve de l'inexistence de tels préjudices.

9. Dans le présent rapport, le terme « efficacité théorique » fait généralement référence à la réduction relative des taux de mortalité par cancer du sein dans le groupe soumis au dépistage par rapport aux taux dans le groupe témoin. Le terme « efficacité pratique » fait parfois référence à l'efficacité telle qu'elle peut être observée dans le « monde réel » [Last, 2001], alors que l'efficacité théorique est mesurée dans le contexte d'un essai. Comme la notion de force du contraste d'un essai met en évidence ces différences entre la théorie et la pratique, nous ne distinguerons pas davantage ces deux types d'efficacité.

3 ANALYSE DES ESSAIS PUBLIÉS ET COMPARAISON AVEC LE CONTEXTE QUÉBÉCOIS

Dans cette section, les études publiées sur l'efficacité des programmes de dépistage mammographique seront évaluées, de même que leur force de contraste et leur validité. Pour ce faire, les échelles d'évaluation de la force du contraste et de la validité seront appliquées à chacun des huit essais publiés sur la mammographie de dépistage, et l'essai du Royaume-Uni (*UK Age Trial*) actuellement en cours sera partiellement évalué. Rappelons que les détails complets sur les échelles d'évaluation et leur application à ces essais se trouvent à l'annexe C (Échelle de la force du contraste), à l'annexe D (Échelle de validité) et à l'annexe E (Analyse des neuf essais). La question de la précision des études sera traitée à la section 5 (Discussion et conclusions), où sera expliquée la pertinence de la combinaison statistique des mesures des résultats des études jugées valides qui ont obtenu des contrastes marqués. Les caractéristiques des plans expérimentaux des divers essais sur la mammographie seront analysées et comparées aux activités de dépistage offertes au Québec, principalement dans le cadre du Programme québécois de dépistage du cancer du sein (PQDCS). La présente section s'amorce par un examen du Programme québécois (section 3.1) et se poursuit avec une description de la structure et des processus des neuf essais sur la mammographie (section 3.2) montrant les ressemblances et les différences entre ces essais et le Programme québécois (section 3.3).

3.1 LE DÉPISTAGE AU QUÉBEC

Depuis les années 1970, la mammographie fait partie de la pratique médicale courante au Québec. Elle est offerte gratuitement aux patientes dans le cadre du régime d'assurance maladie universel en vigueur. Le recours à la mammographie a lentement augmenté : en 1987, seulement 44 % des femmes de 50 à 59 ans avaient déjà subi une mammographie, dont 24 % dans les 24 mois précédents. En 1997-98, 56 % des femmes de 50 à 69 ans avaient subi au moins une mammographie, bien qu'il soit difficile de savoir dans quelles proportions ces examens visaient le dépistage ou l'exploration diagnostique de signes ou de symptômes de maladies mammaires. Le PQDCS, instauré en 1998, invite activement toutes les femmes du Québec âgées de 50 à 69 ans qui n'ont jamais eu de diagnostic de cancer du sein à subir une mammographie tous les deux ans. Le dépistage chez les femmes de moins de 50 ans ou de plus de 70 ans, bien qu'il ne fasse pas officiellement partie du PQDCS, est offert par les mêmes centres, sur ordonnance d'un médecin. La structure, les processus et les objectifs du Programme sont décrits dans un document de référence du ministère de la Santé et des Services sociaux [MSSS, 1996] et s'inspirent à plusieurs égards des programmes de dépistage du cancer du sein en place en Suède, au Royaume-Uni et en Australie.

Le dépistage est offert dans environ 80 centres. Dans les régions urbaines, ce sont surtout des cliniques privées, et dans les régions moins peuplées, des centres hospitaliers. Ces centres doivent se conformer aux normes de qualité provinciales, qui comprennent l'inspection périodique et l'évaluation du matériel, la formation professionnelle, l'utilisation des technologies de l'information du PQDCS et la participation à l'évaluation de la performance du programme. Les centres de dépistage offrent un examen de dépistage initial (mammographie sous deux incidences) et, habituellement, les explorations radiologiques additionnelles requises. De plus, chaque région désigne un nombre restreint de centres de diagnostic auxquels sont adressées les femmes ayant eu un résultat positif à l'examen de dépistage. Ces centres complètent l'investigation diagnostique et, si un cancer est découvert, ils peuvent amorcer le traitement. La coordination des services s'effectue à l'échelle régionale et comporte

l'envoi de l'invitation par la poste, la transmission des résultats des examens de dépistage et le suivi des femmes ayant eu un résultat positif. Les centres régionaux coordonnent aussi l'utilisation du système d'information, l'évaluation et l'assurance de la qualité. À l'échelle provinciale, le soutien technique, la maintenance et le développement des technologies de l'information ainsi que la liaison avec les corps professionnels et les politiques ministérielles sont pris en charge par des équipes du ministère de la Santé et des Services sociaux et de l'Institut national de santé publique du Québec [MSSS, 2004].

3.2 ESSAIS PUBLIÉS SUR LA MAMMOGRAPHIE DE DÉPISTAGE

Au cours des 40 dernières années, neuf essais sur la mammographie de dépistage ont été publiés. Comme il a été mentionné à la section 2, les éléments en jeu dans la présente évaluation sont la pertinence du contraste étudié dans chaque essai, la force de ce contraste et l'absence de biais à direction inconnue. Dans cette section, les caractéristiques importantes de ces éléments seront définies et comparées à leur équivalent québécois pour déterminer si les résultats des essais s'appliquent au contexte du Québec. Une description plus détaillée de ces essais est présentée à l'annexe E.

3.2.1 New York (HIP)

Amorcé en 1963, le premier essai structuré sur la mammographie a été mené auprès des clientes d'un régime d'assurance maladie privé de New York, le *Health Insurance Plan* (HIP). L'intervention de dépistage était constituée de trois à quatre mammographies à un an d'intervalle avec examen clinique des seins, alors que les femmes du groupe témoin n'ont rien changé à leur façon d'obtenir leurs soins médicaux. Sur l'échelle d'évaluation de la validité, cet essai obtient une note de 1,5 sur 4. Il partage le principal défaut de la plupart des autres études, soit l'absence quasi totale d'informations sur les caractéristiques de la cohorte témoin. De plus, l'information sur les suivis était très fragmentaire à cause du rythme rapide des départs et du manque de données sur l'état de santé des personnes qui ont quitté le régime, rendant difficile l'exclusion des cas de cancer du sein diagnostiqués avant la mise en contraste du dépistage.

Cette étude obtient 12 % sur l'échelle de la force du contraste, la plus basse note de tous les essais publiés. Comme l'essai HIP a débuté avant l'avènement d'appareils conçus expressément pour la mammographie, les différences avec le contexte du Québec d'aujourd'hui sont considérables. Toutefois, les mesures d'assurance de la qualité relatives à la lecture des clichés étaient beaucoup plus sévères que dans le Programme québécois. La lecture était en effet confiée à quelques radiologistes et tous les clichés, que le résultat soit positif ou négatif, faisaient l'objet d'une deuxième lecture [Strax *et al.*, 1973].

3.2.2 Malmö (MMST)

Le *Malmö Mammographic Screening Trial* (MMST) a été mis sur pied à Malmö au début de 1976. Il s'agit de la première de quatre études suédoises comparant la mammographie seule aux soins usuels. Dans cet essai, les patientes du groupe témoin n'ont pas eu de dépistage, mais celles qui avaient des symptômes mammaires « ont reçu les services médicaux usuels ». Sur l'échelle d'évaluation de la validité, cet essai obtient 3 sur 4, la plus haute note après les études canadiennes. Il partage, avec toutes les autres études sauf les études canadiennes, un manque quasi total de renseignements sur la population témoin, ce qui rend difficile de vérifier si la randomisation a réussi à assurer l'égalité de départ entre les groupes et à exclure efficacement les cas de cancer du sein diagnostiqués avant le dépistage.

Sur l'échelle d'évaluation de la force du contraste, cette étude obtient 27 %. L'essai de Malmö diffère considérablement du Programme québécois en ce qui a trait à l'expertise des lecteurs. Comme pour toutes les études suédoises, le dépistage dans l'essai de Malmö était dirigé par un service hautement

centralisé, les clichés étant réalisés dans des unités fixes et mobiles et interprétés par des spécialistes qui lisaient entre 5 000 et 20 000 clichés par année. Par comparaison, les lecteurs du Programme québécois ont un volume médian d'environ 600 lectures de clichés par année, et seulement 20 % de ces clichés sont interprétés par des radiologistes qui lisent plus de 2 000 clichés par année [INSPQ, 2003]. De plus, les lecteurs suédois étaient responsables des interventions diagnostiques lorsque les clichés de dépistage révélaient des anomalies, tandis que le Programme québécois recommande que les femmes qui ont un résultat positif soient adressées à des centres de diagnostic désignés, généralement distincts des centres de dépistage.

3.2.3 *Two-County* (TCS)

L'essai *Two-County*, le deuxième des quatre essais suédois, a commencé dans le comté de Kopparberg en 1977, et en 1978 dans celui d'Östergötland. Tous deux sont situés dans le centre sud-est de la Suède et représentent ensemble environ 8 % de la population totale de ce pays. Conçu comme une étude unique, le protocole de dépistage était le même à Kopparberg et à Östergötland. Les données contradictoires et imprécises des nombreux articles publiés dans les revues scientifiques, un processus de randomisation par blocs compliqué et parfois arbitraire, le manque de renseignements sur les critères d'exclusion et un âge inégal d'entrée dans l'étude entre les cohortes de dépistage et témoin compromettent la validité de l'essai. Sur l'échelle d'évaluation de la validité, cette étude obtient 0,5 sur 4.

L'essai mettait en contraste le dépistage mammographique à une seule incidence, sans examen clinique des seins, avec les soins usuels. Comme les autres études suédoises, l'essai *Two-County* avait une unité centrale et plusieurs unités mobiles, et la lecture de tous les clichés était centralisée. Tous les clichés étaient lus par deux radiologistes, Fagerberg et Tabár, qui étaient aussi les chercheurs principaux de l'étude. Le taux de participation moyen était d'environ 87 %, et le taux de contamination était relativement bas. L'intervalle entre les examens de dépistage était plus long que celui du Programme québécois, soit tous les deux ans pour les femmes âgées de 40 à 49 ans au début du dépistage, et tous les trois ans pour les femmes âgées de 50 à 69 ans. La mise en contraste était de courte durée. Même si les comptes rendus ne concordent pas tout à fait, il semble qu'une période d'intervention de cinq ans était prévue, bien qu'on ait tardé à offrir un dépistage à la cohorte témoin, jusqu'à huit ans après la randomisation dans certains cas. La pleine réduction de la mortalité en régime établi ne peut être observée avec une période aussi courte de mise en contraste. Toutefois, l'effet de dilution causé par l'inclusion des premières années du suivi dans les données sur les taux de mortalité cumulatifs est moins important que dans d'autres essais ayant une mise en contraste courte. En effet, l'essai TCS était analysé à partir du modèle d'évaluation suédois, où seules les femmes qui avaient eu un diagnostic de cancer pendant la mise en contraste de l'étude étaient suivies pour la mesure des taux de mortalité, alors que les autres essais ont utilisé le modèle d'évaluation traditionnel. De plus, le suivi dans l'essai *Two-County* était exceptionnellement long (20 ans). Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 42 %.

3.2.4 Édimbourg

L'*Edinburgh Randomised Trial of Screening for Breast Cancer* a été amorcé en 1979. Sa validité est compromise par une randomisation par blocs réalisée à partir de la clientèle d'omnipraticiens d'Édimbourg ajustée de diverses façons par les chercheurs, l'exclusion largement inégale des cas de cancer du sein préalablement diagnostiqués, et des différences de base considérables dans le statut socioéconomique et les taux de mortalité globaux. Cet essai obtient 0,5 sur 4 pour la validité.

L'intervention de dépistage dans le groupe expérimental comprenait une série de quatre cycles de mammographies et un examen clinique des seins tous les deux ans, avec, pendant les années intermédiaires, un examen

clinique des seins seulement. La mammographie était prise sous deux incidences au premier cycle, et sous une seule incidence par la suite. Pour les femmes du groupe témoin, aucun protocole spécifique n'est mentionné. Contrairement aux autres études et au Programme québécois, la plupart des clichés n'ont pas été lus par des radiologistes, mais plutôt par des « médecins spécialement formés ». Les clichés présentant des anomalies et 5 % de tous les clichés pris au hasard étaient cependant lus par des radiologistes. Les taux de participation étaient bas, avec un taux moyen de 52 % et, bien qu'ils n'aient pas été mesurés, les taux de contamination étaient probablement faibles aussi. Sur l'échelle d'évaluation de la force du contraste, cette étude obtient une note de 26 %.

3.2.5 Canada-1 (NBSS-1)

L'étude canadienne *National Breast Screening Study* (NBSS) a commencé en 1980. Deux différents protocoles étaient définis, l'un pour les femmes âgées de 40 à 49 ans à l'entrée (NBSS-1), et l'autre pour les femmes de 50 à 59 ans (NBSS-2). Les deux essais ont fait l'objet de nombreuses critiques, certaines sur la validité, d'autres sur la mise en contraste.

En ce qui a trait à la validité, une révision indépendante du processus de randomisation concluait qu'elle assurait l'égalité de départ [Bailar et MacMahon, 1997]. Les cas de cancer dépistés au premier examen n'ont pas été exclus du suivi ni de l'analyse, et les exclusions postrandomisation s'équivalaient entre les groupes. Cette étude a réussi à obtenir des groupes comparables pour 10 caractéristiques importantes [Miller *et al.*, 2000b; Baines, 1994]. Un plus grand nombre de petites tumeurs avec atteinte ganglionnaire ont été détectées chez les femmes du groupe de dépistage, mais la mammographie est censée déceler de telles tumeurs [Bailar et MacMahon, 1997]. Peu d'études ont été scrutées d'aussi près, et moins encore ont pu profiter d'une révision indépendante aussi poussée, qui a récusé toutes les critiques importantes sur sa validité. Sur l'échelle d'évaluation de la validité, cette étude obtient 4 sur 4.

La mise en contraste du dépistage dans l'essai NBSS-1 pourrait être résumée ainsi : l'intervention témoin était constituée d'un examen physique général avec enseignement de l'auto-examen des seins, la mammographie étant seulement offerte au besoin dans un contexte diagnostique; l'intervention de dépistage, quant à elle, débutait par une mammographie au cycle 0, puis une mammographie, un examen physique et l'incitation à pratiquer l'auto-examen des seins étaient ajoutés aux cycles subséquents. Le contraste technique était donc plus fort que celui du Programme québécois et de la plupart des programmes canadiens et européens, où la mammographie est offerte sans être systématiquement accompagnée d'un examen physique. Certains auteurs ont critiqué cette étude parce que la participation était essentiellement fondée sur le volontariat plutôt qu'offerte à toute la population. La seule conséquence importante de cette stratégie serait le taux élevé de participation (92 %). Cette forte participation tend à produire un contraste plus marqué, mais en revanche, celui-ci se trouve affaibli par d'autres caractéristiques, notamment sa courte durée de 4,6 années. Sur l'échelle d'évaluation de la force du contraste, cette étude reçoit la meilleure note des essais sur la mammographie, soit 45 %.

3.2.6 Canada-2 (NBSS-2)

La deuxième partie de l'étude canadienne sur le dépistage du cancer du sein, qui portait sur des femmes âgées de 50 à 59 ans à l'entrée, a aussi commencé en janvier 1980. Les questions de validité sont les mêmes que celles qui ont été soulevées sur l'étude NBSS-1, et pour les mêmes raisons, cette étude obtient une note de 4 sur 4.

Le NBSS-2 est différent de toutes les autres études et du Programme québécois, car il mettait en contraste deux interventions de dépistage. La première comprenait un examen physique systématique effectué par des infirmières et des médecins, ainsi que

l'enseignement de l'auto-examen des seins et l'incitation à le pratiquer. La deuxième comprenait les mêmes interventions, avec en plus une mammographie annuelle sous deux incidences. Ce contraste est considérablement plus faible parce que les femmes du groupe témoin sont soumises à un protocole de dépistage (examen physique et auto-examen des seins) qui pourrait avoir une certaine efficacité en soi. Comme dans l'étude NBSS-1, le taux de participation était élevé dans le groupe de dépistage, soit 91 %, mais le taux de contamination dans le groupe témoin atteignait 6 %. De plus, la mise en contraste était de courte durée, avec une moyenne de 4,6 années. Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 31 %. Puisque cet essai ne permet pas de répondre à la question décisionnelle sur l'utilité du dépistage mammographique, il est exclu de la présente analyse.

3.2.7 Stockholm

Le troisième des quatre essais suédois, le *Stockholm Mammographic Screening Trial*, a commencé en 1981. Les patientes n'étaient pas réparties au hasard, mais plutôt selon leur date de naissance, ce qui rendait l'étude vulnérable au biais de sélection en fonction du risque de cancer du sein. La validité de cet essai est aussi compromise par l'arrondissement des données quantitatives, l'absence de données sur la comparabilité de base des groupes de dépistage et témoin et l'exclusion tardive des cas de cancer du sein diagnostiqués avant l'étude. Sur l'échelle d'évaluation de la validité, cet essai obtient 2 sur 4. L'intervention de dépistage consistait en seulement deux cycles de mammographies sous une seule incidence (oblique), sans examen physique, à 28 mois d'intervalle. Ni la mammographie ni l'examen physique n'ont été offerts aux femmes du groupe témoin, bien qu'elles aient été invitées au dépistage mammographique à la fin du deuxième cycle.

En raison de la période extrêmement courte de mise en contraste (deux cycles seulement avec un intervalle unique de 28 mois) et du fait que la mammographie était prise sous une seule incidence aux deux cycles, le contraste est plus faible que celui du Programme québécois. De plus, même avec un taux de participation relativement élevé au premier cycle, la contamination était aussi relativement élevée, car 25 % de toutes les femmes de Stockholm avaient subi une mammographie au cours des trois années précédant l'essai [Frisell *et al.*, 1991]. Sur l'échelle d'évaluation de la force du contraste, cette étude obtient 28 %.

3.2.8 Göteborg

Le *Gothenburg Breast Screening Trial*, amorcé en 1982, est le plus récent des huit essais sur lesquels des résultats ont été publiés. Sa validité est compromise par les mêmes facteurs que les autres essais suédois, dont des données fragmentaires sur la cohorte témoin et l'exclusion tardive des cas de cancer du sein diagnostiqués avant le dépistage. Cet essai obtient une note de 1,5 sur 4 sur l'échelle d'évaluation de la validité.

L'intervention de dépistage consistait en un maximum de cinq cycles de mammographies de dépistage réalisés à des intervalles de 18 mois dans le groupe expérimental, alors que les femmes du groupe témoin (qui recevaient les soins usuels) n'étaient soumises à aucune mammographie, sauf pour un dernier cycle offert en même temps aux deux groupes. Il est à noter que la mise en contraste comportait un cycle de moins pour les femmes de 50 ans et plus, car le dépistage était effectué systématiquement dans ce groupe d'âge. Comme pour le Programme québécois, on a eu recours à deux incidences (dans la plupart des cas), et l'examen clinique des seins n'était pas offert. Les taux de participation et de contamination étaient élevés, soit de 79 % et de 19 % respectivement. Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 37 %.

3.2.9 Royaume-Uni (*UK Age Trial*)

En 1991, le *United Kingdom Coordinating Committee on Cancer Research* a mis sur pied un essai contrôlé randomisé multicentrique à l'échelle nationale appelé « Age Trial »

[Moss *et al.*, 2005a]. Une brève description de cet essai sera présentée ici afin d'établir une base de comparaison avec les autres essais et de lui accorder une note préliminaire. Bien que les résultats de l'essai ne soient pas encore publiés, il sera vraisemblablement l'un des plus forts au chapitre de la validité, de l'usage de techniques modernes de randomisation, de la mesure des caractéristiques de départ et du suivi des résultats. Sur l'échelle d'évaluation de la validité, cet essai obtient un score anticipé de 4 sur 4 sur la foi du plan expérimental publié. Cette évaluation se fonde toutefois sur des conjectures qui devront être confirmées à la lumière des nouvelles données qui seront publiées.

Cet essai aura la plus longue période de mise en contraste, soit jusqu'à neuf cycles à 12 mois d'intervalle. En effet, comme les chercheurs veulent évaluer si le dépistage est valable à partir de l'âge de 40 ou 41 ans pour une période de 10 années complètes, il sera le seul à avoir une durée de mise en contraste comparable à celle du programme qui pourrait un jour être mis en place. En d'autres mots, l'étude évalue le parcours complet d'un programme potentiel au lieu d'évaluer quelques cycles de dépistage et d'extrapoler leurs résultats. L'intervention témoin n'a pas été définie, mais elle ne comporte probablement aucune invitation au dépistage et s'en tient aux « soins usuels ». La contamination par dépistage dans la cohorte témoin sera problématique, car il sera difficile de l'éviter dans le contexte actuel, où le dépistage du cancer du sein est largement recommandé. Enfin, la structure du programme du Royaume-Uni diffère considérablement de celle du Programme québécois, car il est hautement centralisé et hiérarchisé, et accorde une importance particulière non seulement à la qualité des clichés, mais aussi à l'expertise des lecteurs, aux deuxièmes lectures des clichés, etc. À partir des conjectures qu'on peut faire sur la façon dont cette étude sera menée et analysée, le score de la force du contraste devrait atteindre 43 %. La continuation du suivi donnera probablement à cette étude la force de contraste la plus élevée de tous les essais réalisés sur la mammographie de dépistage.

Le tableau 1 résume les caractéristiques importantes des neuf essais sur la mammographie de dépistage et les principaux éléments des protocoles.

Le tableau 2 récapitule les notes accordées aux essais sur la mammographie pour les caractéristiques qui contribuent à réduire les biais à direction inconnue. Bien que toutes ces études soient désignées comme des essais randomisés, la randomisation n'était pas toujours faite dans les règles de l'art, ce qui explique que la comparabilité de base entre les groupes de dépistage et témoin ne peut être établie. Dans plusieurs essais, les cas de cancer du sein préalablement diagnostiqués n'ont pu être exclus adéquatement, et il a été impossible de faire un suivi complet des résultats sur la mortalité dans chaque groupe.

Enfin, on trouvera au tableau 3 un récapitulatif des points accordés aux caractéristiques relatives à la force du contraste des essais. Les éléments de structure et de processus tels le contraste technique, l'époque de la technologie et les normes de qualité sont relativement forts, tandis que le facteur temps (choix du moment pour mesurer les effets du dépistage sur la mortalité) a affaibli de façon marquée les contrastes dans la plupart des essais.

TABLEAU 1

Description des neuf essais

ESSAIS ET ANNÉE D'INSTAURATION	ÂGE À L'ENTRÉE (ANS)	GROUPE DE DÉPISTAGE				GROUPE TÉMOIN		DURÉE (ANNÉES)
		MODALITÉ	PARTICIPATION (CYCLES INITIAL ET SUBSÉQUENTS) (%)	INTERVALLE (MOIS)	NOMBRE MOYEN DE CYCLES	MODALITÉ	TAUX DE CONTAMINATION ESTIMÉ (%)	
HIP (New York), 1963	40-64	2 incidences + ECS	65; 52; 48; 45	12	4	Pratique habituelle	5	3,5
MMST (Malmö), 1976	45-69	2 incidences, puis 1 (CC + O / puis O seul.)	74; 70 subs.	18-24	6	Services médicaux habituels	10	8,8
TCS (*Two-County*), 1977	40-74	1 incidence (MLO)	91; 86; 84	24-33	2,5	Soins usuels	10	6,5
Édimbourg, 1979	45-64	2 incidences, puis 1 (CC + O / puis O seul.)	61; nd; nd; 44	24	4	Soins usuels	10	6,4
NBSS-1 (Canada), 1980	40-49	2 incidences, ECS, AES	100; 89,4; nd; nd; 85,6	12	4,5	ECS initial, puis soins usuels	12,5	4,6
NBSS-2 (Canada), 1980	50-59	2 incidences, ECS, AES	100; 90,4; nd; nd; 86,7	12	4,5	ECS, AES	6,5	4,6
Stockholm, 1981	40-64	1 incidence	81; 80	28	2	Soins usuels	20	4,6
Göteborg, 1982	39-59	2 incidences, puis 1	85; 78; 79; 77; 75	18	5	Soins usuels	15	6,4
UK Age Trial, 1991	40-41	2 incidences, puis 1	61	12	9	Aucune mammographie	4	11,0

AES : auto-examen des seins; CC : incidence craniocaudale; ECS : examen clinique des seins; MLO : incidence médiolatérale oblique; O : incidence oblique.
nd : non disponible.

TABLEAU 2

Scores de validité

ESSAIS ET ANNÉE D'INSTAURATION	RANDOMISATION	COMPARABILITÉ DE BASE	EXCLUSIONS	SUIVI	SCORE TOTAL (SUR 4)
HIP (New York), 1963	1	0,5	0	0	1,5
MMST (Malmö), 1976	1	0,5	0,5	1	3
TCS (*Two-County*), 1977	0	0	0	0,5	0,5
Édimbourg, 1979	0	0	0	0,5	0,5
NBSS-1 (Canada), 1980	1	1	1	1	4
NBSS-2 (Canada), 1980	1	1	1	1	4
Stockholm, 1981	0,5	0,5	0,5	0,5	2
Göteborg, 1982	0,5	0	0,5	0,5	1,5
UK Age Trial, 1991	1	1	1	1	4

TABLEAU 3

Scores de la force du contraste

ESSAIS	CONTRASTE TECHNIQUE (C)	ÉPOQUE (E)	QUALITÉ (Q)	SOUS-INDICE, STRUCTURE ET PROCESSUS (C x E x Q)	PARTICIPATION (P)	FACTEUR TEMPS (T)	TOTAL (C x E x Q x P x T)
HIP (New York)	1,25	0,70	1,00	0,88	0,49	0,28	0,12
MMST (Malmö)	1,00	0,80	0,85	0,68	0,61	0,64	0,27
TCS (*Two-County*)	1,00	0,80	0,85	0,68	0,82	0,76	0,42
Édimbourg	1,25	0,80	0,90	0,90	0,43	0,67	0,26
NBSS-1 (Canada)	1,25	0,90	0,95	1,07	0,75	0,56	0,45
NBSS-2 (Canada)	0,80	0,90	0,95	0,68	0,82	0,56	0,31
Stockholm	1,00	0,90	0,80	0,72	0,60	0,65	0,28
Göteborg	1,00	0,90	0,88	0,79	0,64	0,73	0,37
*UK Age Trial**	1,00	1,00	0,98	0,98	0,57	0,73	0,41

* Résultats anticipés après 13 ans de suivi.

3.3 COMPARAISON DE LA STRUCTURE ET DES PROCESSUS DES ESSAIS AVEC LE PQDCS

Pour apprécier la valeur d'un programme de dépistage mammographique, il faudrait examiner les statistiques de mortalité par cancer du sein dans une cohorte de femmes québécoises âgées de 50 ans à l'entrée dans le programme, à qui le dépistage serait offert tous les deux ans pendant 20 ans au total. Ces chiffres seraient comparés aux statistiques enregistrées dans une autre cohorte similaire n'ayant eu aucun dépistage. Cependant, pour plusieurs raisons, il est impossible d'étudier ce contraste. Premièrement, un programme de dépistage de masse n'est pas un essai scientifique et ne dispose pas d'une cohorte témoin de comparaison. La formation d'une telle cohorte ne serait ni réalisable ni admissible du point de vue éthique à cause du consensus clinique actuel sur l'efficacité du dépistage mammographique. Deuxièmement, comme il a été mentionné, un tel essai devrait s'étendre sur 30 ou 40 ans et ses résultats, après un délai aussi long, ne seraient probablement plus pertinents pour l'orientation des politiques. Troisièmement, il serait impossible de trouver une cohorte suffisamment similaire à la population soumise au dépistage, ni avant le programme, ni ailleurs dans le monde. Il y a trop de différences dans la répartition des facteurs de risque de cancer du sein (âge à l'apparition des premières règles, nombre de grossesses, allaitement, prise d'hormones, etc.) et dans les taux de mortalité (accès aux services diagnostiques, protocoles thérapeutiques, etc.) pour que l'on puisse faire des comparaisons utiles avec d'autres populations en dehors d'un contexte expérimental. Bien que le PQDCS entende fonder en partie l'évaluation de sa performance sur sa capacité à réduire la mortalité par cancer du sein de 25 % en 10 ans [MSSS, 1996], une telle réduction ne pourrait être attribuée seulement au dépistage. Ainsi, une amélioration de l'efficacité d'un traitement pourrait réduire les taux de mortalité par cancer du sein sans égard au dépistage [Miller *et al.*, 2002].

Il faudra plutôt comparer le contraste obtenu dans le contexte des essais publiés avec le contraste obtenu par le Programme québécois. Dans la mesure où la mise en contraste de ce dernier est semblable à celles des essais, il peut espérer obtenir des résultats similaires. Cette notion s'apparente à celle de la « validité externe » de ces essais, sauf que l'on n'évalue pas la validité externe des essais, mais plutôt la pertinence des résultats de chaque étude par rapport au contexte québécois.

3.3.1 Structure et processus

Nous considérons les mêmes éléments constitutifs de la force de contraste qui ont été appliqués aux différents essais examinés dans la présente analyse.

Contraste technique : comme la plupart des études analysées, le Programme québécois propose une mammographie de dépistage sans examen clinique des seins ni autoexamen des seins. Il obtient donc un score de 1 pour cet aspect.

Époque : Les appareils de mammographie sont généralement plus perfectionnés que les appareils utilisés dans les essais, qui peuvent dater de 40 ans. Le Programme obtient une note de 1 pour cet aspect.

Qualité de la mammographie : le PQDCS obtient de bons résultats pour certains aspects, et réussit moins bien pour d'autres. Les mesures d'assurance de la qualité sont rigoureusement appliquées, puisque tous les centres de dépistage doivent être agréés par l'Association canadienne des radiologistes et être en mesure de produire des clichés de qualité supérieure. Cependant, peu de normes ont été définies pour la lecture des clichés. Alors que la plupart des programmes et des études ont veillé à assurer que les lecteurs aient l'expertise voulue en établissant des normes sur le volume de lecture de mammographies – généralement un minimum de 2 000 à 5 000 clichés par lecteur par an – les radiologistes du Programme québécois n'ont qu'à maintenir un volume de lecture annuel de 480 clichés. En règle générale, leur volume est de beaucoup inférieur à celui des

lecteurs de la plupart des études. En 2002, 45 % des clichés ont été lus par des radiologistes qui avaient lu moins de 1 000 clichés de dépistage dans l'année, 36 % par des lecteurs ayant lu entre 1 000 et 1 999 clichés, et 19 % par des radiologistes ayant lu 2 000 clichés et plus [INSPQ, 2003]. En outre, alors que d'autres programmes et études ont souvent exigé une deuxième lecture de tous les clichés révélant une anomalie radiologique et d'une proportion des clichés ayant un résultat négatif, le Programme québécois n'exige pas une deuxième lecture des clichés présentant des résultats positifs ou négatifs. Dans le Programme québécois, la mammographie est systématiquement prise sous deux incidences, alors que plusieurs études utilisaient une seule incidence ou une association de simple et de double incidences. L'intervalle de 24 mois entre les mammographies du Programme québécois lui mérite un score de 1, note comparable à la moyenne des études sur le dépistage, dont les intervalles vont de 12 à 33 mois. Combinés, ces trois éléments de contraste, qui peuvent être considérés comme des éléments de la structure et des processus, obtiennent une note de 0,90, ce qui se compare favorablement aux scores des essais (entre 0,68 et 1,07). Le tableau 4 récapitule les scores alloués aux essais et au Programme québécois actuel pour ces éléments.

3.3.2 Participation

Le Programme québécois a été mis en place en 1998, période où les taux de dépistage étaient déjà beaucoup plus élevés qu'à l'époque des essais. Au cours de la période de 24 mois allant de janvier 1997 à décembre 1998, 56 % des femmes âgées de 50 à 69 ans avaient subi au moins une mammographie, qu'elle soit de dépistage ou de diagnostic. Pour la période des deux plus récentes années sur lesquelles on dispose de données, soit entre 2001 et 2002, ces taux avaient atteint 63 % chez les femmes de 50 à 69 ans. Plusieurs ajustements doivent être apportés à ces chiffres. Premièrement, étant donné que les règles qui régissent actuellement la mammographie de dépistage et de diagnostic sont les mêmes et que les conditions de facturation de la mammographie diagnostique sont moins restrictives, un certain nombre de mammographies facturées pour le diagnostic sont en fait des mammographies de dépistage. Deuxièmement, à mesure que les taux de dépistage augmenteront, il devrait y avoir moins de mammographies de diagnostic, puisque les lésions seront découvertes plus tôt, avant qu'elles soient symptomatiques. Troisièmement, la participation au Programme tend à augmenter depuis quelques années (environ 2 % par année). À la fin de 2003, le Programme était implanté dans 16 des 18 régions, et des mesures étaient prises pour offrir des services de dépistage à toutes les Québécoises par la création de services itinérants de mammographie [MSSS, 2004]. Il est plus utile d'évaluer l'efficacité du dépistage mammographique de masse par rapport à l'absence de dépistage. Dans cette perspective, et en présumant que le Programme atteigne un jour son objectif d'un taux de participation de 70 %, ce taux se trouverait dans la tranche supérieure de ceux obtenus dans les huit études, qui variaient entre 43 et 82 %. Le PQDCS, vu sous cet angle, permet d'assurer une structure et des processus optimaux au Québec ainsi que des taux de participation aussi élevés que possible, et ce, même si de nombreux éléments de structure et de processus étaient déjà en place et s'il y avait des taux de participation importants avant l'instauration du programme.

TABLEAU 4

Scores pour la structure et les processus des essais et du PQDCS

ESSAIS ET PQDCS	CONTRASTE TECHNIQUE (C)	ÉPOQUE (E)	QUALITÉ (Q)	SOUS-INDICE, STRUCTURE ET PROCESSUS (C x E x Q)
HIP (New York)	1,25	0,70	1,00	0,88
MMST (Malmö)	1,00	0,80	0,85	0,68
TCS (*Two-County*)	1,00	0,80	0,85	0,68
Édimbourg	1,25	0,80	0,90	0,90
NBSS-1 (Canada)	1,25	0,90	0,95	1,07
NBSS-2 (Canada)	0,80	0,90	0,95	0,68
Stockholm	1,00	0,90	0,80	0,72
Göteborg	1,00	0,90	0,88	0,79
UK Age Trial	1,00	1,00	0,98	0,98
PQDCS, 2004	1,00	1,00	0,90	0,90

Enfin, il est important de souligner qu'une femme qui participe régulièrement à un programme de dépistage en tirera de plus grands bienfaits que la moyenne, puisque celle-ci est la valeur intermédiaire entre le faible bienfait qu'en tirent les femmes qui subissent un dépistage de façon irrégulière ou n'en ont pas du tout et le bienfait plus élevé qu'en tirent les femmes qui se soumettent au dépistage. Dans ce sens, le calcul de la réduction de la mortalité modulé par le taux de participation sous-estime le bienfait qu'une femme donnée obtient du dépistage. Inversement, la population étant la somme des femmes qui la constituent, il est évident que l'effet à l'échelle populationnelle sera plus important si un grand nombre de femmes participent régulièrement au dépistage.

3.3.3 Dilution temporelle

La dilution temporelle est problématique dans les études ayant une mise en contraste et un suivi courts, mais ce problème ne s'applique pas à la réduction de la mortalité qu'atteindra éventuellement un programme en régime établi. La valeur de 1,0 est donc accordée au facteur de dilution temporelle d'un tel programme de dépistage. Dans les essais sur la mammographie de dépistage, les dilutions temporelles allaient de 0,28 à 0,76 (tableaux 3 et B-4). Bien que la grandeur de ces indices soit influencée en bonne partie par la structure, les processus et les taux de participation qui caractérisent ces essais, la nature des données, leur mesure et les méthodes d'analyse ont probablement dilué davantage les effets examinés. Un programme de dépistage comme le PQDCS a donc plus de chances d'obtenir de meilleurs résultats que les essais, tous handicapés par une dilution temporelle considérable.

3.3.4 Force du contraste

L'échelle de la force du contraste a été conçue pour mesurer dans quelle proportion chaque essai a obtenu le plein potentiel de la mammographie de dépistage de masse, c'est-à-dire l'efficacité réelle qu'on obtiendrait avec un programme « idéal » : taux de participation optimal, normes de qualité modernes, intervalles courts entre les mammographies, etc. Selon ce paradigme, le PQDCS obtiendrait un score de 0,63 pour le contraste avec un taux de participation de 70 %. Cette

note se compare très favorablement aux scores qu'ont obtenus les essais sur la mammographie, qui allaient de 0,12 à 0,45. Cela indique que la performance du dépistage en vigueur aujourd'hui sera très probablement meilleure que celle des essais sur la mammographie. D'éventuelles améliorations dans la qualité du dépistage mammographique et dans les taux de participation augmenteraient probablement cet avantage.

4 SYNTHÈSE DES RÉSULTATS

4.1 ÉVALUATIONS RÉALISÉES PAR D'AUTRES CHERCHEURS

Bien que la plupart des recensions systématiques et des méta-analyses aient tenu compte de toutes les études sans discrimination, plusieurs chercheurs ont exclu certains essais à cause de problèmes de validité. Le tableau 5 montre que les chercheurs qui ont analysé la qualité des études ont évalué assez sévèrement l'étude HIP et celle d'Édimbourg. De plus, l'analyse d'Olsen et Gøtzsche [2001] a montré que les études HIP et d'Édimbourg étaient certes faibles, mais que les trois études suédoises les plus récentes étaient aussi de mauvaise qualité. Quelques auteurs ont exclu les données de l'étude canadienne NBSS-2 parce que son contraste technique (mammographie *versus* examen clinique des seins) différait, d'un point de vue qualitatif, des autres études.

TABLEAU 5

Recensions précédentes : essais inclus et exclus

MÉTA-ANALYSES	ESSAIS EXCLUS	ESSAIS INCLUS, MAIS DE FAIBLE QUALITÉ	ESSAIS INCLUS, QUALITÉ MOYENNE OU BONNE, OU QUALITÉ NON ÉVALUÉE
CETS, 1993; 1990	HIP	–	Tous sauf HIP
Fletcher *et al.*, 1993	NBSS-2 (ECS* dans le groupe témoin)	–	Tous sauf NBSS-2
Kerlikowske *et al.*, 1995	Aucun	–	Tous
Glasziou, 1997	Aucun	–	Tous
Hendrick *et al.*, 1997	Aucun	–	Tous
Olsen et Gøtzsche, 2001 (Cochrane)	HIP; Édimbourg	TCS; Stockholm; Göteborg	NBSS 1 et 2; Malmö
US Preventive Services Task Force Humphrey *et al.*, 2002	Édimbourg	–	Tous, sauf Édimbourg
Health Council of the Netherlands, 2002	Édimbourg	–	Tous sauf Édimbourg
Centre International de Recherche sur le Cancer, 2002	Édimbourg (facteurs de confusion); NBSS-2 (ECS dans le groupe témoin); HIP; NBSS-1 (ECS* dans le groupe de dépistage)	–	Essais suédois
Le présent rapport : AETMIS, 2005	NBSS-2 (ECS* dans le groupe témoin)	Göteborg; Stockholm; Édimbourg; TCS; HIP	NBSS-1; Malmö; (*UK Age Trial*)

* ECS : examen clinique des seins.

En plus de ces recensions systématiques, plusieurs chercheurs ont analysé et commenté la revue controversée qu'Olsen et Gøtzsche [2001] ont réalisée pour le groupe de collaboration nordique Cochrane. Leurs conclusions varient mais, en général, ils sont plutôt d'accord avec les critiques soulevées par Olsen et Gøtzsche tout en tempérant ou en réfutant les conclusions de ces derniers. Des révisions importantes de la critique d'Olsen et Gøtzsche ont depuis été faites par le *Health Council of the Netherlands* [2002], l'Agence Nationale d'Accréditation et d'Évaluation en Santé de France [ANAES, 2002], Santé Canada [Kakuma, 2002] le *US Preventive Services Task Force* [Humphrey *et al.*, 2002] et un groupe du Centre International de Recherche sur le Cancer [2002] de l'Organisation mondiale de la santé (OMS).

Le groupe français a souligné que les résultats d'Olsen et Gøtzsche contredisent neuf méta-analyses menées entre 1993 et 1998. Cela n'a pourtant rien d'étonnant, car ces méta-analyses incluaient généralement toutes les études sans égard à leur qualité. L'ANAES était aussi d'avis qu'Olsen et Gøtzsche accordaient trop d'importance à la mortalité globale comme critère principal d'évaluation, et n'était pas convaincue que ces derniers avaient adéquatement justifié leur évaluation de la qualité méthodologique des essais qu'ils avaient recensés. L'agence française a conclu que la méta-analyse d'Olsen et Gøtzsche n'apportait pas d'arguments assez solides pour remettre en question ses recommandations en faveur du dépistage mammographique chez les femmes de 50 à 69 ans.

Le *Health Council of the Netherlands* était d'accord avec une grande part des critiques sur les essais, surtout sur les études HIP et d'Édimbourg. Quant aux trois essais suédois qu'Olsen et Gøtzsche jugeaient de mauvaise qualité, le groupe hollandais était d'accord avec plusieurs des conclusions de ces deux méta-analystes, mais doutait que les problèmes relevés soient suffisamment graves pour justifier leur classement comme études de mauvaise qualité. Il a aussi reproché à Olsen et Gøtzsche de ne pas avoir tenu compte des différences entre les études canadiennes et suédoises, ni surtout du fait que le contraste était plus faible dans l'une des études canadiennes (NBSS-2). Il a noté qu'Olsen et Gøtzsche avaient donné préséance à la validité (interne) des essais contrôlés randomisés et, bien qu'étant d'accord avec cette approche en principe, il a critiqué la façon dont les critères de qualité ont été choisis et appliqués. Il a conclu que les arguments présentés dans cette recension Cochrane ne réfutent pas de façon convaincante la preuve que le dépistage du cancer du sein améliore la survie des femmes de plus de 50 ans [HCN, 2002].

Dans un rapport préparé pour Santé Canada, Kakuma [2002] a révisé la qualité des essais sur la mammographie de dépistage et relevé plusieurs des mêmes problèmes énoncés par Olsen et Gøtzsche. Elle a aussi révisé les méta-analyses de Fletcher et ses collaborateurs [1993] et de Kerlikowske et ses collègues [1995]. Fletcher avait relevé de nombreux problèmes relatifs à la validité des essais sur le dépistage et avait donc choisi de ne pas procéder à une méta-analyse en bonne et due forme, mais elle a tout de même accepté les résultats de tous les essais, sauf celui du Canada. Kerlikowske, par ailleurs, n'avait pas évalué la qualité des essais avant d'en regrouper les données. Kakuma a conclu que la recension d'Olsen et Gøtzsche étant plus exhaustive et que les lacunes de chacun des essais étant suffisamment nombreuses pour compromettre la validité de leurs résultats, la combinaison des données de tous les essais aurait tendance à fausser les résultats.

Le *US Preventive Services Task Force* a critiqué la méta-analyse d'Olsen et Gøtzsche parce qu'elle ne tenait pas compte de l'âge des femmes étudiées, du nombre, du type et de la qualité des dépistages, des intervalles entre les mammographies, des taux d'observance et de contamination, ni des différences qualitatives dans la mise en contraste des études. Le groupe de travail américain a conclu que les données scientifiques sur l'efficacité du dépistage, même si elles sont

de qualité moyenne, indiquent une réduction de la mortalité de 20 à 35 %.

Enfin, le Centre International de Recherche sur le Cancer [2002] de l'OMS, dans un chapitre de son document sur l'efficacité du dépistage par mammographie, a passé en revue tous les essais et a exclu les études HIP et NBSS-1 parce que les femmes de la cohorte de dépistage avaient eu des examens cliniques des seins, celle d'Édimbourg en raison de facteurs de confusion, et l'essai NBSS-2 à cause de la nature différente de son plan expérimental (examen clinique et auto-examen des seins dans le groupe témoin). Ces chercheurs ont jugé que les quatre essais suédois étaient valides, malgré les critiques soulevées par Olsen et Gøtzsche, et ont conclu que le dépistage avait réduit la mortalité par cancer du sein d'environ 25 % chez les femmes âgées de 50 à 69 ans, et de 19 % chez les femmes de moins de 50 ans.

4.2 APPROCHE D'ANALYSE DE LA PREUVE

La présente analyse des essais publiés sur la mammographie a relevé des différences marquées dans la qualité des plans et de l'exécution des études. Les recensions systématiques traditionnelles concluent généralement qu'il y a un bon niveau de preuve à l'appui du dépistage mammographique, puisqu'il y a au moins un essai randomisé contrôlé rigoureux ou une méta-analyse portant sur plus d'un essai. Une telle analyse est simpliste et surestime la qualité de la preuve. D'une part, les études sur la mammographie n'ont pas atteint les normes modernes de qualité en vigueur pour les essais. Certaines n'étaient pas randomisées du tout, les participantes étant choisies par les chercheurs pour des raisons pratiques. Les comptes rendus de la plupart des essais sont incomplets. Notamment, la plupart ne fournissaient même pas les caractéristiques de départ des femmes des groupes de dépistage et témoin, souvent parce qu'on ne disposait pas d'informations sur la cohorte témoin. L'exclusion des cas de cancer préalablement diagnostiqués n'a pas été systématique, et aucun essai n'a été mené à l'insu, ni des patientes, ni du personnel soignant. Ces lacunes ont amené un chercheur [Baum, 2004] à dénoncer « […] un système de deux poids, deux mesures, car les essais cliniques sur le dépistage sont jugés acceptables à un degré de qualité qui invaliderait totalement des essais portant sur la prévention ou le traitement du cancer du sein » [trad.].

D'autre part, aucune étude n'a été conçue ni menée de manière à obtenir le plein potentiel du dépistage mammographique, et cet aspect de la validité mérite aussi d'être soumis à un examen détaillé. La présente analyse, plutôt que de qualifier le faisceau d'arguments de faible ou de fort, examine chaque étude séparément en fonction de sa pertinence pour l'évaluation d'un programme de dépistage moderne, et évalue la force de son contraste et les autres aspects de la validité qui contrôlent les biais systématiques. Cette approche peut se résumer ainsi :

1re étape : sélectionner les essais pertinents en fonction de la nature du contraste étudié;

2e étape : trier ces essais par degré de contrôle des biais à direction inconnue;

3e étape : combiner les essais dans cet ordre, en ajoutant successivement les essais de moindre validité et en comptabilisant la force du contraste de ces combinaisons ainsi que les résultats relatifs à la mortalité par cancer du sein.

Comme nous l'avons déjà mentionné, pour des raisons de pertinence de la mise en contraste, l'essai NBSS-2 sera exclu de la présente analyse et examiné séparément. Les essais pertinents, classés par ordre chronologique, sont présentés au tableau 6.

Ces essais pertinents ont ensuite été classés par score de validité (tableau 7), car les essais ayant de faibles scores peuvent avoir des résultats biaisés, bien qu'il soit impossible de connaître la direction de ce biais. L'échelle d'évaluation indique que la validité de deux essais, Édimbourg et *Two-County*, est particulièrement problématique. L'essai HIP

pourrait aussi être considéré comme inutilisable, car même s'il a obtenu une note moyenne sur l'échelle de validité, la moitié des patientes ont été perdues de vue au suivi, ce qui compromet la validité de l'étude. Par ailleurs, la plupart des auteurs qui se sont prononcés sur la qualité des essais individuels ont exclu l'un ou l'autre de ces trois essais à cause de leur mauvaise qualité méthodologique.

L'essai du Royaume-Uni (*UK Age Trial*) a été inclus dans ces tableaux à des fins de comparaison, mais il ne sera pas utilisé dans les analyses subséquentes, car aucun résultat définitif sur la mortalité n'a encore été publié.

Enfin, la force du contraste de chacun de ces essais a été évaluée sur une échelle quantitative. Les résultats de cette évaluation sont présentés au tableau 8 dans le même ordre de validité décroissante. En appliquant l'inverse de la variance comme facteur de pondération, la valeur cumulative de la force du contraste de ces essais est aussi établie au fur et à mesure que les essais de moindre validité sont progressivement inclus.

TABLEAU 6

1re étape : essais pertinents pour l'évaluation d'un programme moderne de dépistage

ANNÉE D'INSTAURATION	ESSAIS
1991	*UK Age Trial*
1982	Göteborg
1981	Stockholm
1980	NBSS-1
1979	Édimbourg
1977	*Two-County* (TCS)
1976	Malmö (MSST)
1963	HIP

TABLEAU 7

2e étape : essais par ordre de validité

CLASSEMENT	ESSAIS	VALIDITÉ
1	NBSS-1	4,0
2	Malmö (MSST)	3,0
3	Stockholm	2,0
4	Göteborg	1,5
5	HIP	1,5
6	*Two-County* (TCS)	0,5
7	Édimbourg	0,5
	*UK Age Trial**	4,0

* Présenté à titre de comparaison.

TABLEAU 8

3e étape : force du contraste des essais

CLASSEMENT	ESSAIS	FORCE DU CONTRASTE	FORCE DU CONTRASTE CUMULATIVE*
1	NBSS-1	0,45	0,45
2	Malmö (MSST)	0,27	0,38
3	Stockholm	0,28	0,36
4	Göteborg	0,37	0,36
5	HIP	0,12	0,29
6	*Two-County* (TCS)	0,42	0,34
7	Édimbourg	0,26	0,32

* Pondérée par l'inverse de la variance du logarithme du rapport de cotes, soit 1/VAR(ln(OR)).

Aucun essai ne s'est approché de la norme d'une étude idéale qui comprendrait plusieurs années d'examens de dépistage réguliers, du matériel moderne, un contrôle de la qualité, des mammographies sous deux incidences à des intervalles de deux ans ou moins, et une participation maximale. Par rapport à cette norme, les sept essais sur le dépistage mammographique n'ont obtenu qu'une force de contraste située entre 12 et 45 % du plein potentiel. Par rapport à cette même norme, les programmes modernes obtiendront probablement une force de contraste très supérieure. Par rapport à l'absence de dépistage, le Programme québécois pourrait atteindre 63 % de son plein potentiel s'il obtenait un taux de participation de 70 %. Une meilleure performance des programmes de dépistage n'est pas nécessairement associée à de plus grandes réductions de la mortalité, puisque la mortalité peut aussi être réduite par une efficacité accrue des traitements des cancers de stade plus avancé. Or, dans le contexte des options thérapeutiques dont on dispose aujourd'hui, un programme moderne ayant un bon score pour la force de contraste permettrait de déceler les lésions plus tôt et d'obtenir une réduction maximale de la mortalité.

Les revues systématiques précédentes suscitent aussi d'autres réflexions. Premièrement, la plupart d'entre elles ont inclus toutes les études, sans égard à leur validité ni, surtout, à leur force de contraste. Si ce procédé offre l'avantage d'éviter les choix arbitraires et de donner des intervalles de confiance plus petits, il suppose néanmoins que la validité et la force du contraste ne sont pas des facteurs importants. Un tel optimisme n'est pas justifié dans le cas des essais sur le dépistage mammographique, car ils laissaient tous beaucoup à désirer sur le plan de la méthodologie et de la qualité, et différaient tous de façon marquée. Deuxièmement, ceux qui ont énoncé des critères d'exclusion ne se sont préoccupés que de certains aspects de la validité susceptibles d'introduire des biais à magnitude et à direction inconnues. Ces aspects, quoique importants, ne suffisent pas lorsque la faiblesse du contraste à l'étude est susceptible de fausser systématiquement les résultats dans le sens de conclusions négatives. C'est ce qui se passe dans les essais sur le dépistage mammographique. De plus, si des critères de validité sont utilisés, ils devraient être explicites pour que les lecteurs soient en mesure de juger de leur pertinence et de la justesse de leur application.

Les sections suivantes présenteront les résultats des essais relatifs aux effets du dépistage sur la mortalité par cancer du sein en fonction de leur force de contraste et de leur validité. Pour ce faire, les résultats sur la mortalité de chaque essai seront encore présentés par ordre décroissant de validité. L'effet cumulatif sur la mortalité est donné

pour toutes les études, quelle qu'en soit la validité, afin de permettre au lecteur de juger jusqu'où il est prêt à descendre sur la liste. Le score de validité permet donc d'organiser l'ordre des essais en fonction du degré de contrôle des biais à direction inconnue, et le score de la force du contraste permet d'évaluer le degré de dilution des résultats de chaque essai ou de chaque combinaison d'essais.

4.3 RÉSULTATS SUR LA MORTALITÉ PAR CANCER DU SEIN DANS TOUS LES GROUPES D'ÂGE

Bien que le présent rapport n'ait pas pour principal objectif d'en arriver à une seule estimation de la capacité de la mammographie à réduire la mortalité par cancer du sein, il est intéressant de comparer les essais selon leur validité et leur force de contraste d'une part, et leurs résultats d'autre part. Dans cette partie et dans les deux suivantes, les résultats publiés sur la mortalité par cancer du sein seront examinés dans tous les groupes d'âge, chez les femmes plus jeunes (moins de 50 ans) et les femmes plus âgées (50 ans et plus). Bien que les résultats de toutes les études soient présentés, l'accent sera mis sur les quatre études les plus valides. Les deux essais qui obtiennent le plus haut score de validité montrent une réduction modeste et non significative de la mortalité par cancer du sein. Le grand essai canadien NBSS-1 a enregistré un taux de mortalité similaire dans le groupe de dépistage et le groupe témoin, avec un risque relatif (RR) de 0,97 (IC 95 % : de 0,74 à 1,27). L'essai de Malmö a aussi enregistré une réduction non significative, avec un risque relatif de 0,82 (IC 95 % : de 0,59 à 1,15). Les résultats de l'essai du Royaume-Uni (*UK Age Trial*), aussi de validité supérieure, n'ont pas encore été publiés.

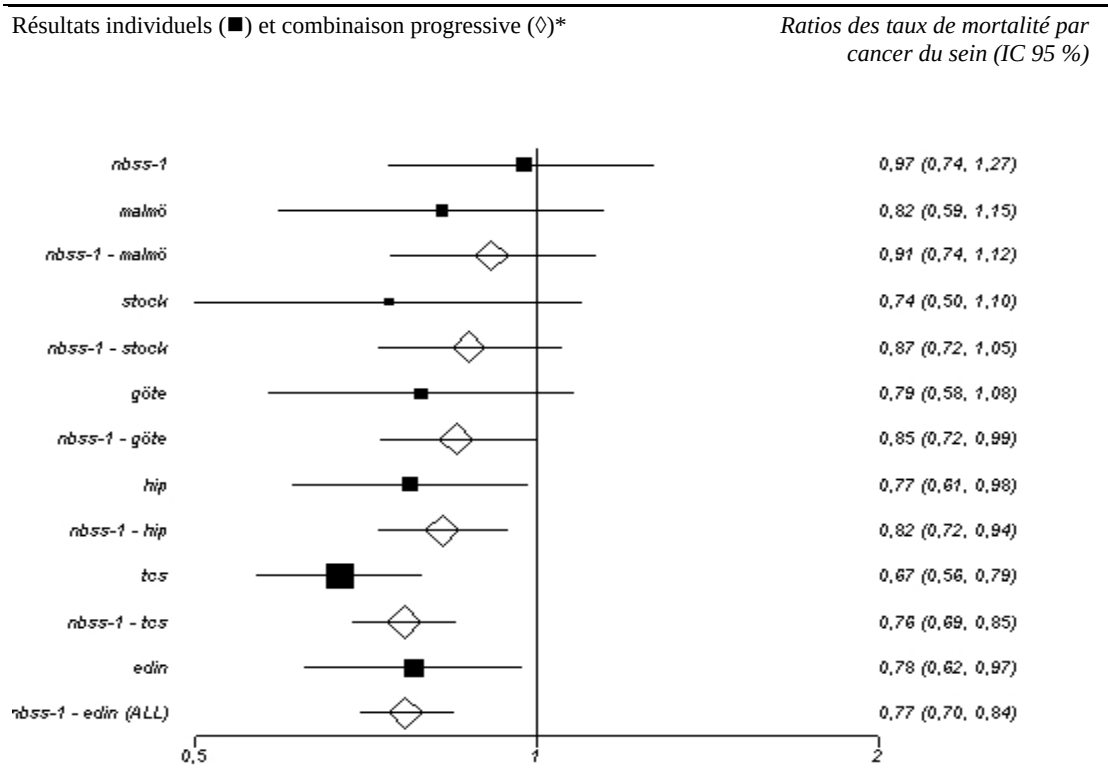
Les essais de Stockholm, de Göteborg et HIP, qui affichent de faibles scores pour la force du contraste et la validité, ont obtenu de plus grandes réductions de la mortalité, avec un RR de 0,74 (IC 95 % : de 0,50 à 1,10), de 0,79 (IC 95 % : de 0,58 à 1,08), et de 0,77 (IC 95 % : de 0,61 à 0,98) respectivement. Enfin, les études *Two-County* et d'Édimbourg, de très mauvaise qualité, ont enregistré de meilleurs risques relatifs, soit 0,67 (IC 95 % : de 0,56 à 0,79) et 0,78 (IC 95 % : de 0,62 à 0,97) respectivement. Certains spécialistes de la méta-analyse sont d'avis que les études de faible validité ont tendance à surestimer l'efficacité des interventions qu'elles évaluent. Les sept études examinées ici tendent à corroborer cette conclusion. Si l'on considérait seulement les aspects traditionnels de la validité, sans tenir compte de la force du contraste, on pourrait conclure qu'il serait justifié de douter sérieusement de l'efficacité du dépistage mammographique.

La figure 1 résume les résultats de chaque essai au chapitre de la mortalité par cancer du sein, par ordre décroissant de validité, avec le résultat cumulatif moyen pondéré à mesure que les essais de moindre validité sont ajoutés.

On pourrait s'attendre à ce que les études aux contrastes plus faibles obtiennent des résultats négatifs, mais c'est l'inverse qui se produit dans les essais de New York (HIP), de Stockholm et d'Édimbourg, qui, malgré des scores inférieurs, ont pourtant enregistré des réductions significatives de la mortalité. Ces résultats contradictoires pourraient s'expliquer par la faible validité de ces études. Ou encore par le fait que, même si les essais les plus anciens, comme celui du HIP, n'avaient pas une mise en contraste marquée du dépistage, les traitements disponibles à l'époque n'étaient pas aussi avancés, et toute amélioration, si minime soit-elle, dans la détection précoce des tumeurs a pu faire une différence plus grande que dans les essais plus récents, maintenant que les traitements sont devenus plus efficaces, tant pour les tumeurs à un stade précoce que plus avancé. Quelle que soit l'explication, ces résultats doivent être traités avec prudence, car ils pourraient ne pas s'appliquer au contexte actuel.

FIGURE 1

Synthèse des résultats sur la mortalité par cancer du sein (tous âges confondus)



* Chaque ligne contenant le symbole ◊ représente l'effet cumulatif de tous les essais individuels précédents.

Globalement, dans les deux études considérées comme les plus valides, le RR cumulatif est de 0,91 (IC 95 % : de 0,74 à 1,12). Si l'on ajoute les données de deux autres études de validité moindre, le RR global est de 0,85 (de 0,72 à 0,99). Les ajouts successifs des résultats d'études indiquent toujours une certaine réduction de la mortalité par cancer du sein, allant de 9 % (études de qualité moyenne), à 15 % (études de qualité moyenne et faible), à 23 % (toutes les études). Autrement dit, les études les plus valides tendent à enregistrer des réductions moindres, et les intervalles de confiance comprennent parfois la valeur nulle.

4.4 RÉSULTATS SUR LA MORTALITÉ PAR CANCER DU SEIN CHEZ LES FEMMES PLUS JEUNES

Les études ont inclus des femmes âgées de 39 à 74 ans à la randomisation, mais la taille de la plupart des essais individuels est trop petite pour que l'on puisse tirer des conclusions sur l'efficacité relative du dépistage dans les différents groupes d'âge. Malgré les nombreuses différences qualitatives des études, il était naturel d'essayer de combiner les résultats des divers groupes d'âge de tous les essais. L'accent a surtout été mis sur l'efficacité de la mammographie dans les

groupes d'âge qui correspondent en gros aux périodes préménopausiques et postménopausiques, l'âge de 50 ans étant habituellement le point de démarcation. En ce qui a trait aux femmes plus jeunes, il faut considérer deux éléments importants avant de s'attaquer à la preuve de l'efficacité. Premièrement, une étude qui recrute des femmes de 40 à 49 ans est tout à fait différente d'une étude qui recrute des femmes à l'âge de 40 ans et qui poursuit le dépistage jusqu'à l'âge de 49 ans. Si l'on veut savoir si le dépistage est efficace à partir de l'âge de 40 ans, la deuxième stratégie, soit celle adoptée par l'étude du Royaume-Uni, est tout indiquée. Cependant, tous les autres essais qui ont inclus des femmes de moins de 50 ans ont recruté des femmes de tous âges dans ce groupe, soit des femmes de 40 ans, mais aussi de 41 ans, de 42 ans, jusqu'à 49 ans. Par exemple, dans l'étude de Göteborg, au moment de l'inclusion dans la cohorte, toutes les catégories d'âge entre 39 et 49 ans étaient également représentées (entre 8 et 11 % par catégorie). Ces femmes ont donc en moyenne cinq ans de plus qu'il ne faudrait pour montrer les effets d'un dépistage qui débuterait à l'âge de 40 ans.

Deuxièmement, au cours de la période de mise en contraste d'une étude, plusieurs femmes qui commencent dans un groupe plus jeune passent au groupe d'âge suivant. Ainsi, la catégorie « femmes plus jeunes » dans l'étude de Malmö était composée de femmes âgées de 45 à 54 ans. Les plus jeunes de ces femmes, qui avaient commencé le dépistage à 45 ans, étaient âgées de 54 ans lorsque la période de dépistage a pris fin, et celles qui avaient 54 ans au début de l'étude avaient 63 ans à la fin de la période de mise en contraste. Comme l'ont observé les chercheurs de l'étude HIP, on ne peut être sûr que l'efficacité de la mammographie de dépistage chez ces femmes n'était pas attribuable au dépistage qu'elles ont eu après qu'elles ont changé de catégorie d'âge. De plus, des sept essais qui incluaient des femmes plus jeunes, deux (Malmö et Édimbourg) ne comptaient aucune femme de moins de 45 ans [Shapiro *et al.*, 1985].

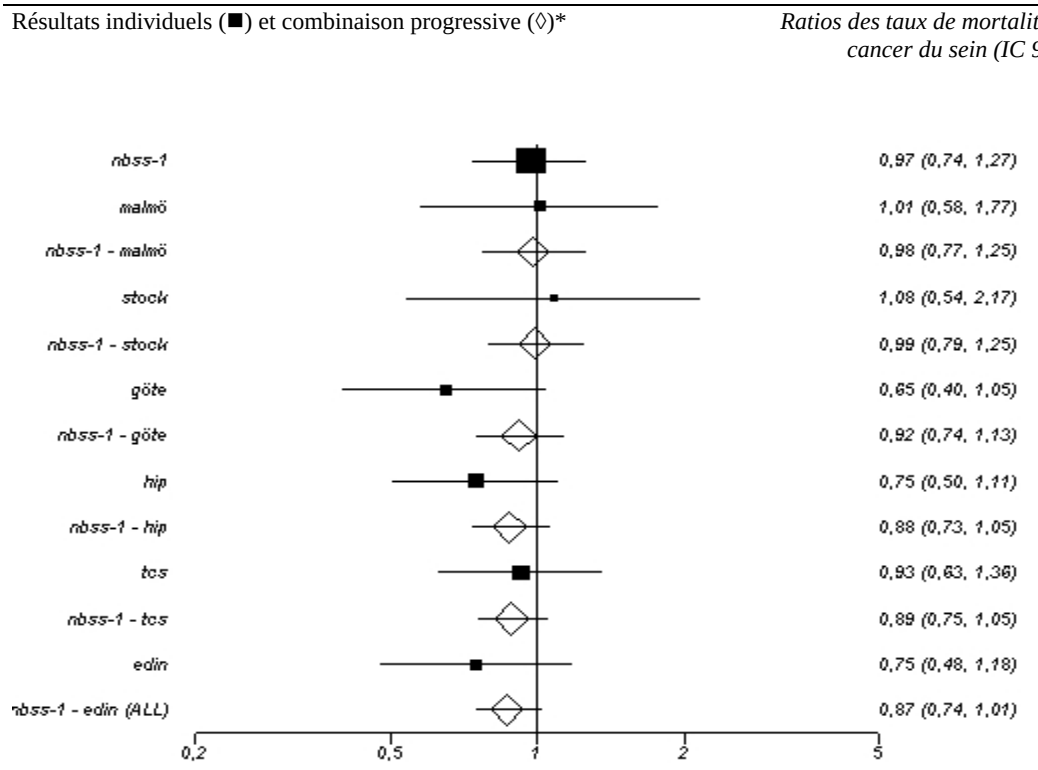
Les intervalles de confiance beaucoup plus grands des résultats de ces sept essais peuvent être attribués tant au fait qu'il y avait moins de femmes dans ces groupes d'âge qu'au fait que moins de décès sont escomptés chez ces femmes, l'incidence du cancer du sein étant plus faible dans ce groupe d'âge.

Encore une fois, ces études sont trop hétérogènes pour qu'on puisse combiner leurs résultats sans procéder à un examen détaillé de leurs différences, mais en règle générale, la réduction de la mortalité est moindre dans ce groupe que chez les femmes plus âgées. Les deux essais qui affichent les meilleurs scores de validité ont tous deux enregistré le même taux de mortalité dans le groupe soumis au dépistage, l'essai NBSS-1 indiquant un RR de 0,97 (IC 95 % : de 0,74 à 1,27) et l'essai de Malmö, un RR de 1,01 (IC 95 % : de 0,58 à 1,77). Les études dont les scores de validité sont plus bas ont obtenu des résultats plus variés, avec un RR de 1,08 pour l'étude de Stockholm (IC 95% : de 0,54 à 2,17) et de 0,65 pour l'étude de Göteberg (IC 95 % : de 0,40 à 1,05). La figure 2 résume les résultats disponibles sur ces groupes de femmes plus jeunes et présente les résultats des essais individuels et les résultats cumulatifs des essais ajoutés par ordre de validité.

Globalement, dans les essais de qualité moyenne, le RR cumulatif est de 0,98 (IC 95 % : de 0,77 à 1,25). Si l'on inclut les données des essais de moindre validité, le RR général serait de 0,92 (IC 95 % : de 0,74 à 1,13). Les meilleures données indiquent ainsi une réduction modeste et non significative de la mortalité par cancer du sein. L'addition successive de résultats d'essais de moindre validité ne montre pas de réduction significative de la mortalité par cancer du sein, ces résultats allant de 2 % (essais de qualité moyenne), à 8 % (essais de qualité moyenne et faible), et à 13 % (toutes les études). Encore une fois, les essais les plus valides tendent à enregistrer des réductions plus faibles. De plus, la réduction de la mortalité est beaucoup plus faible chez les femmes plus jeunes, et les intervalles de confiance comprennent la valeur nulle pour toutes les études et combinaisons d'études.

FIGURE 2

Synthèse des résultats sur la mortalité par cancer du sein (femmes plus jeunes)



* Chaque ligne contenant le symbole ◊ représente l'effet cumulatif de tous les essais individuels précédents.

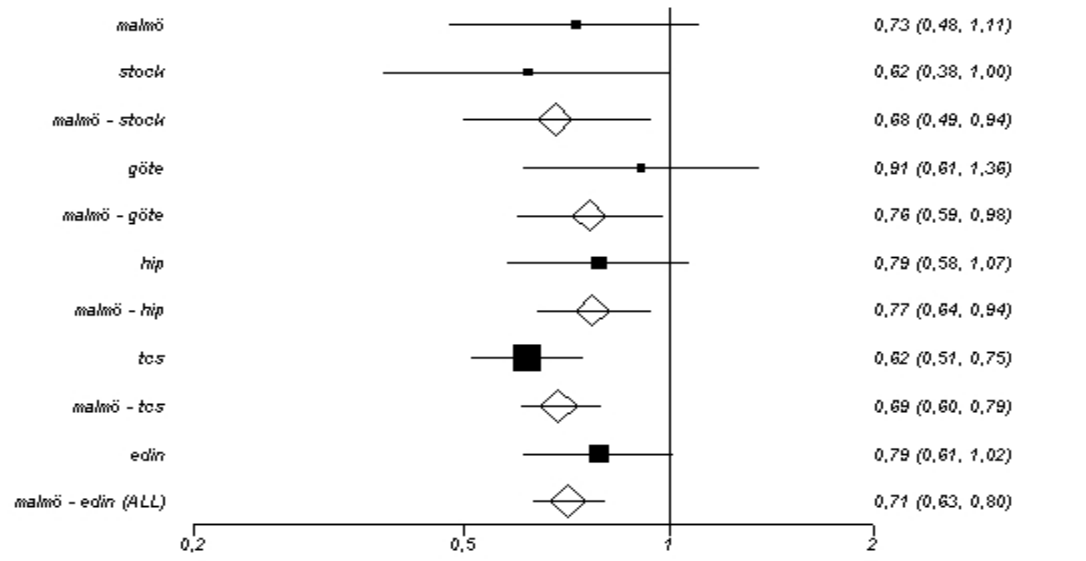
4.5 RÉSULTATS SUR LA MORTALITÉ PAR CANCER DU SEIN CHEZ LES FEMMES PLUS ÂGÉES

Une deuxième analyse par sous-groupes peut être effectuée à partir des données sur les femmes âgées de 50 ans et plus à l'entrée dans l'étude (55 ans dans l'essai de Malmö). Encore une fois, les essais sont probablement trop hétérogènes pour que la combinaison de leurs résultats soit appropriée. Les essais de Stockholm et de Göteborg, de validité moindre, ont enregistré des RR de 0,62 (IC 95 % : de 0,38 à 1,00) et de 0,91 (IC 95 % : de 0,61 à 1,36) respectivement. La figure 3 récapitule les résultats disponibles sur ces groupes plus âgés, tant ceux des études individuelles que les résultats cumulatifs des études ajoutées par ordre décroissant de validité.

Globalement, dans l'essai de Malmö, le seul essai de validité moyenne, le RR est de 0,73 (IC 95 % : de 0,48 à 1,11). Si l'on ajoute les résultats des études de faible qualité, le RR global est de 0,76 (IC 95 % : de 0,59 à 0,98), et si l'on inclut tous les essais sans égard à la validité, le RR global se situe à 0,71 (IC 95 % : de 0,63 à 0,80). Les données indiquent donc des réductions de la mortalité allant de 24 à 29 %, réductions considérablement plus importantes que celles enregistrées pour tous les groupes d'âge considérés ensemble. Les intervalles de confiance dans ce sous-groupe sont naturellement plus grands, mais aucune combinaison d'essais n'inclut la valeur nulle.

FIGURE 3

Synthèse des résultats sur la mortalité par cancer du sein (femmes plus âgées)



* Chaque ligne contenant le symbole ◊ représente l'effet cumulatif de tous les essais individuels précédents.

Bien que les résultats de l'essai NBSS-2 n'aient pas été combinés avec ceux des autres essais à cause de son contraste technique différent, celui-ci n'a montré aucune réduction de la mortalité par cancer du sein dans le groupe soumis au dépistage mammographique, avec un RR de 1,02 (IC 95 % : de 0,78 à 1,33). Autrement dit, cet essai n'a pas démontré que la mammographie de dépistage offrait une plus-value supérieure au dépistage intensif par des examens cliniques des seins et des auto-examens des seins périodiques et de bonne qualité. Il n'est pas du ressort de ce rapport de déterminer si ces résultats s'expliquent plutôt par l'efficacité de l'examen clinique des seins, de l'auto-examen des seins ou des deux, ou encore, comme le croit le chercheur principal, par de meilleurs traitements des cancers à un stade plus avancé [Miller *et al.*, 2000a], ou par tout autre facteur.

5 DISCUSSION ET CONCLUSIONS

Tous les essais publiés sur la mammographie de dépistage ont été passés en revue, et leurs forces et leurs faiblesses méthodologiques ont été examinées en détail. Les conclusions qui en découlent se situent à mi-chemin entre celles de la plupart des auteurs de revues systématiques, qui estiment que l'efficacité réelle du dépistage mammographique est bien démontrée, et celles d'Olsen et Gøtzsche, qui affirment que la preuve de l'efficacité du dépistage reste à faire.

Après analyse de la validité des essais, le plus haut score est accordé aux études du Canada et de Malmö, qui n'appuient pas l'hypothèse de l'efficacité de la mammographie de dépistage. Les essais qui ont obtenu un score de validité moindre en raison de leur plan expérimental et de leur déroulement ayant généralement formulé des conclusions plus favorables au dépistage, leur inclusion dans l'analyse donne évidemment des réductions plus importantes de la mortalité, mais tend à affaiblir la fiabilité de ces conclusions. De plus, plusieurs essais avaient aussi un faible contraste, y compris ceux du Canada et de Malmö, ce qui rend plus difficile de démontrer le plein potentiel de la mammographie lorsqu'elle est réalisée dans des conditions optimales, c'est-à-dire par un personnel compétent, avec des appareils modernes et sur une période adéquate.

Les essais sur le dépistage mammographique présentent aussi d'autres lacunes, dont la dilution temporelle. Bien que celle-ci puisse être partiellement compensée par le retrait des cas de mortalité des premières années du suivi, cet ajustement ne modifierait pas fondamentalement nos conclusions. Les autres lacunes dans l'établissement ou le maintien d'une mise en contraste ne peuvent être corrigées par l'analyse, mais devraient néanmoins être notées, car elles peuvent avoir masqué le potentiel de réduction de la mortalité de la mammographie. En guise de comparaison, un programme moderne de dépistage comme le PQDCS, avec un taux de participation de 70 %, obtiendrait un contraste plus fort par rapport à l'absence de dépistage que n'importe lequel des essais.

Il est donc probable que le PQDCS permettra de dépister le cancer du sein à un stade plus précoce que les essais. Cela ne suffira pas à donner de plus grandes réductions de la mortalité, puisque les progrès thérapeutiques auront peut-être annulé une partie de cet avantage. Il n'en demeure pas moins qu'une bonne performance de dépistage et de diagnostic constitue une condition nécessaire à l'obtention d'un résultat optimal.

La présente recension a été entreprise pour répondre à trois questions.

Question 1 : Quelle est la solidité des arguments scientifiques à l'appui des programmes de dépistage mammographique ?

La validité de la plupart des essais qui appuient le dépistage mammographique suscite de sérieuses réserves. Certains de ces essais méritent à peine le nom de « randomisés », et le degré de contrôle dans l'application de la mise en contraste prévue rend tout aussi discutable l'usage du terme « contrôlé ». Dans la plupart des études, dont les quatre études suédoises, la liste des problèmes méthodologiques est longue : on savait peu de choses sur la population témoin, excepté l'âge de ses membres, plusieurs études ont eu recours à une répartition non aléatoire entre les groupes de dépistage et témoin, l'exclusion des cas de cancer du sein déjà diagnostiqués était presque toujours faite des années après la fin de la période de mise en contraste de l'étude à partir de dossiers médicaux, la mise à l'insu n'a jamais été tentée, etc. Des aspects importants des essais contrôlés randomisés n'étaient généralement pas consignés selon les normes modernes, les techniques de randomisation, la comparabilité de base des cohortes de dépistage et témoin, les critères d'exclusion, etc., étant souvent

passés sous silence. De plus, dans la plupart des études, les méthodes et les analyses étaient mal décrites. Les critiques relatives à ces aspects des études sur la mammographie [Olsen et Gøtzsche, 2001] n'ont pas incité les chercheurs principaux des essais à élucider certains détails ou à fournir des explications convaincantes [Nyström *et al.*, 2002].

Outre ces problèmes de validité, les huit essais sont hautement hétérogènes en ce qui a trait à la force du contraste qu'ils ont étudié. La mammographie est une technique complexe dont le succès repose sur une longue chaîne d'interventions qui doivent être exécutées avec précision. Les nombreuses lacunes présentes dans toutes les études importantes pourraient signifier que le plein potentiel de la mammographie de dépistage n'a pas été réalisé. Il y a de bonnes raisons de croire qu'un essai valide et bien conçu sur la mammographie moderne, dans des conditions qui garantissent une exécution technique de grande qualité, sur une longue période, avec des taux de participation élevés et de faibles taux de contamination, avec un suivi et une analyse des résultats adéquats, pourrait prouver de façon convaincante l'efficacité de la mammographie de dépistage. L'essai du Royaume-Uni (*UK Age Trial*) pourrait remplir la plupart de ces conditions, mais les résultats complets ne seront probablement pas disponibles avant 2008 ou 2010. Il y a aussi de bonnes raisons de croire qu'un programme moderne de mammographie de dépistage, mené dans les mêmes conditions de qualité supérieure, obtiendrait des résultats considérablement meilleurs que les études.

L'analyse de la littérature scientifique montre que le dépistage mammographique tel qu'il a été réalisé dans les meilleurs essais a entraîné une réduction modérée de la mortalité par cancer du sein de l'ordre de 9 à 15 % pour l'ensemble des groupes d'âge étudiés. Les données restreintes aux femmes âgées de plus de 50 ans montrent des réductions un peu plus importantes, de l'ordre de 24 à 29 %. Par contre, la présente analyse a montré que la mammographie moderne, effectuée dans des circonstances qui maximisent sa performance, a le potentiel de déceler les lésions cancéreuses à un stade plus précoce, ce qui pourrait réduire davantage la mortalité. Les avantages pour la femme qui participe régulièrement à un programme de dépistage devraient en outre être supérieurs aux avantages moyens dont font état les essais populationnels.

Il y a aussi des raisons de nature qualitative de croire que le dépistage mammographique comporte des effets indésirables, mais il y a très peu de données quantitatives qui pourraient aider un décideur à évaluer si les préjudices sont suffisamment importants pour en annuler les avantages. Il est à noter que même si les taux de faux positifs étaient de 2 à 5 % environ dans les études et les programmes de dépistage suédois [Nyström *et al.*, 2002; Elmore *et al.*, 1998], ils sont généralement deux ou trois fois supérieurs en Amérique du Nord [INSPQ, 2003; Santé Canada, 2003; Elmore *et al.*, 1998], rendant le rapport entre les effets potentiellement positifs et négatifs du dépistage moins favorable. Les désagréments de la mammographie, l'anxiété provoquée par les faux positifs, les explorations diagnostiques supplémentaires et le coût du dépistage sont plus certains que l'efficacité espérée. Bien qu'il y ait peu de données sur l'anxiété à long terme que peut provoquer le dépistage [Lampic *et al.*, 2001], il faut comprendre qu'il est difficile de mesurer ce genre d'effets secondaires et qu'il est tout aussi difficile de quantifier leur gravité sur la même échelle que les bienfaits potentiels de la mammographie. À la lumière des données disponibles, il est impossible d'examiner adéquatement l'association éventuelle entre la mammographie et de sérieux effets indésirables, les études publiées jusqu'à maintenant ayant généralement été conçues dans l'unique but de vérifier l'hypothèse de la réduction de la mortalité par cancer du sein. Même si le consensus qui s'est formé autour du dépistage mammographique rend peu probable que de nouvelles études randomisées soient réalisées sur le sujet, il devrait être possible d'évaluer les effets indésirables de la mammographie à partir d'un plan non expérimental [Miettinen, 1983], puisque le risque de subir ces effets

secondaires n'est probablement pas lié à la participation à un programme de dépistage. Les effets indésirables de la mammographie constituent un problème de magnitude inconnue qui pourrait faire contrepoids à la réduction de la mortalité par cancer du sein due à la mammographie.

Les prochaines mises à jour des huit essais ne modifieront probablement pas les résultats déjà publiés, la plupart des périodes pertinentes où le dépistage a pu avoir un effet sur la mortalité étant déjà passées. Le *UK Age Trial*, dont les résultats n'ont pas encore été publiés, constitue l'exception majeure à cette règle. En raison de sa taille et de sa qualité méthodologique, les résultats de cet essai pourraient bien avoir un effet notable sur le corpus général de la preuve. Il n'y aura probablement pas d'autres essais sur la mammographie de dépistage pour les femmes de 50 à 69 ans, car un tel essai ne serait pas réalisable dans le contexte actuel, où le dépistage est généralisé. De plus, un nouvel essai ne saurait répondre aux exigences éthiques habituelles en raison du consensus clinique en faveur de la mammographie. Les prochaines améliorations apportées au dépistage mammographique devront probablement être évaluées d'après des mesures de performance diagnostique associées à des données sur la survie, et en fonction du degré de malignité et du stade de la tumeur.

Conclusion : *Les études scientifiques existantes, malgré leurs lacunes, peuvent justifier les programmes de dépistage mammographique. De plus, il y a de bonnes raisons de croire que des programmes de dépistage modernes et bien menés pourraient permettre de détecter et de diagnostiquer le cancer du sein à un stade plus précoce et peut-être d'obtenir des réductions de la mortalité par cancer du sein plus importantes que celles qu'ont enregistrées les essais sur le dépistage.*

Question 2 : Quelles sont les preuves à l'appui de la mammographie de dépistage pour les femmes de 40 à 49 ans ?

Pour ce groupe d'âge, les données se font plus rares. En effet, la plupart des études portent sur les femmes de plus de 50 ans, bien qu'un certain nombre de femmes dans certaines études aient commencé le dépistage quelques années avant leur cinquantième anniversaire. Si l'efficacité du dépistage avait été démontrée pour le sous-groupe plus jeune de ces études, on aurait seulement pu conclure que le dépistage est efficace chez les femmes de 45 à 54 ans. Les meilleures données disponibles ne montrent aucune réduction significative de la mortalité par cancer du sein chez les femmes soumises à un dépistage avant l'âge de 50 ans. Nous concluons qu'il n'y a pratiquement aucune preuve que le dépistage est efficace chez les femmes de 40 à 49 ans, et même qu'il y a quelques éléments de preuve de son inefficacité. Heureusement, le *UK Age Trial* fournira de nombreuses données sur les femmes de ce groupe d'âge. Qui plus est, si ce dernier montre une certaine efficacité, l'argument qui défend l'efficacité de la mammographie chez les femmes plus âgées sera aussi substantiellement renforcé.

Les effets secondaires sont particulièrement inquiétants chez les femmes plus jeunes, étant donné que les taux de faux positifs tendent à être plus élevés dans la plupart des programmes. Cela pourrait être dû au fait qu'il est plus difficile d'évaluer les femmes plus jeunes en raison de la plus grande densité de leurs seins et de leur réponse physiologique aux hormones pendant le cycle menstruel. En l'absence de données convaincantes sur l'efficacité de la mammographie dans ce groupe d'âge, il est possible que les effets néfastes dépassent les effets bénéfiques.

Conclusion : *Les résultats des essais publiés jusqu'à ce jour ne fournissent pas de justification scientifique permettant de recommander le dépistage chez les femmes de moins de 50 ans. Toutefois, ces résultats n'excluent pas la possibilité que le dépistage individuel, fondé sur une évaluation personnalisée du risque de ces femmes, puisse être bénéfique. Il sera opportun de revoir ces conclusions lorsque les résultats de l'essai du Royaume-Uni seront publiés.*

Question 3 : Quelles sont les implications des données scientifiques sur les possibilités de maximiser l'efficacité de programmes modernes comme le Programme québécois de dépistage du cancer du sein (PQDCS) ?

Une certaine incertitude entoure encore l'efficacité de la mammographie de dépistage, mais des réductions significatives de la mortalité par cancer du sein ont été notées, en particulier chez les femmes de plus de 50 ans. Il se peut que la faible qualité des plans d'étude, les lacunes dans la mise en contraste et une analyse inadéquate des données aient pu compromettre les résultats des essais. La participation sur plusieurs années à un programme moderne comme le PQDCS offre, par rapport à l'absence de dépistage, un contraste à plusieurs égards plus marqué que celui qu'ont obtenu les huit essais sur le dépistage. Il est donc possible que ce programme donne des résultats supérieurs à ceux des essais scientifiques.

Bien que le PQDCS exerce déjà un contrôle rigoureux de la qualité des clichés produits, certains aspects de la structure et des processus des essais qui ont été examinés sous la rubrique de la force du contraste peuvent être transposés en normes de qualité additionnelles. Parmi ces aspects, citons la double lecture des clichés et un volume de lecture annuel suffisant pour permettre à chaque radiologiste d'acquérir et de maintenir l'expertise nécessaire pour détecter les cancers du sein à un stade précoce. Ces deux aspects contribuent à réduire les taux de faux positifs et les taux subséquents d'interventions diagnostiques inutiles.

Un taux de participation élevé aux différents cycles de dépistage permettrait aussi d'atteindre et même de dépasser la réduction de la mortalité obtenue dans les essais.

Conclusion : *Les programmes modernes de dépistage du cancer du sein comme le PQDCS pourront atteindre des résultats comparables, voire supérieurs à ceux des essais si leur fonctionnement assure une qualité de mammographie analogue à celle des essais ou meilleure. La qualité des clichés mammographiques, l'expertise des lecteurs et l'instauration de mesures comme la deuxième lecture d'une certaine proportion des clichés constituent des moyens susceptibles de réduire les taux de faux positifs et d'assurer la qualité du dépistage. Bien qu'il faille viser des taux de participation aussi élevés que possible, les efforts déployés en ce sens ne devraient pas surestimer les avantages de la mammographie ni minimiser les incertitudes et les risques éventuels.*

ANNEXE A

LA MISE EN CONTRASTE DU DÉPISTAGE ET SA DURÉE

INTRODUCTION

L'intervention expérimentale dans les essais sur le dépistage mammographique était le plus souvent constituée de deux à sept cycles de mammographies, séparés par des intervalles de 12 à 28 mois. Les résultats de ces essais sont utiles aux décideurs dans la mesure où ils permettent d'extrapoler à partir de programmes courts (habituellement de 4 à 10 ans de dépistage) l'effet de programmes éventuellement beaucoup plus longs. Les programmes de dépistage actuels sont le plus souvent offerts aux femmes de 50 à 69 ans, et parfois aussi aux femmes plus jeunes (de 40 à 49 ans) et plus âgées (70 ans et plus). Ils peuvent durer 20, 30 ou même 40 ans pour une participante donnée. L'analyse des résultats des essais sur le dépistage doit tenir compte de la durée de l'intervention expérimentale, car les essais plus longs obtiennent habituellement un contraste plus fort et sont, par conséquent, plus pertinents pour la prise de décisions sur les programmes de dépistage. À titre d'exemple extrême, un essai comportant deux cycles de dépistage à un an d'intervalle aurait un contraste faible, et l'effet sur la mortalité par cancer du sein ne serait pas aussi important que celui d'un programme de 20 ou 30 ans. Un essai de 10 ans, adéquatement mené, a de bien meilleures chances de mettre en évidence le potentiel d'un programme de dépistage, et cette différence qualitative doit être prise en compte lorsqu'on examine les effets d'essais de durées substantiellement différentes.

Les essais publiés sur la mammographie de dépistage avaient des durées d'expérimentation variant de trois ou quatre ans (HIP) à huit ou neuf ans (Malmö). Le *UK Age Trial* actuellement en cours prévoit une durée d'expérimentation de 10 à 11 ans. Malheureusement, la durée exacte d'application expérimentale du dépistage n'est pas clairement définie dans les comptes rendus des essais publiés. Le présent rapport tente de définir les concepts nécessaires pour analyser la durée de l'intervention expérimentale de façon à mieux évaluer les effets du facteur temps (choix du moment pour évaluer les effets du dépistage sur la mortalité) sur les résultats escomptés d'un essai sur le dépistage.

DÉFINITION DES CONCEPTS

Les essais cliniques comportent la répartition de sujets dans au moins deux groupes comparables qui seront soumis à des interventions différentes pendant une période donnée. Les neuf essais sur la mammographie décrits dans la littérature publiée ont comparé une cohorte soumise au dépistage à une cohorte témoin. La randomisation est recommandée pour obtenir des cohortes comparables, c'est-à-dire des cohortes composées de personnes possédant un profil similaire de facteurs de risque connus et inconnus du paramètre à l'étude. Dans le cas qui nous intéresse, les deux cohortes doivent avoir les mêmes antécédents de dépistage pour assurer que toute différence dans les résultats ne puisse être attribuée qu'à l'intervention étudiée. Il est important de noter qu'après la période d'expérimentation, les deux groupes devront avoir accès à la même intervention de dépistage pour que l'on puisse attribuer un quelconque écart des résultats à l'intervention de dépistage étudiée. La mise en contraste du dépistage est donc définie comme la création d'une divergence temporaire entre l'intervention (dite expérimentale) offerte à la cohorte de dépistage et celle offerte à la cohorte témoin.

Le début de la mise en contraste est conceptuellement assez clair, puisqu'il s'agit de la première mammographie offerte à une femme de la cohorte de dépistage. La date de

la randomisation est souvent utilisée pour marquer le début de la mise en contraste, le principe de l'intention de traiter stipulant que la mise en contraste s'étend sur toute la durée de l'étude, sans égard à la participation ou non d'un sujet à tous les cycles de dépistage. Il faut noter qu'une mammographie récente devrait constituer un critère d'exclusion pour l'entrée immédiate dans l'étude, la mammographie n'étant plus nécessaire d'un point de vue clinique. Dans un tel cas, il n'y aurait pas de différence (contraste) entre la femme du groupe de dépistage et son homologue du groupe témoin, aussi soumise à un dépistage récent, en principe, et le début de la mise en contraste serait ambigu. Cependant, aucun des essais examinés ne semble avoir employé un tel critère d'exclusion.

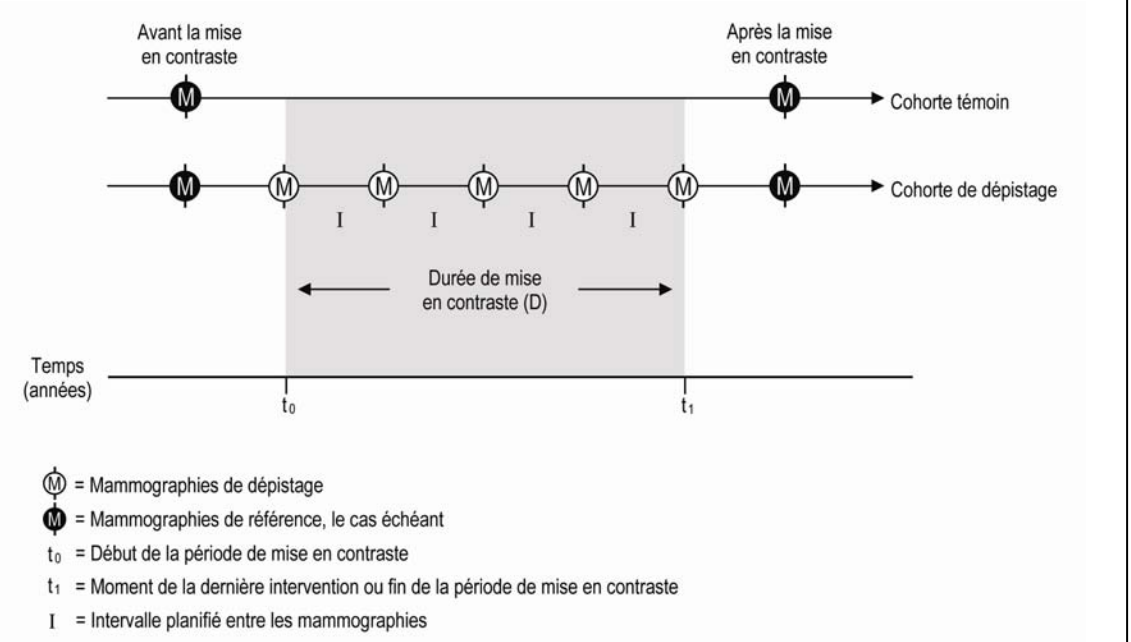
À partir du début de la mise en contraste, une femme de la cohorte de dépistage se verra offrir une série de **n** mammographies de dépistage, séparées par n - 1 intervalles de durée **I**. La mise en contraste se terminera le jour où son homologue de la cohorte témoin commencera le même programme de dépistage, c'est-à-dire au point de convergence des expériences de dépistage. La durée de la mise en contraste du dépistage (D) fera référence à la période totale de mise en contraste.

DURÉE

La durée **D** est la période de mise en contraste du dépistage. Elle sera, en principe, égale au nombre d'intervalles entre deux mammographies à l'intérieur de la période de mise en contraste, multiplié par la durée moyenne de ces intervalles. Cependant, la définition de la fin de la mise en contraste est un peu plus subtile, car elle dépend du protocole de dépistage de la cohorte témoin. On distingue trois cas, selon que la fin de la mise en contraste du dépistage se fasse de façon passive, active et en même temps que le cycle final, ou active et subséquente au cycle final.

FIGURE A-1

Modèle de base de la mise en contraste du dépistage



Dans le premier cas, le dernier cycle de dépistage **n** est offert aux femmes du groupe de dépistage et non à celles du groupe témoin, et aucun dépistage n'est offert par la suite ni à l'un, ni à l'autre groupe. Ce fut le cas, par exemple, dans l'essai HIP. On présume que, après la mise en contraste, les femmes des groupes de dépistage et témoin auront des mammographies à la même fréquence, mais que les femmes du groupe de dépistage, qui viennent d'avoir une mammographie, n'en auront généralement pas d'autre avant la fin d'un autre intervalle **I**. Les femmes du groupe témoin, pour leur part, n'ont pas cette contrainte, et elles auront leur mammographie à n'importe quel moment à l'intérieur de l'intervalle, en moyenne, à la moitié. Comme le premier cycle de dépistage se trouve au temps zéro et le énième cycle à (n - 1)•I, la durée totale D de mise en contraste du dépistage sera donc (n - 1)•I + (1/2)•I = (n - 1/2)•I.

Dans le deuxième cas, la mise en contraste du dépistage peut être terminée activement par l'offre systématique de dépistage à toutes les femmes du groupe témoin, comme ce fut le cas dans tous les essais suédois, sauf celui de Malmö. La mise en contraste se terminera alors un peu plus tôt; la durée sera (n - 1)•I si l'examen de dépistage est offert au groupe témoin au moment du dernier cycle du groupe de dépistage.

Dans le dernier cas, la mise en contraste du dépistage peut être terminée activement par l'offre systématique de dépistage à toutes les femmes du groupe témoin, mais seulement à la fin du énième cycle offert au groupe de dépistage. Dans ce cas, le groupe témoin recevrait son examen en même temps que le $(n + 1)^{e}$ examen du groupe de dépistage. La durée de mise en contraste du dépistage serait ainsi n•I. Cette stratégie n'a été utilisée par aucun des huit essais publiés sur la mammographie.

Trois autres éléments sont pertinents à l'évaluation de la durée de mise en contraste du dépistage dans les essais. Premièrement, c'est le calendrier de l'étude, et non le calendrier civil, qui est en jeu. La durée de l'étude se compte à partir du temps zéro, qui est en principe le moment où une femme du groupe de dépistage est invitée à subir une série de mammographies, alors que son homologue du groupe témoin ne l'est pas. Bien que les femmes puissent joindre l'essai à différentes dates du calendrier, elles commencent toutes leur mise en contraste au temps zéro de l'étude. La durée de mise en contraste du dépistage est évidemment la durée de l'étude entre le temps zéro et la fin de la mise en contraste, et non pas la durée au calendrier entre la première randomisation d'une femme et le dernier examen effectué dans le cadre de l'étude. En d'autres mots, une étude peut s'étendre sur une période substantiellement plus longue que la période au cours de laquelle une participante donnée est soumise à la mise en contraste du dépistage si toutes les participantes ne commencent pas la mise en contraste au début de l'étude.

Deuxièmement, il est important de noter que souvent, la durée de la mise en contraste n'est pas la même pour toutes les participantes. En effet, dans plusieurs essais, de nouveaux sujets se sont joints jusque tard dans l'étude, parfois si tard que l'étude était terminée avant que les dernières participantes aient pu subir tous les cycles de dépistage auxquels avaient eu droit les premières participantes. Ces cas sont habituellement consignés, et pour ceux-ci, la durée moyenne de mise en contraste, pondérée par le nombre de femmes pour chaque nombre de cycles, a été calculée.

Enfin, c'est la longueur prévue des intervalles entre chaque cycle de dépistage, et non pas l'intervalle moyen entre deux mammographies effectivement subies, qui a été utilisée ici. Bien qu'un intervalle unique ou variable soit toujours proposé dans le protocole de l'essai, les intervalles observés entre les examens réels de dépistage peuvent être soit plus courts (si cela est permis) soit, le plus souvent, plus longs, puisque les participantes jouissent d'une assez grande liberté et peuvent subir les examens à leur convenance. Tout délai entre les examens ou tout examen sauté devraient transparaître dans les taux de participation. Si l'on utilisait

la moyenne des intervalles réels entre les cycles de dépistage, ce phénomène serait compté en double.

CONCLUSION

Afin d'apprécier adéquatement la durée de mise en contraste du dépistage telle qu'elle a été conçue et appliquée dans un essai, il est important de définir les notions de début et de fin de la période de mise en contraste et d'obtenir les informations relatives au nombre d'intervalles, à la durée de chaque intervalle et à la répartition des femmes selon les différentes durées de mise en contraste. Heureusement, les huit essais publiés fournissent cette information. La méthode choisie par les chercheurs pour terminer la mise en contraste du dépistage peut modifier la durée totale de la mise en contraste par une période aussi longue qu'un intervalle I. On a formulé l'hypothèse que les taux de dépistage dans les groupes témoin et expérimental étaient identiques avant et après la mise en contraste. Pour assurer la validité et la justesse de l'analyse, cette information devrait figurer dans les publications des chercheurs, mais cela n'a pas été fait dans les articles publiés jusqu'ici.

En ce qui concerne le taux de dépistage préalable à la mise en contraste de l'étude, on peut présumer que les taux de mammographies chez les femmes plus jeunes étaient bas, mais ils devraient être similaires dans les groupes de dépistage et témoin si la randomisation a été faite adéquatement. On peut aussi présumer que les femmes qui avaient subi un examen de dépistage immédiatement avant le début de l'essai en étaient exclues, quoique cette information n'ait pas été fournie non plus. Si cette exclusion n'a pas été faite, la durée réelle de mise en contraste en serait légèrement réduite.

Quant au dépistage subséquent à la mise en contraste, il se peut que les taux de mammographies n'aient pas été similaires, par exemple si les femmes du groupe de dépistage avaient pris l'habitude de subir des mammographies. Cela aurait tendance à augmenter la durée réelle de mise en contraste. Comme cette information n'est pas disponible, et n'aurait probablement pas d'effet majeur de toute façon, les taux de dépistage ont été considérés comme similaires avant et après la mise en contraste du dépistage.

ANNEXE B

FACTEUR TEMPS : CHOIX DU MOMENT POUR ÉVALUER LES EFFETS DU DÉPISTAGE SUR LA MORTALITÉ PAR CANCER DU SEIN

INTRODUCTION

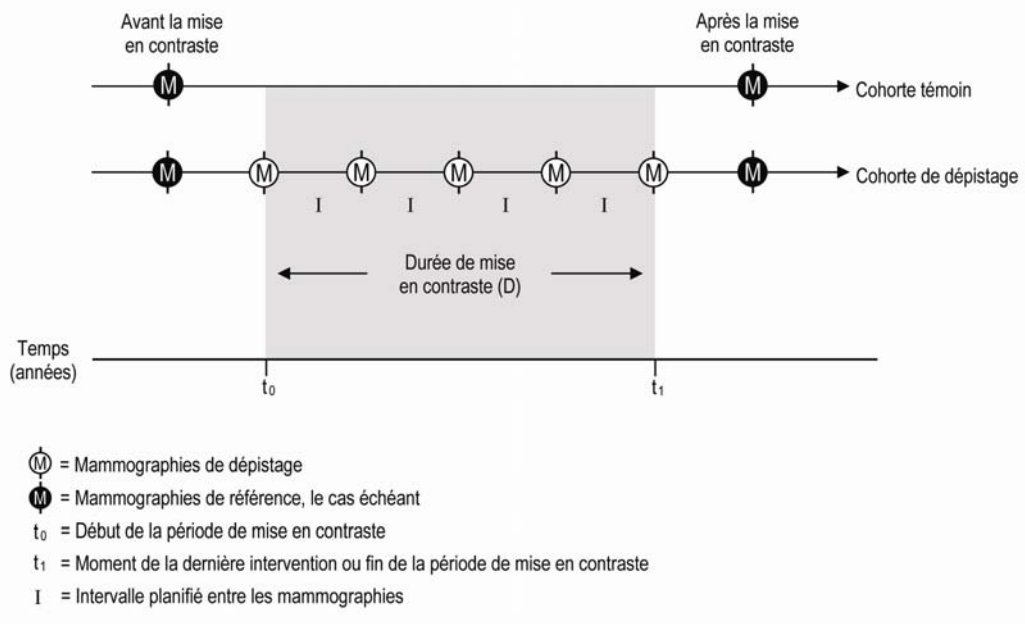
Le contraste a été défini comme la divergence temporaire entre la cohorte de dépistage et la cohorte témoin dans le contexte d'un dépistage (voir l'annexe A). Une randomisation réussie implique que les deux cohortes auraient eu la même expérience de dépistage avant le début de la mise en contraste. À partir du début de la mise en contraste, une femme de la cohorte de dépistage se verra offrir une série de **n** mammographies, séparées par l'intervalle **I**. La mise en contraste se terminera le jour où son homologue de la cohorte témoin commencera un programme de dépistage similaire au sien, c'est-à-dire au point de convergence des interventions de dépistage. La durée de mise en contraste du dépistage (D) fera référence à la période totale pendant laquelle le dépistage a été mis en contraste avec l'intervention témoin (voir la figure B-1).

Modèle de base

Le dépistage du cancer du sein ne sera efficace que s'il permet de découvrir un nombre significatif de tumeurs à la fois mortelles lorsqu'elles sont détectées au stade symptomatique, et curables si le diagnostic et le traitement sont précoces.

FIGURE B-1

Modèle de base de la mise en contraste du dépistage



Dans ce cas, le contraste type entre deux séries d'événements hypothétiques serait le suivant :

1) Groupe témoin : une tumeur potentiellement mortelle commence à grossir, devient symptomatique, le diagnostic est établi, la tumeur peut être traitée, mais la mort survient néanmoins un certain nombre d'années plus tard.
2) Groupe de dépistage : une tumeur potentiellement mortelle se met à grossir, devient dépistable avant même qu'elle soit symptomatique, elle est détectée au cycle de mammographies suivant, le diagnostic et le traitement se font à ce stade précoce, et la mort est évitée.

Le **décalage L** (pour *lag-time*) est défini comme le temps écoulé entre le moment où un cas potentiellement mortel pourrait être détecté par dépistage et, en l'absence de dépistage, la mort par cancer du sein. Cela ne peut être observé directement dans un essai sur le dépistage qui soit acceptable d'un point de vue éthique, les cas diagnostiqués précocement devant toujours être traités. Il est cependant possible de déduire ce temps à partir d'un essai contrôlé randomisé, car on peut présumer que les cancers seront les mêmes dans les groupes soumis et non soumis au dépistage, et que les différences dans les taux de mortalité sont attribuables aux interventions thérapeutiques rendues possibles grâce au dépistage. Si le décalage L entre la détectabilité par dépistage et la mort était connu et s'il était le même pour chacun des cas, on pourrait mesurer la mortalité dans la période située entre les années L après le début du dépistage et les années L après la fin du dépistage. Selon Miettinen et ses collaborateurs [2002], pour atteindre la pleine réduction de la mortalité, un essai devrait s'étendre sur une période au moins aussi longue que ce décalage. L'inclusion des données sur les décès préalables à l'année L ou subséquents à l'année D + L aurait des effets de dilution, car certains des décès survenus à l'intérieur de ce laps de temps n'auraient pas eu de lien avec le dépistage[10].

Par exemple, si le décalage était toujours L = 5 ans avec un dépistage s'étendant sur D = 10 ans, le profil temporel de la mortalité suivrait le schéma présenté à la figure B-2.

Par ailleurs, si le décalage était toujours de 10 ans, la réduction du taux de mortalité serait déplacée progressivement de cinq ans dans l'avenir, comme l'illustre la figure B-3.

En fait, le cancer du sein est un ensemble de plusieurs maladies qui s'attaquent au même organe mais proviennent de structures différentes, dans des circonstances hormonales, immunologiques et génétiques singulières, qui se développent chez des personnes différentes et peuvent être détectées et traitées selon différentes modalités. Le décalage entre la détection d'un cancer par dépistage et le moment où celui-ci devient mortel variera donc considérablement. Si L_{min} est défini comme le décalage le plus court, et L_{max} le plus long, les différents décalages devraient varier continuellement de L_{min} à L_{max}. On peut donc supposer que toute réduction dans les taux de mortalité ne se produira que L_{min} années après le début du dépistage, et que la pleine réduction ne sera pas atteinte avant L_{max} années après le début du dépistage.

Lorsque le dépistage prend fin, on continue à noter une réduction du taux de mortalité pour un certain temps, les nouveaux décès résultant des cas décelés pendant la période de dépistage. Après L_{min} années, cependant, les décès commencent à se produire chez les cas dont le décalage est L_{min}. Comme ces cas sont apparus ou ont été détectés après la période de mise en contraste D, le taux de létalité entre les cohortes de dépistage et témoin devrait être le même. Dès L_{max} années après le dépistage, tous les nouveaux décès se produiront chez les cas apparus après la fin de la période de dépistage D, et les taux de

10. Bien que le dépistage puisse procurer certains avantages en détectant des cancers plus avancés, cet effet est probablement négligeable.

mortalité chez les personnes soumises et non soumises au dépistage devraient être identiques. Encore une fois, ce retour aux taux normaux ne sera pas soudain, mais graduel. Si $L_{min} = 5$ et $L_{max} = 10$, la période de réduction maximale de la mortalité aura le profil illustré à la figure B-4.

Selon la durée de la période de dépistage et la durée du suivi, le temps de suivi peut donc être divisé logiquement en cinq périodes :

1. Les premières L_{min} années, avec peu ou pas de différence dans les taux de mortalité;
2. Les $L_{max} - L_{min}$ années suivantes, avec l'effet du dépistage qui augmente graduellement;
3. Les $D - L_{max} + L_{min}$ années suivantes, avec le plein effet;
4. Les $L_{max} - L_{min}$ années suivantes, avec l'effet qui diminue graduellement;
5. Les années subséquentes à $D + L_{max}$, avec peu ou pas d'effet.

FIGURE B-2

Modèle d'un décalage de 5 ans

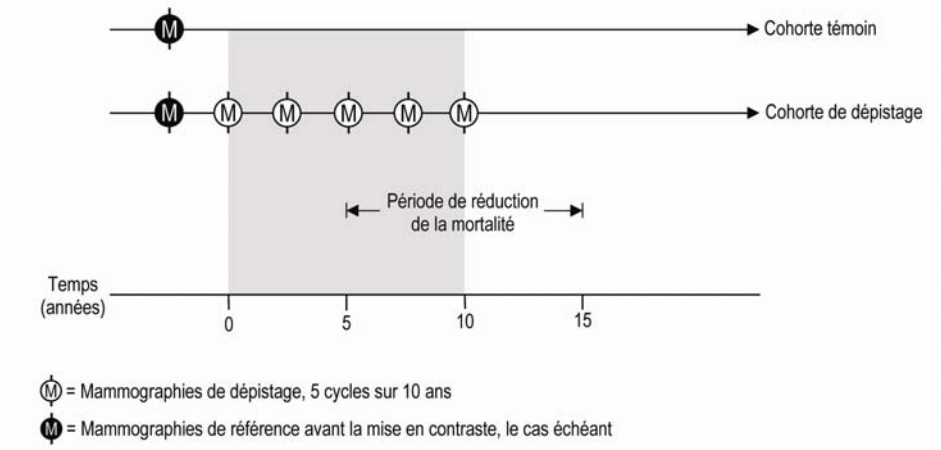


FIGURE B-3

Modèle d'un décalage de 10 ans

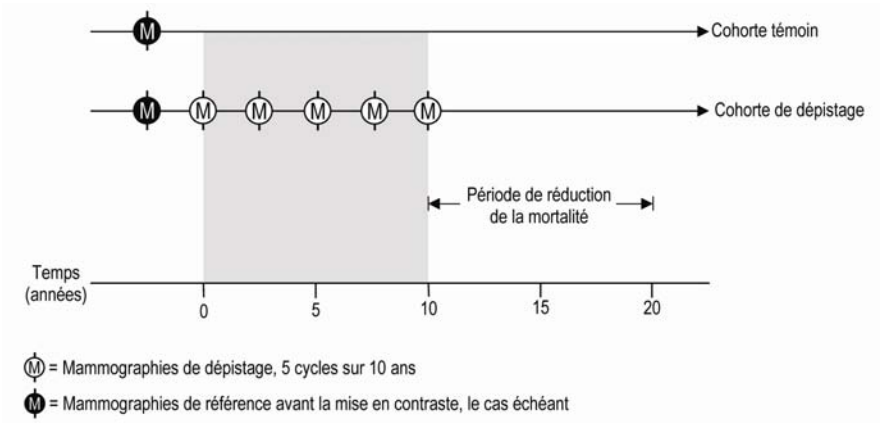
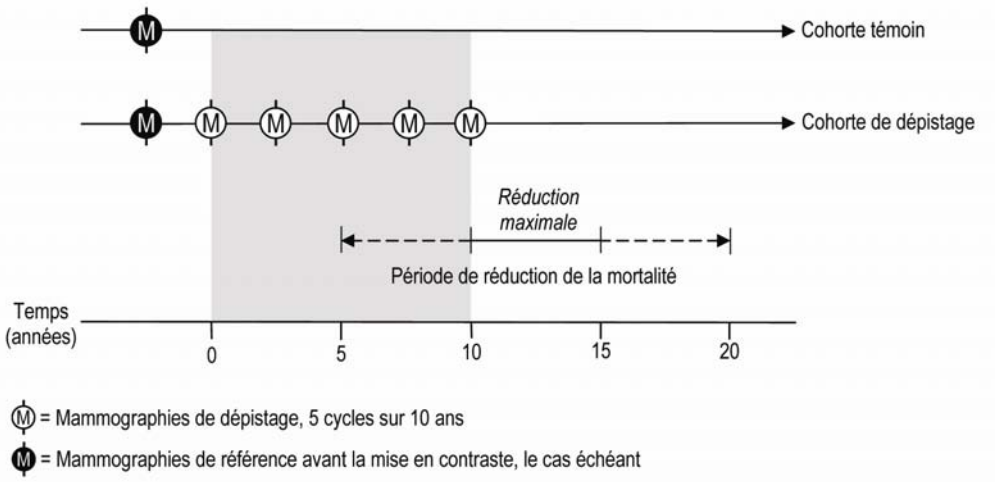


FIGURE B-4

Profil de mortalité, décalage variant continuellement entre 5 et 10 ans



À la lumière des analyses précédentes, il est évident que la pleine réduction ne sera pas atteinte avant L_{max} années après le début du dépistage, même si on peut observer une certaine réduction des taux de mortalité entre les années L_{min} et L_{max}. C'est ce que le rapport précédent du CETS [1993] appelait le rythme de croisière, c'est-à-dire le programme en régime établi *(steady-state)*. De plus, la pleine réduction des taux de mortalité (en régime établi) prendra fin dans $D + L_{min}$ années, soit L_{min} années après la fin du dépistage. Comme le régime établi se situe entre $D + L_{min}$ et L_{max}, la durée de cette période peut être calculée simplement par $D + L_{min} - L_{max}$.

On peut faire une approximation assez juste de D à partir de la description du protocole de recherche, mais L_{max} et L_{min} sont des caractéristiques propres au cancer du sein dans la population à qui le dépistage a été offert. On peut en faire l'estimation en notant le seuil de diminution des taux de mortalité relatifs chez les femmes soumises au dépistage (L_{min}) et la différence entre les deux cohortes à son point le plus bas (L_{max}).

ESTIMATION DU DÉCALAGE

Un examen des courbes de survie au cancer du sein en l'absence de dépistage indique que des décès ont lieu dès la première année et continuent jusqu'à 20 ans après le diagnostic initial. Cependant, comme le décalage L fait référence à l'intervalle entre la détection par mammographie et la mort par maladie évitable par dépistage, la source d'information pertinente pour l'estimation de L doit provenir d'études qui comparent les cas détectés par mammographie et leur équivalent dans un groupe témoin, c'est-à-dire des données provenant des essais, car on ne connaît pas l'évolution naturelle de la maladie.

Un certain nombre d'études fournissent des informations utiles pour l'estimation de L_{min} et de L_{max}. Des données provenant de l'étude HIP montrent une divergence dans les taux de décès par cancer du sein à partir de la quatrième année; l'écart entre les courbes se creuse jusqu'à la sixième ou septième année, puis demeure relativement constant [Shapiro *et al.*, 1982, p. 351]. Cela correspondrait à $L_{min} = 4$ et $L_{max} = 6$ ou 7. Fait intéressant, l'effet favorable du dépistage est apparu plus tard chez les femmes âgées de 40 à 49 ans à l'entrée que chez les femmes plus âgées

[Shapiro, 1997, p. 30]. Un résultat similaire a été observé dans l'étude de Malmö, où la divergence dans les taux de mortalité par cancer du sein chez les femmes âgées de plus de 50 ans à leur entrée dans l'étude (MMST-II) a été observée dès la deuxième année, et où l'écart entre les taux s'est stabilisé vers la septième année. Par ailleurs, pour les femmes de moins de 50 ans à l'entrée (MMST-I), les taux n'ont commencé à diverger que vers la sixième année, et l'écart était constant vers la onzième année. Les chercheurs de l'essai *Two-County* constatent que « les courbes de mortalité se séparent à compter de la cinquième ou sixième année après la randomisation » [Tabár *et al.*, 1999], ce qui correspondrait à L_{min} = 5 ou 6. L'étude de Stockholm [Frisell *et al.*, 1997, figure 1] montre des courbes de mortalité pratiquement identiques jusqu'à cinq ans après la randomisation, qui divergent par la suite jusqu'à la dixième ou onzième année, donc que L_{min} = 5 et L_{max} = 10 ou 11. Dans la plus récente revue des essais suédois, la divergence dans les courbes de mortalité se produit vers la cinquième ou la sixième année, et l'effet absolu s'accentue jusqu'à 12 ans après la randomisation, puis est maintenu par la suite [Larsson *et al.*, 1997]. Cela correspondrait à L_{min} = 5 et L_{max} = 12. Dans ce cas, même si la divergence dans les courbes de mortalité est moins importante chez les femmes plus jeunes, il n'y a pas de tendance claire indiquant si les décalages précédant la divergence se situent plus tôt ou plus tard que chez les femmes plus âgées.

Parce que le nombre de décès enregistré dans chaque étude était assez petit (moins de 10 décès par année dans chaque groupe, dans certains cas), l'estimation du décalage ne devrait pas se fonder entièrement sur une seule étude.

Il est difficile d'établir avec précision le décalage minimum, mais il se situe probablement entre 2 et 6 ans, et L_{max} se trouve probablement entre 6 et 12 ans. Cette imprécision est toutefois préférable à l'approche traditionnelle, qui ne tient pas compte du décalage et relève tous les cas de mortalité apparus depuis la randomisation, ce qui suppose implicitement que L_{min} est 0. De plus, l'approche traditionnelle, où tous les cas de décès apparus pendant le suivi sont comptabilisés, suppose implicitement que L_{max} est indéfini dans le temps. En fait, en accordant la même importance à chaque période, de la randomisation à la fin du suivi, le modèle traditionnel accorde la même importance à chaque année de suivi, ce qui suppose implicitement que les réductions de la mortalité dans les années 1, 5, 10 et même 15 sont de même magnitude. Selon ce modèle traditionnel, la dilution est inévitable, à moins que ces suppositions extrêmes et improbables correspondent à la réalité. Il est plus réaliste d'estimer L_{min} et L_{max} à partir des meilleures données disponibles plutôt que de présumer que L_{min} = 0 et que L_{max} est indéfiniment long. Pour les besoins de l'analyse qui suit, on présumera que L_{min} = 5 et L_{max} = 10.

INDICE DE DILUTION TEMPORELLE

À la lumière de ces considérations, il est possible d'estimer dans quelle mesure l'analyse des résultats de chaque étude a dilué ces derniers. La présente analyse exige de nombreux postulats sur la répartition des effets du dépistage ainsi que l'estimation de la durée de chaque mise en contraste, le nombre d'années de suivi dans le dernier rapport publié de chaque essai, et la technique d'analyse utilisée pour calculer les résultats. Ces aspects sont décrits ci-dessous.

Le décalage entre le dépistage et ses effets sur la mortalité constitue un postulat important pour estimer les effets de dilution. On postule donc que L_{min} = 5 et L_{max} = 10, et que les cas sont également répartis entre ces deux extrêmes. Un certain nombre de considérations sur la durée de la mise en contraste de l'étude ont été exposées à l'annexe A. Elles seront prises en compte pour chaque étude pour l'estimation de sa durée de mise en contraste. Des nombreuses publications des résultats de chaque étude, seules seront retenues celles qui ont indiqué les suivis les plus longs.

La proportion des effets sur la mortalité pouvant être attribuée au dépistage sera

calculée pour chaque année de suivi. Ainsi, pour les années de suivi préalables au décalage L, aucun effet n'est prévu, de même que pour les années de suivi postérieures à D + L. Le calcul est un peu compliqué en raison de la variation de L entre L_{min} = 5 et L_{max} = 10. Enfin, pour calculer la proportion des décès qui se sont produits au cours des années où ils auraient pu être évités, il faut postuler de combien d'années le dépistage a accéléré le diagnostic de cancer du sein. Cette période a été estimée à deux ans. En plus de l'analyse pour chacune des années, la dilution cumulative qui se serait produite du temps zéro à l'année en question est calculée pour chaque année de suivi.

Deux principales méthodes d'analyse des données ont été employées par les chercheurs : la première, rudimentaire, tient compte de tous les cas de mortalité dès le temps zéro, ce qui produit inévitablement une dilution temporelle considérable. On peut raffiner partiellement cette méthode en laissant tomber les données des cinq premières années du suivi lorsque les études présentent les cas de mortalité selon l'année. La deuxième méthode d'analyse, adoptée dans les trois études suédoises postérieures à celle de Malmö, ne tient compte que des décès des cas diagnostiqués au cours de la période de mise en contraste de l'étude. La dilution temporelle est moindre, particulièrement après la mise en contraste du dépistage, bien qu'une dilution considérable demeure à cause du diagnostic de cas de cancer avancés pendant la mise en contraste du dépistage. Les tableaux B-1 et B-2 illustrent cette analyse de la dilution temporelle.

Paramètres généraux du calcul	
Décalage temporel (L)	
L_{min}	5 ans
L_{max}	10 ans
Devancement du diagnostic (A)	2 ans

TABLEAU B-1

Paramètres particuliers des essais

ESSAIS	DURÉE DE LA MISE EN CONTRASTE (D) EN ANNÉES	DURÉE DU SUIVI (EN ANNÉES)	MÉTHODE D'ANALYSE*	RÉSULTAT DU CALCUL DE LA DILUTION
HIP	3,5	18	R	0,28
Malmö (MSST)	8,8	11	R	0,64
TCS	6,5	20	S	0,76
Édimbourg	6,4	14	R	0,67
NBSS-1	4,6	11-16	R	0,56
NBSS-2	4,6	13	R	0,56
Stockholm	4,6	11,4	S	0,65
Göteborg	**6,4**	**14**	**S**	**0,73**
UK Age Trial	11,0	13	R?	0,73

* R : analyse rudimentaire, inclusion de tous les cas de mortalité; S : analyse suédoise (modèle d'évaluation), qui n'inclut que la mortalité des cas diagnostiqués pendant la période de dépistage.

TABLEAU B-2

Calcul de la dilution temporelle (exemple de l'essai de Göteborg)

	ANNÉES DE SUIVI PERTINENTES							ANNÉES DE SUIVI RÉELLEMENT COMPTÉES							INDICE DE DILUTION	
	DÉCALAGE TEMPOREL (L)*							DÉCALAGE TEMPOREL MOINS DEVANCEMENT DU DIAGNOSTIC (D - A)							PAR ANNÉE DE SUIVI†	CUMULATIF‡
Année de suivi	5	6	7	8	9	10	Total	3	4	5	6	7	8	Total		
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	1	1	0	0	0	0	2	0	0
6	1	0	0	0	0	0	1	1	1	1	0	0	0	3	0,33	0,17
7	1	1	0	0	0	0	2	1	1	1	1	0	0	4	0,50	0,30
8	1	1	1	0	0	0	3	1	1	1	1	1	0	5	0,60	0,40
9	1	1	1	1	0	0	4	1	1	1	1	1	1	6	0,67	0,48
10	1	1	1	1	1	0	5	1	1	1	1	1	1	6	0,83	0,56
11	1	1	1	1	1	1	6	1	1	1	1	1	1	6	1,00	0,64
12	0	1	1	1	1	1	5	1	1	1	1	1	1	6	0,83	0,68
13	0	0	1	1	1	1	4	1	1	1	1	1	1	6	0,67	0,71
14	**0**	**0**	**0**	**1**	**1**	**1**	**3**	**1**	**1**	**1**	**1**	**1**	**1**	**6**	**0,50**	**0,73**
15	0	0	0	0	1	1	2	1	1	1	1	1	1	6	0,33	0,74
16	0	0	0	0	0	1	1	1	1	1	1	1	1	6	0,17	0,75
17	0	0	0	0	0	0	0	1	1	1	1	1	1	6	0	0,75
18	0	0	0	0	0	0	0	1	1	1	1	1	1	6	0	0,75
19	0	0	0	0	0	0	0	1	1	1	1	1	1	6	0	0,75
20	0	0	0	0	0	0	0	1	1	1	1	1	1	6	0	0,75

* 0 : dilution totale; 1 : aucune dilution.
† L'indice annuel correspond au quotient du total (L) divisé par le total (D - A) pour chaque année de suivi.
‡ L'indice cumulatif correspond au quotient du total (L) cumulatif divisé par le total (D - A) cumulatif; par exemple, à l'année de suivi 6, on divise (0 + 0 + 0 + 0 + 0 + 1) par (0 + 0 + 0 + 1 + 2 + 3), ce qui donne 0,17.

Remarque : les chiffres de cet exemple proviennent de l'essai de Göteborg.

ANNEXE C

ÉCHELLE DE LA FORCE DU CONTRASTE

INTRODUCTION

La mammographie de dépistage ne peut être efficace qu'à un niveau au moins minimal de qualité, avec du matériel adéquat et un personnel formé. Elle doit être offerte à des intervalles appropriés et obtenir la plus grande participation possible des femmes à qui elle a été offerte. Trois conditions supplémentaires s'appliquent au contexte d'un essai sur le dépistage : la population témoin ne doit pas subir de mammographie (contamination) et ne doit pas être soumise à un autre type de dépistage, et les résultats doivent être calculés par une analyse des données sur la mortalité relevées au cours d'une période de suivi adéquate. Dans le présent rapport, ces aspects ont été regroupés en cinq catégories : le contraste technique, l'époque, la qualité du dépistage, les taux de participation et de contamination, et le facteur temps. Ces concepts seront davantage développés afin d'expliquer les éléments d'une échelle d'évaluation mise au point pour mesurer la force du contraste de chacun des neuf essais sur la mammographie de dépistage. Enfin, les résultats de l'application de cette échelle aux neuf essais seront présentés.

Contraste technique (C)

Les essais parus sur la mammographie comportent trois formes de mise en contraste différentes du point de vue qualitatif, qui ont été notées en fonction de l'intensité de leur contraste. Les trois niveaux sont :

1) examens mammographique et clinique *versus* aucun dépistage (contraste le plus fort);
2) mammographie seule *versus* aucun dépistage (contraste moyen);
3) dépistage mammographique *versus* dépistage par examen clinique (contraste faible).

Époque (E)

Le dépistage mammographique a été introduit dans les années 1950, et le premier appareil spécialisé pour la mammographie date de 1966. Depuis, la technologie n'a cessé d'évoluer. Certains experts sont d'avis que les progrès qualitatifs les plus importants ont eu lieu au début des années 1980 [Sickles, 1997]. On considère quatre catégories : 1) avant 1970; 2) 1970-1979; 3) 1980-1989; 4) depuis 1990.

Qualité (Q)

Pour évaluer la qualité du contraste technique, quatre composantes peuvent être dégagées à partir des renseignements publiés par les chercheurs. Ce sont :

1) les mesures de contrôle de la qualité mises en place par les chercheurs (comprenant l'agrément des centres de dépistage et une deuxième lecture des clichés présentant des anomalies);
2) l'expertise des lecteurs de mammographies selon leur volume de lecture (expertise définie par l'interprétation de plus de 3 000 clichés par année);
3) le nombre d'incidences par mammographie de dépistage, soit une, deux, ou une association d'une et de deux incidences, par exemple lorsque les clichés initiaux se font sous deux incidences et les clichés subséquents sous une seule quand le cliché est jugé adéquat à l'examen initial;
4) la longueur de l'intervalle entre les cycles (trois catégories : moins de 18 mois; entre 18 mois et moins de 24 mois; et 24 mois et plus).

Le score de qualité a été calculé selon l'échelle présentée au tableau C-1.

Participation et contamination (P)

La participation et la contamination s'expriment naturellement en pourcentages, et leur différence, aussi exprimée en pourcentage, constitue une estimation raisonnable du pourcentage de contraste entre les deux cohortes. Par exemple, si la moitié de la cohorte de dépistage y participait et que la moitié de la cohorte témoin avait aussi une mammographie de dépistage à l'extérieur de l'essai, on pourrait dire qu'il n'y a eu aucun contraste (50 % - 50 % = 0 %). Un contraste de 100 % nécessiterait la pleine participation de toutes les femmes de la cohorte de dépistage, sans aucune contamination dans la cohorte témoin. Pour toute combinaison intermédiaire, la différence produirait un contraste se situant entre 0 et 100 %, selon le cas. Cette approche cadre avec une proposition de Glasziou [1992], qui suggère essentiellement une formule (attribuée à Newcombe) selon laquelle l'estimation de l'intention de traiter peut être ajustée par un facteur $1/\Delta$, où $\Delta = p_1 + p_2$, p_1 et p_2 étant les taux d'observance dans les groupes de dépistage et témoin. Puisque p_2 est le taux d'observance dans le groupe témoin et $1 - p_2$ est le taux de contamination, $p_1 + p_2 - 1 = p_1 - (1 - p_2)$. Le facteur d'ajustement peut être reformulé en $1/(p - c)$, où p est le taux de participation dans le groupe de dépistage et c, le taux de contamination dans la cohorte témoin. Un ajustement similaire pour la non-participation et la contamination a également été appelé estimation « causale » [Baker *et al.*, 2002].

TABLEAU C-1

Échelle de qualité et score attribué

SOUS-COMPOSANTE	SCORE
Contrôle de qualité (double lecture)	
Adéquat	1
Inadéquat	0
Expertise des lecteurs	
Gros volume	1
Petit volume	0
Nombre de clichés	
2 (chaque cycle)	1
2 ou 1 (selon le cycle)	0,5
1 (chaque cycle)	0
Intervalle entre les mammographies	
< 18 mois	1
18 < 24 mois	0,5
> 24 mois	0
SCORE TOTAL	**(entre 0 et 4)**

Facteur temps (T)

L'indice de dilution temporelle représente le degré de réduction de la mortalité escompté, exprimé en proportion de la pleine réduction potentielle en régime établi *(steady-state)* de la mise en contraste du dépistage. Cet indice se situe aussi entre 0 et 1, 0 signifiant qu'aucune réduction potentielle de la mortalité ne serait observable dans la période mesurée (durant les premières années de suivi, par exemple), et 1 signifiant que la pleine réduction serait observée (durant la période située entre L_{max} et D - L_{max} + L_{min}). On peut considérer les valeurs intermédiaires comme une proportion de la réduction potentielle de la mortalité diluée par les décès qui ne peuvent être reliés à la mise en contraste du dépistage en raison du moment où ils se produisent. L'annexe B, qui traite du décalage des effets du dépistage, explique comment cette dilution peut être estimée à partir de postulats sur le décalage entre le dépistage et ses effets sur la mortalité.

COMBINAISON DES SCORES

Si le choix de la façon de noter chacune de ces cinq composantes exige un arbitrage, il ne devrait toutefois pas être arbitraire. Comme l'exercice vise à faire la différence entre les essais, avec un intérêt particulier pour les plus pertinents (ceux dont le contraste ressemble le plus au programme de dépistage éventuel), il est utile d'attribuer un score qui pourrait correspondre à la capacité de chaque essai d'atteindre le potentiel d'efficacité d'un programme de dépistage mammographique en régime établi dans le contexte actuel.

Les taux de participation et de dilution temporelle s'expriment déjà naturellement sous forme de pourcentages. De toute évidence, un taux de participation de 0 % aurait zéro effet, et un taux de participation de 100 % aurait un effet maximal. De même, une dilution temporelle maximale (si l'on utilisait seulement les données des premières années du suivi, par exemple) n'aurait aucun effet, et l'absence de dilution temporelle produirait un effet maximal. Il est donc naturel d'exprimer les trois autres composantes en pourcentages et d'élaborer une évaluation de la force du contraste qui soit le produit des scores des cinq composantes. Sur une échelle additive, toute composante notée 0 ferait baisser le score général de 1 point sur 5, ou de 20 %. Sur une échelle multiplicative, toute composante notée 0 signifierait que le score général serait également 0. Pour maintenir la différenciation entre les scores, il est préférable de ne permettre que la variation du même pourcentage pour chaque composante, de 20 % par exemple.

La valeur 1 est donc attribuée à une composante qui correspond à la mise en contraste proposée dans la question de recherche initiale, c'est-à-dire pour un programme de dépistage mammographique sans examen clinique, avec du matériel moderne, des mesures de contrôle de la qualité et des mammographies sous deux incidences à 24 mois d'intervalle, un taux de participation de 100 %, sans dilution temporelle (effet maximal en régime établi). Les scores obtenus selon cette formule pour les composantes susmentionnées sont présentés au tableau C-2.

TABLEAU C-2

Évaluation de la force du contraste

COMPOSANTES	CRITÈRES	MULTIPLICATEUR
Contraste technique (C)	Mammographie + examen clinique *vs* aucun dépistage	1,25
	Mammographie *vs* aucun dépistage	1
	Mammographie *vs* examen clinique	0,8
	DATE DU DÉBUT DE L'ESSAI	
Époque (E)	Avant 1970	0,7
	1970-1979	0,8
	1980-1989	0,9
	Depuis 1990	1
	TOTAL DU SCORE DE QUALITÉ	
Qualité (Q) (voir le tableau C-1)	0	0,8
	0,5	0,825
	1	0,85
	1,5	0,875
	2	0,9
	2,5	0,925
	3	0,95
	3,5	0,975
	4	1
	DIFFÉRENCE ENTRE LES COHORTES	
Participation - contamination (P)	Taux de participation au dépistage (taux dans l'étude, voir le tableau C-4)	Différence observée
	Taux de contamination de la cohorte témoin (taux dans l'étude, voir le tableau C-4)	
	INDICE DE DILUTION TEMPORELLE	
Facteur temps (T)	Taux calculé selon la méthode présentée aux tableaux B-1 et B-2	Voir le tableau B-1

ÉVALUATION DES ESSAIS

À l'aide de l'échelle et du système de notation présenté au tableau C-2, deux chercheurs (WD et RK) expérimentés dans l'analyse des études épidémiologiques et qui connaissent bien le domaine des essais sur la mammographie ont attribué indépendamment une note à chaque étude. Tous les articles faisant référence aux neuf essais publiés ont été soigneusement passés en revue. Les informations ont été puisées dans les introductions (conditions générales de l'étude, notamment les taux initiaux de mammographies), les sections sur les méthodes de recherche (matériel et techniques de mammographie et mesures de contrôle de la qualité, intervalles de dépistage) et les sections sur les résultats (taux de participation). En règle générale, la note la plus basse était donnée lorsqu'il n'y avait aucune information sur une catégorie donnée. Le taux de participation moyen a été calculé à partir des taux de participation aux différents cycles de dépistage. On a eu recours à l'interpolation lorsque les taux n'étaient consignés que pour les premiers et les derniers cycles. Habituellement, les taux de participation dans la cohorte témoin n'étaient pas fournis, car les chercheurs de la plupart des essais n'avaient

obtenu aucune information sur les femmes non soumises au dépistage, à l'exception de leur âge. Cependant, certains renseignements sur les taux de dépistage dans la population générale étaient souvent disponibles. Dans d'autres cas, les taux de participation (contamination) dans la cohorte témoin étaient estimés à partir des informations générales sur les taux de dépistage dans la population au moment de l'essai et à l'endroit où il avait lieu.

Les notes attribuées par les deux chercheurs ont été comparées, et ils ont résolu les divergences par consensus en retournant aux rapports de recherche originaux. Les tableaux C-3, C-4 et C-5 montrent les résultats de l'évaluation de chaque essai.

TABLEAU C-3

Calcul du score de chaque essai : qualité

ESSAIS	HIP	Malmö	TCS	Édimbourg	NBSS-1	NBSS-2	Stockholm	Göteborg	*UK Age*	**Comparaison PQDCS**
Contrôle de qualité	1	0	0	1	1	1	0	0,5	1	0,5
Expertise des lecteurs	1	0	1	0	0	0	0	0	1	0
Nombre de clichés	1	0,5	0	0,5	1	1	0	0,5	0,5	1
Intervalle entre les mammographies	1	0,5	0	0,5	1	1	0	0,5	1	0,5
Score total de qualité	4	1	1	2	3	3	0	1,5	3,5	2
Qualité (Q)	**1**	**0,85**	**0,85**	**0,90**	**0,95**	**0,95**	**0,80**	**0,88**	**0,98**	**0,90**

TABLEAU C-4

Différence entre les cohortes : participation - contamination

ESSAIS	HIP	Malmö	TCS	Édimbourg	NBSS-1	NBSS-2	Stockholm	Göteborg	*UK Age*	**Comparaison PQDCS**
Participation (groupe de dépistage)	0,54	0,74	0,87	0,527	0,875	0,883	0,805	0,788	0,61	0,70
Contamination (groupe témoin)	0,05	0,10	0,05	0,10	0,125	0,065	0,20	0,15	0,04	0
Différence	**0,49**	**0,64**	**0,82**	**0,43**	**0,75**	**0,82**	**0,61**	**0,64**	**0,57**	**0,70**

TABLEAU C-5

Calcul du score de chaque essai : force du contraste

ESSAIS	HIP (1963)	Malmö (1976)	TCS (1977)	Édimbourg (1979)	NBSS-1 (1980)	NBSS-2 (1980)	Stockholm (1981)	Göteborg (1982)	*UK Age* (1991)	**Comparaison PQDCS**
Contraste technique (C)	1,25	1	1	1,25	1,25	0,80	1	1	1	1
Époque (E)	0,70	0,80	0,80	0,80	0,90	0,90	0,90	0,90	1	1
Qualité (Q)	1	0,85	0,85	0,90	0,95	0,95	0,80	0,88	0,98	0,90
Participation - contamination (P)	0,49	0,61	0,82	0,43	0,75	0,82	0,61	0,64	0,57	0,70
Facteur temps (T)	0,28	0,64	0,76	0,67	0,56	0,56	0,65	0,73	0,73	1
Total (C x E x Q x P x T)	**0,12**	**0,27**	**0,42**	**0,26**	**0,45**	**0,31**	**0,29**	**0,37**	**0,41**	**0,63**

ANNEXE D

ÉCHELLE DE VALIDITÉ

INTRODUCTION

Les principales composantes du plan expérimental des essais contrôlés randomisés sur la mammographie de dépistage nécessaires pour assurer la validité de leurs résultats sont : une randomisation réalisée dans les règles de l'art, l'égalité de départ des groupes de dépistage et témoin, l'égalité dans les exclusions subséquentes dans l'une et l'autre des cohortes, et l'égalité des suivis sur la mortalité par cancer du sein. Ces critères sont les mêmes que ceux que l'on utilise habituellement dans les revues systématiques de la littérature. Par exemple, Jadad et ses collaborateurs [1996] ont proposé une échelle en cinq points portant sur la randomisation, le double insu et l'égalité des perdus de vue par rapport au résultat étudié. De même, Olsen et Gøtzsche [2001], dans leur examen des essais sur le dépistage mammographique, ont accordé des points en fonction de la méthode de randomisation, de la comparabilité de base, des exclusions, de l'évaluation non biaisée des résultats et du taux de dépistage dans la population témoin [Mocharnuk, 2002].

Sur l'échelle de Jadad, largement utilisée, 2 des 5 points possibles sont accordés pour le double insu. Comme aucune étude sur la mammographie n'a été tentée à l'insu, ni des patientes, ni des chercheurs, elles récoltent toutes la note de 3 ou moins sur cette échelle. De même, il est difficile de déterminer l'égalité des abandons et des retraits, car la plupart des essais sur la mammographie sont avares d'informations sur la cohorte témoin, les études canadiennes faisant figure d'exception. La plupart des études sur la mammographie obtiendraient ainsi la note de 0 ou 1 sur 5. L'usage de cette échelle limiterait donc la capacité d'établir des différences entre les essais. Il est cependant utile de retenir qu'aucun essai sur la mammographie ne répond aux plus hautes normes de qualité généralement appliquées aux essais randomisés.

Le principal objectif de la mise à l'insu dans les essais cliniques est de contraindre l'affectation initiale des participantes aux interventions de dépistage ou témoin de façon à les répartir au hasard, c'est-à-dire sans égard à leur risque [Chalmers, 2001]. À vrai dire, certains estiment que l'insu constitue le déterminant le plus important de la qualité d'un essai contrôlé randomisé [Chalmers *et al.*, 1981]. Un essai ouvert a pour conséquence principale de permettre aux patientes et à leur médecin d'opérer une auto-sélection par rapport à la mammographie en fonction, peut-être, de leur perception du risque de cancer du sein ou des bienfaits escomptés de la mammographie. Cela pourrait expliquer les niveaux élevés de non-participation dans le groupe de dépistage et de contamination dans le groupe témoin, deux phénomènes observés dans la plupart des essais. La question du dépistage dans le groupe témoin a été analysée sous l'angle de la force du contraste aux annexes A et C. Les quatre autres volets soulevés ici peuvent aussi fausser les résultats d'une étude mais, contrairement aux biais liés à la participation, il est difficile de juger si ces biais mènent ou non dans la direction de l'hypothèse nulle. Ces problèmes de biais à direction inconnue sont regroupés ici sous la rubrique de la validité (même si les biais systématiques vers l'hypothèse nulle, et principalement la faiblesse du contraste, compromettent aussi la validité). Ces volets de la validité et leur inclusion dans l'outil d'évaluation sont décrits ci-dessous.

Randomisation (R)

Trois niveaux de validité ont été définis (adéquate, faible et inadéquate ou non documentée) et associés à des notes de 1, de 0,5 et de 0 respectivement. Il est généralement admis qu'il est préférable de randomiser des individus plutôt que des groupes ou des collectivités. Si cette dernière méthode est

employée, l'analyse et les intervalles de confiance refléteront cette caractéristique du plan expérimental. De plus, la randomisation implique par définition une répartition aléatoire (tables de nombres aléatoires, pile ou face) plutôt qu'une répartition systématique (selon la date de naissance, par exemple).

Comparabilité de base (B)

Trois niveaux de validité ont été établis (adéquate, faible et inadéquate ou non documentée) et associés à des notes de 1, de 0,5 et de 0, respectivement. Pour évaluer la comparabilité de base, il faut mesurer les caractéristiques des cohortes de dépistage et témoin associées au risque de cancer du sein ou aux avantages potentiels de la mammographie. Dans la plupart des études, on n'a pour ainsi dire obtenu aucune information sur la cohorte témoin, à part la date de naissance (donc l'âge) des femmes qui en faisaient partie. L'égalité de l'âge uniquement est une mesure de comparabilité de base nécessaire mais insuffisante, d'autant plus que la mise à l'insu n'a pas été tentée et que les participantes avaient *de facto* un contrôle considérable sur l'intervention qu'elles ont reçue. Olsen et Gøtzsche [2001] ont noté des écarts dans l'âge moyen allant jusqu'à trois mois entre les cohortes de dépistage et témoin, et ont jugé que la randomisation n'avait pas été faite dans les règles de l'art dans certains essais en se fondant sur ce critère. Leurs détracteurs estiment qu'on ne peut attribuer les écarts importants dans les taux de cancer du sein à ces différences d'âge, et même que ces différences d'âge tendraient à masquer l'efficacité du dépistage [Law *et al.*, 2000, par exemple]. Cependant, cette critique ne tient pas compte du fait que l'âge moyen est la seule preuve du succès ou de l'échec de la randomisation à assurer la comparabilité de base. Si l'âge n'est pas égal entre les groupes, il est peu probable que la randomisation ait pu équilibrer adéquatement d'autres facteurs de risque importants entre les deux cohortes. De même, des écarts importants dans la situation socioéconomique des cohortes de dépistage et témoin indiquent que les groupes n'étaient pas comparables au départ.

Égalité dans les exclusions (E)

Trois niveaux de validité (adéquate, faible et inadéquate ou non documentée) ont été établis et associés à des notes de 1, de 0,5 et de 0, respectivement. Dans les essais, l'exclusion des cas de cancer du sein diagnostiqués avant la mise en contraste du dépistage est particulièrement préoccupante. En raison des visites régulières des femmes de la cohorte de dépistage, les chercheurs avaient plus souvent l'occasion d'obtenir des informations cliniques sur les diagnostics préalables de cancer du sein. Les femmes des cohortes témoins étaient aussi exclues si elles avaient déjà eu un diagnostic de cancer du sein, mais puisque cette information était souvent découverte 20 ans plus tard, il est fort possible que les données cliniques sur les femmes de cette cohorte aient été moins complètes, ce qui pourrait produire un biais important en faveur du dépistage. Par exemple, Olsen et Gøtzsche [2001] ont calculé que dans l'étude HIP, 853 femmes ont été exclues du groupe de dépistage pour cette raison, contre seulement 336 femmes du groupe témoin. D'après eux, puisque la différence entre les groupes n'était que de 44 décès, la différence entre les groupes aurait été encore moindre si les auteurs avaient été capables de repérer tous les cas préalables de cancer du sein dans le groupe témoin.

Suivi (S)

Trois niveaux de validité (adéquate, faible et inadéquate ou non documentée) ont été établis et associés à des notes de 1, de 0,5 et de 0, respectivement. Les facteurs pris en compte pour cette évaluation incluent l'existence d'une base de données nationale sur la mortalité à des fins de suivi et la validation des indicateurs de résultats par autopsie ou révision des cas.

COMBINAISON DES SCORES

Une fois ces dimensions de la validité évaluées, reste le défi de combiner les résultats pour obtenir le score total qui sera utilisé dans l'analyse. Faute de meilleure solution, une simple addition des quatre composantes sera utilisée à cette fin. Cette formule rejoint les méthodes fréquemment utilisées, comme celles de Jadad et ses collaborateurs [1996].

APPLICATION DE L'ÉCHELLE D'ÉVALUATION DE LA VALIDITÉ AUX ESSAIS

Une fois les composantes de l'échelle de validité établies, il faut mesurer la validité de chaque essai. Bien sûr, un certain degré de subjectivité demeure lorsque vient le moment de décider si un essai répond ou non à un critère donné, par exemple, si la répartition aléatoire, la comparabilité de base ou le traitement des exclusions sont « adéquats », ou si les descriptions des méthodes, souvent sommaires, sont suffisamment explicites pour être qualifiées de « documentées ». Malgré sa part de subjectivité, cette analyse est néanmoins préférable à l'approche habituelle, qui consiste à présumer que tous les essais ont la même validité, ou à exclure arbitrairement certains essais sans critères explicites.

À l'aide de l'échelle et du système de notation précités, deux chercheurs (WD et RK) expérimentés dans l'analyse des études épidémiologiques et qui connaissent bien le domaine des essais sur la mammographie ont attribué indépendamment une note à chaque étude. Tous les articles se référant aux neuf essais publiés ont été soigneusement passés en revue. Les informations ont été puisées dans les sections portant sur les méthodes (randomisation) et les résultats (comparabilité de base). En règle générale, la note la plus basse était donnée lorsqu'il n'y avait aucune information sur une catégorie donnée.

Les notes attribuées par les deux chercheurs ont été comparées, et ils ont résolu les divergences par consensus en retournant aux rapports de recherche originaux. Le tableau D-1 présente les résultats de cet exercice de classement.

TABLEAU D-1

Scores de validité attribués aux essais publiés sur la mammographie

CRITÈRES D'ÉVALUATION	ESSAIS								
	HIP	Malmö	TCS	Édimbourg	NBSS-1	NBSS-2	Stockholm	Göteborg	*UK Age*
Randomisation (R)	1	1	0	0	1	1	0,5	0,5	1
Comparabilité de base (B)	0,5	0,5	0	0	1	1	0,5	0	1
Exclusions (E)	0	0,5	0	0	1	1	0,5	0,5	1
Suivi (S)	0	1	0,5	0,5	1	1	0,5	0,5	1
R + B + E + S	1,5	3	0,5	0,5	4	4	2	1,5	4
Selon Olsen et Gøtzsche [2001]	"*Flawed*"	"*Medium*"	"*Poor, likely flawed*"	"*Flawed*"	"*Medium*"	"*Medium*"	"*Poor*"	"*Poor*"	
Selon le *US Preventive Services Task Force* [Humphrey *et al.*, 2002]	"*Fair*"	"*Fair*"	"*Fair*"	"*Poor*"	"*Fair +*"	"*Fair +*"	"*Fair*"	"*Fair*"	

1 = adéquate
0,5 = faible
0 = inadéquate ou non documentée

ANNEXE E

ANALYSE DES NEUF ESSAIS

1. NEW YORK (HIP)

Le premier essai en bonne et due forme sur la mammographie, mieux connu sous le nom de l'étude HIP, a été réalisé auprès de clientes d'une compagnie d'assurance maladie privée, le *Health Insurance Plan* (HIP) de New York. Il a été amorcé en décembre 1963, et le recrutement s'est poursuivi jusqu'en juin 1966. Les derniers dépistages ont été effectués en juin 1970, et le suivi des résultats s'est prolongé jusqu'en 1986, année de la fin officielle de l'étude. Des rapports importants sur le plan expérimental, le déroulement et les résultats de l'étude ont été publiés entre 1966 et 1997 [Shapiro, 1997; 1994; 1989; Shapiro *et al.*, 1988; 1985; 1982; Shapiro, 1977; Strax *et al.*, 1973; Fink *et al.*, 1972; Shapiro *et al.*, 1966].

Mise en contraste

L'intervention de dépistage était constituée de trois à quatre mammographies annuelles prises sous deux incidences, avec examen clinique des seins. Différents taux de participation au premier cycle de dépistage ont été communiqués : 64 % [Shapiro *et al.*, 1966], 65 % [Fink *et al.*, 1972] et 67 % [Shapiro *et al.*, 1985]. Trois cycles subséquents à un an d'intervalle étaient prévus. Seules les femmes ayant participé au premier cycle étaient invitées aux cycles subséquents, ce qui a fait baisser les taux de participation dans ces cycles à 52 %, 48 % et 45 %, pour une moyenne de 52 %. Les femmes du groupe témoin n'ont rien changé à leur façon d'obtenir leurs soins médicaux [Shapiro, 1989]. Les taux de mammographies ou d'examens des seins n'ont pas été mesurés dans ce groupe, mais les auteurs affirment que le recours à la mammographie de dépistage dans le groupe témoin était faible, car à l'époque, cette intervention n'était pas couverte par la compagnie d'assurances et la mammographie ne faisait pas partie des soins de santé usuels dans les années 1960.

Conscients de l'effet de dilution noté dans les suivis successifs, les chercheurs de l'essai HIP offrent, dans leur plus récent rapport, des analyses restreintes aux cas diagnostiqués dans les cinq ou sept années suivant le dépistage. Cela élimine une grande part de l'effet de dilution des années les plus éloignées du dépistage, où les cas diagnostiqués après la période de mise en contraste évoluent jusqu'à la mort. Cependant, à cause des cancers diagnostiqués pendant la période de mise en contraste mais trop avancés pour que la mammographie ait pu avoir un quelconque effet préventif, les effets de dilution des premières années demeurent.

L'essai a obtenu 12 % sur l'échelle de la force du contraste, le score le plus bas des essais publiés. Des appareils de mammographie vétustes, antérieurs à l'introduction des appareils de mammographie spécialisés, la courte durée de mise en contraste, les effets de dilution temporelle et les faibles taux de participation ont contribué à ce faible contraste.

Validité

La randomisation était basée sur un appariement selon l'âge, la taille de la famille assurée et la catégorie d'emploi [Shapiro *et al.*, 1966]. Les méthodes employées pour définir la séquence des appariements ne sont pas claires : les chercheurs décrivent une méthode selon laquelle, à l'intérieur de chaque clinique médicale, « deux échantillons randomisés systématiquement » étaient sélectionnés, et « chaque énième femme était placée dans le groupe de dépistage, et la femme appariée (n + 1), dans le groupe témoin ». Les paires de patientes d'une

clinique médicale soumises au dépistage et témoins étaient choisies au hasard, et les femmes soumises au dépistage étaient sélectionnées dans une liste de façon séquentielle [Shapiro *et al.*, 1985]. L'exclusion postrandomisation la plus importante concernait les femmes ayant déjà été victimes d'un cancer du sein, mais cette exclusion n'a pas été appliquée uniformément, car ce critère était plus rigoureusement vérifié chez les femmes soumises au dépistage [Shapiro *et al.*, 1985]. Olsen et Gøtzsche [2001] ont calculé que 853 femmes ont été exclues du groupe de dépistage pour cette raison, contre seulement 336 dans le groupe témoin. D'après eux, puisque la différence entre les groupes n'était que de 44 décès, si les auteurs avaient été capables de repérer tous les cas préalables de cancer du sein dans le groupe témoin, la différence entre les deux groupes aurait été moindre. Miller [2001], l'un des chercheurs, réplique que les patientes du groupe témoin ayant eu un diagnostic de cancer du sein (d'après les dossiers médicaux) avant la randomisation ont été exclues à la 18^{e} année du suivi. Il ne mentionne toutefois pas le nombre d'exclusions faites en fonction de ce critère, et concède que cette exclusion tardive n'aurait pas éliminé l'inégalité relevée par Olsen et Gøtzsche. Les auteurs eux-mêmes ont reconnu le problème du suivi, admettant que « l'utilité des dossiers médicaux du HIP diminue à mesure que les personnes assurées quittent la compagnie d'assurances en raison d'un changement d'emploi, d'un déménagement à l'extérieur du territoire couvert ou du changement pour d'autres types de couverture » [trad.]. Au bout de 15 ans, plus de la moitié des clientes avaient quitté le HIP [Shapiro *et al.*, 1985].

Gøtzsche et Olsen [2000] ont relevé d'autres différences de base, notamment la présence d'une masse déjà présente dans un sein, la ménopause et la scolarité. De nombreuses réactions à la méta-analyse de Gøtzsche et Olsen ont fait valoir que de petites différences de base ne peuvent expliquer les réductions importantes du taux de mortalité, ce qui est probablement vrai [McTiernan, 2002]. Cependant, la question n'est pas de savoir si oui ou non les différences de base faussent la comparaison, mais plutôt si elles révèlent des failles dans le processus de randomisation, auquel cas il est douteux qu'il ait permis de circonscrire adéquatement d'autres facteurs de risque plus importants de cancer du sein plus difficilement mesurables, comme les antécédents familiaux de cancer du sein, les antécédents génésiques, la prise d'hormones, etc.

Sur l'échelle de la validité, l'essai HIP obtient 1,5 sur 4, principalement parce qu'il n'a pu exclure uniformément les cas préalables de cancer du sein ni faire un suivi adéquat des cas à partir d'une base de données sur les clientes d'une compagnie d'assurances.

2. MALMÖ (MMST)

Le MMST (*Malmö Mammographic Screening Trial*), mené à Malmö, en Suède, a commencé le recrutement des patientes à l'automne de 1976. Des rapports sur le plan expérimental, son déroulement et ses résultats ont été publiés en 1979, 1981, 1988 et, pour les femmes de moins de 50 ans, en 1997 [Andersson et Janzon, 1997; Andersson *et al.*, 1988; Andersson, 1981; Andersson *et al.*, 1979]. Le premier rapport indique que la population admissible à l'étude était composée de femmes âgées de 50 à 69 ans résidant à Malmö [Andersson *et al.*, 1979], mais le deuxième rapport indique que la participation à l'étude était offerte aux femmes de 45 à 69 ans. Une publication ultérieure sur le MMST mentionne que les données sur les femmes de 55 ans et plus à l'entrée et les données sur les femmes plus jeunes ont fait l'objet d'analyses séparées [Andersson *et al.*, 1988]. Le dernier rapport [Andersson et Janzon, 1997] fait référence à deux cohortes différentes, le MMST-I (femmes de moins de 50 ans à l'entrée, soit une cohorte née entre 1927 et 1932) et le MMST-II (femmes nées entre 1933 et 1945). Enfin, une revue suédoise parue récemment fait référence à la « continuation de l'essai de Malmö (MMST-II) » [Nyström *et al.*, 2002], expliquant que le MMST-I constituait la partie de l'essai dont le recrutement s'était terminé en 1978, alors

que le MMST-II a été mené de 1978 à 1990 et comptait des femmes qui, dès qu'elles atteignaient 45 ans, étaient continuellement randomisées selon le même protocole. Le MMST-II n'a pas suivi de protocole rigoureux [Olsen et Gøtzsche, 2001], et ses résultats n'ont jamais été publiés séparément. Bien que la différence entre les MMST I et II demeure confuse, il semble qu'il ne s'agisse pas d'essais distincts, mais bien de deux cohortes distinguées longtemps après la fin de la mise en contraste de l'étude. Le MMST-I comportait un plus grand nombre de cycles de dépistage que le MMST-II, simplement parce que la cohorte plus jeune du MMST-II ne pouvait commencer le dépistage qu'à l'âge de 45 ans. Ainsi, la cohorte née en 1945 ne pouvait commencer le dépistage qu'en 1990, année où l'essai a pris fin, et n'a donc participé qu'à un seul cycle de dépistage.

Mise en contraste

L'intervention de dépistage comprenait une mammographie prise sous deux incidences, craniocaudale et oblique, aux deux premiers cycles. Aux cycles subséquents, les patientes dont la mammographie révélait une trame parenchymateuse translucide n'ont eu que des incidences obliques. L'intervention de dépistage ne comprenait aucun examen médical. Les taux de participation étaient de 74 % au premier cycle et de 70 % dans les cycles subséquents, sans distinction [Andersson *et al.*, 1988]. Un autre rapport indique que le taux de participation variait entre « 75 et 80 % » [Andersson et Janzon, 1997]. Aux fins du présent rapport, un taux de participation moyen de 71 % sera postulé. Les patientes du groupe témoin n'ont pas eu de dépistage, mais celles qui avaient des symptômes mammaires « ont reçu les services médicaux usuels » [Andersson, 1981]. Les taux de mammographies dans ce groupe n'ont pas été mesurés, mais on présume qu'ils étaient bas, le dépistage mammographique n'étant pas pratique courante en Suède à l'époque. Comme 24 % des patientes du groupe témoin ont eu au moins une mammographie au cours de la période à l'étude, une seule dans la plupart des cas [Andersson *et al.*, 1988], on estime à 10 % le taux de dépistage dans ce groupe. Les femmes de la première cohorte (année de naissance entre 1927 et 1932) ont eu six cycles de dépistage; celles de la deuxième cohorte (1933-1945) en ont eu moins, et les plus jeunes n'ont eu qu'un cycle. L'intervalle entre les cycles de dépistage était de 18 à 24 mois, et la durée de mise en contraste du dépistage était de 8,8 années en moyenne. Le compte rendu des résultats relatifs à la mortalité pour chaque année de l'étude permet une analyse plus poussée des résultats sans dilution pour ces années [Andersson *et al.*, 1988, tableau IX].

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 27 %. Il est dommage que les résultats obtenus avec le suivi le plus long n'aient pas été publiés séparément, car la mise en contraste relativement longue de l'essai aurait permis un suivi plus long et une dilution moins importante.

Validité

Toutes les femmes de 45 à 69 ans résidant à Malmö étaient réparties individuellement au hasard par ordinateur par année de naissance [Andersson, 1981]. La moitié était affectée au dépistage (21 242), et l'autre moitié (21 240) au groupe témoin. Les femmes ayant eu un cancer du sein diagnostiqué avant la randomisation étaient exclues, ce qui a réduit de 393 le groupe soumis au dépistage et de 412 le groupe témoin [Gøtzsche et Olsen, 2000]. Les données de départ sur l'âge étaient similaires dans les groupes de dépistage et témoin, ce qui peut indiquer une randomisation adéquate, mais les données de départ sur les autres facteurs de risque ne sont pas disponibles. Jusqu'en 1988, la détermination des causes de décès des femmes ayant eu un diagnostic de cancer du sein se faisait à l'insu des chercheurs [Gøtzsche et Olsen, 2000] et, selon le rapport de 1988, les taux d'autopsies étaient relativement élevés (76 %), bien qu'un rapport ultérieur ait fait état d'un taux inférieur (40 %) [Andersson et Janzon, 1997].

Le MMST-II, mené sans protocole strict, n'a pas fait l'objet d'un rapport séparé, et il est

beaucoup plus difficile de déterminer la validité de cet essai semi-officiel. De graves erreurs administratives ont été signalées, tant dans les rapports sur l'essai [Andersson et Janzon, 1997] que dans des communications personnelles avec les auteurs [Gøtzsche et Olsen, 2000]. Au départ, cet essai n'était pas inclus dans la revue suédoise de Nyström et ses collaborateurs [1993], mais on le retrouve dans une revue plus récente [Nyström *et al.*, 2002], avec la mention que le protocole, quoique identique pour les deux cohortes, n'avait pu être strictement observé pour la cohorte du MMST-II. Aucune donnée de départ n'est disponible, et il n'y a pas d'informations détaillées sur les exclusions ou la validation du résultat final, comme les taux d'autopsies. Seule la partie MMST-I de l'essai, sur laquelle toutes les informations sont disponibles, a donc été incluse dans la présente évaluation.

Sur l'échelle d'évaluation de la validité, cet essai obtient une note de 3 sur 4. Il s'agit de l'un des deux seuls essais dont la validité a été jugée de « qualité moyenne » par Gøtzsche et Olsen, la plus haute note qu'ils aient accordée.

3. TWO-COUNTY (TCS)

L'essai *Two-County* a d'abord été désigné comme « l'essai suédois » par Tabár et Gad en 1981, mais il est en fait le deuxième de quatre essais menés en Suède. Amorcé en 1977 dans le comté de Kopparberg, l'essai s'est poursuivi dans le comté d'Östergötland en 1978. Ces deux comtés du centre sud-est de la Suède représentent environ 8 % de la population totale du pays. On appelle aussi parfois cet essai l'étude WE, Kopparberg étant situé à l'ouest (W) et Östergötland à l'est (E). Les auteurs insistent sur le fait que cet essai a été conçu comme un tout et, en effet, les protocoles de dépistage sont identiques quant à tous les aspects importants. Des rapports sont parus entre 1981 et 2001 [Tabár *et al.*, 2001; 2000; 1999; 1997; 1995a; 1995b; 1992a; 1992b; 1992c; 1989; Fagerberg et Tabár, 1988; Tabár *et al.*, 1985a; 1985b; Fagerberg *et al.*, 1985; Tabár et Gad, 1981].

Mise en contraste

Tout comme les autres essais suédois, cette étude visait à évaluer le dépistage mammographique sous une seule incidence, soit médiolatérale oblique, sans examen clinique des seins. Aussi comme les autres essais suédois, celui-ci a utilisé une unité centrale et plusieurs unités mobiles. La lecture des clichés était centralisée et confiée à deux radiologistes qui étaient aussi les chercheurs principaux de l'étude (Fagerberg et Tabár). La population admissible à l'essai était composée de femmes de 40 à 79 ans résidant dans les deux comtés. Cependant, en raison de la faible participation des femmes de plus de 75 ans, un rapport subséquent a restreint l'analyse aux femmes de 40 à 74 ans [Fagerberg *et al.*, 1985], et cette restriction a par la suite été étendue aux femmes de 40 à 69 ans [Tabár *et al.*, 2000]. Le taux de participation des femmes du groupe de dépistage était élevé, 91 % des femmes de 40 à 74 ans (les deux comtés ensemble) ayant participé au programme au premier cycle, 86 % au deuxième cycle, et 84 % au troisième cycle [Fagerberg et Tabár, 1988]. La participation moyenne était donc d'environ 87 %. La mammographie n'était pas offerte aux femmes du groupe témoin, et les taux de mammographies dans la population témoin n'ont pas été vérifiés. Les taux de contamination ont pu atteindre 13 % dans cet essai, car « 13 % des femmes du groupe témoin avaient eu une mammographie dans le cadre de leur examen médical courant jusqu'à la fin de 1984 » [Tabár *et al.*, 1985a], c'est-à-dire avant la mise en contraste du dépistage. L'intervalle de dépistage était de deux ans pour les femmes âgées de 40 à 49 ans à leur entrée dans l'étude, et de trois ans pour les femmes plus âgées. Bien que les rapports divergent un peu, il semble qu'une période de mise en contraste de cinq ans était prévue, avec une règle d'interruption basée sur des épreuves de signification statistique effectuées tous les six mois.

La controverse entourant les détails du protocole de dépistage montre bien à quel point la méthodologie est insuffisamment décrite dans les publications. À la lumière des rapports publiés, il est impossible d'établir avec certitude le début ou la fin des programmes de Kopparberg et d'Östergötland. Il semble cependant que le programme de Kopparberg ait débuté en octobre 1977 pour se poursuivre jusqu'en septembre 1982, moment où le dépistage a été offert au groupe témoin. Celui d'Östergötland a été amorcé en mai 1978 et s'est aussi poursuivi pendant cinq ans. Le dépistage a été offert au groupe témoin en 1984, mais peut-être aussi tard que 1986. Il est difficile de faire d'autres analyses temporelles, car les résultats sont donnés par année civile, malgré que la branche d'Östergötland ait débuté un an après celle de Kopparberg. Les résultats, quel que soit le moment du suivi, proviennent donc d'un composite des deux branches de l'étude décalées d'un an. Quoi qu'il en soit, la période entre la randomisation et le dépistage de la cohorte témoin se situe entre cinq et huit ans. Comme pour l'essai HIP, on ne peut observer la pleine réduction de la mortalité d'un programme en régime établi dans un essai dont la mise en contraste du dépistage est aussi courte. Cependant, comme l'essai a été analysé selon le modèle d'évaluation suédois, où seulement les femmes ayant eu un diagnostic de cancer du sein pendant la mise en contraste de l'étude ont été suivies pour la mesure des résultats sur la mortalité, et comme le suivi est exceptionnellement long (20 ans), l'effet de dilution qu'entraîne l'inclusion des premières années de suivi dans les données sur les taux de mortalité cumulatifs est moins important que dans les autres essais ayant une mise en contraste aussi courte.

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 42 %.

Validité

La répartition des femmes entre les groupes expérimental et témoin est définie par les auteurs comme une randomisation, mais le procédé utilisé était plutôt une affectation par blocs selon une formule compliquée qui se révèle plus créative qu'aléatoire. Les deux comtés de Kopparberg et d'Östergötland ont d'abord été divisés en blocs par région de résidence, et chacun de ces blocs a été divisé à nouveau en unités plus petites, par municipalité et par paroisse. Ces unités plus petites ont été constituées de façon à assurer l'homogénéité de toutes les unités à l'intérieur de chaque bloc, à savoir « des circonstances géographiques et socioéconomiques similaires » [Tabár et Gad, 1981]. Les plus petites unités étaient « réparties de façon aléatoire » entre un groupe de dépistage et un groupe témoin selon un procédé qui n'a pas été défini par les auteurs, la méthode pile ou face ayant peut-être été utilisée pour la formation des groupes à Östergötland [Olsen et Gøtzsche, 2001]. Les règles d'exclusion n'ont pas été décrites dans les rapports publiés, ce qui pourrait expliquer les divergences entre les différents rapports quant au nombre de femmes randomisées dans les deux comtés. Les données sur la comparabilité de départ des deux cohortes n'ont pas été publiées. Les calculs effectués par Olsen et Gøtzsche à partir des différents chiffres figurant dans les divers rapports indiquent des différences mineures dans les caractéristiques de départ, comme les cancers du sein préalables aux essais (un plus grand nombre d'exclusions dans le groupe témoin) et l'âge (groupe de dépistage plus vieux de 0,27 à 0,45 an). Ces deux différences tendraient à réduire l'effet du dépistage sur la réduction de la mortalité, mais elles révèlent aussi de sérieux problèmes dans l'exécution de l'essai et la façon dont les données sont rapportées.

Sur l'échelle d'évaluation de la validité, cet essai obtient la note de 0,5 sur 4.

4. ÉDIMBOURG

Édimbourg était l'un des centres faisant partie d'un essai non randomisé échelonné sur sept ans, le *UK Seven-year Trial of Breast Screening*. Au début de l'essai, en 1979, Édimbourg et un autre district ont été désignés pour offrir le dépistage mammographique, tandis que les autres centres participants offraient un examen clinique ou n'offraient aucune intervention. La partie de l'essai réalisée à Édimbourg a subséquemment été transformée en un essai randomisé ayant sa propre population témoin sélectionnée dans la ville [Roberts *et al.*, 1984], et a reçu le nom d'*Edinburgh Randomised Trial of Screening for Breast Cancer*. Des publications décrivant l'essai et ses résultats sont parues entre 1981 et 1999 [Alexander *et al.*, 1999; Alexander, 1997; Alexander *et al.*, 1994; Roberts *et al.*, 1990; UK Trial of Early Detection of Breast Cancer Group, 1988; Roberts *et al.*, 1984; UK Trial of Early Detection of Breast Cancer Group, 1981].

Mise en contraste

L'intervention expérimentale de dépistage comportait une invitation à une série de quatre cycles de dépistage mammographique accompagnés d'un examen clinique des seins tous les deux ans, et d'un examen clinique des seins seul aux années d'intervalle. Sur une période de 2,5 années entre 1979 et 1981, le dépistage a été offert à toutes les femmes de 45 à 64 ans faisant partie du groupe expérimental. Les nouvelles arrivantes étaient invitées à se joindre au programme en cours de route dès leur 45^{e} anniversaire de naissance, à leur arrivée à Édimbourg ou lorsqu'elles se joignaient à la clientèle d'une clinique médicale faisant partie de l'étude (ces dernières ont participé à moins de dépistages). La mammographie était réalisée sous deux incidences au premier dépistage (oblique et craniocaudale), et avec une incidence oblique seulement aux cycles subséquents. Contrairement aux autres essais, la lecture des clichés n'était pas confiée à des radiologistes, mais à des « médecins formés à cette fin ». Tous les clichés montrant des résultats anormaux et 5 % des clichés pris au hasard étaient cependant lus par des radiologistes. Les femmes du groupe de dépistage subissaient un examen clinique des seins pratiqué par un médecin ou une infirmière « conformément à un protocole standard », et on leur apprenait comment pratiquer l'auto-examen des seins, les incitant à s'y prêter une fois par mois, entre les consultations médicales. Pour les femmes du groupe témoin, aucun protocole spécifique n'est mentionné, mais elles pouvaient recevoir un dépliant sur l'auto-examen des seins. Les taux de participation dans le groupe de dépistage n'étaient que de 61 % au premier cycle, et seulement 44 % des participantes se sont présentées à la dernière mammographie [Alexander *et al.*, 1999]. Les taux de participation aux cycles intermédiaires n'ont pas été publiés. Le taux de participation moyen est estimé à 52 %. Les taux de dépistage chez les femmes du groupe témoin n'ont pas été mesurés, mais les chercheurs étaient d'avis que « très peu de femmes du groupe témoin ont pris l'initiative d'avoir un dépistage », bien qu'ils n'aient eu « aucun moyen de confirmer cette hypothèse » [Alexander *et al.*, 1999]. Les années 0, 2, 4 et 6 étaient prévues au calendrier des dépistages mammographiques, mais comme certaines femmes se sont jointes à l'étude en cours de route, la longueur moyenne pondérée de la mise en contraste était de 6,4 ans, et le suivi pouvait aller jusqu'à 14 ans après la randomisation [Alexander *et al.*, 1999].

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 26 %.

Validité

Le recrutement des participantes s'est fait par l'entremise de 348 omnipraticiens d'Édimbourg qui ont accepté de participer à l'essai. Les cabinets médicaux ont été stratifiés par district, puis « sélectionnés au hasard » par l'un des chercheurs. Les ajustements subséquents ont consisté notamment à allouer le même statut (groupe expérimental ou témoin) aux cabinets médicaux partageant les mêmes locaux et à maintenir les affectations erronées dues à des erreurs de communication avec les

médecins. Certains cabinets choisis initialement dans le groupe « témoin » ont changé de statut afin d'augmenter la population soumise au dépistage. L'exclusion des cas de cancer du sein préalablement diagnostiqués semble avoir été très inégale, 338 femmes ayant été exclues du groupe de dépistage contre seulement 177 du groupe témoin. La comparaison initiale des groupes est impossible, car les facteurs épidémiologiques et démographiques « n'étaient consignés que pour les cas de cancer et pour les femmes soumises au dépistage » [Roberts *et al.*, 1984]. Cependant, lorsqu'en 1983 les chercheurs ont observé un déséquilibre dans les taux de mortalité générale (toutes causes confondues) entre les groupes de dépistage et témoin, ils ont procédé à une estimation des situations socioéconomiques à partir de données de recensement par codes postaux [Alexander *et al.*, 1999]. Cette démarche a révélé des différences majeures entre les deux groupes, 26 % des femmes du groupe témoin et 53 % des femmes du groupe de dépistage faisant partie de la classe socioéconomique la plus élevée [Alexander *et al.*, 1994]. De plus, les taux de mortalité générale étaient inférieurs de 15 % chez les femmes du groupe de dépistage, ce qui pourrait indiquer qu'elles avaient un meilleur état de santé.

Sur l'échelle d'évaluation de la validité, cet essai obtient 0,5 sur 4, la note la plus basse de tous les essais avec celui de *Two-County*.

5. CANADA-1 (NBSS-1)

C'est en janvier 1980, dans dix centres de six provinces canadiennes, que l'essai *National Breast Screening Study* (NBSS) a vu le jour. Deux protocoles différents avaient été définis pour les femmes âgées de 40 à 49 ans à l'entrée dans l'étude (NBSS-1) et pour celles qui étaient âgées de 50 à 59 ans à l'entrée (NBSS-2). Des articles sur le NBSS-1 ont été publiés entre 1981 et 2002 [Miller *et al.*, 2002; 1997; Bailar et MacMahon, 1997; Miller *et al.*, 1992a; 1991; Baines *et al.*, 1990; Miller, 1988; Baines *et al.*, 1986; Miller *et al.*, 1981].

Mise en contraste

La mise en contraste du dépistage dans l'essai NBSS-1 était définie comme suit : l'intervention de dépistage était constituée d'un examen physique initial (P) pratiqué par des infirmières spécialement formées ou des médecins, immédiatement suivi d'une mammographie prise sous deux incidences (M). Les examens subséquents, offerts à un an d'intervalle, pour une série totale de quatre ou cinq cycles de dépistage, comprenaient un examen physique et une mammographie. La cohorte à qui cette intervention a été offerte a été appelée « groupe MP ». Les femmes du groupe témoin subissaient un examen physique initial (antérieur à la randomisation) et avaient un suivi annuel par l'entremise d'un questionnaire postal auto-administré. Cette cohorte a été appelée « groupe UC » (pour *usual care*). De plus, l'auto-examen des seins a été enseigné aux femmes des deux groupes. Pour celles qui retournaient aux centres de dépistage pour leurs examens (vraisemblablement les femmes du groupe MP), l'enseignement de l'auto-examen des seins était consolidé [Miller *et al.*, 1992b]. La mise en contraste du dépistage peut donc se résumer ainsi : examen physique de base et enseignement de l'auto-examen des seins, avec des mammographies pratiquées seulement au besoin dans un contexte diagnostique (groupe témoin), *versus* une mammographie au cycle 0, et une mammographie, un examen physique et la consolidation de l'auto-examen des seins aux cycles subséquents (groupe expérimental).

La proportion des femmes du groupe MP ayant participé à au moins une partie de l'intervention proposée (mammographie et examen physique) était de 100 % au premier cycle et de 89,4 % au deuxième cycle, diminuant à 85,6 % au cinquième cycle. Par ailleurs, de 2 à 3 % des femmes du groupe MP ont accepté l'examen physique mais refusé la mammographie. Entre 93 et 95 % des participantes du groupe UC ont retourné leur questionnaire annuel. De plus, 26,4 % des femmes de ce groupe ont subi au moins une mammographie en dehors du protocole

expérimental pendant toute la période de l'étude; ce taux est passé de 7,0 % entre la première et la deuxième année à 18,1 % entre la quatrième et la cinquième année. Des femmes qui se sont jointes à l'étude au tout début, 62 % ont eu le temps de participer à un programme de quatre ans (cinq cycles), et les autres (38 %) se sont vu offrir un programme de trois ans. On peut ainsi estimer la durée moyenne de mise en contraste du dépistage à 4,6 ans. La durée moyenne de suivi la plus longue ayant fait l'objet d'une publication est de 13 ans (entre 11 et 16 ans) [Miller *et al.*, 2002].

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 45 %. Les données sont présentées de telle sorte qu'elles permettent de faire une analyse comparable à celle du modèle suédois, où seuls les cas diagnostiqués au cours des cinq premières années sont suivis pour la mesure des taux de mortalité par cancer du sein. Lorsque les données sont tronquées de la sorte, la dilution temporelle est beaucoup moindre. Avec cette analyse, la force de contraste totale serait de 75 %. Ce modèle pourrait toutefois causer un biais lié au devancement du diagnostic *(lead-time bias)*, car les femmes du groupe de dépistage sont plus susceptibles d'avoir un diagnostic de cancer du sein parce qu'elles ont eu plus de mammographies. C'est pourquoi cette dernière analyse n'a pas été retenue ici.

Validité

Des volontaires recrutées par différents moyens ont été réparties individuellement au hasard entre les groupes de dépistage MP et témoin UC. Un questionnaire initial remis préalablement à la randomisation a permis d'exclure les femmes ayant déjà eu un diagnostic de cancer du sein. Une révision indépendante du processus de randomisation a conclu que celui-ci avait réussi à assurer l'égalité de départ [Bailar et MacMahon, 1997]. Les cas de cancer décelés à l'examen initial n'ont pas été exclus du suivi ni de l'analyse, et les exclusions postrandomisation s'équivalaient dans les deux groupes. Les groupes étaient comparables pour 10 facteurs de risque importants [Miller *et al.*, 2000a; 1992a]. L'étude a été critiquée parce qu'il y avait un plus grand nombre de petites tumeurs avec atteinte ganglionnaire dans le groupe soumis au dépistage, mais la mammographie est censée déceler de telles tumeurs [Bailar et MacMahon, 1997]. Les femmes atteintes d'un cancer du sein détecté lors d'un examen clinique des seins préalable à la randomisation étaient exclues des deux groupes, mais les femmes dont le cancer du sein avait été détecté subséquemment par mammographie n'auraient pas dû être exclues du groupe de dépistage, la mammographie étant pratiquée précisément dans le but de déceler des cancers, et n'importe quel critère d'exclusion doit être appliqué uniformément à chacun des groupes pour ne pas fausser la comparaison [Kakuma, 2002]. Le suivi effectué pendant la période de mise en contraste de l'étude se faisait lors des examens (groupe MP) ou par l'entremise d'un questionnaire envoyé par la poste (groupe UC). Dans les deux groupes, les femmes qui avaient un diagnostic de cancer du sein étaient suivies annuellement par le bureau central du NBSS, et les nouveaux diagnostics chez les autres femmes étaient repérés par recoupement avec les registres provinciaux des cancers et les bases de données sur la mortalité. Tous les cas de décès par cancer étaient révisés à l'insu par un groupe d'experts à partir de rapports d'anatomopathologie, de dossiers d'hospitalisation et de rapports d'autopsie.

Sur l'échelle d'évaluation de la validité, cet essai obtient la note de 4 sur 4.

6. CANADA-2 (NBSS-2)

La deuxième partie de l'essai *National Breast Screening Study* (NBSS), qui porte sur des femmes âgées de 50 à 59 ans à l'entrée dans l'étude, a aussi débuté en janvier 1980. L'essai NBSS-2 a été mené dans les mêmes localités et sur la même période que le NBSS-1. Des articles détaillés sur le NBSS-2 sont parus en 1981, 1991, 1992 et 2000 [Miller *et al.*, 2000b; 1992b; 1991; 1981].

Mise en contraste

À la différence du NBSS-1, le plan expérimental du NBSS-2 mettait en contraste deux interventions de dépistage. Comme dans le NBSS-1, le groupe expérimental recevait un examen physique initial (P), immédiatement suivi d'une mammographie de dépistage (M). Les examens subséquents, offerts à un an d'intervalle, pour une série totale de quatre ou cinq cycles de dépistage, comprenaient un examen physique et une mammographie réalisée sous deux incidences. La cohorte à qui cette intervention a été offerte a été appelée « groupe MP ». Les femmes du groupe témoin ont eu un examen physique initial (avant la randomisation) et des examens physiques annuels subséquents, mais sans mammographie. Cette cohorte a été appelée « groupe PO » *(Physical examination Only)*. De plus, les femmes des deux groupes ont appris à pratiquer l'auto-examen des seins et, contrairement au NBSS-1, l'enseignement de l'auto-examen des seins a été consolidé dans les deux groupes lors des examens physiques annuels [Miller *et al.*, 1992b]. La mise en contraste du dépistage peut se résumer ainsi : examen physique initial et enseignement de l'auto-examen des seins, avec un examen physique annuel de grande qualité dans les deux groupes, les mammographies n'étant pratiquées qu'au besoin dans un contexte diagnostique (groupe témoin), *versus* l'ajout d'une mammographie de dépistage à chaque cycle (groupe expérimental). Comparativement au NBSS-1, le contraste est considérablement plus faible, les femmes du groupe témoin étant soumises à un protocole de dépistage qui peut être efficace en soi.

La proportion des femmes du groupe MP ayant participé aux deux interventions proposées (mammographie et examen physique) était de 100 % au premier cycle et de 90,4 % au deuxième cycle, diminuant à 86,7 % au cinquième cycle. Par ailleurs, de 2 à 3 % des femmes du groupe MP ont accepté l'examen physique, mais refusé la mammographie. Dans le groupe PO, le taux de participation à l'examen physique annuel était de 100 % au premier cycle et de 89,1 % au deuxième cycle, diminuant à 85,4 % au cinquième cycle. De plus, 16,9 % des femmes de ce groupe ont subi au moins une mammographie en dehors du protocole pendant la période de l'étude. Ce taux est passé de 5,3 % entre la première et la deuxième année, à 8,0 % entre la quatrième et la cinquième année. La durée moyenne de mise en contraste du dépistage était identique à celle du NBSS-1, soit 4,6 ans, ce qui a entraîné une dilution considérable dans le suivi.

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient 31 %. Les données sont présentées de telle sorte qu'elles permettent une analyse comparable à celle du modèle suédois, où seulement les cas diagnostiqués au cours des cinq premières années sont suivis pour la mesure des taux de mortalité par cancer du sein. Lorsque les données sont tronquées de la sorte, la dilution temporelle est beaucoup moindre. Avec cette analyse, la force de contraste totale serait de 54 %. Cependant, ce modèle peut causer un biais lié au devancement du diagnostic *(lead-time bias)*, car les femmes du groupe de dépistage sont plus susceptibles d'avoir un diagnostic de cancer du sein parce qu'elles ont eu plus de mammographies. C'est pourquoi cette dernière analyse n'a pas été retenue ici.

Validité

La randomisation, les critères d'exclusion, la comparabilité des groupes au départ et la méthode d'obtention des résultats étaient pratiquement identiques à ceux de l'essai NBSS-1, et la même analyse que pour le NBSS-1 s'applique au NBSS-2 à tous égards.

Sur l'échelle d'évaluation de la validité, cet essai obtient aussi la note de 4 sur 4.

7. STOCKHOLM

Le *Stockholm Mammographic Screening Trial* a débuté en mars 1981, à Stockholm, en Suède. Des articles sur cet essai ont été publiés entre 1986 et 1997 [Frisell *et al.*, 1997; Frisell et Lidbrink, 1997; Frisell *et al.*, 1991; 1989; 1986].

Mise en contraste

L'intervention de dépistage consistait en seulement deux cycles de mammographies sous une seule incidence (oblique), sans examen physique, à 28 mois d'intervalle. Le taux de participation était de 81 % au premier cycle, et de 80 % au deuxième cycle. Ni la mammographie ni l'examen physique n'ont été offerts aux femmes du groupe témoin, mais elles ont été invitées à un dépistage mammographique à la fin du deuxième cycle de mammographies du groupe de dépistage. Le taux de participation à ce cycle témoin de mammographies était de 77 % [Frisell *et al.*, 1997]. Le pourcentage des femmes du groupe témoin ayant subi une mammographie de dépistage pendant la mise en contraste de l'étude n'est pas connu, mais il pourrait être d'au moins 25 % (taux de mammographies de dépistage pratiquées à Stockholm au cours des trois années qui ont précédé l'essai) [Frisell *et al.*, 1991]. Le premier cycle de dépistage a débuté le 9 mars 1981 pour se terminer 2,5 années plus tard. Démarré le 1er septembre 1983, le deuxième cycle a duré 2,1 années [Frisell *et al.*, 1991]. Au cours de l'année 1986, le groupe témoin a été invité à subir une mammographie de dépistage. La mise en contraste du dépistage était donc de 4,6 années.

Il est difficile d'analyser le degré de dilution, car « aucun cas additionnel n'a été introduit après le 31 décembre 1986 », c'est-à-dire entre 1 et 3,5 ans après la fin de la mise en contraste du dépistage. Cela signifie que le total cumulatif des cas de cancer a été tronqué à une certaine date (un peu plus d'un an après la fin du deuxième cycle de dépistage), mais en temps d'étude, cela équivaut à une période pouvant varier entre un an après la deuxième mammographie (pour les femmes soumises au dépistage à la fin du deuxième cycle) et 3,5 ans (pour les femmes soumises au dépistage au début du deuxième cycle). Cette troncature évite la plupart des problèmes de dilution causés par les cas tardifs de cancer du sein, soit les cas apparus après la fin de la période de mise en contraste du dépistage. Cette méthode, que les chercheurs suédois appellent « méthode d'évaluation » [Nyström *et al.*, 1993], a aussi été utilisée dans deux revues des études suédoises. Par conséquent, un suivi plus poussé ne devrait pas causer davantage de dilution des résultats relatifs à la mortalité.

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient une note de 28 %.

Validité

Les sujets ont été répartis systématiquement entre les groupes de dépistage et témoin selon leur date de naissance, les femmes nées dans les 10 premiers jours du mois ou entre le 21e et le 30e (ou 31e)[11] jour étant affectées au dépistage. Techniquement parlant, ce n'est donc pas un essai randomisé, même si tous les rapports le désignent ainsi. Bien que de toute évidence, la date de naissance ne soit pas un facteur de risque de cancer du sein, un protocole de répartition aussi transparent se prête à l'auto-sélection, car il se peut que la date en question n'ait pas été vérifiée adéquatement et que les femmes nées entre le 11 et le 20 du mois qui voulaient subir un dépistage n'aient pas participé à l'étude si elles ou leur médecin étaient au courant de ce plan de répartition. Comme les chercheurs de la plupart des essais ont conclu que le risque de cancer du sein est plus élevé chez les femmes qui participent aux programmes de dépistage, ce biais potentiel jouerait probablement en faveur de l'hypothèse nulle.

11. Selon Frisell et ses collaborateurs [1986], les femmes nées entre le 1er et le 10 de chaque mois ou entre le 21 et le 31 étaient affectées au groupe de dépistage, mais un autre article indique plutôt que ce sont les femmes nées entre le 21 et le 30 de chaque mois qui ont été affectées au dépistage [Frisell *et al.*, 1989]. Cette erreur pourrait expliquer en partie le léger surplus de femmes dans le groupe de dépistage (40 318) noté par Olsen et Gøtzsche [2001], par rapport aux 40 000 auxquelles on pourrait s'attendre si cette dernière stratégie de répartition avait été appliquée.

Il est difficile d'analyser les données de cet essai, car les chiffres semblent avoir été arrondis. Par exemple, selon Frisell et ses collaborateurs [1986], l'étude incluait la totalité des 60 000 femmes de la partie sud-est du grand Stockholm. On lisait toutefois dans le même article que la population à l'étude comptait 40 318 femmes au total. La seule mention de la taille du groupe témoin se trouve à la figure 1, où les femmes qui n'avaient pas été invitées au dépistage étaient inscrites sous la mention « *control group c. 20 000* » (groupe témoin env. 20 000). Un article ultérieur [Frisell *et al.*, 1989] précise que « l'essai de Stockholm comporte au total 60 261 femmes », et que 19 943 femmes composent la population témoin. Plus loin (tableau 6) figure le nombre de nouveaux cas par rapport à la taille de la population pour différentes années postrandomisation, et il se trouve que les 20 dénominateurs sont tous des multiples de 100. Dans un troisième rapport [Frisell *et al.*, 1991], il y a toujours un groupe de dépistage de 40 318 femmes, mais la population témoin est descendue sans explication à 19 343 femmes. Il peut s'agir d'une erreur typographique, car le chiffre 19 943 figure au tableau 1. S'il est acceptable d'arrondir des données quantitatives pour son usage personnel ou pour transmettre les grandes lignes d'une information là où une précision absolue n'est pas nécessaire, l'arrondissement des chiffres dans le compte rendu des résultats d'une étude ne permet pas de faire un examen approfondi des données publiées sur des aspects comme la qualité de la randomisation, la participation, l'intégrité du suivi et les résultats obtenus.

La comparabilité de départ entre les groupes de dépistage et témoin n'a pas été indiquée. Les femmes ayant eu un diagnostic de cancer du sein avant le début de l'essai et après 1986 ont été exclues après la randomisation [Frisell *et al.*, 1997], mais leur nombre dans le groupe de dépistage et le groupe témoin n'a pas été fourni. Comme dans l'essai *Two-County*, la mortalité par cancer du sein est définie ainsi : décès d'une personne atteinte d'un cancer du sein à sa mort. Cet indicateur semble fausser les résultats en défaveur du dépistage, car les cancers ont plus probablement été diagnostiqués chez des femmes qui avaient eu une mammographie, que ces cancers soient ou non la cause du décès. Le taux d'autopsies de 23 % est le plus bas des essais suédois, mais se compare favorablement à celui des essais canadiens.

Sur l'échelle d'évaluation de la validité, cet essai obtient 2 sur 4.

8. GÖTEBORG

Le *Gothenburg Breast Screening Trial*, mené à Göteborg, en Suède, a débuté en décembre 1982. Les résultats relatifs aux femmes de 39 à 49 ans ont paru en 1997, et des résultats plus récents sur les femmes de 39 à 59 ans ont été publiés en 2003 [Bjurstam *et al.*, 2003; 1997].

Mise en contraste

L'intervention de dépistage était constituée d'un maximum de cinq cycles de mammographies de dépistage, avec des intervalles prévus de 18 mois entre les cycles. Les mammographies ont été réalisées sous deux incidences dans 70 % des cas, sous une seule incidence dans 30 % des cas où un cliché préalable pris sous deux incidences indiquait que la mammographie sous une incidence serait adéquate [Bjurstam *et al.*, 1997]. Aucun examen clinique des seins, aucune formation ni aucune incitation à l'autoexamen des seins n'ont été offerts. Les taux de participation étaient de 85 %, 78 %, 79 %, 77 % et 75 % respectivement aux cinq cycles. Les femmes du groupe témoin n'étaient pas invitées à subir une mammographie, mais un sondage effectué auprès de 1 655 femmes du groupe de dépistage révèle que 28 % d'entre elles avaient eu une mammographie avant le début de l'étude, et 13 % au cours des deux années précédentes. Dans un sondage similaire réalisé auprès de 1 641 femmes du groupe témoin, 51 % ont dit avoir subi une mammographie avant le début de l'étude, et 19 % dans les deux années précédentes. À la fin de l'essai, soit approximativement sept ans après le début du dépistage du groupe

expérimental, on a offert une mammographie de dépistage aux femmes du groupe témoin, et 66 % d'entre elles s'y sont soumises. Cependant, il y a des doutes quant à la date exacte du dépistage des femmes du groupe témoin : il est possible qu'il n'ait pas été réalisé en décembre 1988 [Bjurstam *et al.*, 1997], mais plutôt en novembre 1987 [Nyström *et al.*, 2002], ou même plus tôt, entre trois et six ans après la randomisation [Nyström *et al.*, 1993] ou entre quatre et cinq ans après la randomisation [Nyström *et al.*, 1995]. On ne sait donc pas si la durée de la mise en contraste du dépistage était de cinq, six ou sept ans. L'analyse la plus récente [Nyström *et al.*, 2002] donne 6,7 ans comme durée médiane de l'essai.

Sur l'échelle d'évaluation de la force du contraste, cet essai obtient une note de 37 %.

Validité

Les femmes ont été randomisées par blocs (18 %) et ensuite individuellement (82 %) en fonction de leur date de naissance, avec un ratio de femmes réparties dans les groupes de dépistage et témoin de 1 sur 1,2 dans le groupe âgé de 40 à 49 ans, et de 1 sur 1,6 dans le groupe âgé de 50 à 59 ans. Les ratios exacts de randomisation étaient en fait différents même à l'intérieur de ces groupes [Olsen et Gøtzsche, 2001]. Selon une communication personnelle citée par Olsen et Gøtzsche [2001], une proportion similaire de femmes (1,2 %) a été exclue de chacun des groupes en raison d'un diagnostic de cancer préalable, bien que cette exclusion ait été faite subséquemment à la mise en œuvre de l'essai [Bjurstam *et al.*, 1997]. Les caractéristiques de départ des groupes de dépistage et témoin ne sont pas disponibles, sauf en ce qui a trait à l'âge, mais là encore, les ratios de randomisation variables appliqués aux différents groupes d'âge rendent difficile de se servir de la moyenne d'âge pour vérifier si la randomisation a réussi à assurer la comparabilité de base. Les autres caractéristiques de départ rattachées aux facteurs de risque n'ont pas été publiées. Dans les deux groupes, les cas de carcinomes canalaires *in situ* (DCIS : *ductal carcinoma in situ*) ont été inclus, mais pas les cas de carcinomes lobulaires *in situ* (LCIS : *lobular carcinoma in situ*). Les cas diagnostiqués entre le moment de la randomisation et le cycle de dépistage offert au groupe témoin ont été suivis afin de repérer les décès par cancer du sein à partir du registre suédois des causes de décès. Le taux d'autopsies était de 31 % [Olsen et Gøtzsche, 2001].

Sur l'échelle d'évaluation de la validité, cet essai obtient 1,5 sur 4.

9. ROYAUME-UNI (UK AGE TRIAL)

En 1991, le *United Kingdom Coordinating Committee on Cancer Research* a mis sur pied un essai contrôlé randomisé multicentrique à l'échelle nationale, appelé *Age Trial* [Moss, 1999]. Cet essai a été décrit dans un article scientifique [Moss, 1999] et dans un article sur les résultats préliminaires [Moss *et al.*, 2005a]. Une première analyse des résultats des 10 premières années de dépistage donne plus de détails sur le déroulement de l'essai et la détection du cancer [Moss *et al.*, 2005a]. Il y a encore trop peu de données sur la mortalité due au cancer du sein pour conclure à une réduction de la mortalité, mais une première analyse, fondée sur des mesures intermédiaires comme la détection du cancer ainsi que la taille et le degré de malignité des tumeurs a permis d'estimer le taux de mortalité chez les 1 287 femmes qui avaient eu un diagnostic de cancer du sein avant le 31 décembre 1999 [Moss *et al.*, 2005b].

Mise en contraste

L'intervention de dépistage consiste en une série de neuf cycles de mammographies commencés dès l'âge de 40 ou 41 ans et réalisés à intervalles d'un an jusqu'à l'âge de 48 ans. Une mammographie est pratiquée sous deux incidences au premier examen, puis sous une seule incidence par la suite, sauf indication contraire [Moss, 1999]. Aucun examen clinique des seins, aucune formation sur l'auto-examen des seins ni aucune incitation à le faire ne sont offerts.

L'intervention témoin n'a pas été définie, mais elle ne comporte probablement aucune invitation au dépistage et s'en tient aux « soins usuels ». Un taux de participation de 70 % était prévu, mais après cinq cycles de dépistage, le taux moyen était plutôt de 61 %. Le taux de dépistage chez les femmes du groupe témoin n'était que de 3,9 %, mais ce chiffre surestime le phénomène de la contamination, puisqu'il représente la proportion des femmes qui ont eu une mammographie dans les trois années précédentes, alors que le dépistage proposé aux femmes de la cohorte de dépistage est annuel. Dès qu'elles auront atteint l'âge de 50 ans, les femmes des groupes de dépistage et témoin seront invitées systématiquement au programme de dépistage déjà en place offrant des mammographies à intervalles de trois ans. La durée totale de la mise en contraste du dépistage serait donc de 11 ans, faisant de cet essai celui dont la mise en contraste est la plus longue, et de loin. Cet essai est donc le moins vulnérable à la dilution temporelle après une période de suivi suffisamment longue.

Sur l'échelle d'évaluation de la force du contraste, il obtient une note préliminaire de 41 % à 13 ans de suivi.

Validité

Les femmes sont randomisées en deux groupes : un groupe expérimental de 65 000 femmes et un groupe témoin de 130 000 femmes qui ne reçoivent aucune intervention. La randomisation est individuelle et stratifiée par cabinet d'omnipraticien. Les détails relatifs à l'exclusion des femmes qui ont eu préalablement un diagnostic de cancer du sein et aux caractéristiques de départ des deux groupes n'ont pas été publiés. Le suivi pour la mesure de l'incidence du cancer du sein et de la mortalité se fait à partir des rapports des centres responsables des essais, des laboratoires de pathologie et des registres des cancers, et les résultats des rapports d'anatomopathologie de tous les cas de cancer du sein dépistés au cours de l'essai seront révisés.

Sur l'échelle d'évaluation de la validité, à partir de conjectures qui devront être confirmées lorsque plus de détails seront publiés, cet essai obtient une note potentielle de 4 sur 4 sur la foi du plan expérimental et des premières informations sur le déroulement et l'analyse de l'essai qui ont été publiées.

RÉFÉRENCES

Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES). Dépistage du cancer du sein par mammographie : évaluation de la méta-analyse de Gøtzsche et Olsen. Paris : ANAES; 2002.

Alexander FE. The Edinburgh randomised trial of breast cancer screening. J Natl Cancer Inst Monogr 1997;(22):31-5.

Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet 1999;353(9168):1903-8.

Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. The Edinburgh randomised trial of breast cancer screening: Results after 10 years of follow-up. Br J Cancer 1994;70(3):542-8.

American Cancer Society (ACS). Cancer facts and figures. Atlanta, GA : ACS; 2004.

Andersson I. Radiographic screening for breast carcinoma. I. Program and primary findings in 45-69 year old women. Acta Radiol Diagn (Stockh) 1981;22(2):185-94.

Andersson I, Aspegren 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(6654):943-8.

Andersson I, Hellstrom L, Bjurstam N, Lundgren B, Fagerberg G, Tabár L. [Breast cancer screening by mammography in Sweden]. Lakartidningen 1983;80(25):2559-62 (article en suédois).

Andersson I, Andren L, Hildell J, Linell F, Ljungqvist U, Pettersson H. Breast cancer screening with mammography: A population-based, randomized trial with mammography as the only screening mode. Radiology 1979;132(2):273-6.

Andersson I et Janzon L. Reduced breast cancer mortality in women under age 50: Updated results from the Malmo Mammographic Screening Program. J Natl Cancer Inst Monogr 1997; (22):63-7.

Bailar JC et MacMahon B. Randomization in the Canadian National Breast Screening Study: A review for evidence of subversion. CMAJ 1997;156(2):193-9.

Baines CJ. A different view on what is known about breast screening and the Canadian National Breast Screening Study. Cancer 1994;74(4):1207-11.

Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders DE, Sickles EA, et al. Canadian National Breast Screening Study: Assessment of technical quality by external review. Am J Roentgenol 1990;155(4):743-9.

Baines CJ, McFarlane DV, Wall C. Audit procedures in the National Breast Screening Study: Mammography interpretation. Can Assoc Radiol J 1986;37(4):256-60.

Baker SG, Kramer BS, Prorok PC. Statistical issues in randomized trials of cancer screening. BMC Med Res Methodol 2002;2(1):11.

Baum M. Commentary: False premises, false promises and false positives—The case against mammographic screening for breast cancer. Int J Epidemiol 2004;33(1):66-73.

Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nyström L, et al. The Gothenburg Breast Screening Trial. Cancer 2003;97(10):2387-96.

Bjurstam N, Bjorneld L, Duffy SW, Smith TC, Cahlin E, Eriksson O, et al. The Gothenburg breast screening trial: First results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. Cancer 1997;80(11):2091-9.

Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. J Natl Cancer Inst 2002;94(3):167-73.

Brown BW, Brauner C, Minnotte MC. Noncancer deaths in white adult cancer patients. J Natl Cancer Inst 1993;85(12):979-87.

Centre International de Recherche sur le Cancer (CIRC). IARC handbooks of cancer prevention. Volume 7: Breast cancer screening. Lyon, France : IARC Press; 2002.

Chalmers I. Comparing like with like: Some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. Int J Epidemiol 2001;30(5): 1156-64.

Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. Control Clin Trials 1981;2(1):31-49.

Conseil d'évaluation des technologies de la santé du Québec (CETS). Le dépistage du cancer du sein chez les femmes de 40 à 49 ans. Montréal, Qc : CETS; 1993.

Conseil d'évaluation des technologies de la santé du Québec (CETS). Dépistage du cancer du sein au Québec : estimations des coûts et des effets sur la santé. Montréal, Qc : CETS; 1990.

DerSimonian R et Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-88.

Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998;338(16): 1089-96.

Fagerberg G, Baldetorp L, Grontoft O, Lundstrom B, Manson JC, Nordenskjold B. Effects of repeated mammographic screening on breast cancer stage distribution. Results from a randomised study of 92 934 women in a Swedish county. Acta Radiol Oncol 1985;24(6): 465-73.

Fagerberg G et Tabár L. The results of periodic one-view mammography screening in a randomized, controlled trial in Sweden. Part 1: Background, organization, screening program, tumor findings. Dans : Day NE, Miller AB. Screening for breast cancer. Toronto, Ont. : Hans Huber Publishers; 1988 : 33-8.

Fink R, Shapiro S, Roester R. Impact of efforts to increase participation in repetitive screenings for early breast cancer detection. Am J Public Health 1972;62(3):328-36.

Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 1993;85(20):1644-56.

Freedman DA, Petitti DB, Robins JM. On the efficacy of screening for breast cancer. Int J Epidemiol 2004;33(1):43-55.

Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years—Update of mortality results in the Stockholm Mammographic Screening Trial. Breast Cancer Res Treat 1997; 45(3):263-70.

Frisell J, Eklund G, Hellstrom L, Lidbrink E, Rutqvist LE, Somell A. Randomized study of mammography screening—Preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat 1991;18(1):49-56.

Frisell J, Eklund G, Hellstrom L, Glas U, Somell A. The Stockholm breast cancer screening trial—5-year results and stage at discovery. Breast Cancer Res Treat 1989;13(1):79-87.

Frisell J, Glas U, Hellstrom L, Somell A. Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. Breast Cancer Res Treat 1986; 8(1):45-54.

Frisell J et Lidbrink E. The Stockholm mammographic screening trial: Risks and benefits in age group 40-49 years. J Natl Cancer Inst Monogr 1997;(22):49-51.

GE Medical Systems. Mammography: History of the mammography 2004. Disponible à : http:// www.gemedicalsystems.com/rad/whc/mswhhis.html (consulté le 28 juillet 2004).

Glasziou PP. Meta-analysis adjusting for compliance: The example of screening for breast cancer. J Clin Epidemiol 1992;45(11):1251-6.

Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355(9198):129-34.

Health Council of the Netherlands (HCN). The benefit of population screening for breast cancer with mammography. La Haye : HCN; 2002.

Hendrick RE, Smith RA, Rutledge JH, Smart CR. Benefit of screening mammography in women aged 40-49: A new meta-analysis of randomized controlled trials. J Natl Cancer Inst Monogr 1997;(22):87-92.

Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: A summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137 (5 Ptie 1):347-60.

Institut national de santé publique du Québec (INSPQ), Brisson J, Hébert-Croteau N, Langlois A. Déterminants du taux de référence lors d'une première mammographie de dépistage – Programme québécois de dépistage du cancer du sein 1999. Québec : INSPQ; 2003.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 1996;17(1):1-12.

Jung H. [Assessment of usefulness and risk of mammography screening with exclusive attention to radiation risk]. Radiologe 2001;41(4):385-95 (article en allemand).

Kakuma R. Screening for breast cancer with mammography among women aged 50-69 years. Report prepared for the Breast Cancer Screening Unit. Ottawa, Ont. : Santé Canada; 2002.

Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA 1995;273(2):149-54.

Lampic C, Thurfjell E, Bergh J, Sjoden PO. Short- and long-term anxiety and depression in women recalled after breast cancer screening. Eur J Cancer 2001;37(4):463-9.

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. J Natl Cancer Inst Monogr 1997;(22):57-61.

Larsson LG, Nyström L, Wall S, Rutqvist L, Andersson I, Bjurstam N, Fagerberg G, Frisell J, Tabár L. The Swedish randomised mammography screening trials: Analysis of their effect on the breast cancer related excess mortality. J Med Screen 1996;3(3):129-32.

Last JM. A dictionary of epidemiology. 4e éd. New York, NY : Oxford University Press; 2001.

Law M, Hackshaw A, Wald N. Screening mammography re-evaluated. Lancet 2000;355(9205): 749-52.

Madlensky L, Goel V, Polzer J, Ashbury FD. Assessing the evidence for organised cancer screening programmes. Eur J Cancer 2003;39(12):1648-53.

McTiernan A. Recent controversies in mammography screening for breast cancer. Medscape Womens Health 2002;7(2):3.

Medical Research Council (MRC). Streptomycin treatment of pulmonary tuberculosis. BMJ 1948: 769-82.

Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983;2(2):267-71.

Miettinen OS, Henschke CI, Pasmantier MW, Smith JP, Libby DM, Yankelevitz DF. Mammographic screening: No reliable supporting evidence? Lancet 2002;359(9304):404-6.

Miller AB. Screening for breast cancer with mammography. Lancet 2001;358(9299):2164, 2167-8.

Miller AB. Screening for breast cancer: A review. Eur J Cancer Clin Oncol 1988;24(1):49-53.

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. A randomized screening trial of mammography in women age 40 to 49 years. Ann Intern Med 2002;137(5 Ptie 1):305-12.

Miller AB, Baines CJ, To T, Wall C. Screening mammography re-evaluated. Lancet 2000a; 355(9205):747-52.

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 2000b;92(18):1490-9.

Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study: update on breast cancer mortality. J Natl Cancer Inst Monogr 1997;(22):37-41.

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 1992a;147(10):1459-76.

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 1992b;147(10):1477-88.

Miller AB, Baines CJ, Turnbull C. The role of the nurse-examiner in the National Breast Screening Study. Can J Public Health 1991;82(3):162-7.

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.

Ministère de la Santé et des Services sociaux (MSSS). Bilan 1998-2003 – Programme québécois de dépistage du cancer du sein. Québec : MSSS; 2004.

Ministère de la Santé et des Services sociaux (MSSS). Programme québécois de dépistage du cancer du sein – Rapport d'activité 2000-2001. Québec : MSSS; 2003.

Ministère de la Santé et des Services sociaux (MSSS). Programme québécois de dépistage du cancer du sein – Cadre de référence. Québec : MSSS; 1996.

Mocharnuk RS. Screening mammography: The controversy continues. 25th Annual San Antonio Breast Cancer Symposium; 2002. Disponible à : www.medscape.com/viewprogram/ 2193_pnt.

Morrison BJ. Dépistage du cancer du sein. Dans : Groupe d'étude canadien sur l'examen médical périodique. Guide canadien de médecine clinique préventive. Ottawa : Santé Canada; 1994: 892-900.

Morrone D, Giorgi D, Ciatto S, Frigerio A, Catarzi S, Rosselli Del Turco M. [Assessment of diagnostic accuracy of mammography carried out for secondary prevention. Results of a test with a sample caseload conducted by 75 Italian radiologists]. Radiol Med (Torino) 2001;101(1-2):44-7 (article en italien).

Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. J Med Screen 1999;6(3):144-8.

Moss S, Thomas I, Evans A, Thomas B, Johns L, for the Trial Management Group. Randomised controlled trial of mammographic screening in women from age 40: Results of screening in the first 10 years. Br J Cancer 2005a;92:949-54.

Moss S, Waller M, Anderson TJ, Cuckle H, for the Trial Management Group. Randomised controlled trial of mammographic screening in women from age 40: Predicted mortality based on surrogate outcome measures. Br J Cancer 2005b;92:955-60.

Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. Lancet 2002;359(9310):909-19.

Nyström L, Larsson LG, Wall S, Rutqvist LE, Andersson I, Bjurstam N, Fagerberg G, Frisell J, Tabár L. An overview of the Swedish randomised mammography trials: Total mortality pattern and the representivity of the study cohorts. J Med Screen 1996;3(2):85-7.

Nyström L, Larsson LG, Rutqvist LE, Lindgren A, Lindqvist M, Ryden S, et al. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. Acta Oncol 1995;34(2):145-52.

Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: Overview of Swedish randomised trials. Lancet 1993;341(8851):973-8.

Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2001;(4):CD001877.

Paci E et Alexander FE. Study design of randomized controlled clinical trials of breast cancer screening. J Natl Cancer Inst Monogr 1997;(22):21-5.

Ringash J. Preventive health care, 2001 update: Screening mammography among women aged 40-49 years at average risk of breast cancer. CMAJ 2001;164(4):469-76.

Roberts MM, Alexander FE, Anderson TJ, Chetty U, Donnan PT, Forrest P, et al. Edinburgh trial of screening for breast cancer: Mortality at seven years. Lancet 1990;335(8684):241-6.

Roberts MM, Alexander FE, Anderson TJ, Forrest AP, Hepburn W, Huggins A, et al. The Edinburgh randomised trial of screening for breast cancer: Description of method. Br J Cancer 1984;50(1):1-6.

Roberts RA et Birch NJ. A comparison of breast cancer secondary prevention activities and satisfaction with access and communication issues in women 50 and over. Prev Med 2001;32(4):348-58.

Santé Canada. Programmes organisés de dépistage du cancer du sein au Canada : rapport de 1999 et 2000. Ottawa, Ont. : Ministre des Travaux publics et Services gouvernementaux Canada; 2003.

Shapiro S. Periodic screening for breast cancer: The HIP randomized controlled trial. Health Insurance Plan. J Natl Cancer Inst Monogr 1997;(22):27-30.

Shapiro S. Screening: Assessment of current studies. Cancer 1994;74(Suppl 1):231-8.

Shapiro S. The status of breast cancer screening: A quarter of a century of research. World J Surg 1989;13(1):9-18.

Shapiro S. Evidence on screening for breast cancer from a randomized trial. Cancer 1977;39 (Suppl 6):2772-82.

Shapiro S, Venet W, Strax P, Venet L. Periodic screening for breast cancer: The Health Insurance Plan project and its sequelae, 1963-1986. Baltimore, MD : Johns Hopkins University Press; 1988.

Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis in the Health Insurance Plan study: A randomized trial with breast cancer screening. Natl Cancer Inst Monogr 1985;67:65-74.

Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982;69(2):349-55.

Shapiro S, Strax P, Venet L. Evaluation of periodic breast cancer screening with mammography. Methodology and early observations. JAMA 1966;195(9):731-8.

Sickles EA. Screening outcomes: Clinical experience with service screening using modern mammography. Dans : National Institute of Health. NIH Consensus Development Conference, breast cancer screening for women ages 40-49, program and abstracts. Bethesda, MD: NIH; 1997 : 105-10.

Strax P, Venet L, Shapiro S. Value of mammography in reduction of mortality from breast cancer in mass screening. Am J Roentgenol Radium Ther Nucl Med 1973;117(3):686-9.

Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: Organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 2001;91(9):1724-31.

Tabár L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am 2000;38(4):625-51.

Tabár L, Vitak B, Chen HH, Prevost TC, Duffy SW. Update of the Swedish Two-County Trial of breast cancer screening: Histologic grade-specific and age-specific results. Swiss Surg 1999;5(5):199-204.

Tabár L, Chen HH, Fagerberg G, Duffy SW, Smith TC. Recent results from the Swedish Two-County Trial: The effects of age, histologic type, and mode of detection on the efficacy of breast cancer screening. J Natl Cancer Inst Monogr 1997;(22):43-7.

Tabár L, Fagerberg G, Chen HH, Duffy SW, Gad A. Screening for breast cancer in women aged under 50: Mode of detection, incidence, fatality, and histology. J Med Screen 1995a; 2(2):94-8.

Tabár L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, Smith RA. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer 1995b; 75(10):2507-17.

Tabár L, Fagerberg G, Day NE, Duffy SW. The Swedish two-county trial of mammographic screening for breast cancer: Recent results on mortality and tumor characteristics. Pathol Biol (Paris) 1992a;39(9):846.

Tabár L, Fagerberg G, Day NE, Duffy SW, Kitchin RM. Natural history of breast cancer. Lancet 1992b;339(8801):1108.

Tabár 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 1992c; 30(1):187-210.

Tabár L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: Recent results and calculation of benefit. J Epidemiol Community Health 1989;43(2):107-14.

Tabár L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985a;1(8433):829-32.

Tabár L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer. Results of the first seven years of mammography screening in Kopparberg, Sweden. Diagn Imaging Clin Med 1985b;54(3-4):158-64.

Tabár L et Gad A. Screening for breast cancer: The Swedish trial. Radiology 1981;138(1):219-22.

UK Trial of Early Detection of Breast Cancer Group. First results on mortality reduction in the UK Trial of Early Detection of Breast Cancer. Lancet 1988;2(8608):411-6.

UK Trial of Early Detection of Breast Cancer Group. Trial of Early Detection of Breast Cancer: Description of method. Br J Cancer 1981;44(5):618-27.

Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of recall rates with sensitivity and positive predictive values of screening mammography. Am J Roentgenol 2001;177(3):543-9.

Zahl PH, Kopjar B, Maehlen J. [Norwegian breast cancer mortality rates and validity of Swedish mammographic studies]. Tidsskr Nor Laegeforen 2001;121(16):1928-31 (article en norvégien).

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

Québec