THE UNIVERSITY OF CALGARY

# Current Use of Hormone Therapy and Screening Mammography Outcomes

by

Chandra Mary Thomas

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

SEPTEMBER, 1999

**Abstract**

Objective: The intent of this study was to determine if current use of hormone therapy (HT) was associated with a decreased performance of screening mammography.

Methods: An historical cohort design was used to assess the sensitivity, specificity, predictive values and likelihood ratios of screening mammography for a cohort of women who were currently using HT and a cohort of non-users. Women between 50 and 69 years screened through Screen Test: The Alberta Program for the Early Detection of Breast Cancer between October 1990 and March 1996 were included.

Results: The sensitivity, specificity, positive and negative predictive values of screening mammography were found to be significantly lower for current users of HT compared to non-users. The likelihood ratio positive was found to be lower with current use of HT while the likelihood ratio negative was found to be higher.

Conclusion: Current use of HT is a associated with a decrease in the sensitivity, specificity and positive predictive value as well as modifications in the likelihood ratios for screening mammography. These differences could result in a reduced ability of screening mammography to reduce mortality from breast cancer and in increased morbidity and financial cost through increased diagnostic follow-up for women who are current users of HT.

**Acknowledgments**

I would like to acknowledge the assistance of the staff at Screen Test – some who aided with my project, others who helped by making the environment where I worked on my thesis more hospitable. I would like to thank Bill Ghali and Sarah Rose whose guidance and feedback on my thesis project ensured that its completion was a positive experience. I would like acknowledge the apt advice of my supervisor Heather Bryant. while she never provided me with the answers she always pointed me in the appropriate direction to seek them out myself.

Finally, I would like to acknowledge the financial support I have received during my master's program through an Alberta Heritage Foundation for Medical Research Health Research Graduate Studentship and a MRC - Burroughs Wellcome Student Research Award.

## Dedication

I would like to dedicate this work to my family and friends who provided me with so much support through the whole process of completing my master's program.

## Table of contents

## List of tables

## List of figures

## List of abbreviations

HT _____ hormone therapy
BDS _____ breast density score
BI-RADS_ American College of Radiologists Breast Imaging Reporting and Data System
CEE _____ conjugated equine estrogen
CI _____ confidence interval
HMO _____ health maintenance organization
IQ _____ interquartile
LR _____ likelihood ratio
ppv _____ positive predictive value
PEPI _____ Postmenopausal Estrogen/Progestin Interventions
npv _____ negative predictive value
RR _____ relative risk
SE _____ standard error
SEER___ National Institutes of Health Surveillance, Epidemiology, End Results program

# 1. Introduction

## I. Study Problem

### A. Statement of the Problem

This study will provide evidence for determining if there is a difference in the performance of screening mammography for women currently taking hormone therapy (HT) compared to non-users in the context of an established, organized breast cancer screening program. In order to assess this problem, the sensitivity, specificity, predictive values and likelihood ratios of screening mammography for women who are current users and non-users of HT were determined and compared.

### B. Overview of the Problem

Studies have found a decreased sensitivity and specificity of mammography in women whose breasts are more radiographically dense (1-3). Studies have also shown that HT is associated with an increase in mammographic parenchymal density (4, 5). There is still controversy as to whether there is a reduction in the sensitivity, specificity and predictive values of screening mammography for women who are currently using HT (6-9).

HT is of particular importance due to the high prevalence of its use among women who would benefit from screening mammography. The literature on the prevalence of HT

use in Canada is limited. Only one study that assessed the prevalence of use of HT in a Canadian setting was identified in the literature (10). This study, which was conducted in 1986, found the prevalence of current use of HT to be 6% in a sample of approximately 2100 women aged 40 to 59 years. The prevalence of HT use amongst post-menopausal women in the study was 5% for women who underwent natural menopause and 16% amongst women who underwent surgical menopause. The results of this study were not reported in such a manner to determine the prevalence of HT use in women aged 50 to 69 years – those women for which screening mammography is known to be effective. Therefore the relevance of the results of this study to this project is limited. Furthermore, the study was conducted using a postal questionnaire achieving only a 68% response rate that leaves the results of the study vulnerable to selection bias.

Brett et al (11) conducted a population-based, prospective cohort study in the United States. They found an adjusted prevalence of 45% (95% CI 43% to 47%) for ever use of HT for at least one month. International studies have reported varying estimates for the unadjusted prevalence in the 1990s of ever use of HT: 5.2% in Japan (12), Great Britain - 19% (13) to 24.9% (14), Sweden - 40% (15), 41% (16). Variability in these estimates could be the result of inconsistent outcome measures or true variation in the use of HT between countries.

In the Alberta Screen Test Program, a population-based, breast-cancer screening program in which this study was conducted, 28% of participants report current use of HT.

Trends show that the prevalence of use of HT will only increase (11, 13, 17-19). Brett et al (11) found that among women who underwent natural menopause, prevalence of HT use increased from 11% for women becoming menopausal prior to 1945 to 46% of women who underwent natural menopause during the years 1990-92. In the same time period the prevalence of use of HT for women who had undergone a bilateral oopherectomy rose from 37% to 71%. Earlier American studies also show the trend of increasing use of HT over time (17, 18). The trend of increasing use of HT has been found in population-based studies conducted in international settings including Finland (19) and Great Britain (13).

In addition to the high and increasing prevalence of use of HT, there is a potential association of long duration of use of HT with an increased breast cancer risk (20-22).

Thus, there is the potential that there may be a reduced performance of screening mammography and potentially a decreased effectiveness of screening mammography for women who use HT. A reduced effectiveness of screening mammography for women using HT is important because it is potentially associated with a decrease in the ability for a screening program to reduce mortality from breast cancer in these women. Additionally,

decreased performance of mammography could lead to increased psychological stress for the women undergoing screening and economic costs for society resulting from both unnecessary diagnostic investigations and managing breast cancer that presents at a more advanced stage.

No study has been carried out to determine whether or not there is a change in the sensitivity, specificity and predictive values of screening mammography in the context of a North American population-based breast cancer screening program that meets the guidelines for an organized screening program. Information from this study will allow an assessment of what impact, if any, HT has on the performance of screening mammography, and will develop some scientific information on which future recommendations or patient information could be based.

## II. Background and Literature Review

### A. HT and increased parenchymal density

Changes in the parenchymal density of the breast in response to HT were first described by Peck and Lowman (23). Since that time several case series have described this effect (24-27) and several descriptive and analytic studies have found a significant association between radiographic breast density and use of HT (4, 5, 28-32).

Stomper et al assessed the parenchymal density of consecutive mammograms (5). They found that among women aged 50 to 79, 26% were using HT and there was a significant difference in the proportion of women with predominantly fatty breasts, defined as a parenchymal density of less than 50%, between women using HT and women not using HT. Seventy-eight per cent of non-users had predominantly fatty breasts compared to 51% of HT users ($p<0.01$).

Laya et al (26) conducted a prospective study and found that 73% of women who commenced use of HT had a quantitative increase in breast parenchymal density. A mean increase of 6.7% from baseline density measurement was observed (95% CI 2.5% - 11.0%; $p=0.003$). Parenchymal density was measured quantitatively by tracing areas of density and then measuring the area using a planimeter. Its small size (41 participants), no control group and the use of a sample of convenience limited this study.

McNicholas et al (30) used convenience sampling to assemble a cohort of HT users ($n=33$) and a cohort of non-users ($n=31$). Changes in density were measured quantitatively by drawing a line from the nipple to the chest wall bisecting the breast and then measuring the density of a circular region immediately above and below the line using a densitometer. Density was measured at baseline and then at follow-up mammography and a ten-percent change in the average of five readings of the density of the circular reading was considered to be a real difference. Change in density was also

measured subjectively and was classified as focal, multifocal or diffuse. A subjective increase in density was seen in 27% of the users of HT compared to none of the non-users ($p=0.002$). A quantitative increase in density occurred in 18% of current users. No information on the quantitative increase in density in non-users was provided.

The most important limitation of this study was the subjective measurement of change in breast density. There is no mention of whether or not the readers of the mammograms were blinded to HT status so it is possible that the measurement of density could be associated with HT status. Thus misclassification bias cannot be ruled out as an alternative explanation of the results.

Leung et al (31) assessed the parenchymal density of consecutive women attending for breast cancer screening at the Kingston General Hospital. The study population included women older than 54 years who were post-menopausal and were either using HT for at least six months or had never used HT. Breast density was determined using the breast density score (BDS). This scale was a derived scale (range 2 to 7) that was the sum of the volume of dense breast parenchyma rated on a scale of 1 to 4 and the density of this parenchyma rated on a scale of 1 to 3.

The study found that the mean BDS was significantly higher in women who were using HT even when stratified by age. Amongst women aged 55 to 65 years the mean

score was 4.7 for HT users compared to 3.7 for non-users ($p<0.001$). For women aged greater than 65 years the mean score was 4.8 for HT users compared to 3.2 for non-users ($p<0.001$). The proportion of women with low BDS scores (BDS=2-3) was greater for non-users compared to users (0.58 compared to 0.25; $p<0.001$) while the proportion of women with high BDS (BDS=6-7) was higher for HT users compared to non-users (0.37 compared to 0.11; $p<0.001$). These proportions were not adjusted by age.

This study was limited by employing a novel measure of mammographic density that had not been previously validated and having only one radiologist read the films and assign a BDS. However, the classification of density occurred independently of HT status so if there were any misclassification of density it would likely be non-differential and would serve to dilute the association between HT and breast density.

A study by Kaufman et al (28) compared the mammograms of 194 women who were currently using HT and had done so for a duration of at least five years and 216 mammograms of women taken prior to commencing use of HT. A modified version of the Wolfe classification method was employed to assess breast parenchymal density. Wolfe had classified mammographic parenchymal patterns into four categories which included primarily fat, ductal patterns occupying less than 25% of the breast volume, more extensive ductal patterns and dense stromal pattern without prominent ductal patterns (33, 34). For the study by Kaufman et al, these categories were collapsed into

two categories. Mammograms were classified as either low density (mammographic pattern composed predominantly of fat with prominent ducts occupying less than 25% of breast volume) or high density (ductal pattern occupying more than 25% of the breast volume). Women who had used HT for a duration of at least five years did not have a significant decrease in mammographic breast density pattern from high density to low density with increasing age ($p=0.088$) while the mammograms of women who had not yet commenced HT showed a significant decrease in density pattern with age ($p<0.001$).

The most convincing evidence of the effect of HT on mammographic parenchymal density was provided by retrospective cohort studies conducted by Persson et al (4) and Greendale et al (32).

**i. Persson et al 1997**

Persson et al found relative risks of an increase in parenchymal density (with 95% confidence intervals) to be 1.5 (0.7-3.6), 3.6 (1.6-7.7), 12.4 (6.3-24.4) and 1.5 (0.6-3.9) for estradiol only, estradiol with cyclically combined progestin, estradiol with continuously combined progestin and vaginal preparations respectively (4). Women who had been screened in the first two rounds of the Uppsala county breast cancer-screening program, a population-based screening program in a Swedish county, were included. Women who did not report use of HT at their initial screening visit but did report use on their subsequent visit were eligible for inclusion into the HT cohort. Of these women, a

random sample of women was selected. In addition, all women who reported that they were still menstruating and those reporting use of a continuously combined estradiol-progestin preparation were included. An age-matched control group was randomly selected from the women who did not report exposure to HT. Since women were selected via random selection and only a small number of women were excluded (n=4) it is unlikely that selection bias is a potential explanation for the results.

Mammograms were then assessed by a radiologist blinded to the participant's identity and HT status. Change in mammographic parenchymal density was determined using a subjective scale that included the categories of a moderate or slight decrease, no change, slight, moderate or substantial increase. Since a subjective measurement of breast density was employed the potential for measurement error exists. However, the determination of changes in density was done without knowledge of HT status so if there were any misclassification of density change it would likely be non-differential which would serve to dilute the association between use of HT and an increase in mammographic density.

Logistic regression modeling was used to adjust for the potentially confounding factors height, weight, parity, menstrual status and previous use of oral contraceptives.

Estradiol-only and vaginal preparations of HT were associated with a much smaller relative risk of an increase in mammographic density and these associations were not statistically significant. Given that the effect size in these subgroups is likely small, it is possible that there was inadequate statistical power in these subgroups to detect an association if one did exist. The paper did not present sufficient detail for a calculation of power to be carried out.

### ii. Greendale et al 1999

Greendale et al conducted a retrospective cohort study assessing change in mammographic parenchymal density with use of HT as a sub-study within the Postmenopausal Estrogen/Progestin Interventions (PEPI) (32). This study selected its participants from the women who were enrolled in the PEPI trial – a randomized, double-blind, placebo-control trial assessing the impact of HT on a variety of health outcomes. Women were randomized to one of five groups – placebo, conjugated-equine estrogen (CEE), or one of three CEE plus progestin combinations. Only women who had not used HT in the year preceding the study, had complete data on HT use, were adherent to treatment, did not have breast implants, had a baseline mammogram and at least one follow-up mammogram were included in the study. The final study population included 307 women.

Mammograms were classified by three radiologists using the Breast Imaging Reporting and Data System (BI-RADS). This system categorized mammograms as being entirely fatty (1), fatty with scattered fibroglandular tissue (2), heterogeneously dense (3) and extremely dense (4). Mammograms were assessed for an increase in BI-RADS grade. Women with an initial BI-RADS grade of 4 could not have a further increase in grade and were, therefore, excluded from the analyses (n=12). The proportion of women with an increase in BI-RADS grade was 0 (0-4.6) for the placebo group and 3.5 (1.0-12.0) for the CEE group and from 16.4 (6.6-26.2) to 23.5 (11.9-35.1) for the CEE plus progestin groups. A logistic regression analysis was carried out using an increase in BI-RADS grade as the outcome variable. This model controlled for clinic and uterine status and baseline BI-RADS grade, age, cigarette smoking, alcohol use, which could act as potentially confounding variables. Parity and body mass index were found not to contribute significantly to the model. Since there were no increases in the placebo group this group was not used in the adjusted analysis and the group using CEE was, therefore, used as the baseline group. The odds ratio of an increase in BI-RADS for women using one of the combined CEE-progestin preparations compared to CEE ranged from 7.2 (95% CI 1.3-40.0) to 13.1 (95% CI 2.4-73.3).

This was a methodologically strong study. However, it is impossible to rule out selection bias as an alternative explanation for the results. Although women in the PEPI

trial were randomized to hormone preparation, women were excluded for a number of reasons from the sub-study sample. If these exclusions were somehow linked to both hormone use and changes in breast density then selection bias could figure in the results. For instance, if women using HT were not adherent to treatment due to severe breast tenderness which resulted from rapid proliferation of breast tissue in response to use of HT then the association between HT use and an increase in breast density would be underestimated. The authors did not indicate which hormone preparation the women who were excluded from the study were using so it is impossible to determine what effect, if any, selection bias could have had on the study results. The radiologists who read the films were blinded to HT status so it is unlikely that differential misclassification could have occurred. However, given the subjective nature of the BI-RADS classification system it is possible that non-differential misclassification could have occurred. Misclassification of this sort would serve to dilute any association between HT use and change in parenchymal density. The results for CEE versus the combined preparations were adjusted for potentially confounding variables so it is unlikely that confounding could have explained these study results.

Additional, indirect evidence of the effect of estrogen on mammographic density has been provided by one study that has shown that a regression of mammographic densities after cessation of HT (35). Harvey et al described a regression of

mammographic densities following cessation of HT for a two-week period. In the study population described, routine cessation of HT was recommended in cases where a focal mass was found on mammography. Repeat mammography was then carried out in two weeks. Seventy-four percent (n=35) of focal masses regressed following the cessation of HT.

One study found no association between use of HT and increased mammographic breast density (36). This study, conducted by Erel et al, had a sample size of 108, thus had limited power to detect an association if one did exist. In addition, it was unclear whether the participants in the study were taking HT at the time the follow-up mammography took place. If HT was ceased prior to that point, changes in mammographic density could potentially have regressed.

Based on the available literature there is sufficient evidence to support that there is an association between HT and radiographic breast density.

## B. Decreased performance of mammography with increased radiographic breast density

A decreased ability to detect breast cancer in mammograms of women with a dense parenchymal pattern was first outlined by Holland in a small descriptive study that assessed mammographically occult breast cancer, breast cancers that are not visible by mammography (37). Subsequently, additional studies have found that radiographically

dense breast tissue has been associated with a decreased ability to detect breast cancer through mammography (1-3). Kerlikowske et al found that sensitivity of first screening mammography was significantly decreased in woman who were over 50 years who had primarily dense breasts compared to women who had primarily fatty breasts (83.7% compared to 98.4%; $p<0.01$) (2). Ma et al performed a case-control study where mammograms of women with histologically confirmed breast cancer over a ten year period were classified as to parenchymal density (3). Mammograms were classified as true positives or false negatives. It was found that increased breast parenchymal density was associated with a decreased ability to detect breast cancer by mammography (false-negative result) (OR=9; 95% CI= 1.8-44.3). Bird et al. in an analysis of cancers missed at screening mammography, found that missed cancers were more likely to occur in women with dense breasts ($p=0.046$) (1).

There is also evidence that confidence of diagnosis is reduced with increased breast density. A study found that radiologists who read mammograms were less certain of their diagnosis if there was a greater parenchymal density (38). The radiologists were asked to rate their certainty of diagnosis using a Likert scale. There was a significant correlation ($p<0.0001$) between increasing complexity of mammographic pattern and increased uncertainty of diagnosis.

One well designed study found that the decrease in cancer screening performance with increased mammographic breast density was not significant when adjusted for early versus later rounds of screening (39). However, this study only had 24.5% power to detect a reduction of 5% in sensitivity assuming a sensitivity of 86% in the cohort of women with radiolucent breast appearance on mammography.

A methodological problem that is common to many of the studies assessing the association between use of HT on the performance of screening mammography, or use of HT, and differences in radiographic breast density is the subjective nature of the measurement of radiographic breast density. In the Screen Test Program radiographic breast density is classified subjectively. While this is an important issue and a significant threat to study validity when radiographic breast density is an outcome measure, subjective measurement of breast density is not a limitation of studies that assess the effect of HT on the performance of screening mammography. These studies have as outcomes measures such as sensitivity, specificity which do not rely on information on radiographic breast density.

Despite the methodological limitations of the studies described, there is compelling evidence to suggest that the performance of mammography is impeded by increased mammographic breast density. Since use of HT is associated with increased

radiographic breast density it could potentially effect the performance of screening mammography.

## C. Potentially confounding factors

When carrying out a study that assesses the association between use of HT and the performance of screening mammography it is important to consider factors which could potentially confound this association. This allows for adjustment of estimates for these factors and ensures that the study results are not biased through confounding. Potentially confounding factors are factors that are independently associated with both the exposure and outcome of interest. In the context of this study, these factors are associated with use of HT and differences in the performance of screening mammography or breast density.

### 1. Factors associated with increased radiographic density on mammograms

Since increased breast density can lead to a decreased performance of screening mammography, it follows that any factors which lead to changes in breast density could potentially, or have been shown to, modify the performance of mammography. Age and menopausal status are two important factors that influence breast density. Following menopause the breast undergoes involution (40). This process is characterized mainly by atrophy of the secretory portions of the breast parenchyma and replacement with adipose tissue. The ducts and the interlobular connective tissue also atrophy but to a lesser extent. After age 30 the proportion of epithelial tissue, fibro-connective tissue and the number of

lobular units is decreased while the proportion of adipose tissue increased (41). These changes lead to a decreased radiographic appearance of the breast. Epidemiological studies have shown that decreased radiographic parenchymal density of the breast is associated with increasing age (5, 42-46) and post-menopausal status (44, 47).

In addition to age and post-menopausal status many other factors have been found to be associated with differences in radiographic breast density. Factors that have also been associated with lower radiographic breast density are increasing weight (adjusted for height) (42, 44, 47), larger breast size (42, 48), and increased parity (5, 44, 45). An increased radiographic parenchymal density has been associated with an increased age at first birth (5, 43, 45, 48, 49), nulliparity (43), a family history of breast cancer (45, 50), a history of previous breast biopsy (43, 45, 48, 51), and use of alcohol (52). Additionally, two studies have compared the parenchymal densities of Asian and Caucasian women and have found Caucasian women to have a significantly higher density (53, 54). A study conducted by White et al (49) found that the proportion of women with radiographically dense breasts was significantly greater for black/African-Americans and Asian/Pacific Islanders at 0.75 (95% CI 0.74-0.76) and 0.85 (95% CI 0.85-0.85) compared to white individuals at 0.67 (95% CI 0.67-0.67). However, this study was conducted in a pre-menopausal population and the results may not apply to post-menopausal women.

Of the factors described above several are also associated with use of HT and could therefore act as potentially confounding factors. Use of HT has been found to be associated with a leaner body mass index (weight in kilograms divided by height in meters squared) (11, 17, 18, 55), decreased parity (12, 55), prior use of oral contraceptives (13, 15, 56), ethnicity or country of origin (11, 14, 57) and the consumption alcohol regularly compared to those who do not drink (11, 18, 19, 58, 59). Also associated with use of HT, but not statistically significant, are increased age at first pregnancy or age at first birth (12, 60), and negative family history of breast cancer (59). These factors described above could potentially confound the association between use of HT and the performance of screening mammography.

## III. Relevance and Significance of the Project

### A. The effects of HT on the performance of screening mammography

Population based screening mammography has been shown to be efficacious in reducing mortality from breast cancer in women aged 50 to 69 (61). In a meta-analysis of studies which assessed the efficacy of mammography, Kerilowske et al found an overall relative risk for breast-cancer mortality for women aged 50 to 74 who underwent mammography compared to those who did not to be 0.74 (95% CI 0.66 to 0.83) (62). Changes in the performance of screening mammography measured via sensitivity,

specificity and predictive value of mammography could potentially lead to a change in effectiveness of screening mammography for women in this age range.

HT is increasingly being employed by women in this age group to prevent the side effects that accompany menopause and prevent the long-term complications of osteoporosis and cardiovascular disease. As described above there is compelling evidence that use of HT is associated with increased radiographic density of breast tissue. The implication of this increased density is the potential for a decrease in the performance of mammography and, subsequently, the failure to diagnose breast cancer and an increase in false-positive results and unnecessary follow-up.

To date five population based studies have attempted to assess whether the use of HT has a negative impact on the performance of screening mammography (6-9, 63).

**i. Laya et al 1996**

The association between use of HT and a decreased performance of screening mammography was assessed by Laya et al (6). They conducted a retrospective cohort study in the context of a health maintenance organization (HMO) in Puget Sound, WA. The study included women who were members of the HMO who did not have any breast symptoms prior to screening, were at least 50 years old and had undergone natural menopause or greater than 55 years old and had undergone a surgical menopause.

Women were required to have been members of the HMO for 12 months preceding and 12 months following screening mammography.

Two-view mammography was used and each woman's first mammogram in the program was included. Mammograms were classified as being negative, indeterminate and suggestive of a malignancy, with the latter two categories being classified as abnormal screening-mammography outcomes. The data from the HMO screening program were linked with the Seattle-Puget Sound Surveillance, Epidemiology and End Results (SEER) cancer registry on a quarterly basis in order to determine whether a diagnosis of breast cancer occurred within 12 months of the screening visit. Information on whether women were using HT was determined from responses to a self-administered survey and the HMO's pharmacy database.

Analysis consisted of comparing characteristics of the two cohorts – HT users and non-HT users – using t-tests for continuous variables and chi-squared tests for categorical variables. Logistic regression was then carried out to adjust sensitivity and specificity for potentially confounding factors. Included as potential confounding factors were age, history of first degree relative with breast cancer, previous mammogram outside the HMO, age at menarche less then 11 years, weight, height, and body mass index, age greater than 29 years at first full-term pregnancy, year of the mammogram in the HMO, current smoking and past oral contraceptive use of longer than one year. They found that

current users of HT were younger, had a lower body mass index, previous mammogram outside of the HMO, previous breast biopsy, surgical menopause, increased prevalence of nulliparity, late age at first birth and early menarche.

The overall breast cancer prevalence was found to be 7.2 per thousand with a higher prevalence amongst users of HT. However, this difference was not found to be statistically significant (p=0.14). In total 63 cancers were diagnosed of which 56 were true-positive results and 7 were false-negative results. The abnormality rate (number of mammograms that required further review) was 15.9%. Seventy-one percent of cancers were node negative (stages 0 or I) and 36% of cancers were less than 15 mm in size at detection. The percentage of women with abnormal results was 15.6%, 14.9% and 17.9% for never, former and current users of HT respectively.

The odds ratio of false-positive outcomes for current users of HT compared to never- or former-users adjusted for age, history of first degree relative with breast cancer and history of a mammogram outside of the HMO was found to be 1.33 (95% CI = 1.15-1.54). An adjusted specificity was calculated by entering the mean value of the important covariates into the logistic regression equation to arrive at a false-positive frequency and then subtracting this value from one. The adjusted specificity of screening mammography was found to be 86%, 86%, and 82% for former, never and current users of HT respectively.

Sensitivity was found to be lower in current users of HT compared to non-users of HT. However, this difference was not statistically significant. The study had limited power to detect a difference in sensitivity between the HT cohorts. If sensitivity was assumed to be 94% amongst non-users then the study had only 33% power to detect a reduction in sensitivity of 15% between the HT cohorts. Additionally, the small number of cancers precluded adjustment for potentially confounding factors.

While this is a robust study with little room for misclassification or selection bias it does have some limitations. Guidelines outlined for organized breast cancer screening programs (64) suggest that the following targets be met in order for the program to achieve maximum effectiveness:

- greater than 50% of screen detected tumors should be less than 15 mm.
- at least 30% of grade 3 tumors should be less than 15 mm.
- greater than 70% of tumors detected should be node negative.

The study by Laya et al was conducted using information from the prevalent screen so the program did not meet all of these guidelines. The guideline that greater than 70% of detected tumors be node negative was met – 71% of the tumors detected were stage 0 or I. However, only 36% of cancers detected were less than 15 mm in diameter. It is likely that the increased proportion of tumors of larger diameter is due to the fact that the prevalent screen for women in the program was used. The first screen captures longer

growing and, therefore, larger prevalent tumors where as subsequent screens would show incident and, therefore, smaller tumors.

In addition, in order to minimize screening costs both in exposing women to additional procedures and the financial cost of the program, it is recommended that the recall rate of a breast cancer screening program should not exceed 8% to 9% (64). The recall rate in this study was 15.9% (1387/8779).

For a program with an abnormality rate within the recommended level, if the same relative risk of false-positive results for current users of HT compared to non-users held, then the absolute difference in specificity between the cohorts would be smaller. In this study the false-positive rate was 14% for non-users and 18% for current users and the relative risk of a false-positive result for current users of HT was 1.3. An organized program should have an abnormality rate of less than 9%, which means that the false-positive rate can be no greater than 9%. If the false-positive rate were 7% for non-users and 9% for current users, the relative risk of a false positive result would be 1.3, which is the same as in the unorganized program. However, the absolute difference is only 2% in the organized program compared to 4% in the unorganized program. If the abnormality rate were lower, as is the case with many organized programs, the absolute difference between the cohorts would be even smaller. The estimate of the impact of current use of

HT on specificity in this study therefore does not apply to an organized screening program.

### ii. Thurfjell et al 1997

A study conducted by Thurfjell et al in the context of an organized screening program found a small decrease in specificity with current use of HT that was moderately significant and found no difference in the sensitivity with use of HT (7). This study included over 20, 000 women who were at least 50 years of age who had participated in the second round of screening in the breast screening program in Uppsala County, Sweden. Use of HT was determined via in person interview. The same definitions for true-positive, false-positive, true-negative, and false-negative results as employed by Laya et al were used.

The method of mammography used in this study differed from that used in the study by Laya et al. Women who were judged to have dense breasts on the first round of screening and those with unknown breast density had two-view mammography while all others had one-view mammography. Since women using HT could potentially have denser breasts, it is possible that women who were current or former users of HT underwent two-view mammography to a greater degree than never users. Fifty-eight percent of women who used HT, 45% of never users and 48% of former users were judged to have dense breasts and therefore received two-view mammography. If the

proportion of women with breasts of unknown density was the same in all of the groups then a greater proportion of women who were current users of HT were screened with two-view mammography than former or never users.

The study found a small difference in specificity between current and never-users of HT. The specificity was 94.0% (95% CI 93.3%-95.0%) for current users and 95.0% (95% CI 94.8%-95.5%) for never users. This difference was of borderline statistical significance (p=0.034).

The study did not find a statistically significant difference in sensitivity between the HT cohorts (p=0.69). Current users of HT had a sensitivity of 96% (95% CI 81.0%-99.1%) while non-users had a sensitivity of 91% (95% CI 85.7%-96.8%). The study had only 60% power to detect a difference in sensitivity of 15% if one did truly exist.

Additionally, no information on the means by which breast cancer diagnoses were determined was described in the study. Therefore, it is impossible to rule out the possibility that interval cancers may have been underascertained and the sensitivity overestimated.

The results of this study are limited due to the lack of adjustment for potential confounding variables including age.

### iii. Rosenberg et al 1998

Rosenberg et al assessed the impact of HT on the sensitivity and cancer stage at diagnosis in the breast cancer screening program in Albuquerque, New Mexico (63). The screening program consisted of several radiology groups in the Albuquerque area.

The study included mammograms from women screened between January 1991 and December 1994. Information on demographic characteristics, risk factors for breast cancer, type of examination performed, results of screening mammography and follow-up recommendations were contained in a common database. Breast-cancer diagnoses were determined through linking the database with the New Mexico Tumor Registry.

The unadjusted estimate of sensitivity was 81% (95% CI 77.6%-84.5%) for women who did not use HT and 74% (95% CI 67.7%-79.1%). These estimates included women aged less than 50 years. Sufficient information was provided in the study that allowed for the calculation of estimates of sensitivity for women aged 50 years of greater. For this age range the sensitivity was 83.2% (78.4%-87.3%) for women who did not use HT and 71.3% (63.6%-78.1%) for current users. Twenty-eight percent of women aged greater than 50 were not included in this estimate due to missing information on HT or breast density. The authors did adjust the sensitivity for potentially confounding factors. However, radiographic breast density was adjusted for as a potentially confounding factor. HT is purported to lead to a decrease in sensitivity through an increase in breast

density. Therefore, breast density is a component in the causal pathway by which HT would lead to a decrease in mammographic sensitivity and should not be adjusted for. Consequently, the adjusted estimate determined in this study is not valid.

This study had several weaknesses. The overall reported sensitivity of the screening program was low – only 80%. The quality of the screening program is, therefore, questionable. The quality of the data is suspect. The database is acknowledged by the authors to be incomplete as not all radiology groups provided information and the information that was provided was heterogeneous. Large proportions of women were excluded due to missing information. Thus, it is impossible to rule out selection bias. There may have been misclassification of HT status, which could result in an underestimation of the difference in sensitivity between the HT groups. Confounding cannot be ruled out. Information came from several radiology groups and no attempt was made to adjust for varying level of quality of interpretation from the groups. If one radiology group had poor sensitivity and a low prevalence of women who used HT then the difference in the sensitivity between the HT groups may be underestimated. The converse is also possible.

Additional concerns were that it was unclear if women with symptoms were excluded from the analysis and the study was not restricted to post-menopausal women and did not make an attempt to adjust for menopausal status.

### iv. Seradour et al 1999

Seradour et al conducted a study in the Bouches de Rhone screening program in France (8). The study included participants in the program who were screened between 1993 and 1996. Participants in the program were screened every three years with one medio-lateral oblique view. Information on HT status was recorded by the radiography technician at the time that mammography was performed. No information on type of HT preparation used or details of use was recorded.

Mammograms were classified as negative or positive and positive mammograms resulted in subsequent review. Cancers detected did not include lobular carcinoma in situ. A false-positive result was defined as no diagnosis of breast cancer in one year following an abnormal mammogram while a true-negative result was defined as the absence of breast cancer in women who had a negative screening result.

The study found that the sensitivity of screening mammography was 92% for non-users and 71% for current users of HT. The odds ratio of being diagnosed with a screen-detected cancer rather than an interval cancer was 5.14 (95% CI 2.5-11.8) for non-users of HT compared to current HT users. An analysis stratifying by age and first versus subsequent screening visit showed that there was no evidence that this association was confounded or modified by these factors. When stratified by age the specificity was found

to be 0.6% lower for HT users compared to non-users for subsequent screening visits but there was no difference at the first screening visit.

A potential limitation of this study is misclassification bias. Use of HT was self-reported. However, it is unlikely that this could be an alternative explanation of the study results. Since the report of HT use status occurred in advance of interpretation of the mammograms it is likely that if there were misclassification then the association between use of HT and a decreased in the performance of screening mammography would be underestimated.

The cancer registry from which interval breast cancer cases were ascertained was only estimated to be 90% complete. Missed interval cancer cases would result in an overestimation of the sensitivity. However, even if ten percent of interval breast cancer cases were missed amongst non-users of HT and none amongst current users there would be no appreciable overestimation of the sensitivity of non-users of HT given the small number of interval breast cancer cases (n=28).

Another weakness was that there was limited information on factors that could potentially confound the association between use of HT and the performance of screening mammography. The study was only able to control for age and first versus subsequent

screening visit. Several other potentially confounding factors of the association between use of HT and the performance of screening mammography have been identified.

### v. Litherland et al 1998

Litherland et al conducted a study within the West of Scotland Breast Screening Service which assessed the sensitivity of screening mammography for current and non-users of HT (9). This program provided one-view screening mammography to women between the ages of 50 to 64 every three years. The study included women screened between May 1988 and December 1995. The radiography technician ascertained current use of HT at the time of screening. The information consisted only of use or non-use of HT; no information on type of preparation or details of use was obtained. An interval cancer was defined as a breast cancer diagnosed within three years of a normal screening mammography result.

Current users of HT were more likely to have breast cancer diagnosed as an interval cancer ($p<0.001$). After using logistic regression to adjust the interval cancer rate for age, year of screening visit and a measure of social deprivation; the relative risk of a women using HT being diagnosed with an interval breast cancer compared to an non-users was 1.79 ($p=0.002$).

The relative risk of an interval cancer for current users of HT compared to non-users when only cancers diagnosed in the one-year interval following screening were included was 2.27 (p=0.003). When stratified by age, women aged 50 to 59 who were using HT were five times more likely to be diagnosed with breast cancer as an interval cancer (RR=4.94 (95% CI 2.42-10.10)) than non-users aged 60 to 64.

The authors provided sufficient information to calculate estimates of sensitivity for the HT cohorts. Non-users had a sensitivity of 96% (95% CI 95%-97%) while current users of HT had a sensitivity of 90% (95% CI 86%-94%).

One of the potential limitations of this study is misclassification bias. The completeness of the cancer registry from which interval cancers were ascertained was not stated. If there was incomplete ascertainment of interval cancers then the sensitivity estimates in this study could be overestimated. If the degree of under-ascertainment were the same in both cohorts then the difference in sensitivity between the cohorts would be underestimated. If the proportion of interval cancers that were not included in the cancer registry varied depending of HT use then the difference in sensitivity between the HT cohorts could have been over- or underestimated. Misclassification of HT status is also possible. If this occurred it would likely be non-differential as the status was obtained prior to the interpretation of the mammogram. Non-differential misclassification would result in an underestimation of the difference in specificity between the HT cohorts.

There was no information collected on menopausal status so it is possible that pre-menopausal women were included in the study population. Since pre-menopausal women would not be taking HT and have denser breasts, inclusion of these women would lead to an underestimation of the sensitivity in the cohort of non-users. Thus the difference in sensitivity between the HT cohorts would be underestimated.

One additional study which purported to assess the performance of screening mammography stated that there was no difference in the proportion of false-negative mammograms between women who used HT and non-users (65). The study included women who were diagnosed with breast cancer at their institution between January 1987 to May 1997 who were between the ages of 55 to 65, had all information regarding the dates of prior mammography and HT status and stage of the disease available (n=178). Of these women 115 had had at least one screening mammogram within 24 months before diagnosis. The authors use as outcomes the detection method for the breast cancer diagnosed which were mammography alone, palpation alone or mammography and palpation together.

These outcomes used by the authors are confusing. It is unclear what types of individuals were included in these groups. The authors state that the palpation alone group is equivalent to interval cancers that presented following normal mammograms. The mammography alone and palpation and mammography groups are unclear. The

authors do not state if the women in the mammography alone group had abnormal mammograms and breast cancer was detected as a result of an abnormal mammogram. For women in the palpation and mammography group the authors do not state if the women had a normal mammogram and then detected a palpable abnormality that was determined to be breast cancer after follow-up with mammography.

Given the confusion with the outcome measures and the uncertainty of who is included in the different categories, it is impossible to make any conclusions about the false-negative proportions from the data presented in this study.

A summary of the results and methodological limitations of the studies of reasonable quality that have assessed the impact of HT on the performance of screening mammography is provided below in Table 1.1.

**Table 1.1: Summary of results and methodological limitations of population based studies assessing the impact of HT on the performance of screening mammography**

| study | methodological limitations | specificity estimates | | | sensitivity estimates | | |
|---|---|---|---|---|---|---|---|
| | | current users | non-users | | current users | non-users | |
| Laya (6) | -small sample size (total=8779; cancers=63) | 82 (81-84) (n=2087) | 86 (85-86) (n=6729) | | 69 (38-91) (n=13) | 94 (87-100) (n=50) | |
| | -abnormality rate exceeds guideline for an organized screening program | | never 85 (84-86) (n=3792) | former 85 (84-87) (n=2837) | | never 94 (80-99) (n=34) | former 94 (69-99) (n=16) |
| Thurtjell (7) | -means of follow-up for breast cancers not described | 94 (93-95) (n=3009) | 95 (94.7-95.4) (n=16759) | | 96 (81-99) (n=27) | 91 (86-96) (n=115) | |
| | -no adjustment of potentially confounding factors including age -type of mammography may have differed between HT cohorts | | 95 (95-96) (n=14703) | 95 (93.7-95.6) (n=2146) | | never 91 (86-97) (n=103) | former 92 (62-100) (n=12) |
| Seradour (8) | -adjusted for only a few potentially confounding factors -cancer registry from which breast cancers were determined was only 90% complete | 96.1 (95.8-96.4) (n=14765) | 96.6 (96.4-96.7) (n=74130) | | 71 (60-82) (n=65) | 92 (89-95) (n=377) | |
| Litherland (9) | -completeness of cancer registry from which breast cancers were determined was not stated -population potentially included pre- and peri-menopausal women | | | | 90 (86-94) (n=243) | 96 (95-97) (n=1158) | |

Indirect evidence which supports an impact of HT on the performance of screening mammography was provided by a study assessing the number of women with abnormal screening results. A study conducted by Litherland et al which analyzed data from the National Breast Screening Study in the UK found that there was a significant decrease in the rate of recall ($p<0.016$) for women who did not use HT in the second round of screening compared to the first round of screening (66). However, there was no significant decrease in the recall rate for women using HT.

The published studies that have assessed the association between use of HT and the performance of screening mammography are insufficient in providing conclusive evidence that there is a change in screening mammography outcomes with use of HT. Some studies have found a significant difference in the specificity between HT users and non-users (6, 8) while another has not (7). A decrease in the sensitivity of screening mammography with current use of HT have been noted in several studies (6-9, 63). Only three of these studies found a significant difference in the sensitivity (8, 9, 63) and one of these studies one was of poor quality (63). One study found no difference in the false-negative proportion between users and non-users of HT (65). However, this study was of extremely poor quality.

Common limitations of these studies assessing the impact of use of HT on the performance of screening mammography are low statistical power to detect a difference in sensitivity between the HT cohorts and insufficient information to adjust for factors which could potentially confound the association between current use of HT and the performance of screening mammography. To date, there is no published information on a study assessing this association that has been conducted in North America in the context of a well established, organized breast cancer screening program with sufficient power to detect a difference in sensitivity between users and non-users of HT and sufficient information to adjust for potentially confounding variables that have been identified.

## B. Potential association between HT and breast cancer

In addition to a possible decreased effectiveness of mammography with use of HT, there is a potentially increased risk of breast cancer for women who use HT. The association between HT and breast cancer has been studied extensively. There is biological evidence that estrogen could potentially play an etiological role in breast cancer. A mitogenic effect of estrogen has been observed in animal models and in human tissue culture (67). In addition, many of the risk factors that have been associated with breast cancer are situations associated with an increased endogenous estrogen exposure – increased age at fist birth, nulliparity, early age at menarche, late age at menopause (68).

The potential for exogenous estrogen to function in the etiology of breast cancer is plausible and has been assessed in numerous epidemiological studies. There has been variability in the results of studies that have assessed the risk of breast cancer associated with the use of HT. However, a recent meta-analysis which re-analyzed individual data from 51 studies of an eligible 63 studies that assessed the association between HT and breast cancer found a significant association between HT and breast cancer (22). The relative risk of breast cancer in woman currently using HT or those who had used HT within five years was found to be 1.02 (95% CI 1.01-1.04) for each year of use. In addition, women who had currently or recently used HT had a relative risk of breast cancer of 1.35 (95% CI 1.21-1.49) if duration of use was greater than 5 years.

Women who use HT are potentially at an increased risk for breast cancer. If screening mammography is less effective for women who take HT there are large public health implications especially considering the already high, and increasing, prevalence of use of HT. A study that effectively examines the association between the current use of HT and changes in outcomes of screening mammography is, therefore, extremely important.

## IV. Study Objectives

### A. Research Objective

The aim of this study was to examine the effect of current use of HT on the performance of screening mammography in the context of a population-based, organized screening program with a high level of ongoing quality assurance.

### B. Research Question

Are differences in the specificity, sensitivity, predictive values and likelihood ratios of screening mammography associated with current HT use?

### C. Hypotheses

Current HT use is associated with a reduction in the specificity, sensitivity and predictive values of screening mammography in the context of a population-based screening program with ongoing quality assurance. Additionally, likelihood ratios for

breast cancer given a particular screening result will differ between users and non-users of HT.

# 2. Research Methods

## I. Study design

An historical cohort design was employed to compare the sensitivity, specificity, predictive values and likelihood ratios of screening mammography for a cohort of post-menopausal women who were currently using HT to that of another cohort who were not currently using HT.

## II. Setting

Alberta is a Canadian province with a population of approximately 2.8 million. In October of 1990, a population based breast cancer-screening program was initiated. This program operates under the auspices of the Alberta Cancer Board, the provincial agency responsible for the operation of cancer facilities and programs in Alberta. The target population for the program is women aged 50 to 69 years and these women are actively recruited to attend biennial screening through personalized letters of invitation signed by the director of the program. Although not actively recruited, women outside of the target population are accepted into the program. Women with a previous diagnosis of breast cancer and women with breast symptoms are not accepted into the program. However, in some instances, women with a history of breast cancer who reside in rural areas with limited access to mammography do receive mammography through the program on a

compassionate basis. Although initially excluded from the program, women with breast implants are accepted for screening.

The screening program operates two fixed sites in Calgary and Edmonton, the major cities in the province, with the remainder of the province being serviced through mobile mammography units. During the study period one mobile mammography unit was in operation.

## III. Screening process

At the time of registration women were asked about breast symptoms. Any women who indicated that they were currently experiencing any symptoms (i.e. lumps or nipple discharge) were encouraged to seek care through their primary care provider. Women who presented for screening were asked to complete a self-administered questionnaire (Appendix). This questionnaire included information on breast symptoms, risk factors for breast cancer, breast-health practices and demographic information. It also included a question on the name of a physician who would coordinate follow-up in the event that the woman was referred on for further diagnostic procedures.

Screening consisted of two-view mammography – craniocaudal and mediolateral-oblique views – using Lorad (Danbury, Conn) M3 units at the fixed sites and a Lorad Transpo unit at mobile sites. At the fixed sites in Calgary and Edmonton, mammograms

were developed on site with Kodak Min RE film and Min R medium screens while women waited. The films were assessed for any technical concerns and additional views or repeat films were taken if necessary. Films from mobile units were sent daily to a fixed site for development and, if necessary, women were recalled for additional views subsequent to development. In order to ensure consistent reading and interpretation, all films were read at the fixed site in Edmonton by radiologists with a dedicated interest in mammography.

The Canadian Association of Radiologists accredited all mammography sites and mobile units and on-going quality assurance took place according to the procedures outlined by the Canadian Association of Radiologists. In addition, regular Radiology Quality Assurance meetings assessed the interpretation results of the program.

The program also met the guidelines for organized screening programs as outlined above. From 1990 to 1996, node-negative breast cancers comprised 77.7% (278/359) of breast cancers detected through the program. Thirty five percent of nuclear grade three tumors were less than 1.5 cm in size (Personal Communication, Zeva Mah, Screen Test). Tumors less than 1.5 cm in size comprised 53.6% of screen detected tumors from April 1995 to March 1997 (69).

## IV. Screening results

Mammograms were classified as normal or "requiring further review." Women with normal results were advised to return at a 1- or 2-year interval. If a women was recommended to be screened at an interval less than one year this was considered an abnormal result. Approximately 6.5% of films read over the study period were determined to be abnormal ("requiring further review") and resulted in subsequent follow up. In addition to the screening results, fibroglandular tissue density was classified as either fatty, less than 25%, 25% to 49%, 50% to 74% or greater than 75%. Information on the screening result and breast density for each visit was recorded and in the event that follow-up occurred, data was obtained from the physician who the woman named on her questionnaire to coordinate care.

## V. Data sources

Information on Screen Test participants including screening results, follow-up diagnostic results and the information collected on the self-administered questionnaire was maintained in a relational database with controls for patient confidentiality. Routine linkage with the Alberta Cancer Registry, carried out quarterly, confirmed cancers, whether or not detected by screening.

Information on current HT use and potentially confounding factors was routinely collected through the self-administered questionnaire completed by each woman screened (Appendix).

## VI. Participants

The study included women who underwent screening through the Screen Test Program from October 1, 1990 to March 31, 1996. Follow-up for the occurrence of breast cancer was truncated at March 31, 1997. Allowing for a two-year period from the end of follow-up to the time that the study was commenced ensured that there was sufficient time for information on cancers associated with screening visits occurring late in the study period to be entered into the Alberta Cancer Registry and subsequently entered into the Screen Test database.

### A. Inclusion and Exclusion Criteria

The study included all women enrolled in the Screen Test program who were aged between 50 to 69 at at least one visit during the study period. The most recent screening visit was used for all women except for women who turned seventy during the study period. For these women the last screening visit before they turned 70 was employed. Fifty to 69 is the age range for which screening mammography has been found to be effective (62).

Women for which there was no information on HT status (n=163, 0.40%) were not included in the study. Women with a prior history of breast cancer were excluded. In addition, women who had a history of breast implants were excluded as the effectiveness of mammography for this group of women is likely different than women without breast implants.

Women who could have potentially been pre-menopausal or peri-menopausal were also excluded from the study population. In the Screen Test database information on menopausal status was available only as the age at menopause. If a women indicated that she had not yet experienced menopause, the age at menopause was indicated as being unknown or the age at menopause was missing then the age at menopause was unavailable. The likelihood that women without information on the age at menopause were pre-menopausal would be higher for younger women. Therefore, women without information on the age at menopause who were less than the age by which 95% of the population would be expected to be post-menopausal were excluded. These women were identified by assessing the upper limit of the 90% prediction interval for the mean age at menopause and the frequency of women with a missing age at menopause was assessed across one-year age strata. It would be expected that 95% of the study population would be post-menopausal by the upper limit of the 90% prediction interval for mean age at menopause (56 years). Women with a missing age at menopause whose age was less than

or equal to 56 years were excluded from the study population as potentially being pre-menopausal. The 95% prediction interval was calculated using the formula:

90% prediction interval = $\overline{X}_{\text{age at menopause}} \pm \left[ z_{\alpha=0.05} * \text{standard deviation} \left( \overline{X}_{\text{age at menopause}} \right) \right]$.

The final study population consisted of 37, 917 individuals.

## B. Definition of HT exposure status

HT use status was based on the woman's response to the question, "Are you currently taking estrogen (hormones for menopause)?" on the self-administered questionnaire (Question 12 of Appendix). Women who provided an affirmative response were classified as current users of HT (n=10, 880). Women who provided a negative response were classified as non-users (n=27, 037).

## VII. Study Outcomes

Screen Test mammograms were classified as being "normal" or "needing further review." Any mammogram that was designated as "needing further review" was considered to be an abnormal mammogram. A positive screening result was defined as any abnormal mammogram.

Breast cancer cases included any diagnosis of breast cancer, including ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) that were diagnosed

within 12 months of the date of the screening visit. A true-positive result was any abnormal mammogram associated with a diagnosis of breast cancer within the follow-up period, while a false-positive result was any abnormal mammogram that was not associated with a subsequent diagnosis of breast cancer within the follow-up period.

A negative result for a mammogram was any mammogram for which additional follow-up was not recommended. A true-negative result was any normal mammogram that was not associated with a diagnosis of breast cancer within the follow-up period. A false-negative screening result was any normal mammogram that was associated with a subsequent diagnosis of breast cancer in the 12 months following the date of the screening visit.

Sensitivity (**Se**) was defined as the probability that an individual with breast cancer would be detected by screening mammography. It is equal to the total number of true-positive results divided by the total number of breast-cancer cases diagnosed in the entire follow-up period. The total number of breast-cancer cases included both true-positive and false-negative results.

$$Se = \frac{\text{true positives}}{(\text{true positives} + \text{false negatives})}$$

Specificity (**Sp**) was defined as the probability that a woman without breast cancer has a normal screening result. It is equal to the number of true negative screens divided by the total number of women who did not have breast cancer. The total number of women without breast cancer is equal to the number of true-negative mammograms plus the false-positive mammograms.

$$Sp = \frac{\text{true negatives}}{(\text{true negatives} + \text{false positives})}$$

Positive predictive value (**ppv**) was defined as the probability that a woman with an abnormal mammogram would have breast cancer. It is equal to the number of true-positive results divided by the total number of positive screens.

$$ppv = \frac{\text{true positives}}{(\text{true positives} + \text{false positives})}$$

Negative predictive value (**npv**) was defined as the probability that a woman who has a normal mammogram did not have breast cancer. It is equal to the number of true-negative results divided by the total number of negative screens

$$npv = \frac{\text{true negatives}}{(\text{true negatives} + \text{false negatives})}$$

Likelihood ratios (LRs) for screening mammography were also calculated. A likelihood ratio was defined as the probability that a person with breast cancer had a specific screening-mammography outcome compared to the probability that a patient without breast cancer had the same screening outcome. A LR positive was the ratio of the probability of having an abnormal result given that the woman had breast cancer to the probability of having an abnormal result given that the woman did not have breast cancer. This was calculated from the equation $LR_{positive} = \frac{Se}{1 - Sp}$ where Se and Sp are the sensitivity and specificity of screening mammography respectively. A LR negative was the ratio of the probability of having a normal result given that the woman had breast cancer to the probability of having a normal result given that the woman did not have breast cancer. The likelihood ratio negative was calculated from the equation $LR_{negative} = \frac{1 - Se}{Sp}$.

## VIII. Data set

A data file including the variables from the Screen Test database that were relevant to the study analysis was obtained from the Coordinator of Evaluation of the Screen Test Program.

## IX. Data cleaning procedure

### A. Anomalous values

Frequencies for all variables were printed out and examined. Anomalous values were identified. Examples of these values included age at menopause greater than 60 years, age at first pregnancy greater than age at menopause, and age at menopause less than or equal to 25 years. The charts for the individuals with anomalous values were examined and any data entry errors were identified and corrected in the database. In instances where there were no data entry errors but the values provided by the woman were not plausible these values were set as missing in the data set.

### B. Age at menopause for women with a history of hysterectomy

Since age at menopause was one of the potentially confounding factors that would be adjusted for in subsequent analyses, it was necessary to assign an age at menopause to women who reported a history of hysterectomy. While many of these women did provide an age at menopause they often reported this as the time at which the hysterectomy took place. This study was concerned with menopause as defined as the end of the climacteric not a the cessation of menses. Since age at hysterectomy only corresponds with age at climacteric if a bilateral oopherectomy was performed, using the age at menopause provided by these women was not necessarily valid. Therefore, it was decided that women who had reported a hysterectomy would be assigned an age of menopause that

was equal to the mean age of menopause for the study population. In order to determine the mean age of menopause, women with a missing age at menopause and women who indicated a history of hysterectomy were excluded from the analysis and the mean was calculated. The mean age at menopause was determined to be 49.56 years (95% CI 49.50-49.62). The potential of a cohort effect existed for age at menopause; that is, women who were younger could have a younger age at menopause. For instance, women aged 50 to 54 years could either be pre- or peri-menopausal or could have experienced menopause only up to 54 years of age. Thus, it was conceivable that the mean age of menopause would be lowest for the youngest age strata and highest for the oldest age strata. The mean age was assessed for five-year age strata (Table 2.1). Women aged 50 to 54 had a mean age at menopause of 48.75 (95% CI 48.65-48.87). This was significantly lower than the other age strata. However, since the difference, although significant, was small the mean age for the study population was assigned to all women who had undergone a hysterectomy.

**Table 2.1: Mean age at menopause for whole sample and 5-year age strata**

| Age strata | n | mean (95% confidence interval) |
|---|---|---|
| whole sample | 20, 962 | 49.56 (49.50-49.62) |
| 50-54 | 4166 | 48.75 (48.65-48.87) |
| 55-59 | 5725 | 49.36 (49.75-49.97) |
| 60-64 | 5390 | 49.93 (49.80-50.05) |
| 65-69 | 5681 | 49.51 (49.38-49.63) |

## X. Analysis

Analyses were carried out using the statistical software packages SAS Version 6.12 and STATA Version 5.

### A. Characteristics of the study population

The demographic make-up of the study population was described. The self-reported ethnic origin of the study participants was collapsed in broad categories including Aboriginal, African, Asian, British, Eastern European, Northern European, Southern European, Western European and all other ethnic origins. Education level was divided into two categories – no post-secondary education and at least some post-secondary education. Proportions with 95% confidence intervals were used to describe self-reported ethnic origin, education level and age distribution in 5-year strata of the study population. The standard error of the estimates was calculated using the normal approximation to the binomial where appropriate and were calculated using the exact

method otherwise. These methods were used in all following instances when 95% confidence intervals for proportions were calculated.

Information from the 1996 Census of Canada was obtained and the self-reported ethnic origin and education level of the study population stratified by 5-year age strata were compared to the population of Alberta (70).

The proportion of women using HT for each of the demographic variables was also described.

## B. Screening visit outcomes

The outcomes of the screening visit including the abnormality rate and the radiographic breast density were described using proportions with 95% confidence intervals for the study population and the two HT cohorts. The abnormality rate was also stratified by age and first versus subsequent visit while radiographic breast density was stratified only by age.

## C. Breast cancers diagnosed

The screen detected and interval cancer rates were described as the number of cancers per 1000 women screened. The screen-detected-cancer rate was defined as the number of individuals with an abnormal mammogram associated with a diagnosis of

breast cancer in a one-year period following screening per thousand women screened. The interval cancer rate was the number of individuals with a normal screening result who had a breast cancer diagnosed within one year of screening per thousand women with a normal screening result. The cancer detection rates were also stratified by age and by screening visit.

The characteristics of the breast cancers diagnosed were described using proportions or means with 95% confidence intervals as appropriate. These characteristics included the proportion of stage 0 or I tumors, proportion of tumors less than 1.5 cm, and the mean tumor size. The cancers were classified as screen-detected (diagnosed within one year of an abnormal screening result) or interval cancers (diagnosed within in one year of a normal screening result). The characteristics of the breast cancers diagnosed in the Screen Test Program were compared to the targets for organized breast cancer screening programs outlined above.

## D. Outcomes measuring the performance of screening mammography

### 1. Unadjusted estimates

The sensitivity, specificity, positive predictive value and negative predictive value of screening mammography were described as proportions with 95% confidence intervals for the entire study population and by HT exposure cohort. Sensitivity, specificity and positive predictive value were then stratified by 5-year age strata and first versus

subsequent screening visit. P-values for the difference in the outcome measures between the HT cohorts were calculated using a chi-square test.

## 2. Adjusting estimates using logistic regression modeling

### (a) Potentially confounding variables

Variables that could potentially confound the association between the performance of screening mammography and current use of HT were identified from the literature. These included variables contained in the Screen Test database that had been associated with the use of HT and either a change in the effectiveness of screening mammography or radiographic breast density. These variables included a history of hysterectomy; history of a sister with breast cancer; maternal history of breast cancer; nulliparity; age at first birth greater than or equal to 30 years; menarche less than or equal to 11 years; previous breast biopsy; previous Screen Test mammogram; previous mammogram; history of oral contraceptive use; and ethnic origin.

Since only variables that were significantly associated with current use of HT in the study population would be adjusted for in the logistic regression model, associations between current use of HT and the potential confounding factors were assessed. Means and 95% confidence intervals, or medians with interquartile (IQ) ranges for variables with a non-normal distribution were used to describe the continuous variables for the entire study population and the two HT exposure cohorts. Proportions with 95% confidence

intervals were used to describe the categorical variables. P-values for the differences in variables between the exposure cohorts were determined using Student's t-test for continuous variables and chi-squared test for categorical variables. Differences in proportions for each of the estimates were also calculated. The standard error of the difference was calculated using the formula: $SE_{p_1-p_2} = \sqrt{\text{var}(p_1) + \text{var}(p_2)}$

$$\text{var}(p_n) = \frac{p_n(1-p_n)}{n_n}.$$

The relative risk (RR) of having the characteristic for current users of HT compared to non-users was calculated for the variables described above. The following formula was employed: $RR = \frac{p_{HT+}}{p_{HT-}}$ where $p_{HT+}$ is the proportion of individuals with the characteristic in the cohort of current users of HT and $p_{HT-}$ is the proportion of individuals with the characteristic in the cohort of non-users. The 95% CI of the estimate of the RR was calculated using the formula $\left(\text{RR} * \exp\left[-z\sqrt{v}\right],\ \text{RR} * \exp\left[z\sqrt{v}\right]\right)$ where

$$v = \text{var}(\ln RR) = \frac{1-p_{HT+}}{n_{HT+}} + \frac{1-p_{HT-}}{n_{HT-}} \quad (71).$$

It was possible that confounding or effect modification by age could account for any association between HT exposure status and the potentially confounding variables. A stratified analysis using 5-year age strata was carried out for all of the potentially

confounding variables that were found to be significantly associated with current use of HT.

**(b) Modeling procedure**

Following crude analysis, sensitivity, specificity, and predictive values were modeled using logistic regression techniques. Since logistic regression requires a dichotomous outcome, sensitivity, specificity and predictive values could not be modeled directly as the dependent variable. Therefore, logistic regression was carried out using the modeling method outlined by Coughlin et al (72). In this method the form $y = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n$ was used where $y$ was the log odds of the screening outcome and $x_1$ was the disease status for modeling sensitivity and specificity and $y$ was the log odds of disease and $x_1$ was the screening outcome for modeling predictive values. A normal screening outcome corresponded to a value of zero and an abnormal screening outcome to a value of one. A diagnosis of breast cancer was be given a value of one where as no diagnosis was be assigned the value zero. Recalling that the dependent variable of modeling sensitivity and specificity was the log odds of the screening outcome it followed that $\text{sensitivity} = \dfrac{1}{1+\exp\left[-\left(\alpha+\beta_1 x_1+\sum_{k=1}^{K}\beta_k x_k\right)\right]}$ where $x_1 = 1$, corresponding to a diagnosis of breast cancer. A 95% CI was calculated for specific values of the covariates ($x_j^*$) using the formula:

95% CI for SE at $x_j^* = \frac{1}{1+\exp\left[-\left(\alpha+\beta_1 x_1+\sum_{k=2}^{K}\beta_j x_j^*+1.96*\sqrt{\sigma}\right)\right]}$ where $\sigma$ was the variance of the log odds of a positive result with the specific values of the covariates. It is equal to the sum of the variance of each term plus two times the covariance of all possible pairwise combinations of terms based on the rules for calculating the variance of a linear combination or related variables (73). Mathematically this can be expressed as

$\text{var}\,\hat{L} = \sum_{j=1}^{k} x_j^2 \,\text{var}\,\beta_j + 2\sum_{all}\sum_{j<j'} x_j x_{j'} \,\text{cov}(\beta_j, \beta_{j'})$ where $\hat{L} = \sum_{j=1}^{k} \beta_j x_j$ is the predicted value of the linear combination of $k$ covariates with $\beta$ being the coefficients from the regression equation.

Specificity was modeled using the same form as sensitivity.

$$\text{Specificity} = 1-\left\{\frac{1}{1+\exp\left[-\left(\alpha+\sum_{k=1}^{K}\beta_k x_k\right)\right]}\right\}$$ where $x_1 = 0$. corresponding to no diagnosis of breast cancer. The

$$95\%\ \text{CI for Sp at}\ X_j^* = 1-\left\{\frac{1}{1+\exp\left[-\left(\alpha+\beta_1+\sum_{k=2}^{K}\beta_j x_j^*\right)\pm 1..96\sqrt{\sigma^2}\right]}\right\}.$$

Recalling that the dependent variable for modeling predictive values was the log odds of breast cancer, it followed that the positive predictive value (ppv)

$$= \frac{1}{1+\exp\left[-\left(\alpha + \sum_{k=1}^{K} \beta_k x_k\right)\right]}$$ where $x_1 = 1$, corresponding to a positive screening

outcome. The 95% confidence interval was calculated using an analogous formula to that used for sensitivity.

Since the negative predictive value for both cohorts was extremely high, greater than 99%, this outcome was not adjusted using logistic regression modeling.

Since Coughlin et al suggest that there is the potential for bias in the sensitivity estimate obtained using this modeling method, the validity of the method for modeling the outcomes in the study population was assessed. In order to assess whether or not there was bias in the estimates, models containing HT status as the only additional covariate were fit for each of sensitivity, specificity and positive predictive value. The estimates obtained via modeling were compared to those derived from the stratified analysis. Since there were differences in the estimates obtained by the two analysis methods it was determined that a correction method would need to be employed.

The correction methods suggested by Coughlin et al were to fit a disease covariate interaction term or to model sensitivity using only individuals with disease. Both of these

correction methods were attempted and the results from these modeling methods were compared to the results obtained from the stratified analysis. It was decided that modeling would proceed using only women with breast cancer to model sensitivity. The reason for this decision was that the results from the model with only women with breast cancer and the breast cancer*HT use interaction term yielded results that were identical to those obtained from the stratified analysis. However, since there were numerous covariates that were to be fit in the model, fitting a disease covariate interaction term for each covariate would make the model quite cumbersome.

In order to ensure that the estimates for all the outcomes were not biased, modeling for each of sensitivity, specificity and predictive values proceeded with fitting a model with all of the potential covariates, using only the appropriate individuals from the study population. That is, only women with breast cancer were used to model sensitivity, only women without breast cancer were used to model specificity and only women with abnormal results were used to model positive predictive value. The forms of the logistic regression equations were modified from those described above. For modeling sensitivity the model remained the same except that the breast cancer covariate term was dropped from the model. For modeling specificity the log odds of a normal screening result was modeled as the dependent variable (instead of an abnormal result). It followed the equations described above for calculating sensitivity could be employed in the same form

to calculate specificity. As was described in regards to sensitivity the breast cancer term was not included. Positive predictive value was modeled as described above except the screening result covariate term was not included.

For each of the outcomes, modeling began with fitting a full model with all of the potential covariates. The p-value of the Wald chi-square statistics and the coefficients of the covariates in the model were examined. Covariates whose coefficients were not significantly different from zero (e.g. had a p-value of greater than 0.05) were dropped sequentially from the model beginning with the least significant. Each time a covariate was dropped from the model the coefficient of the HT status term was examined to ensure that removal of a non-significant covariate did cause large changes in its value. If there were large changes in the coefficient the covariate was kept in the model. At the point when a reduced model containing only significant terms (or terms whose removal from the model led to a large change in the coefficient of the HT term) was reached, interaction terms which contained covariates which contributed significantly to the model were fit into the model. Interaction terms considered were interactions between age and each of previous Screen Test mammogram, ethnic origin, hysterectomy, history of a sister with breast cancer, history of previous breast biopsy and HT use. Once again the p-value for the Wald chi-square statistics and the coefficients were examined and the procedures described above were applied to arrive at a parsimonious model.

The fit of the final model was assessed using the Pearson chi-square and the Hosmer-Lemeshow test. If the test statistic was not statistically significant (e.g. p-value>0.05) it was concluded that there was insufficient evidence to suggest that there was lack of fit between the model and the data. Since the Hosmer-Lemeshow test may have a significant result even when the model is well fit when the sample size is large the observed and expected values used to calculate the test were examined.

Before accepting a model, regression diagnostics were assessed. These included leverage of covariate patterns, change in Pearson residuals, change in deviance residuals and the Δbeta influence statistic.

Diagnostic plots including plots were also examined in order to assess if there were any covariate patterns for which there was lack of fit.

Covariate patterns that were poorly fit were assessed to determine if the patterns were plausible. These generally included patterns with a leverage greater than 0.1, change in Pearson residuals greater than 5, change in deviance residuals greater than 5 and Δbeta influence statistic greater than 0.5. However, the criteria were loosely applied. The proportion of individuals in the data contained in the poorly fit patterns was assessed. If this proportion was low (less than 15 to 20%), even if there were significant scores for the

Pearson chi-square or Hosmer Lemeshow test, the model was considered to be sufficiently well fit.

After a final model was adopted, conditional effect plots showing the effect of HT use status for all possible covariate patterns were plotted.

## E. Sensitivity analyses to quantify impact of misclassification

In order to quantify the potential impact of misclassification on the sensitivity estimates, sensitivity analyses assessing the impact of misclassification of false-negative status and HT use status were carried out.

### 1. Impact of misclassification of false negative results on sensitivity

The sensitivity estimates in this study were calculated using what is known as the detection method where false-negative results are cancers diagnosed in one year following a normal screening result. The sensitivity obtained using this method may be underestimated. This underestimation may occur because the individuals classified as having false-negative results may include two types of individuals. The first group is individuals who had a breast cancer that was of the size and character that it could have been detected mammographically at the time of mammography but this cancer was not detected. The second group is women who did not have a breast cancer that was of the size and character that it could have been detected mammographically at the time of

mammography but had a tumor grow quickly in the interval proceeding screening to the point where it was clinically detectable by other means.

In order to assess for the impact of the latter group of women being classified as women having false-negative results, a sensitivity analysis assessing the impact of misclassification was undertaken. Corrected sensitivities were calculated for each of the HT cohorts.

It was assumed that all of the true-positive results in the study were correctly classified and the misclassification of new cancers as false-negative results varied from zero up to 25%. The corrected number of false-negative results was determined using the equation: $FN_{corrected} = (1 - p) * FN_{estimated}$ where $p$ is the proportion of false-negative results misclassified and $FN_{estimated}$ is the number of individuals with a diagnosis of breast cancer in the one year following a normal screening-mammography result. The corrected sensitivity was then calculated using the equation: $Se_{corrected} = \frac{TP}{TP + FN_{corrected}}$ where $TP$ is the number of individuals who had a diagnosis of breast cancer in the one-year interval following an abnormal screening mammography result.

**2. Impact of misclassification of HT status on sensitivity**

Since, it was unlikely that the classification of HT exposure status was completely accurate, a sensitivity analysis was undertaken to quantify the impact of misclassification

of HT exposure status on sensitivity. The methods described by Rothman were employed (74). In all cases misclassification was assumed to be independent of outcome. Thus, for sensitivity the overall number of true-positive results and the number of false-negative results remained constant. The corrected number of true-positive results for each of the HT cohorts were determined using the formulae:

$$TP_{HT+\,corrected} = \frac{\left[TP_{HT+\,estimated} - (1-Sp)TP_{total\,estimated}\right]}{(Se+Sp-1)} \text{ and}$$

$TP_{HT-\,corrected} = TP_{total\,estimated} - TP_{HT+\,corrected}$ where $TP_{HT+\,estimated}$ is the number of individuals with a true-positive result in the HT cohort. $TP_{total\,estimated}$ is the total number of women with a true-positive result and Se and Sp are the sensitivity and specificity of the classification of HT exposure status. The corrected number of false-negative results was calculated using analogous formulae. The corrected sensitivity for each of the cohorts was then calculated using the equation: $Se_{corrected} = \frac{TP_{corrected}}{TP_{corrected} + FN_{corrected}}$. The sensitivity analysis was carried under two scenarios. The first was the assumption that there was no misclassification of false-negative status and the second was that there was 25% misclassification among current HT users and no misclassification in the non-users.

## F. Likelihood ratios and post-test probabilities

Likelihood ratios (LRs), for positive and negative results, were first calculated for the study population and the two HT cohorts. Stratum specific LRs were calculated following stratification by age and first versus subsequent visits. The confidence intervals for the estimates of the LRs were calculated using the methods described above for calculating for confidence intervals for relative risks.

LRs were then employed to determine the risk of being diagnosed with breast cancer following mammography – the post-test probability. The risk of breast cancer prior to screening mammography was assumed to be equal to the cancer detection rate in the one year following screening in the study population. this included both screen-detected and interval cancers. The risks were stratified by age and first versus subsequent screening visit. The prior probabilities were converted to prior odds using the formula: $\text{prior odds} = \frac{\text{prior probability}}{1 - \text{prior probability}}$. The post-test odds of being diagnosed with breast cancer within one year of screening were determined by multiplying the pre-test odds by the LR for both positive and negative screening results. The post-test odds were then converted to a probability using the formula: $\text{posterior probability} = \frac{\text{posterior odds}}{1 + \text{posterior odds}}$. The 95% confidence interval for the post-test probability was calculated by calculating

the post-test probability using the lower and upper bounds of the 95% confidence interval which had been calculated for the LR.

# 3. Results

## I. Overview of results

This results section is divided into three sections. The first section describes the study population and HT cohorts. This section includes a description of the demographic characteristics of the study population, the screening visit outcomes and the breast cancers diagnosed.

The main study outcomes, those measuring the performance of screening mammography, are described in the next section that starts on page 94. The description includes, firstly, unadjusted estimates and estimates adjusted using stratification for age and first versus subsequent screening visit. The association between current use of HT and potentially confounding variables and the estimates for outcomes measuring the performance of screening mammography adjusted for these factors using logistic regression modeling are then presented. A sensitivity analysis examining the effect of misclassification of the false-negative results and HT status on the sensitivity estimates is also presented.

Finally, likelihood ratios for the two HT cohorts, unadjusted and those adjusted using stratification, are presented. The post-test probabilities of breast cancer that were determined using the likelihood ratios are then described.

## II. Description of Study population and HT cohorts

### A. Definition of study population and HT cohorts

There were 40, 676 age eligible women with no history of breast implants or breast cancer who were screened through the Screen Test Program during the study period. Of these women, 163 women were excluded because they had incomplete information on HT status due to missing responses (n=146) or a response that the status was unknown (n=17).

An additional 2, 596 women were excluded because they could have potentially been pre-menopausal. Since, the intent of this study was to examine the impact of current use of HT on the performance of screening mammography in post-menopausal women it was important that pre-menopausal women were not included. Examining the mean age of menopause and the proportion of women for which an age at menopause was not available identified these potentially pre-menopausal women. Since the mean age at menopause in the study population was 49.56 with a standard deviation of 4.45 years, it was expected that 95% of the population would be post-menopausal by age 57 as this was the upper limit of the 90% prediction interval for the mean age at menopause. The proportion of women with an unavailable age at menopause was also examined by one-year age strata (Table 3.1). In the Screen Test database information on menopausal status was available only as the age at menopause. If a women indicated that she had not yet

experienced menopause, the age at menopause was indicated as being unknown or the age at menopause was missing then the age at menopause was not available. The proportion of women for which the age at menopause was unavailable was high for younger women. Fifty-five percent of women aged 50 years did not have information available. With each increasing year of age the proportion unavailable decreased significantly until 56 years where only 13% of women had unavailable information. For women aged 57 years or older there was no significant difference in the proportion unavailable across age strata. For younger women the reason for the unavailable information on age at menopause was likely to be due to the fact that these women had not yet experienced menopause. The unavailable information is more likely to be due to a missing or an unknown age at menopause for the older women. Interestingly, the age after which there were no significant decreases with increasing age in the proportion of women for which the age at menopause was unavailable, 57 years, was the age by which 95% of the study population would be expected to be post-menopausal.

**Table 3.1: Proportion of unavailable values for age at menopause by one-year age strata**

| age stratum | n | proportion unavailable (95% confidence interval) |
|---|---|---|
| 50 | 884 | 0.55 (0.52-0.59) |
| 51 | 994 | 0.47 (0.43-0.50) |
| 52 | 1473 | 0.37 (0.34-0.39) |
| 53 | 1587 | 0.28 (0.26-0.31) |
| 54 | 1473 | 0.20 (0.18-0.22) |
| 55 | 1317 | 0.14 (0.12-0.16) |
| 56 | 1263 | 0.13 (0.11-0.15) |
| 57 | 1301 | 0.09 (0.08-0.11) |
| 58 | 1253 | 0.09 (0.07-0.10) |
| 59 | 1266 | 0.08 (0.06-0.09) |
| 60+ | 12033 | 0.08 (0.08-0.08) |

The number of women excluded is described below in Figure 3.1.

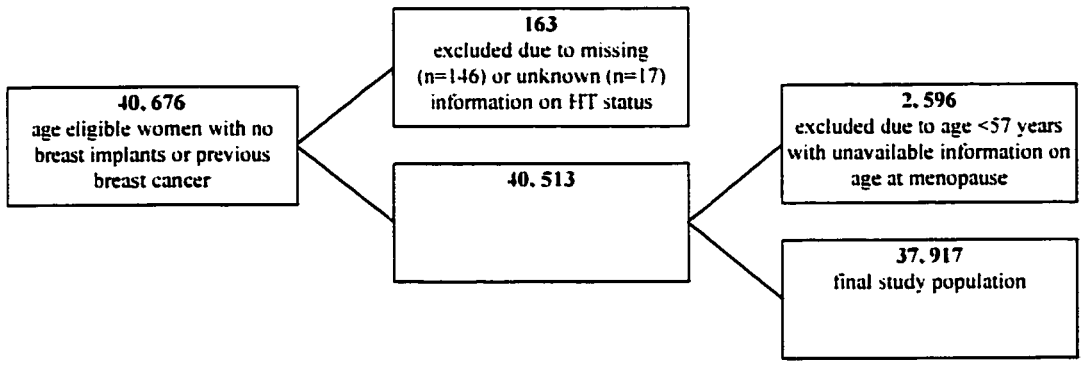


**Figure 3.1: Numbers of study participants excluded and reasons for exclusion**

Of the 37, 917 women in the study population 10, 880 (28.7%) were currently using HT. These women comprised the HT cohort.

## B. Characteristics of the study population

### 1. Demographic characteristics

Demographic characteristics of the study population including age, education and ethnic origin were examined and compared with the general population of Alberta using information from the 1996 Census of Canada. The HT use across demographic characteristics was also examined.

#### (a) Age

The mean age of the study population was 60 years. The distribution of the study population across 5-year age strata is described in Table 3.2. Twenty percent of the study population was aged 50 to 54 years with the rest of the population being evenly distributed over the age strata from 55 to 69 years.

**Table 3.2: Age distribution of study population**

| Age | number of women | % of study population (95% CI) |
|---|---|---|
| 50-54 | 7,651 | 20.18 (19.77-20.58) |
| 55-59 | 10,095 | 26.62 (26.18-27.07) |
| 60-64 | 9,907 | 26.13 (25.69-26.57) |
| 65-69 | 10,264 | 27.07 (26.62-27.52) |

A comparison of the proportion of women in 5-year age strata for the Screen Test population and the population of Alberta is presented in Table 3.3. There were

significantly fewer women aged 50 to 54 years in the Screen Test population compared to the general population and significantly more women aged 55 to 69 years.

**Table 3.3: Age of study population compared to population of Alberta**

| Age | % of study population (95% CI) | % of Alberta population (95% CI) |
|---|---|---|
| 50-54 | 20.18 (19.77-20.58) | 32.50 (31.30-33.71) |
| 55-59 | 26.62 (26.18-27.07) | 24.53 (23.43-25.65) |
| 60-64 | 26.13 (25.69-26.57) | 22.24 (21.20-23.33) |
| 65-69 | 27.07 (26.62-27.52) | 20.74 (19.70-21.80) |

The mean age of current users of HT was 59 years while the mean age of non-users was 61 years. This difference, although small, was highly significant ($p<0.0001$). The proportion of women using HT for 5-year age strata was also examined and is presented in Table 3.4. Younger women were more likely be current users of HT. The percent of users was 39% (95% CI 38.27%-40.46%) and 35% (95% CI 34.11%-35.97%) for women aged 50 to 54 years and 55 to 59 years respectively while only 14.05% (95% CI 13.54%-14.56%) of women aged 60 to 64 years and 17.91% (95% CI 17.17%-18.65%) of women aged 65 to 69 years were current users of HT.

The difference in use of HT with age is important to consider. Since younger age is associated with increased use of HT and a potentially poorer performance of screening mammography, age could act as a potentially confounding factor for this association.

Therefore, the effect of age on the performance of screening mammography was considered in subsequent analyses.

**Table 3.4: Hormone use by 5-year age strata in study population**

| Age strata | number of women | % currently using HT (95% CI) |
|---|---|---|
| 50-54 | 3012 | 39.37 (38.27-40.46) |
| 55-59 | 3537 | 35.04 (34.11-35.97) |
| 60-64 | 2493 | 14.05 (13.54-14.56) |
| 65-69 | 1838 | 17.91 (17.17-18.65) |

### (b) Self-reported ethnic origin

The self-reported ethnic origin of the study population is described in Table 3.5. Approximately 92% of the study population was comprised of women with British or European ancestry. Aboriginal women, a group that is considered to have limited access to health care resources compared to the general population (75), comprised 1.2% of the study population.

**Table 3.5: Self-reported ethnic origin for the study population**

| Self-reported ethnic origin | number of women | % of study population (95% CI) |
|---|---|---|
| Aboriginal | 441 | 1.18 (1.08-1.30) |
| African | 205 | 0.55 (0.48-0.63) |
| Asian | 2, 416 | 6.48 (6.23-6.74) |
| British | 14, 952 | 40.11 (39.61-40.61) |
| Eastern European | 5, 816 | 15.60 (15.23-15.97) |
| Northern European | 1, 999 | 5.36 (5.14-5.60) |
| Southern European | 733 | 2.07 (1.93-2.22) |
| Western European | 9, 903 | 26.56 (26.12-27.02) |
| other | 773 | 2.08 (1.93-2.23) |

A comparison of the screening population with the general population for selected ethnic origins by 5-year age strata is presented in Table 3.6. Compared to the general population there were significantly more women of British and European descent in the Screen Test population for all age strata. There did not appear to be evidence that Aboriginal women were under-represented in the Screen Test population. The proportion of Aboriginal women in the screening population was similar to the general population except for women age 55 to 59 years where there were significantly fewer Aboriginal women in the screening than in the Alberta population.

**Table 3.6: Comparison of selected self-reported ethnic origin in study population to Alberta population**

| Ethnic origin | % of study population (95% CI) | % of Alberta population (95% CI) |
|---|---|---|
| British and European | | |
| 50-54 | 88.63 (87.89-89.34) | 76.37 (74.39-78.26) |
| 55-59 | 88.58 (87.94-89.20) | 75.12 (72.80-77.34) |
| 60-64 | 89.51 (88.89-90.11) | 77.71 (75.35-79.94) |
| 65-69 | 91.82 (91.27-92.35) | 77.82 (75.38-80.13) |
| Aboriginal | | |
| 50-54 | 1.74 (1.46-2.06) | 1.42 (0.94-2.06) |
| 55-59 | 1.43 (1.21-1.69) | 2.64 (1.88-3.61) |
| 60-64 | 1.06 (0.86-1.28) | 1.85 (1.19-2.73) |
| 65-69 | 0.64 (0.50-0.82) | 0.99 (0.51-1.72) |

The proportion of use of HT across self-reported ethnic origins is described in Table 3.7. The proportion of use was highest for women of British, 30.95% (95% CI 30.21%-31.69%), and Western European, 31.14% (95% CI 30.23%-32.05%), descent. This is the group that comprised the majority of the study population. Aboriginal women had a proportion of use of HT of 26.76% (95% CI 22.63%-30.89%) which is in line with the proportion using HT for most of the other ethnic groups. Asian women, with a proportion of use of HT of 11.88% (95% CI 10.59%-13.17%), were much less likely to use HT than all of the other ethnic groups.

The significant differences in the use of HT across ethnic groups are important. There is evidence that radiographic breast density, and potentially the performance of screening mammography, differs in women of different ethnic groups. Since ethnic origin is associated with both use of HT and radiographic breast density it could potentially confound any association between use of HT and the performance of screening mammography. Therefore, outcomes measuring the performance of screening mammography were adjusted for ethnic origin in subsequent analyses.

**Table 3.7: Hormone use by self-reported ethnic origin**

| self-reported ethnic origin | number of women | % currently using HT (95% CI) |
|---|---|---|
| Aboriginal | 118 | 26.76 (22.63-30.89) |
| African | 53 | 25.85 (19.86-31.85) |
| Asian | 287 | 11.88 (10.59-13.17) |
| British | 4628 | 30.95 (30.21-31.69) |
| Eastern European | 1604 | 27.58 (26.43-28.73) |
| Northern European | 585 | 29.26 (27.27-31.26) |
| Southern European | 169 | 21.86 (18.95-24.78) |
| Western European | 3084 | 31.14 (30.23-32.05) |

### (c) Education level

In the Screen Test population 29% of the women had at least some post-secondary education. A comparison of the education level for the Screen Test population and the Alberta population is presented in Table 3.8. For women of all ages, there were fewer women in the Screen Test population with at least some post-secondary education

compared to the general population of Alberta. However, the differences between the two populations were only significant for women aged 50 to 59 years.

**Table 3.8: Comparison of percentage of women with at least some post-secondary education in study population compared to Alberta population**

| Age | % of study population (95% CI) | % of Alberta population (95% CI) |
|---|---|---|
| 50-54 | 37.67 (36.58-38.76) | 46.19 (43.93-48.46) |
| 55-59 | 31.12 (30.22-32.03) | 36.10 (33.61-38.64) |
| 60-64 | 26.72 (25.85-27.61) | 29.90 (27.42-32.47) |
| 65-69 | 24.01 (23.18-24.85) | 27.10 (24.63-29.72) |

Women who had at least some post-secondary education were significantly more likely to use HT than women with no post-secondary education. Thirty-five per cent (95% CI 34.41%-36.21%) of women with some post secondary education used HT compared to 25.98% (95% CI 25.16%-26.81%) of women with no post-secondary education.

**Table 3.9: Hormone use by education level in study population**

| Education level | number of women | % currently using HT (95% CI) |
|---|---|---|
| no post-secondary | 6940 | 25.98 (25.16-26.81) |
| at least some post-secondary | 3922 | 35.31 (34.41-36.21) |

## C. Screening visit outcomes

### 1. Abnormality rate and radiographic breast density

There were significant differences between the cohorts for the outcomes of the screening visit. These results are described in Table 3.10. Overall, 6.20% (95% CI

5.95%-6.44%) of mammograms performed were determined to be abnormal and resulted in subsequent follow-up. Women who were current users of HT were more likely to have an abnormal mammography result ($p<0.001$). The relative risk of an abnormal mammogram for current users of HT compared to non-users was 1.23 (95% CI 1.13-1.34). The relative risk of having a fatty radiographic breast appearance which was defined as having less than 25% fibroglandular tissue density was 0.7 (95% CI 0.6-0.7) for current users compared to non-users. The absolute difference between the HT cohorts was 16% ($p<0.001$).

**Table 3.10: Results of screening visit**

| screening outcome | non-users % (95% CI) | current users % (95% CI) | difference between cohorts* % (95% CI) | p-value | relative risk# |
|---|---|---|---|---|---|
| abnormality rate | 5.8 (5.5-6.1) | 7.2 (6.7-7.6) | 1.3 (0.8-1.9) | <0.001 | 1.2 (1.1-1.3) |
| fatty breast density | 48.1 (47.5-48.7) | 31.8 (31.0-32.7) | -16.2 (-17.3-(-15.2)) | <0.001 | 0.7 (0.6-0.7) |

*The difference between the cohorts was the proportion of women with the outcome amongst current users of HT minus the proportion amongst non-users.
#The relative risk was calculated using non-users of HT as the referent group.

## 2. Abnormality rate stratified by first versus subsequent screen

The abnormality rate of screening mammography is lower for the subsequent screening visit compared to the first visit and in some populations women who use HT have been found to be more likely to engage in preventive health practices (12, 14, 18,

60). Thus it was possible that the differences in abnormality rate between the HT cohorts could have been the result of confounding or effect modification by previous mammographic screening. Therefore, an analysis stratifying by screening visit was carried out. These results are presented in Table 3.11.

Confounding and effect modification were assessed for by examining the crude and stratum specific relative risks. If the crude relative risk, the risk in the unstratified study population, differed from the stratum specific relative risks and the stratum specific relative risks were homogenous, then it was concluded that confounding by age was operating. If there was heterogeneity across the stratum specific relative risks then it was concluded that age was acting as an effect modifier.

The stratified analysis showed that screening visit modified the risk of an abnormal mammogram. This effect modification did not explain the elevated risk of an abnormal mammogram amongst current HT users as the relative risk of an abnormal mammogram was significantly higher than the null value of one for current users of HT compared to non-users for both first and subsequent screening visits. The relative risk of an abnormal mammogram for current users of HT compared to non-users increased from 1.2 (1.1-1.3) at the first screening visit to 1.7 (1.4-1.9) for the subsequent visit.

**Table 3.11: Risk of abnormal mammogram by HT status stratified by first or subsequent screening visit**

| | HT cohort | | difference between cohorts* | p-value | Relative Risk# |
|---|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | | (95% CI) |
| study population | 5.8 (5.5-6.1) | 7.2 (6.7-7.6) | 1.3 (0.8-1.9) | <0.001 | 1.2 (1.1-1.3) |
| first visit | 7.9 (7.5-8.4) | 9.3 (8.5-10.1) | 1.4 (0.5-2.3) | 0.002 | 1.2 (1.1-1.3) |
| subsequent visit | 3.2 (2.9-3.6) | 5.4 (4.8-6.0) | 2.1 (1.5-2.8) | 0.001 | 1.7 (1.4-1.9) |

*The difference between the cohorts was the proportion of women with an abnormal mammogram amongst current users of HT minus the proportion amongst non-users.
#The relative risk was calculated using non-users of HT as the referent group.

### 3. Abnormality rate stratified by age

Since the abnormality rate of screening mammography is higher for younger women and younger women are more likely to use HT, the association between abnormality rate could have been confounded or modified by age. Therefore an analysis stratifying by age, in addition to screening visit was done. The results of this analysis are presented in Table 3.12.

For the first screening visit, the risk of an abnormal screening result for current users of HT compared to non-users was modified by age. The relative risk of having an abnormal mammogram with current use of HT increased from 0.87 (95% CI 0.71-1.06) in women aged 50 to 54 years to 1.53 (95% CI 1.23-2.05) in women aged 65 to 69 years. In

non-users of HT the risk of an abnormal mammogram decreased from 9.5% (95% CI 8.4%-10.6%) in women aged 50 to 54 years to 6.5% (95% CI 5.8%-7.3%) in women aged 65 to 69 years while there was no significant change in abnormality rate with age for current users of HT. For the subsequent screening visit the risk of an abnormal screening result was not modified or confounded by age. There were no statistically significant differences between the crude and the stratum specific relative risks of an abnormal mammogram nor were there differences in the relative risks across age strata. For all age strata the relative risk of an abnormal mammogram for current users of HT compared to non-users was significantly greater than the null value of one.

**Table 3.12: Risk of abnormal mammogram by HT status stratified by age at most recent mammogram and first versus subsequent screen**

| age stratum | HT cohort | | difference between cohorts* | p-value | relative risk# (95% CI) |
|---|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | | |
| first screen | | | | | |
| study population | 7.9 (7.5-8.4) | 9.3 (8.5-10.2) | 1.4 (0.5-2.3) | 0.002 | 1.2 (1.1-1.3) |
| 50-54 | 9.5 (8.4-10.6) | 8.2 (6.9-9.7) | -1.2 (-3.0-0.46) | 0.157 | 0.87 (0.71-1.06) |
| 55-59 | 8.4 (7.5-9.4) | 9.1 (7.7-10.7) | 0.71 (-1.02-2.44) | 0.416 | 1.08 (0.88-1.35) |
| 60-64 | 8.1 (7.2-9.0) | 10.8 (9.0-12.9) | 2.78 (0.69-4.88) | 0.005 | 1.35 (1.10-1.65) |
| 65-69 | 6.5 (5.8-7.3) | 10.0 (8.0-12.2) | 3.44 (1.27-5.61) | <0.001 | 1.53 (1.23-2.05) |
| Subsequent screen | | | | | |
| study population | 3.3 (2.9-3.6) | 5.4 (4.8-6.0) | 2.1 (1.5-2.8) | <0.001 | 1.7 (1.4-1.9) |
| 50-54 | 2.9 (2.1-3.7) | 4.7 (3.6-5.8) | 1.84 (0.51-3.17) | 0.006 | 1.65 (1.15-2.35) |
| 55-59 | 3.3 (2.7-4.0) | 4.7 (3.8-5.7) | 1.44 (0.33-2.55) | 0.009 | 1.44 (1.10-1.89) |
| 60-64 | 3.1 (2.5-3.7) | 6.6 (5.4-8.0) | 3.58 (2.19-4.97) | <0.001 | 2.17 (1.66-2.84) |
| 65-69 | 3.6 (3.0-4.2) | 5.9 (4.5-7.5) | 2.32 (0.76-3.89) | <0.001 | 1.65 (1.24-2.21) |

*The difference between the cohorts was the proportion of women with an abnormal mammogram amongst current users of HT minus the proportion amongst non-users.
#The RR was calculated using non-users of HT as the referent group.

## 4. Radiographic breast density stratified by age

Since the radiographic appearance of the breast is generally denser for younger women compared to older women and current users of HT tend to be younger than non-

users this variable was stratified by age in order to assess for confounding or effect modification by age. These results are presented in Table 3.13. There was evidence that the relative risk of having a fatty radiographic breast density for current users compared to non-users was modified by age. The relative risk of having a fatty radiographic breast density decreased from 0.83 (95% CI 0.77-0.88) in women aged 50 to 54 years to 0.66 (95% CI 0.61-0.70) in women aged 65 to 69 years. For all age strata the relative risk of having a fatty radiographic breast density for current HT users compared to non-users was significantly less than one.

The increase in relative risk resulted because the probability of having a fatty radiographic breast appearance increased with age for non-users but remained constant with increasing age for current users of HT. In non-users the proportion of women with fatty breast density increased from 36.4% (95% CI 35.0-37.8) in women aged 50 to 54 years to 55.0% (95% CI 53.9%-56.0%) in women aged 65 to 69 years. In current users the proportion of women with fatty breast density was much lower. Thirty percent (95% CI 28.4%-31.7%) of women aged 50 to 54 years had fatty radiographic breast density. There was no significant increase in this proportion until women were aged between 65 to 69 years where 36.0% (95% CI 33.8%-38.2%) of women had a fatty radiographic breast appearance.

**Table 3.13: Risk of having predominantly fatty radiographic breast density stratified by age**

| age stratum | HT cohort | | Difference between HT cohorts* | p-value | relative risk# (95% CI) |
|---|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | | |
| study population | 48.1 (47.5-48.7) | 31.8 (30.9-32.7) | 16.24 (15.18-17.30) | <0.001 | 0.66 (0.64-0.68) |
| 50-54 | 36.4 (35.0-37.8) | 30.0 (28.4-31.7) | 6.38 (4.23-8.53) | <0.001 | 0.83 (0.77-0.88) |
| 55-59 | 44.5 (43.3-45.7) | 31.1 (29.6-32.7) | 13.34 (11.40-15.29) | <0.001 | 0.70 (0.66-0.74) |
| 60-64 | 50.8 (49.6-51.8) | 31.9 (30.1-33.7) | 18.79 (16.63-20.95) | <0.001 | 0.63 (0.59-0.67) |
| 65-69 | 55.0 (53.9-56.0) | 36.0 (33.8-38.2) | 18.98 (16.54-21.42) | <0.001 | 0.66 (0.61-0.70) |

*The difference between the cohorts was the proportion of women with a fatty breast appearance amongst non-users of HT minus the proportion amongst current-users.
#The relative risk was calculated using non-users of HT as the referent group.

## D. Breast cancers diagnosed

### 1. Cancer detection rates

During the study period there were 327 breast cancers diagnosed in the 37. 917 eligible study participants. Eighty-four of these cancers were diagnosed amongst HT users. Of these breast cancers 64 were diagnosed as the result of an abnormal screening mammogram while 20 of the cancers were diagnosed in women who had a normal screening mammogram through some other means during the one-year interval that proceeded screening. In non-users there were 243 breast cancers diagnosed of which 224 were screen detected and 19 were interval cancers.

There were significant differences between the HT cohorts in the rates of both screen-detected and interval cancers. These results are presented in Table 3.14. The rate of screen detected cancers in the study population was 7.6 per thousand. There were significantly fewer screen detected cancers diagnosed amongst current users of HT than the non-users (p=0.015). The relative risk of being diagnosed with a screen detected cancer for current users of HT compared to non-users was 0.71 (95% CI 0.54-0.94).

The interval cancer rate for the study population was 1.04 per 1000 women who had a normal screening result. The interval-cancer rate was significantly higher in current users of HT at 2.0 per 1000 (95% CI 1.11-2.85) compared to 0.75 per 1000 (95% CI 0.41-1.08) in non-users (p=0.002). The relative risk of being diagnosed with an interval cancer was 2.65 (95% CI 1.41-4.97) for current users of HT compared to non-users.

**Table 3.14: Cancer detection rate and relative risk of breast cancer for HT cohorts**

| cancer type | HT cohort | | difference between cohorts* | p-value | relative risk# (95% CI) |
|---|---|---|---|---|---|
| | non-users rate per 1000 (95% CI) | current users rate per 1000 (95% CI) | rate per 1000 (95% CI) | | |
| screen detected | 8.28 (7.20-9.37) | 5.88 (4.45-7.32) | 2.40 (0.60-4.20) | 0.015 | 0.71 (0.54-0.94) |
| interval | 0.75 (0.41-1.08) | 2.00 (1.11-2.85) | -1.23 (-2.16-(-0.30)) | 0.002 | 2.65 (1.41-4.97) |

*The difference between the cohorts was calculated as the cancer detection rate for non-users minus the rate for current users.
#The relative risk was calculated using non-users of HT as the referent group.

### (a) Cancer detection rate stratified by first versus subsequent visit

Cancer detection rate varies depending on first or subsequent screening visit and in some populations use of HT is associated with use of preventive health practices (12, 14, 18, 60). Thus it is possible that the association between current use of HT and the cancer detection rate could be confounded or modified by previous mammographic screening. In order to assess whether or not there was confounding or effect modification by screening visit the cancer detection rates were stratified by first versus subsequent screening visit. The results are presented in Table 3.15.

The relative risk of a screen detected breast cancer was not modified or confounded by screening visit as there was no significant difference between the crude or the stratum specific risks. For both first and subsequent visits there were fewer screen-

detected cancers in the HT cohort. However, these differences were not statistically significant.

There was no evidence that the relative risk of an interval cancer for current users compared to non-users was modified or confounded by screening visit. There was no significant difference between the crude of the stratum specific relative risks. The interval cancer rate was higher for current users of HT for both first and subsequent visits. However, the difference was only significant for subsequent screening visits.

**Table 3.15. Breast cancer detection rate per 1000 women screened and relative risk of breast cancer for HT cohorts stratified by first versus subsequent visit**

| cancer type | HT cohort | | difference between cohorts* (95% CI) | p-value | relative risk# (95% CI) |
|---|---|---|---|---|---|
| | non-users (95% CI) | current users (95% CI) | | | |
| screen detected | | | | | |
| study population | 8.28 (7.20-9.37) | 5.88 (4.45-7.32) | 2.40 (0.60-4.20) | 0.015 | 0.71 (0.54-0.94) |
| first visit | 10.8 (9.1-12.4) | 9.0 (6.4-11.7) | 1.8 (-1.4-4.9) | 0.293 | 0.84 (0.60-1.17) |
| subsequent visit | 5.3 (4.0-6.6) | 3.3 (1.9-4.8) | 2.0 (0.02-3.9) | 0.067 | 0.63 (0.38-1.04) |
| interval | | | | | |
| study population | 0.75 (0.41-1.08) | 2.00 (1.11-2.85) | -1.14 (-2.16-(-0.30)) | 0.002 | 2.65 (1.41-4.97) |
| first visit | 1.03 (0.56-1.73) | 1.81 (0.78-3.56) | -0.78 (-2.14-0.58) | 0.150 | 1.73 (0.72-4.11) |
| subsequent visit | 0.42 (0.14-0.98) | 2.10 (1.09-3.69) | -1.68 (-2.93-(-0.43)) | <0.001 | 4.90 (1.73-13.91) |

*The difference between the cohorts was calculated as the cancer detection rate for non-users minus the rate for current users

#The relative risk was calculated using non-users of HT as the referent category

### (b) Cancer detection rate stratified by age

Younger women are more likely to be current users of HT and the cancer detection rate varies with age so the rates were stratified by age in order to assess for confounding or effect modification by age.

### i. Screen detected cancers

There was no evidence that the relative risk of a screen detected breast cancer for current users of HT compared to non-users was confounded or modified by age (Table 3.16). For all age strata, except one, point estimates of relative risk of a screen detected breast cancer for current users of HT compared to non-users were close to the null value of one and were not statistically significant. Women aged 60 to 64 years who were current users of HT were about 50% less likely than non-users to be diagnosed with a screen detected cancer. The relative risk was 0.52 (95% CI 0.30-0.92) which was significantly less than the null value of one.

**Table 3.16: Screen detected cancer rate per 1000 women screened stratified by age at most recent mammogram and first versus subsequent screen**

| age stratum | difference between HT cohorts* (95% CI) | p-value | relative risk# (95% CI) |
|---|---|---|---|
| study population | 2.40 (0.60-4.20) | 0.015 | 0.71 (0.54-0.94) |
| 50-54 | 0.33 (-2.61-3.33) | 0.828 | 0.92 (0.48-1.88) |
| 55-59 | 2.10 (-3.61-3.20) | 0.905 | 1.03 (0.63-1.68) |
| 60-64 | 5.17 (1.41-8.93) | 0.021 | 0.52 (0.30-0.92) |
| 65-69 | 2.30 (-2.05-6.65) | 0.847 | 0.98 (0.46-1.88) |

*The difference between the cohorts was calculated as the screen detected cancer rate for non-users minus the rate for current users.
#The relative risk was calculated using non-users of HT as the referent category.

### ii. Interval Cancers

There was no evidence that the relative risk of an interval cancer was confounded or modified by age as there were no significant differences between the crude and stratum specific relative risks (Table 3.17). The relative risk of an interval cancer was less than the null value of one for women aged 50 to 54 years. However, this estimate is quite imprecise due to the small number of interval cancers diagnosed in this age group (n=7). For women aged 55 to 59 years women who were current users of HT were three times more likely to have an interval breast cancer diagnosed following a normal screening result than non-users. This difference was not significant. However, the confidence interval of the estimate overlaps the null value of one by only a small amount. The relative risk of an interval breast cancer was 4.59 (95% CI 1.30-16.26) and 3.93 (1.20-12.85) for women aged 50 to 64 years and 65 to 69 years respectively.

**Table 3.17: Relative risk and absolute differences between HT cohorts for interval cancer rate per 1000 women screened stratified by age**

| Age stratum | Difference between HT groups* (95% CI) | p-value | relative risk (95% CI)# |
|---|---|---|---|
| study population | 1.23 (0.30-2.16) | 0.002 | 2.65 (1.42-4.97) |
| 50-54 | -0.45 (-1.86-0.97) | 0.711 | 0.61 (0.12-3.16) |
| 55-59 | 1.47 (-0.22-3.16) | 0.058 | 3.26 (0.96-11.13) |
| 60-64 | 2.05 (-0.12-4.23) | 0.018 | 4.59 (1.29-16.26) |
| 65-69 | 2.20 (-0.45-4.85) | 0.03 | 3.93 (1.20-12.85) |

*The difference between the cohorts was calculated as the interval breast cancer rate for current users of HT minus the rate for non-users.
#The relative risk was calculated using non-users of HT as the referent category.

## 2. Characteristics of breast cancers

The characteristics of the breast cancers diagnosed were examined to determine whether or not the Screen Test Program met the targets described above for organized screening programs and to examine whether or not there was a difference between the two HT cohorts.

### (a) Screen detected breast cancers

The characteristics of the 286 screen detected breast cancers were examined and are presented in Table 3.18. Early stage tumors (Stage 0 or I) comprised 67.7% (95% CI 61.97%-73.08%) of the screen detected breast cancers. Forty-seven percent (95% CI

40.6%-53.61%) of the invasive breast cancers were less than 1.5 centimeters in size. The mean tumor size was 1.8 cm (95% CI 1.6-2.0). There were no significant differences between the HT cohorts with respect to any of these tumor characteristics.

**Table 3.18: Tumor characteristics for HT cohorts for screen detected breast cancers**

| tumor characteristic | n | study population | HT cohort | | p-value |
|---|---|---|---|---|---|
| | | | non-users | current users | |
| Stage 0 or I, % (95% CI) | 286 | 67.7 (61.97-73.08) | 67.0 (60.39-73.08) | 70.3 (57.84-81.09) | 0.613 |
| size <1.5 cm*, % (95% CI) | 238 | 47.1 (40.6-53.61) | 48.1 (40.8-55.5) | 42.9 28.8-57.8) | 0.469 |
| Mean tumor size*, cm (95% CI) | 238 | 1.8 (1.6-2.0) | 1.8 (1.6-2.1) | 1.7 (1.3-2.1) | 0.586 |

*The size was calculated for invasive breast cancers only.

A comparison of the characteristics of the screen detected breast cancers to the guidelines for organized breast cancer screening programs are described in Table 3.19. The screen detected breast cancers detected in the Screen Test program were slightly below the guidelines for organized screening programs. However, there were no statistically significant differences between the characteristics in the Screen Test program and the guidelines.

**Table 3.19: Comparison of characteristics of screen detected breast cancers in Screen Test population to guidelines for organized screening programs**

| tumor characteristic | Screen Test population | guideline |
|---|---|---|
| Stage 0 or I, % (95% CI) | 67.7 (61.97-73.08) | 70 |
| size <1.5 cm*, % (95% CI) | 47.1 (40.6-53.61) | 50 |

*includes only invasive breast cancers

### (b) Interval breast cancers

For completeness, the characteristics of the 39 interval breast cancers are described in Table 3.20. The interval cancers were very different from the screen-detected cancers. Only 28.9% (95% CI 15.4%-45.9%) were stage 0 or I. Twenty-six percent (95% CI 24.9%-43.3%) were less than 1.5 centimeters in size and the mean tumor size was 2.97 cm (95% CI 1.91-3.46). As with, the screen detected cancers, there were no significant differences between the HT cohorts.

**Table 3.20: Tumor characteristics of interval breast cancers**

| tumor characteristic | n | study population | HT cohort | | p-value |
|---|---|---|---|---|---|
| | | | non-users | current users | |
| Stage 0 or I, % (95% CI) | 38 | 28.9 (15.4-45.9) | 27.8 (9.7-53.5) | 30.0 (11.9-54.2) | 0.644 |
| size <1.5 cm*, % (95% CI) | 35 | 25.7 (24.9-43.3) | 23.5 (6.81-49.9) | 27.8 (9.7-53.5) | 0.749 |
| Mean tumor size*, cm (95% CI) | 35 | 2.69 (1.91-3.46) | 2.41 (1.70-3.12) | 2.94 (1.59-4.30) | 0.501 |

*The size was calculated for invasive breast cancers only.

## III. Outcomes measuring performance of screening mammography

### A. Unadjusted estimates

There were significant differences between the HT cohorts with respect to sensitivity, specificity, positive and negative predictive values. These results are described in Table 3.21. Current use of HT was associated with a significant decrease in sensitivity. This decrease was substantial, an absolute decrease of 15.99% (95% CI 6.28%-25.70%). There was a significant reduction in specificity with current use of HT. This reduction was smaller that that found with sensitivity – an absolute difference of only 1.59% (95% CI 1.05%-2.12%). Positive predictive value was reduced 6% from 14.26% (95% CI 12.53%-16.00%) in non-users to 8.23% (95% CI 6.30%-10.16%) in current users of HT. The relative decrease in positive predictive value was about 40%. There was a significant but very small difference (0.13%) in negative predictive value between the HT cohorts.

**Table 3.21: Sensitivity, specificity and predictive values of screening mammography with one-year follow-up**

| screening outcome | study population % (95% CI) | HT cohort non users % (95% CI) | HT cohort current users % (95% CI) | p-value |
|---|---|---|---|---|
| sensitivity | 88.07 (84.56-91.59) | 92.18 (88.81-95.56) | 76.19 (67.08-85.30) | <0.001 |
| specificity | 94.52 (94.29-94.75) | 94.97 (94.71-95.23) | 93.39 (92.92-93.86) | <0.001 |
| positive predictive value | 12.26 (10.93-13.59) | 14.26 (12.53-16.00) | 8.23 (6.30-10.16) | <0.001 |
| negative predictive value | 99.89 (99.76-100.00) | 99.93 (99.89-99.96) | 99.80 (99.71-99.89) | 0.002 |

## B. Estimates adjusted using stratification

### 1. Stratified analysis adjusting for age

Since age was identified as a potentially confounding factor of the association between current use of HT and the performance of screening mammography age-stratified analyses for sensitivity, specificity and predictive values were carried out. The results for these analyses are described in Table 3.22 through Table 3.25.

There was no evidence that the association between current use HT and the sensitivity of screening mammography was confounded or modified by age (Table 3.22). There were no differences between the crude and stratum specific differences in sensitivity between the HT cohorts. The sensitivity was substantially reduced for current

users of HT compared to non-users in all age strata except women aged 50 to 54 where current users had a higher sensitivity (Table 3.22). The difference between the HT cohorts was only significant for women aged 60 to 64 and 65 to 69 years where current users had a sensitivity that was 25% (95% CI 4.64-45.83) and 20.72% (95% CI -0.68%-42.12%) respectively. For all age strata the confidence intervals of the difference in sensitivity between the HT cohorts are wide due to the small number of cancers.

**Table 3.22: Sensitivity of screening mammography stratified by age**

| age | HT cohort | | difference between cohorts* | p-value |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | |
| study population | 92.18 (88.81-95.57) | 76.19 (67.08-85.30) | 15.99 (6.28-25.70) | <0.001 |
| 50-54 | 80.00 (59.30-93.17) | 85.71 (57.19-98.22) | -5.71 (-29.84-18.41) | 1.00 |
| 55-59 | 91.84 (80.40-97.73) | 78.13 (60.03-90.72) | 13.71 (-2.53-29.96) | 0.102 |
| 60-64 | 95.24 (88.25-98.69) | 70.00 (45.72-88.10) | 25.24 (4.64-45.83) | 0.003 |
| 65-69 | 92.94 (85.27-97.37) | 72.22 (46.52-90.31) | 20.72 (-0.68-42.12) | 0.022 |

*The difference was calculated as the sensitivity for non-users minus the sensitivity for current users of HT.

There was evidence that the difference in specificity of screening mammography between the HT cohorts was modified by age. The specificity was reduced for current users of HT compared to non-users for all age strata except women aged 50 to 54 years

where it was higher (Table 3.23). The stratum specific differences were significantly different from the crude difference and there were significant differences across age strata. The difference in specificity between the HT cohorts increased from -0.39% (95% CI (-1.50%)-0.73%) in women aged 50 to 54 years to 2.82% (95% CI 1.56-4.08) in women aged 65 to 69 years. The increase in the difference in specificity between the two cohorts occurred because specificity increased with age in non-users but remained constant for current users of HT. The specificity for non-users increased from 93.48% (95% CI 92.76%-94.19%) in women aged 50 to 54 years to 95.73% (95.30%-96.17%) for women aged 65 to 69 while there was no significant change in specificity with increasing age for women who were currently using HT.

**Table 3.23: Specificity of screening mammography stratified by age**

| age | HT cohort | | difference between cohorts* | |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | p-value |
| study population | 94.97 (94.71-95.23) | 93.39 (92.92-93.86) | 1.59 (1.05-2.12) | <0.001 |
| 50-54 | 93.48 (92.76-94.19) | 93.86 (93.00-94.72) | -0.39 (-1.50-0.73) | 0.500 |
| 55-59 | 94.62 (94.07-95.17) | 94.09 (93.61-94.87) | 0.53 (-0.42-1.48) | 0.271 |
| 60-64 | 95.36 (94.88-95.84) | 92.16 (91.10-93.22) | 3.21 (2.04-4.37) | <0.001 |
| 65-69 | 95.73 (95.30-96.17) | 92.91 (91.73-94.09) | 2.82 (1.56-4.08) | <0.001 |

*The difference was calculated as the specificity for non-users minus the specificity for current users of HT.

There was a suggestion that the difference in positive predictive value between the HT cohorts was modified by age. While there were no statistically significant differences among the crude or stratum specific differences between the HT cohorts a trend of an increasing difference between the HT cohorts with increasing age could be seen. The difference between the cohorts increased from 0.11% (95% CI -4.16-4.38) in women aged 50 to 54 years to 9.01% (95% CI 3.04%-14.97%) in women aged 65 to 69 years. While there was no statistically significant difference between these differences the point estimates are quite different and there is only a small overlap of the confidence intervals. Additionally, when the positive predictive value amongst current and non-users were examined changes with age could be seen. For non-users of HT positive predictive value

increased with age. The positive predictive value increased from 6.23% (95% CI 3.59%-8.87%) amongst women aged 50 to 54 years to 18.16% (95% CI 14.54%-21.78%) amongst women aged 65 to 69 years. There was no change in positive predictive value with age for current users of HT.

**Table 3.24: Positive predictive value of screening mammography stratified by age**

| age | HT cohort | | Difference between cohorts* | |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | p-value |
| study population | 14.26 (12.53-15.99) | 8.23 (6.30-10.16) | 6.03 (3.44-8.62) | <0.001 |
| 50-54 | 6.23 (3.59-8.87) | 6.12 (2.77-9.48) | 0.11 (-4.16-4.38) | 0.961 |
| 55-59 | 11.39 (8.26-14.53) | 10.78 (6.79-14.77) | 0.62 (-4.46-5.69) | 0.813 |
| 60-64 | 19.05 (15.29-22.80) | 6.73 (3.32-10.14) | 12.32 (7.25-17.39) | <0.001 |
| 65-69 | 18.16 (14.54-21.78) | 9.15 (4.41-13.89) | 9.01 (3.04-14.97) | 0.011 |

*The difference was calculated as the positive predictive value for non-users minus the positive predictive value for current users of HT.

Negative predictive value was extremely high for both cohorts across all age strata. For all age strata the differences between the HT cohorts were extremely small and non-significant and there were no appreciable differences between the crude or stratum specific differences between the cohorts. Since the negative predictive value was so high for both cohorts no subsequent adjustment of this outcome was carried out.

**Table 3.25: Negative predictive value of screening mammography stratified by age**

| age | HT cohort | | Difference between cohorts* | |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | p-value |
| study population | 99.93 (99.89-99.96) | 99.80 (99.49-100) | 0.22 (-0.67-1.11) | 0.002 |
| 50-54 | 99.88 (99.78-99.99) | 99.93 (99.56-100) | -0.04 (-0.43-0.34) | 0.555 |
| 55-59 | 99.94 (99.87-100) | 99.79 (99.20-100) | 0.15 (-0.45-0.74) | 0.045 |
| 60-64 | 99.94 (99.89-100) | 99.73 (99.04-100) | 0.21 (-0.49-0.90) | 0.009 |
| 65-69 | 99.92 (99.86-99.99) | 99.71 (98.81-100) | 0.22 (-0.67-1.11) | 0.015 |

*The difference was calculated as the negative predictive value for non-users minus the negative predictive value for current users of HT.

**2. Stratified analysis adjusting for age and first versus subsequent screening visit**

Sensitivity, specificity and positive predictive value were stratified on first versus subsequent screening visit in addition to age. These results are presented in Table 3.26 through Table 3.28.

Since there were no significant differences across age strata, sensitivity was stratified only on screening visit. While there were no significant differences in the difference in sensitivity between the HT cohorts, there was a suggestion that the difference in sensitivity was greater for the subsequent screening visit compared to the

first screening visit – 60.36% (95% CI 12.53-48.19%) versus 7.29% (95% CI -3.33%-17.91%) (Table 3.26).

**Table 3.26: Sensitivity of screening mammography stratified by screening visit**

| Screening visit | HT cohort | | Difference between cohorts* | |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | p-value |
| study population | 92.18 (88.81-95.57) | 76.19 (67.08-85.30) | 15.99 (6.28-25.70) | <0.001 |
| First screen | 91.91 (86.80-95.50) | 84.62 (71.92-93.12) | 7.29 (-3.33-17.91) | 0.121 |
| Subsequent screen | 92.86 (84.11-97.64) | 62.5 (43.69-78.90) | 30.36 (12.53-48.19) | <0.001 |

*The difference was calculated as the sensitivity for non-users minus the sensitivity for current users of HT.

There were differences in specificity across strata of age and first versus subsequent visit for current users of HT compared to non-users (Table 3.27). In all cases the differences between the cohorts were small – less than 5%. For the first screen, age appeared to modify the association between current use of HT and specificity. Among non-users of HT specificity increased from 91.08% (95% CI 89.66%-92.50%) in women aged 50 to 54 to 94.58% (95% CI 93.02%-96.14%) for women aged 65 to 69 years while it appeared to stay constant with age for current users. There appeared to be a trend of an increasing difference in the specificity between the two cohorts. The difference in specificity increased from -1.31% (95% CI -3.00%-0.38%) in women aged 50 to 54 years

to 3.58 (95% CI 1.51%-5.66%) in women aged 65 to 69 years. There were no significant differences between the crude and stratum specific differences across age strata. The difference between the HT cohorts was only significant for the age strata from 60 to 69 years. For the subsequent screen, there was a significant reduction in specificity with current use of HT for all age strata. The changes in specificity with age that were seen with the first screening visit were not apparent for the subsequent visit.

**Table 3.27: Specificity of screening mammography stratified by screening visit and by age**

| screening visit and age | HT cohort | | difference between cohorts* | p-value |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | |
| First screen | | | | |
| study population | 94.48 (94.15-94.81) | 91.48 (90.70-92.27) | 2.99 (2.14-3.85) | <0.001 |
| 50-54 | 91.08 (89.66-92.50) | 92.39 (91.41-93.37) | -1.31 (-3.00-0.38) | 0.136 |
| 55-59 | 92.34 (90.97-93.70) | 91.85 (90.95-92.76) | 0.48 (-1.17-2.14) | 0.563 |
| 60-64 | 93.37 (91.83-94.91) | 89.94 (88.98-90.89) | 3.43 (1.41-5.45) | <0.001 |
| 65-69 | 94.58 (93.02-96.14) | 91.00 (90.16-91.84) | 3.58 (1.51-5.66) | <0.001 |
| Subsequent screen | | | | |
| study population | 97.27 (96.98-97.55) | 94.92 (94.37-95.48) | 2.34 (1.71-2.97) | <0.001 |
| 50-54 | 97.26 (96.42-98.10) | 95.44 (94.48-96.41) | 1.82 (0.51-3.13) | 0.005 |
| 55-59 | 97.23 (96.52-97.95) | 95.69 (94.97-96.42) | 1.54 (0.48-2.60) | 0.003 |
| 60-64 | 97.52 (96.73-98.32) | 93.67 (92.86-94.47) | 3.86 (2.51-5.20) | <0.001 |
| 65-69 | 97.06 (96.01-98.10) | 94.45 (93.73-95.17) | 2.61 (1.10-4.12) | 0.001 |

*The difference was calculated as the specificity for non-users minus the specificity for current users of HT.

For positive predictive value there appeared to be modification of the difference between the cohorts by age (Table 3.28). For both the first and subsequent screening visits the positive predictive value increased with age for non-users while it remained

constant with increasing age amongst current users of HT. There was a trend of an increasing difference between the cohorts with increasing age. However, the increase in the magnitude of the differences between the cohorts was not statistically significant.

**Table 3.28: Difference in positive predictive value between HT cohorts stratified by age and screening mammogram**

| Screening visit and age | HT cohort | | Difference between HT cohorts* | |
|---|---|---|---|---|
| | non-users % (95% CI) | current users % (95% CI) | % (95% CI) | p-value |
| First screen | | | | |
| study population | 13.57 (11.61-15.53) | 9.67 (6.95-12.39) | 3.90 (0.55-7.25) | 0.033 |
| 50-54 | 6.67 (2.35-10.99) | 7.81 (4.61-11.01) | -1.15 (-6.66-4.67) | 0.676 |
| 55-59 | 9.83 (4.80-14.85) | 11.85 (8.16-15.54) | -2.02 (-8.45-4.40) | 0.525 |
| 60-64 | 18.91 (11.59-26.23) | 8.18 (5.14-11.22) | 10.73 (4.01-17.45) | 0.008 |
| 65-69 | 17.97 (9.66-26.28) | 10.98 (7.41-14.54) | 6.99 (-1.07-15.05) | 0.131 |
| Subsequent screen | | | | |
| study population | 16.29 (12.67-19.91) | 6.19 (3.56-8.82) | 10.10 (5.62-14.58) | <0.001 |
| 50-54 | 3.92 (-0.69-8.54) | 2.94 (-1.70-7.58) | 0.98 (-5.57-7.65) | 0.769 |
| 55-59 | 16.00 (8.70-23.30) | 9.28 (3.59-14.96) | 6.72 (-2.49-15.94) | 0.156 |
| 60-64 | 19.44 (11.61-27.28) | 5.10 (0.95-9.25) | 14.34 (5.70-22.99) | 0.002 |
| 65-69 | 18.57 (8.73-28.41) | 6.67 (2.53-10.80) | 11.90 (2.89-20.92) | 0.031 |

*The difference was calculated as the positive predictive value for non-users minus the positive predictive value for current users of HT.

## C. Estimates adjusted using logistic regression

### 1. Description of potentially confounding variables

Characteristics which have been shown previously to be associated with use of HT and an increase or decrease in radiographic breast density or a change in the performance of screening mammography were considered to be potentially confounding factors of the association between current use of HT and the performance of screening mammography. Since only those factors that were associated with current use of HT in the study population would be used to adjust the performance measures for screening mammography. current use of HT among women with these characteristics was examined.

A description of the study population and the two HT cohorts with respect to these potentially confounding factors is provided in Table 3.29. The mean age of the study population at most recent screening mammogram in the Screen Test Program was 60. There was a significant difference in the mean age between the two cohorts. 61 years in non-users compared to 59 for current users. The difference in means was small. only 2 years. However. current users of HT were 50% more likely to be aged less than 60 years and as described previously the proportion of use was significantly higher for younger women.

There were no statistically significant differences between the cohorts with respect to parity, gravidity, the percent of women who were nulliparous or had a maternal history of breast cancer. There was a statistically significant difference between the HT cohorts for the mean age at first birth. However, when rounded to the nearest year the mean age at first birth was 23 years for both of the cohorts. The same was true for mean age at menopause where there was no difference between the cohorts when the value was rounded to the nearest year. There was a statistically significant difference between the two cohorts with respect to mean time since menopause (p-value<0.0001). When rounded to the nearest year current users of HT had a mean time since menopause of 11 years compared to 9 years in non-users.

Women who used HT were significantly more likely to have an age of menarche less than 12 years, to have a history of hysterectomy, a history of oral contraceptive use, have had a previous breast biopsy, and had a previous mammogram. They were significantly less likely to have a history of a sister with breast cancer and an age at first birth greater than 29 years of age.

**Table 3.29: Description of study population with respect to potentially confounding factors**

| factor | study population | HT cohort | | p-value |
|---|---|---|---|---|
| | | non-users | current users | |
| Age at most recent mammogram, mean | 60.14 (60.08-60.20) | 60.80 (60.73-60.86) | 58.52 (58.42-58.62) | <0.0001 |
| Age less than 60 years, % | 46.80 (46.30-47.30) | 41.41 (40.83-42.00) | 60.19 (59.27-61.11) | <0.001 |
| Age at menopause, mean | 49.75 (49.72-49.78) | 49.69 (49.64-49.73) | 49.90 (49.86-49.95) | <0.0001 |
| Time since menopause, mean | 10.3 (10.2-10.4) | 8.6 (8.5-8.7) | 11 (10.9-11.1) | <0.0001 |
| History of hysterectomy, % | 41.45 (40.96-41.95) | 32.62 (32.06-33.18) | 63.37 (62.46-64.28) | <0.001 |
| History of sister with breast cancer, % | 5.96 (5.72-6.20) | 6.40 (6.11-6.69) | 4.89 (4.46-5.27) | <0.001 |
| Maternal History of breast cancer, % | 5.27 (5.04-5.49) | 5.30 (5.30-5.17) | 5.17 (4.75-5.59) | 0.613 |
| Nulliparous, % | 9.15 (8.86-9.44) | 9.07 (8.73-9.42) | 9.34 (8.79-9.89) | 0.414 |
| Gravidity, median (IQ range) | 3 (2-5) | 3 (2-5) | 3 (2-4) | |
| Parity, median (IQ range) | 3 (2-4) | 3 (2-4) | 3 (2-4) | |
| Age at first birth, mean | 23.24 (23.19-23.28) | 23.36 (23.30-23.41) | 22.93 (22.85-23.02) | <0.0001 |
| Age at first birth >29, % | 8.10 (7.82-8.37) | 8.66 (8.33-9.00) | 6.70 (6.23-7.17) | <0.001 |
| Menarche <11, % | 16.14 (15.77-16.51) | 15.37 (14.93-15.80) | 18.07 (17.34-18.79) | <0.001 |
| previous breast biopsy, % | 14.86 (14.50-15.22) | 14.08 (13.66-14.49) | 16.81 (16.11-17.52) | <0.001 |
| Previous Screen Test Mammogram, % | 48.23 (47.73-48.73) | 45.44 (44.84-46.03) | 55.17 (54.23-56.10) | <0.001 |
| History of previous mammogram (Screen Test or other), % | 79.31 (78.90-79.72) | 75.65 (75.14-76.16) | 88.40 (87.80-89.00) | <0.001 |
| History of oral contraceptive use, % | 47.93 (47.42-48.43) | 43.38 (42.79-43.97) | 59.21 (58.29-60.14) | <0.001 |

**(a) Clinical significance of differences between HT cohorts with respect to potentially confounding factors**

In order to assess whether or not the significant associations were large enough to be considered clinically relevant, the differences between proportions and relative risks of having the potentially confounding factor for current users of HT compared to non-users for the categorical variables that were significantly associated with use of HT were examined (Table 3.30). It was important to examine the relative risks as well as the absolute difference in risk. Since the size of the study population was large, there were several small differences which were found to be highly statistically significant (e.g. p-value <0.001) between the study cohorts. With this, there is the concern that the magnitude of some of these differences between the cohorts may be small and not clinically significant.

While the absolute differences between the cohorts for some of these variables were small (1%-3%) there were moderate to large increases or decreases in the relative risks for all of the variables. Women currently using HT were almost twice as likely to have a history of a hysterectomy and 40% more likely to have used oral contraceptives in the past. Women who used HT were 20% more likely to have had menarche at an age younger than 12 years, a history of a previous breast biopsy, and a history of a previous mammogram, within or outside the Screen Test Program. Women using currently using

HT were 25% less likely to have had a sister with breast cancer and 22% less likely to have an age at first birth of greater than 29 years.

Considering the magnitude of the relative differences between the cohorts, it was determined that it would be necessary to consider all of the variables that were identified as having a statistically significant association with current use of HT as potentially confounding variables.

**Table 3.30: Risk differences and relative risks for HT cohorts with respect to potentially confounding factors**

| potentially confounding factor | difference between HT cohorts* % (95% CI) | p-value | relative risk# (95% CI) |
|---|---|---|---|
| Age less than 60 years | 18.78 (17.69-19.87) | <0.001 | 1.45 (1.42-1.48) |
| History of hysterectomy | 30.75 (29.68-31.81) | <0.001 | 1.94 (1.90-1.99) |
| History of sister with breast cancer | 1.53 (1.03-2.03) | <0.001 | 0.76 (0.69-0.84) |
| Age at first birth >29 | 1.97 (1.32-2.61) | <0.001 | 0.78 (0.72-0.84) |
| Menarche <12 | 2.70 (1.86-3.55) | <0.001 | 1.18 (1.12-1.24) |
| previous breast biopsy | 2.74 (1.92-3.55) | <0.001 | 1.19 (1.14-1.26) |
| previous Screen Test mammogram | 9.73 (8.62-10.83) | <0.001 | 1.21 (1.19-1.24) |
| previous mammogram | 12.75 (11.96-13.54) | <0.001 | 1.17 (1.16-1.18) |
| History of oral contraceptive use | 15.83 (14.73-16.93) | <0.001 | 1.37 (1.34-1.39) |

* Difference between the cohorts is the proportion of women with the characteristic in the HT cohort minus the proportion amongst non-users.
#The relative risk was calculated using non-users as the referent category.

### (b) Age stratified analysis of potentially confounding variables

Since some of the potentially confounding variables are associated with age and HT use is higher among younger women, confounding or effect modification by age could have explained the significant results for risk differences and risk ratios for the potentially confounding factors between the HT cohorts. In order to assess this

possibility, the variables described in Table 3.30 were stratified by 5-year age strata. The results of these stratified analyses are summarized in Table 3.31. This analysis shows that confounding or interaction by age could not have explained the significant differences between the HT cohorts.

There was evidence of effect modification by age for the association between current use of HT and the history of a hysterectomy, past use of oral contraceptives and early age at menarche. Women between the ages of 55 to 69 years who used HT were twice as likely to have had a hysterectomy than non-users. However, women aged 50 to 54 years using HT were only about 80% more likely than non-users to have had a hysterectomy. A similar pattern was seen with a history of oral contraceptive use. Women aged 50 to 54 years who used HT were only about 10% more likely than non-users to have a history of oral contraceptive use. This relative difference is significantly less than that for the age strata 55 to 69 where users of HT were from 20% to 35% more likely to have a history of oral contraceptive use. There was a trend of a greater likelihood of an early age at menarche for users of HT compared to non-users with increasing age.

There was no evidence of effect modification by age for any of the other variables that were significantly associated with current use of HT nor was there any evidence of confounding by age for these variables. An interesting finding was the difference in the direction of the association between use of HT and a history of breast cancer in a sister.

For all age strata, except 60 to 64 years, users of HT were significantly less likely than non-users to have had a sister with breast cancer. Women who were aged 60 to 64 years were more likely to have had a sister with breast cancer. This difference was not statistically significant.

**Table 3.31: Age stratified analysis for potentially confounding variables**

| Variable | crude relative risk[#] | Stratum specific relative risk[#] | | | | combined relative risk[#] |
|---|---|---|---|---|---|---|
| | | 50-54 | 55-59 | 60-64 | 65-69 | |
| Hysterectomy | 1.9 (1.9-2.0) | 1.8 (1.7-1.9) | 2.0 (1.9-2.1) | 2.0 (1.9-2.1) | 2.1 (2.0-2.2) | [$] |
| sister with breast cancer | 0.8 (0.6-0.8) | 0.7 (0.6-0.9) | 0.8 (0.7-1.0) | 1.1 (0.9-1.3) | 0.7 (0.6-0.9) | 0.8 (0.8-0.9) |
| first birth >30 | 0.7 (0.7-0.8) | 0.8 (0.7-1.0) | 0.9 (0.7-1.0) | 0.9 (0.7-1.0) | 0.7 (0.6-0.9) | 0.8 (0.8-0.9) |
| menarche <12 | 1.2 (1.1-1.2) | 1.0 (1.0-1.1) | 1.1 (1.0-1.2) | 1.1 (1.0-1.3) | 1.3 (1.2-1.5) | 1.1 (1.1-1.2) |
| previous breast biopsy | 1.2 (1.1-1.3) | 1.3 (1.1-1.4) | 1.2 (1.1-1.4) | 1.1 (1.0-1.3) | 1.3 (1.2-1.4) | 1.2 (1.2-1.3) |
| previous mammogram | 1.2 (1.2-1.2) | 1.2 (1.1-1.2) | 1.2 (1.2-1.2) | 1.2 (1.2-1.2) | 1.2 (1.1-1.2) | 1.2 (1.2-1.2) |
| past oral contraceptive use | 1.4 (1.3-1.4) | 1.1 (1.1-1.1) | 1.2 (1.2-1.3) | 1.3 (1.2-1.3) | 1.4 (1.3-1.5) | [$] |

[#]The relative risk was calculated using non-users as the referent group.
[$]The combined relative risk was not calculated due to evidence of effect modification.

### (c) Potentially confounding variables adjusted for in subsequent analyses

Based on the above examinations of the potentially confounding factors, age, time since menopause, history of a hysterectomy, history of a sister with breast cancer, age at first birth greater than 29 years, age at menarche less than 12 years, previous breast

biopsy, previous Screen Test mammogram, and a history of past oral contraceptive use were considered as potential covariates for the adjusted estimates of outcomes measuring the performance of mammography. These factors, which are associated with current use of HT in this study population, have been found to be associated with both a change in the performance of mammography or mammographic density and associated with use of HT.

**2. Assessing the validity of estimates obtained using logistic regression modeling**

Coughlin et al (72) suggest that bias could potentially affect the sensitivity estimate obtained using the method they have described. Before proceeding with adjusting the performance measures of screening mammography for potentially confounding factors using this method, the validity of the model was assessed. Coughlin et al suggest that when the prevalence of disease in strata of a covariate is markedly different (e.g. a ten fold difference) then bias may operate. In this study there was no significant difference in the prevalence of breast cancer between the HT cohorts. The prevalence of breast cancer was 7.72 per 1000 (95% CI 6.16-9.55) for current users of HT and 8.99 per 1000 (95% CI 7.90-10.19) for non-users. However, in order ensure the validity of the estimates obtained from the logistic regression modeling, results for the estimates obtained from logistic regression and stratification methods were compared.

A model with screening result as the outcome and HT status and breast cancer as the only covariates was fit for sensitivity and specificity and a model with breast cancer status as the outcome variable and HT status and screening result as covariates was fit for positive predictive value. There were differences in the estimates of sensitivity and positive predictive value obtained by the two adjustment methods. These results are presented in Table 3.32.

The differences between the sensitivity estimates between the HT cohorts that were obtained from the logistic regression analysis were in the opposite direction from those obtained from the stratified analysis. That is. sensitivity was higher for current users of HT compared to non-users in the logistic regression. Because these differences indicated that bias could have potentially affected the estimates obtained via logistic regression modeling, the correction techniques suggested by Coughlin et al were employed.

**Table 3.32: Comparison of screening outcome estimates based on adjustment method**

| | Current users of HT | | non-users | |
|---|---|---|---|---|
| parameter | stratified analysis | logistic regression modeling | stratified analysis | logistic regression modeling |
| sensitivity | 76.19 | 90.03 | 92.18 | 87.40 |
| specificity | 93.39 | 93.49 | 94.97 | 94.93 |
| ppv | 8.23 | 9.72 | 14.26 | 13.52 |

Sensitivity was used to assess whether application of the correction techniques resulted in estimates that were the same as those obtained through the stratified analysis. These results are presented in Table 3.33. Both techniques resulted in estimates that were equal to the estimates obtained from the stratified analysis. However, since modeling sensitivity using only women with breast cancer was less complicated to employ, this method was chosen for subsequent logistic regression analyses.

**Table 3.33: Sensitivity estimates from stratified analysis and other modeling methods**

| HT | Stratified analysis | model* | model$ | model& |
|---|---|---|---|---|
| users | 76.19% | 90.03% | 76.19% | 76.19% |
| non-users | 92.18% | 87.40% | 92.18% | 92.18% |

*model with breast cancer and HT status as covariates
$ model with breast cancer, HT status and breast cancer-HT interaction term
&model with only individuals with breast cancer with HT status as only covariate

### 3. Sensitivity adjusted using logistic regression modeling

#### (a) Model

In order to adjust for factors identified above which could potentially confound the association between current use of HT and sensitivity a logistic regression model was fit. This model included only women with breast cancer with an abnormal screening result modeled as the outcome and HT status and the other potentially confounding factors included as covariates. None of the potentially confounding factors contributed significantly to the model. The model fit is described below in Table 3.34.

**Table 3.34: Logistic regression model output for sensitivity**

| Variable | coefficient | standard error | z | p>\|z\| | 95% CI of coefficient |
|---|---|---|---|---|---|
| hormone use | -1.30 | 0.35 | -3.72 | 0.000 | -1.99, 0.62 |
| constant | 2.47 | 0.23 | 10.33 | 0.000 | 2.00, 2.94 |

**(b) Fit of model**

The likelihood ratio test for the fitted model compared to a model with only a constant was 13.48 (p-value=0.0002) indicating that a model with HT use as a covariate explains significantly more of the variation.

The model was well fit. The Pearson chi-square had a value of zero. Because there were only two distinct covariate patterns the Hosmer-Lemeshow goodness of fit test could not be employed.

Since there were only 327 individuals included in the model, it was possible that important covariates could have been excluded due to limited power. Therefore, the coefficient for the HT use term for the reduced model was compared to a model that included all of the potential covariates. The coefficient for the HT use term was -1.32 with a standard error of 0.44 for the full model compared to -1.30 with a standard error 0.35. Since there was minimal difference between the coefficients of the HT use terms between the two models it was determined that the reduced model was acceptable.

### (c) Plot of sensitivity estimates of HT cohorts

A graph of the sensitivity estimates for the HT cohorts is presented in Figure 3.2. The sensitivity was lower for current users of HT at 76.2% (95% CI 66.1%-84.0%) compared to 92.2% (95% CI 88.1%-95.0%) for non-users.

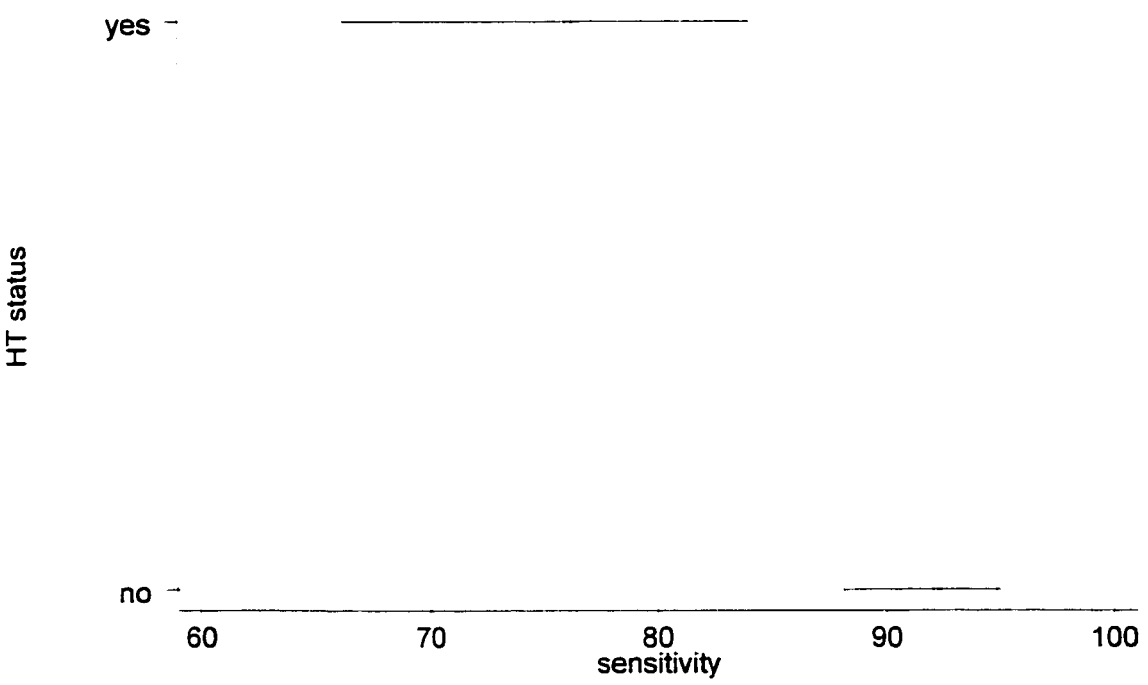


**Figure 3.2: Sensitivity of screening mammography by HT status**

## 4. Specificity adjusted using logistic regression modeling

### (a) Model

A logistic regression model adjusting for the potentially confounding factors outlined above was fit to the data. This model included only women who did not have

breast cancer and a normal screening result was modeled as the dependent variable with HT use and the other potentially confounding factors included as potential covariates. In addition to HT use status, the final model included as covariates age, previous Screen Test mammogram, previous biopsy, aboriginal origin, and age*HT use status and age*previous mammogram interaction terms. None of the other potentially confounding variables or interaction terms contributed significantly to the model. The final model is described below in Table 3.35.

**Table 3.35: Logistic regression model output for specificity**

| variable | coefficient | standard error | z | p>\|z\| | 95% CI of coefficient |
|---|---|---|---|---|---|
| HT use | 2.329 | 0.525 | 4.439 | 0.000 | 1.301, 3.358 |
| age | 0.036 | 0.006 | 6.252 | 0.000 | 0.025, 0.047 |
| age*HT use | -0.045 | 0.009 | -5.110 | 0.000 | -0.062, -0.028 |
| previous ST mammogram | 2.238 | 0.555 | 4.032 | 0.000 | 1.150, 3.326 |
| previous biopsy | -0.398 | 0.057 | -6.947 | 0.000 | -0.510, -0.286 |
| aboriginal origin | .691 | .274 | 2.526 | 0.012 | 0.155, 1.228 |
| age*previous ST mammogram | -0.024 | .009 | -2.580 | 0.010 | -0.0418, -0.0057 |
| constant | 0.539 | 0.345 | 1.560 | 0.119 | -0.138, 1.216 |

**(b) Fit of model**

The likelihood ratio test for this model compared to a model with only an intercept was 423.48 (p-value<0.00001). This indicated that the fitted model explained the variation in the data significantly better that a model with only a constant. When the full model was compared to a model with HT use as the only covariate the likelihood

ratio test yielded a value of 384.62 (p-value<0.00001) which indicated the additional covariates, besides HT status, were important in explaining a significantly amount of the variance in the data. When a complete model with all covariates except HT status was compared to the full model the likelihood ratio test had a value of 19.68 (p<0.00001) indicating that the covariates included in addition to HT status did explain a significant amount of variation in the data.

The Hosemer-Lemeshow test for fit yielded a value of 22.30 (p=0.0044) which indicates a lack of fit. However, this was not unexpected given the size of the sample that was used. In order to assess the fit the expected and observed number of individuals without breast cancer were examined. These results are presented in Table 3.36. In all cases the absolute difference between the observed and expected numbers as a percent of the total was one percent or less.

**Table 3.36: Observed and expeceted numbers for deciles of fitted specificity used to calculate Hosmer-Lemeshow goodness of fit test**

| group | total | observed number | expected number | absolute difference between observed and expected | difference as % of total | Expected specificity |
|---|---|---|---|---|---|---|
| 1 | 3750 | 3366 | 3366.3 | 0.3 | 0.008 | 0.898 |
| 2 | 3922 | 3627 | 3594 | 33 | 0.841 | 0.916 |
| 3 | 3650 | 3386 | 3386.5 | 0.5 | 0.014 | 0.928 |
| 4 | 3592 | 3350 | 3372.1 | 22.1 | 0.615 | 0.939 |
| 5 | 3748 | 3524 | 3551.7 | 27.7 | 0.739 | 0.948 |
| 6 | 3794 | 3629 | 3618.6 | 10.4 | 0.274 | 0.954 |
| 7 | 3902 | 3735 | 3753.5 | 18.5 | 0.474 | 0.962 |
| 8 | 4007 | 3894 | 3882.8 | 11.2 | 0.280 | 0.969 |
| 9 | 3480 | 3406 | 3380.5 | 25.5 | 0.733 | 0.971 |
| 10 | 3414 | 3313 | 3324 | 11 | 0.322 | 0.974 |

To further examine the fit of the model plots of regression diagnostics were examined. These plots indicated that there were a number of poorly fit covariate patterns. Poorly fit patterns were defined as any pattern with a change in Pearson residual of greater than 5, a change in deviance residual greater than 5, a leverage value of greater than 0.1 or a Δbeta of greater than 0.5. In total 26 of the 265 covariate patterns were poorly fit according to these criteria. Upon closer examination it was seen that 6 of the 26 patterns had values for change in Pearson residual and change in deviance residual of less than one but were considered to be poorly fit due only to leverage being between 0.1 and 0.12. These patterns, therefore, were not considered to be poorly fit. In the end there were 4969 individuals, 13% of those included in the model, in 20 poorly fit covariate patterns. Given that 87% of the data was well fit by the model and that no implausible patterns

were detected among the poorly fit covariate patterns, it was concluded that the model fit the data sufficiently well that it could be accepted.

### (c) Specificity estimates adjusted for age, screening visit, previous biopsy and ethnic origin

Conditional effect plots showing the effect of HT use on specificity for all of the possible covariates were also generated. For women of non-Aboriginal and Aboriginal origin, plots of specificity versus age by HT status were plotted for women who had neither a previous biopsy nor a previous Screen Test mammogram, no previous biopsy and a previous Screen Test mammogram, previous biopsy and no Screen Test mammogram and a previous biopsy and previous biopsy. These plots are presented in Figure 3.3 and Figure 3.4.

For non-Aboriginal women the difference in specificity between the two cohorts increased with increasing age for all of the covariate combinations. For women with no previous Screen Test mammogram there was no significant difference in specificity with increasing age for current users of HT while it increased with age for non-users. The difference between the cohorts was significant starting at 57 years for women with no previous biopsy and 59 years for women with a previous biopsy. For women with no previous biopsy there was no significant difference between the cohorts at age 50, a difference of about 2.6% age 60 and a difference of about 5% at age 69. The differences

at age 60 and 69 were statistically significant. as there was no overlap of the 95% confidence intervals of the specificity estimates for the two cohorts.

The specificity of screening mammography was higher for women with a history of a Screen Test mammogram than that for woman who had no previous Screen Test mammogram for both cohorts. The difference in specificity between the HT cohorts was smaller for women who with a previous Screen Test mammogram compared to those who did not. For these women there was a non-significant decrease in specificity with age for current users of HT while there was a non-significant increase with age for non-users. The difference between the HT cohorts was significant at age 65 years for women who had no previous breast biopsy and aged 59 for women who had. For women aged 69 years the difference between the cohorts was about 3%.

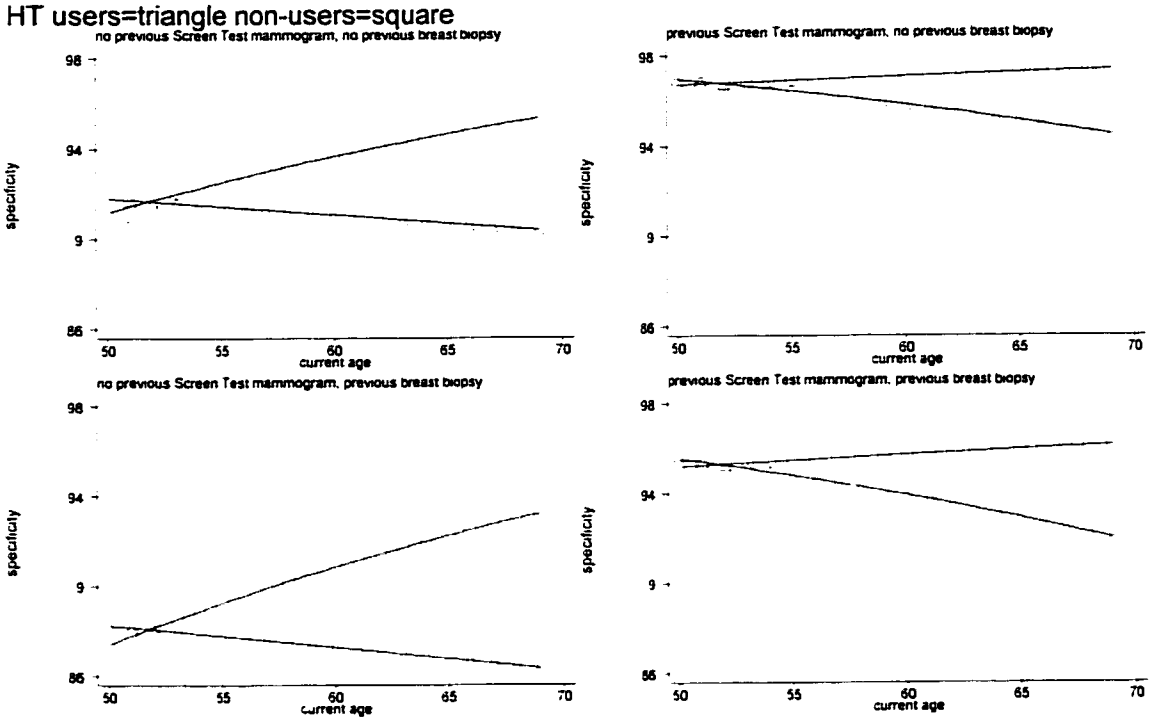


**Figure 3.3: Conditional effect plots for non-Aboriginal women showing effect of HT use on specificity adjusting for age, previous biopsy and previous Screen Test mammogram**

Similar patterns to those described above were seen with Aboriginal women (Figure 3.4). However, there were no significant differences between the HT cohorts for women of Aboriginal origin.

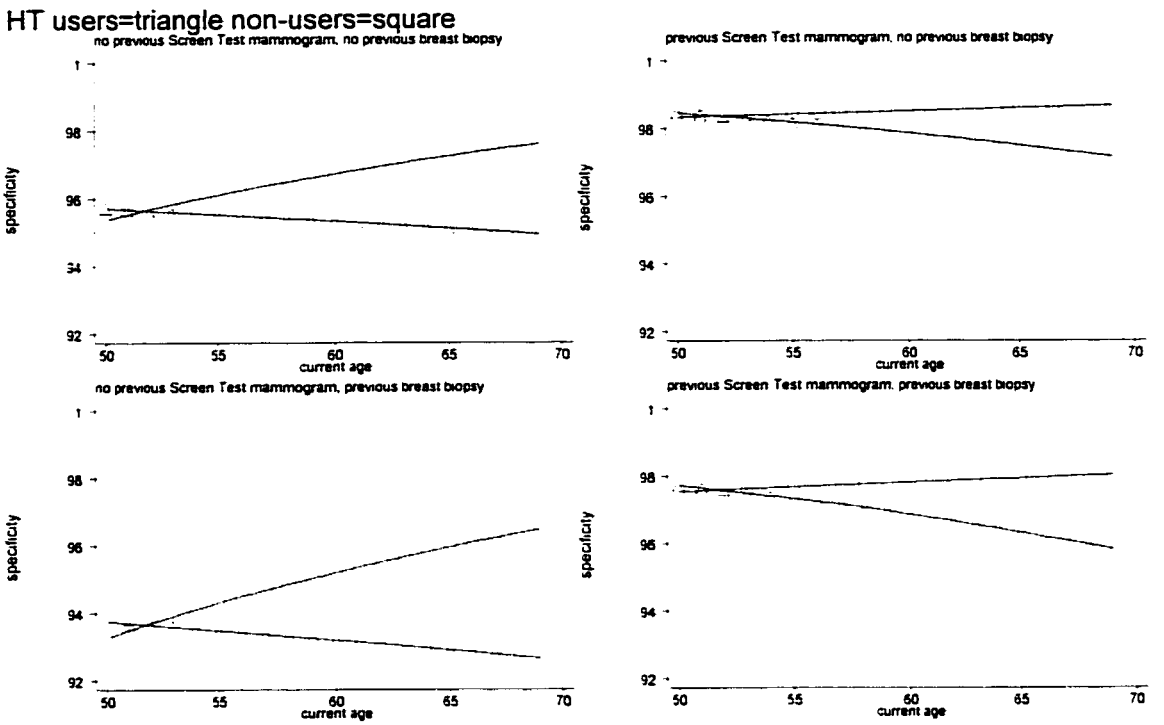


**Figure 3.4: Conditional effect plots for Aboriginal women showing the effect of HT use on specificity adjusting for age, previous biopsy and previous Screen Test mammogram**

Conditional effect plots showing the specificity of screening mammography by HT status and ethnic origin for women with no history of a breast biopsy are included for women with no previous Screen Test mammogram (Figure 3.5) and women with a previous Screen Test mammogram (Figure 3.6). For both women with or without a previous Screen Test mammogram. Aboriginal women had a significantly higher specificity. As was described above, the graphs show that for women with no previous Screen Test mammogram specificity did not change significantly with age for current

users of HT while it increased with age for non-users. For women with a previous Screen Test mammogram the specificity was higher and there was a non-significant decrease for current HT users and a non-significant increase for non-users.

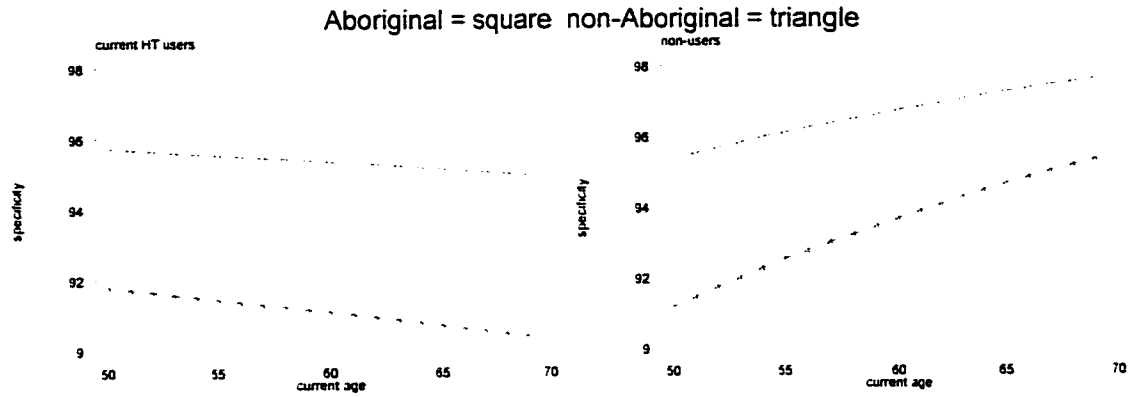


**Figure 3.5: Specificity by HT status and ethnic origin for women with no previous Screen Test mammogram and no previous breast biopsy**

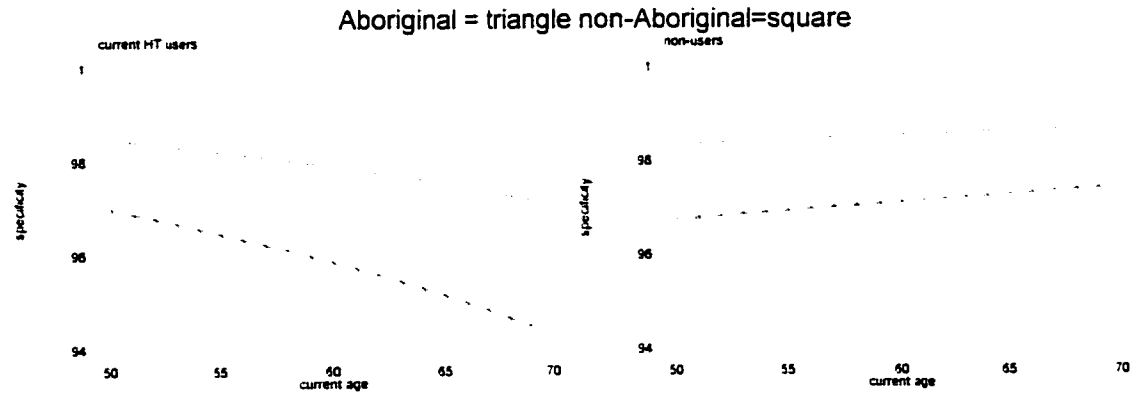


**Figure 3.6: Specificity by HT status and ethnic origin for women with a previous Screen Test mammogram and no previous breast biopsy**

## 5. Positive predictive value adjusted using logistic regression modeling

### (a) Model

In order to adjust positive predictive value for potentially confounding factors, a logistic regression model including only women who had abnormal screening results was fit. A diagnosis of breast cancer was modeled as the dependent variable and HT status

and the other potentially confounding factors were included as potential covariates. The only potential covariates that contributed significantly to the model were age and HT status and an age*HT interaction term. The final model is presented in Table 3.37.

**Table 3.37: Logistic regression model output for positive predictive value**

| variable | coefficient | standard error | z | p>\|z\| | 95% Confidence interval |
|---|---|---|---|---|---|
| HT use | 3.34 | 1.67 | 2.002 | 0.045 | 0.070-6.618 |
| current age | 0.072 | 0.014 | 5.254 | 0.000 | 0.045, 0.098 |
| age*HT use | -0.065 | 0.028 | -2.326 | 0.020 | -0.119, -0.010 |
| constant | -6.155 | 0.844 | -7.296 | 0.000 | -7.808, -4.501 |

**(b) Fit of model**

When the fitted model was compared to a model with only a constant the likelihood ratio test yielded a value of 47.82 (p-value<0.00001) indicating that the fitted model explained more of the variation in the data. Compared to a model with only HT use as the only covariate the likelihood ratio test was 29.13 (p-value<0.00001) indicating that the fitted model explained more variation.

The Hosmer-Lemeshow test for the fitted model had a value of 18.33 (p-value=0.019) indicating some evidence of lack of fit.

To further examine the fit of the model diagnostic plots were examined. Seventeen covariate patterns were identified as potentially being poorly fit. Upon

examination of the diagnostic values it was determined that the criteria used to classify these patterns as poorly fit was overly conservative. Twelve of the covariate patterns had low values for change in residuals and Δbeta but were classified as poorly fit because the leverage value was greater than 0.1. In the end 294 individuals, 12.5% of the individuals in the model, contained in five covariate patterns were considered to be poorly fit. That meant that 87.5% of the data were well fit by the model and there was insufficient evidence to suggest that the model should not be accepted.

### (c) Positive predictive value estimates adjusted by age

A conditional effect plot of positive predictive value versus age by the two HT cohorts is included as Figure 3.7. The difference in the positive predictive value between the two HT cohorts increased with age. The positive predictive value for current users of HT remained at about 8% for women of all ages while the positive predictive value for non-users increased from 7% at age 50 to 23% at age 69. The difference between the cohorts was significant starting at age 59.

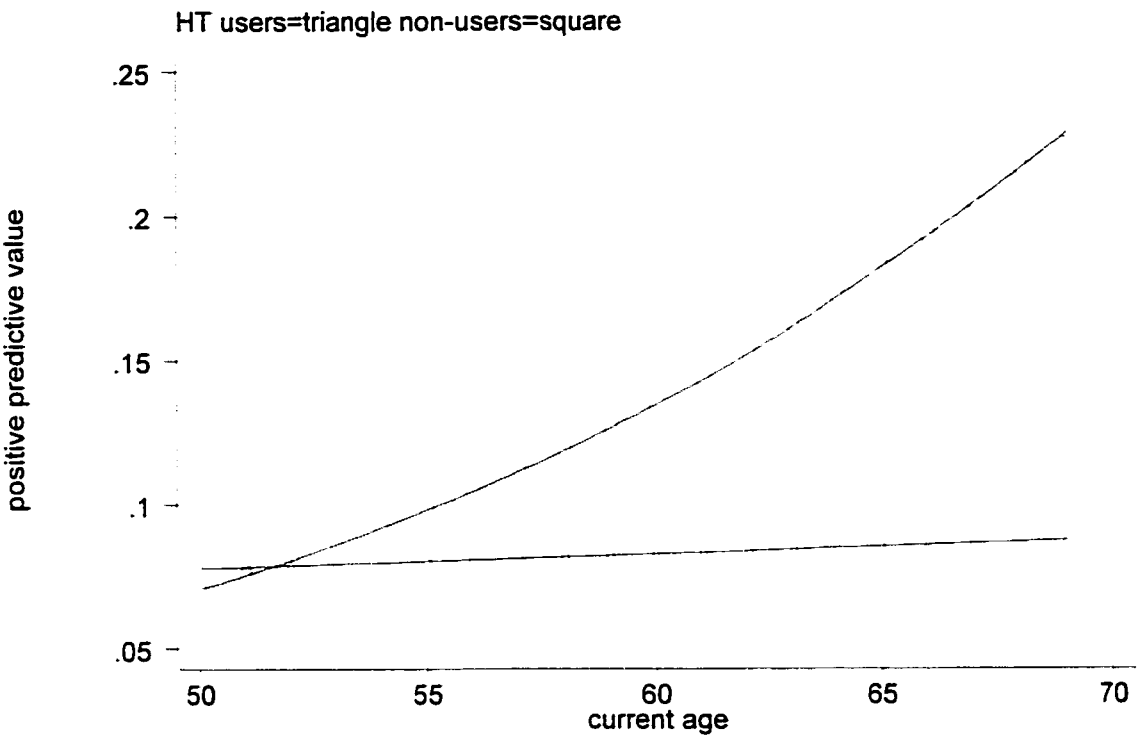


**Figure 3.7: Positive predictive value of screening mammography versus age by HT cohort**

## IV. Sensitivity analysis assessing impact of misclassification

In order to quantify the potential impact of misclassification sensitivity analyses assessing the impact of misclassification of false-negative status and HT status on the sensitivity estimates obtained were carried out.

### A. Misclassification of false-negative results

This study employed what is known as the detection method to determine estimates for sensitivity. In this method the total number of cancers is considered to be

equal to those cancers detected as a result of an abnormal screening result and cancers diagnosed by a method other than screening during a one-year period following an normal screening mammography result, false-negative results. This method can potentially underestimate the sensitivity of screening as those results that are classified as false-negative results could potentially include both breast cancer cases truly missed at mammography and new cancers that grew quickly in the time following screening to the point where they became clinically apparent. In order to assess the impact of misclassifying those new breast cancers as false-negative results a sensitivity analysis was carried out. The intent of this analysis was to assess the impact of misclassification of false negative results on the estimates of sensitivity obtained.

The results from these analyses are presented in Table 3.38. The impact non-differential misclassification, the same degree of misclassification in each of the HT cohorts, is indicated in the cells that are boxed off. In all instances where there is non-differential misclassification of the false-negative results between the cohorts the sensitivity estimates are significantly different. When five percent of false-negatives were actually new cancers the sensitivity for current users of HT is 77% (95% CI 67%-86%) compared to a sensitivity of 93% (95% CI 88%-96%). In the most extreme group when 25% of true cancers in both of the HT cohorts were misclassified as false-negative results

the sensitivity for current HT users was 81% (95% CI 71%-89%) while non-users had a significantly higher estimate at 94% (90%-96%).

When the degree of misclassification was greater in current users of HT compared to non-users, the sensitivity estimates for current users are either significantly lower than those of non-users or there is minimal overlap of the 95% confidence intervals.

**Table 3.38: Corrected sensitivity estimates for current HT users and non-users under several assumptions for the percentage of false-negative results wrongly classified**

| | % of false-negative results wrongly classified | | | | | |
|---|---|---|---|---|---|---|
| | non-users | | | | | |
| current HT users | 0 | 5 | 10 | 15 | 20 | 25 |
| 0 | 92 (88-95)* | 93 (88-96) | 93 (89-96) | 93 (89-96) | 94 (90-96) | 94 (90-97) |
| | **76 (66-84)** | **76 (66-84)** | **76 (66-84)** | **76 (66-84)** | **76 (66-84)** | **76 (66-84)** |
| 5 | 92 (88-95) | 93 (88-96) | 93 (89-96) | 93 (89-96) | 94 (90-96) | 94 (90-97) |
| | **77 (67-86)** | **77 (67-86)** | **77 (67-86)** | **77 (67-86)** | **77 (67-86)** | **77 (67-86)** |
| 10 | 92 (88-95) | 93 (88-96) | 93 (89-96) | 93 (89-96) | 94 (90-96) | 94 (90-96) |
| | **78 (68-86)** | **78 (68-86)** | **78 (68-86)** | **78 (68-86)** | **78 (68-86)** | **78 (68-86)** |
| 15 | 92 (88-95) | 93 (88-96) | 93 (89-96) | 93 (89-96) | 94 (90-96) | 94 (90-96) |
| | **79 (69-87)** | **79 (69-87)** | **79 (69-87)** | **79 (69-87)** | **79 (69-87)** | **79 (69-87)** |
| 20 | 92 (88-95) | 93 (88-96) | 93 (89-96) | 93 (89-96) | 94 (90-96) | 94 (90-96) |
| | **80 (70-88)** | **80 (70-88)** | **80 (70-88)** | **80 (70-88)** | **80 (70-88)** | **80 (70-88)** |
| 25 | 92 (88-95) | 93 (88-96) | 93 (89-96) | 93 (89-96) | 94 (90-96) | 94 (90-96) |
| | **81 (71-89)** | **81 (71-89)** | **81 (71-89)** | **81 (71-89)** | **81 (71-89)** | **81 (71-89)** |

*For all cells the sensitivity for current HT users is located at the bottom and is in bolded. The effect of non-differential misclassification is shown in the boxed cells.

## B. Misclassification of HT status

### 1. Effect on sensitivity estimates

A sensitivity analysis quantifying the impact of misclassification of HT status on the sensitivity estimates was carried out. The results of this analysis are presented in Table 3.39. The degree of misclassification of HT status was assumed to be the same in the two cohorts. The impact of this misclassification when all false-negative results were assumed to be correctly classified and when 25% of false-negative results in the HT cohort and zero in the non-users were assumed to be misclassified was assessed. Under both assumptions for the misclassification of false-negative status, misclassification of HT status would result in an underestimation of the difference in sensitivity between the HT cohorts.

**Table 3.39: Corrected sensitivity estimates for current HT users (bold) and non-users under several assumptions for the sensitivity and specificity of HT status classification assuming no misclassification of false-negative status**

| accuracy of HT status classification | | assumptions for misclassification of false-negative status | |
|---|---|---|---|
| sensitivity | specificity | none in either cohort | 25% among current users / none among non-users |
| 100 | 100 | 92 (89-96) | 92 (89-96) |
| | | **77 (67-85)** | **81 (71-89)** |
| 90 | 90 | 92 (89-96) | 92 (89-96) |
| | | **69 (56-80)** | **75 (62-85)** |
| 80 | 80 | 92 (89-96) | 92 (89-96) |
| | | **69 (56-80)** | **75 (63-84)** |
| 90 | 80 | 92 (89-96) | 92 (89-96) |
| | | **34 (17-54)** | **44 (22-66)** |
| 80 | 80 | 92 (89-96) | 92 (89-96) |
| | | **34 (19-55)** | **44 (26-67)** |

## V. Likelihood ratios

The estimates for sensitivity and specificity described above were used in order to calculate likelihood ratios (LRs) for screening mammography.

### A. Unadjusted likelihood ratios

The LR positive and negative for the study population and the two cohorts are presented below in Table 3.40. The LR positive was lower for current users of HT compared to non-users – 11.52 (95% CI 10.36-12.81) compared to 18.64 (95% CI 17.67-19.03) for current users and non-users respectively. This indicates that current users of

HT are less likely to have a diagnosis of breast cancer given an abnormal screening mammography result. The LR negative was higher for current users of HT, 0.25 (95% CI 0.22-0.31), compared to non-users, 0.08 (95% CI 0.073-0.093). This indicates that current users of HT are more likely than non-users to have a diagnosis of breast cancer given a negative screening mammography result.

**Table 3.40: Likelihood ratios for screening mammography**

| | study population | HT cohort | |
|---|---|---|---|
| | | non-users | current users |
| LR positive | 16.06 (15.45-16.70) | 18.64 (17.67-19.03) | 11.52 (10.36-12.81) |
| LR negative | 0.13 (0.11-0.14) | 0.08 (0.07-0.09) | 0.25 (0.22-0.31) |

The likelihood ratios were used to determine a post-mammography risk for breast cancer. These risks are outlined in Table 3.41. Current HT users with an abnormal result had a risk of breast cancer following mammography of 91.1 per 1000 (95% CI 82.7-100.2). This post-test risk was significantly less than that of non-users who had a risk of 137.6 per 1000 (95% CI 133.2-142.0). Following a normal screening result, current HT users had a risk of breast cancer of 2.21 per 1000 (95% CI 1.84-2.67) which was significantly higher than the risk in non-users which was 0.72 per 1000 (95% CI 0.62-0.81).

**Table 3.41: Risk of breast cancer per 1000 women based on HT status and mammography result**

| HT status | breast cancer risk prior to mammography | breast cancer risk following mammography (95% CI) | |
|---|---|---|---|
| | | abnormal result | normal result |
| non-user | 8.62 | 137.6 (133.2-142.0) | 0.72 (0.62-0.81) |
| current user | 8.62 | 91.1 (82.7-100.2) | 2.21 (1.84-2.67) |

## B. Adjusted likelihood ratios

### 1. Likelihood ratios stratified by age

In order to assess the effect of age on the likelihood ratios, an age-stratified analysis was carried out. The results are presented below in Table 3.42 for LR positive and in Table 3.43 for LR negative.

**Table 3.42: Likelihood ratio positive stratified by age**

| age | study population | HT cohort | |
|---|---|---|---|
| | | non-users | current users |
| study population | 16.06 (15.45-16.70) | 18.64 (17.67-19.03) | 11.52 (10.36-12.81) |
| 50-54 | 12.88 (11.25-14.74) | 12.26 (10.27-14.65) | 13.97 (11.42-17.08) |
| 55-59 | 15.54 (14.31-16.87) | 17.08 (15.71-18.56) | 13.22 (11.21-15.60) |
| 60-64 | 16.59 (15.59-17.66) | 20.53 (19.50-21.62) | 8.92 (7.00-11.38) |
| 65-69 | 18.71 (17.52-19.99) | 21.76 (20.50-23.13) | 10.19 (7.96-13.05) |

The LR positive was significantly greater for non-users of HT for all ages except women aged 50 to 54 years. The LR positive appeared to decrease with age for current users of HT while it increased with age for non-users. The likelihood of having a

diagnosis of cancer given an abnormal screening result increased significantly with age for non-users of HT while there was a non-significant decrease with increasing age for current users of HT.

**Table 3.43: Likelihood ratio negative stratified by age**

| age | study population | HT cohort | |
|---|---|---|---|
| | | non-users | current users |
| study population | 0.13 (0.11-0.14) | 0.08 (0.07-0.09) | 0.25 (0.22-0.31) |
| 50-54 | 0.19 (0.14-0.25) | 0.21 (0.15-0.30) | 0.15 (0.09-0.25) |
| 55-59 | 0.14 (0.12-0.18) | 0.09 (0.07-0.11) | 0.23 (0.17-0.32) |
| 60-64 | 0.10 (0.09-0.12) | 0.050 (0.04-0.06) | 0.33 (0.23-0.47) |
| 65-69 | 0.11 (0.09-0.13) | 0.074 (0.06-0.09) | 0.30 (0.20-0.44) |

The LR negative was significantly greater current HT users compared to non-users for all women except those aged 50 to 54 years. Non-users aged 50 to 54 had a LR negative that was higher than current users. However, this difference was not statistically significant. Unlike with the LR positive there did not appear to be a change in LR negative with age in either HT cohort.

## 2. Likelihood ratios stratified by age and screening visit

A stratified analysis adjusting LR positive for screening visit in addition to age was also carried out. The results of this analysis are presented in Table 3.44. For the first visit LR positive was less in current users of HT compared to non-users for all women except those aged 50 to 54 years. The difference between the cohorts was significant only

for women aged 60 to 69. For women aged 50 to 54 years, current users had a significantly higher LR positive than non-users. The LR positive appeared to decrease with age for current HT users and while it increased with age for non-users. For the subsequent screening visit the LR positive was significantly less for current users of HT compared to non-users for women of all ages. The pattern of an increase or decrease of the LR positive with increasing age described above did not appear for women who had a previous Screen Test mammogram.

**Table 3.44: Likelihood ratio positive stratified by age and screening visit**

| screening visit and age | HT cohort | |
|---|---|---|
| | non-users | current users |
| first visit | | |
| 50-54 | 8.77 (6.54-11.75) | 13.14 (12.68-13.61) |
| 55-59 | 11.83 (10.18-13.74) | 10.91 (9.68-12.30) |
| 60-64 | 14.35 (12.50-16.48) | 7.46 (6.56-8.47) |
| 65-69 | 17.47 (15.08-20.24) | 8.33 (7.29-9.53) |
| subsequent visit | | |
| 50-54 | 36.53 (34.72-38.43) | 10.97 (4.11-29.26) |
| 55-59 | 34.03 (29.76-38.91) | 14.93 (11.22-19.88) |
| 60-64 | 38.55 (32.98-45.07) | 9.87 (7.62-12.77) |
| 65-69 | 30.47 (23.39-39.69) | 12.01 (9.71-14.85) |

LR negative was stratified by screening visit in addition to age. These results are presented in Table 3.45. For women with no history of a previous Screen Test mammogram the LR negative was greater for current users of HT compared to non-users for all age strata except women aged 50 to 54 years. The difference was significant for all women except those aged 55 to 59 years. For women age 50 to 54 years there were no

false negative results in the HT cohort. As a result the sensitivity in this cohort was 100% and therefore the LR negative was zero so comparisons between the cohorts for this age strata could not take place. The LR negative seemed to increase with age for current users of HT. However, there did not appear to be any significant difference in the LR negative with age for non-users.

For women with a history of a previous Screen Test mammogram the LR negative was greater for current users of HT compared to non-users for all age strata. Because there were no false negative results among non-users aged 50 to 54 the LR negative was zero, the differences between the cohorts for these women could not be examined. There did not appear to be an association with age and LR negative for women with a history of a previous Screen Test mammogram.

**Table 3.45: Likelihood ratio negative by age and screening visit**

| age | HT cohort | |
|---|---|---|
| | non-users | current users |
| first visit | | |
| 50-54 | 0.24 (0.14-0.41) | 0 |
| 55-59 | 0.10 (0.065-0.16) | 0.12 (0.087-0.17) |
| 60-64 | 0.052 (0.030-0.090) | 0.28 (0.22-0.34) |
| 65-69 | 0.057 (0.033-0.098) | 0.27 (0.22-0.34) |
| subsequent visit | | |
| 50-54 | 0 | 0.82 (0.20-1.40) |
| 55-59 | 0.060 (0.036-0.10) | 0.37 (0.25-0.55) |
| 60-64 | 0.047 (0.024-0.092) | 0.40 (0.29-0.56) |
| 65-69 | 0.11 (0.050-0.23) | 0.35 (0.26-0.48) |

The likelihood ratios described above were used to arrive at post-mammography probabilities for breast cancer. These probabilities are presented in Table 3.46 and Table 3.47. The differences between the HT cohorts with respect to the post-screening risk of breast cancer mirror those for likelihood ratios.

**Table 3.46: Pre- and post-mammography risk of breast cancer per 1000 women with no previous Screen Test mammogram stratified by age**

| age | pre-test risk | risk following mammography (95% CI) | | | |
|---|---|---|---|---|---|
| | | abnormal result | | normal result | |
| | | non-users | current HT users | non-users | current HT users |
| 50-54 | 7.49 | 62.1 (47.0-81.5) | 90.2 (87.3-93.1) | 1.80 (1.04-3.11) | # |
| 55-59 | 10.03 | 107.1 (93.6-122.2) | 99.6 (89.4-110.9) | 1.03 (0.66-1.60) | 1.22 (0.88-1.70) |
| 60-64 | 15.12 | 180.1 (161.0-201.9) | 102.7 (91.5-115.1) | 0.80 (0.46-1.38) | 4.25 (3.43-5.27) |
| 65-69 | 12.72 | 183.7 (163.7-206.8) | 96.9 (85.6-109.3) | 0.73 (0.42-1.26) | 3.53 (2.81-4.42) |

#Post-test risk was not calculated because likelihood ratio negative was zero.

**Table 3.47: Pre- and post-mammography risk of breast cancer per 1000 women with a previous Screen Test mammogram stratified by age**

| age | pre-test risk | risk following mammography (95% CI) | | | |
|---|---|---|---|---|---|
| | | abnormal result | | normal result | |
| | | non-users | current HT users | non-users | current HT users |
| 50-54 | 1.84 | 63.3 (60.4-66.5) | 1.99 (0.76-5.14) | # | 0.97 (0.36-2.58) |
| 55-59 | 6.06 | 171.9 (153.7-191.8) | 83.5 (64.0-108.2) | 0.37 (0.22-0.61) | 2.27 (1.55-3.32) |
| 60-64 | 5.98 | 188.4 (165.7-213.4) | 56.1 (43.9-71.4) | 0.28 (0.14-0.55) | 2.40 (1.73-3.34) |
| 65-69 | 7.12 | 179.24 (146.6-221.5) | 79.27 (65.1-96.2) | 0.76 (0.35-1.63) | 2.52 (1.88-3.39) |

#Post-test risk was not calculated because likelihood ratio negative was zero.

Additional factors were found to significantly confound or modify the association between current use of HT and specificity, including previous breast biopsy and ethnic

origin. However, the small number of breast cancers that were used to calculate sensitivity precluded stratification on these factors.

# 4. Discussion

## I. Overview of discussion section

This discussion is divided into five parts. First, an overview of the findings and possible explanations for these findings are described. Comparisons are made to other studies assessing the association between use of HT and the performance of screening mammography and explanations of any disparate results have also been included in this section. The next section highlights the strengths of this study and outlines the weaknesses of the study. Potential threats to internal validity are then addressed. The next section discusses the implications of the results of this study and the advantages and disadvantages of potential recommendations based on the findings. Finally, areas for further research are described.

## II. Overview of findings

This was the first study assessing the association between current use of HT and the performance of screening mammography that was conducted in a North American, organized breast cancer screening program which utilized 2-view mammography. The study found significant and potentially important decreases in the sensitivity, specificity, and positive predictive value of screening mammography for current users of HT compared to non-users even after adjustment for many of the factors that could have

potentially confounded the association. Likelihood ratios for screening mammography were also found to be significantly different between the HT cohorts.

## A. Sensitivity

There was a large and statistically significant decrease in sensitivity even after adjustment for potentially confounding factors.

### 1. Comparison to current evidence

The results obtained in this study are not inconsistent with those of previous studies that have assessed the association between HT and the sensitivity of screening mammography (6-9). In the literature there are four high quality studies which have assessed the impact of HT on the sensitivity of screening mammography. A summary of the sensitivity estimates and the limitations of the published studies in the literature was presented above in Table 1.1 on page 34. Three of these studies found a decrease in the sensitivity of screening mammography with current use of HT. Of the studies that found a decrease, two found one that was of a similar magnitude to that found in this study. The other study that found a decrease that was statistically significant but of a smaller magnitude (6%). One study found a non-significant increase in the sensitivity with current use of HT.

The estimates for sensitivity obtained in this study for both current and non-users of HT and the difference in sensitivity between the HT cohorts were very similar to those found by Laya (6) and Seradour (8). However, the difference found by Laya et al was not statistically significant.

A study conducted by Litherland et al found estimates for sensitivity that were higher than those obtained from this study and found a statistically significant but much smaller difference in the sensitivity between current and non-users of HT (9). Possible explanations for the higher sensitivity in the study conducted by Litherland et al could be that the program truly had a higher sensitivity. However, an alternative explanation could be incomplete ascertainment of interval cancers from the cancer registry. If there was incomplete ascertainment then the sensitivity could have been overestimated. However, the completeness of the registry was not reported so the impact of this could not be assessed.

The smaller difference in sensitivity between the HT cohorts could be due to the fact that the program was superior in detecting breast cancers. However, with this one would expect that, although the absolute difference was smaller, the relative decrease would be the same. This was not the case. In the current study the sensitivity among current users was 83% of that of non-users compared to 94% in the study conducted by Litherland et al. The population screened was different than that included in the current

study. It could be that HT had a different impact on breast density in this population, which translated to a different impact on the sensitivity of screening mammography. The current study did not find that self-reported ethnic origin was an independent predictor of sensitivity but had limited power to assess this association. Other potential explanations could be that the smaller difference between the cohorts is due to an artifact such as misclassification bias. If HT status was misclassified this would serve to dilute the impact of current use of HT on the sensitivity of screening mammography and lead to an underestimation of the difference in sensitivity between the HT cohorts. However, the study used self-report of HT status with which good agreement has be shown with pharmacy records so this explanation is not likely (76).

Thurfjell et al found a non-significant increase in the sensitivity of screening mammography (7). The estimate for current users of HT was higher than that found in this study. There are two potential explanations for this difference. The study population consisted of only women who had been screened previously through the program. It is possible that the sensitivity among current users of HT is higher for the subsequent screening visit. The current study had limited power to assess if this was the case. Another potential explanation is that the higher sensitivity found by Thurfjell et al is an artifact due to bias. If there were underascertainment of interval cancers and this took place to a greater degree amongst current users of HT then the sensitivity would be

overestimated. The method by which follow-up for diagnosis of breast cancer was carried out was not described and it is, therefore, impossible to rule out this possibility.

**2. Potential explanations for decrease in sensitivity with current use of HT**

There are several possible explanations for the observed decrease in sensitivity with current use of HT. Use of HT has been associated with an increased radiographic breast density (4, 5, 28-32). It is possible that this increase resulted in tumors being obscured in dense breast tissue in the HT cohort. For both HT cohorts, few of the interval cancers occurred in fatty breast tissue. Only 2 of the 20 interval cancers in the HT cohort and 4 of the 19 interval cancers that were diagnosed among non-users were diagnosed in women whose radiographic breast density was classified as fatty. Additionally, the decrease in sensitivity with current use of HT that was seen in this study is similar to the decrease in sensitivity with increased breast density that was observed by Kerlikowske et al (2). They found a decrease in the sensitivity of 16% for women with radiographically dense breasts compared to women with fatty breasts.

False-negative results were considered to include all cancers that were diagnosed in a one-year interval following a normal screening mammography result. These cancers could include not only cancers that could not be detected at screening, but also cancers that were not detectable by mammography at the time of screening but grew quickly during the interval following screening to the point were they became clinically apparent.

Day et al have proposed a method for estimating sensitivity which takes this into account (77). Use of this method in the study was considered. However, in order to employ the method it is necessary to know the cumulative distribution of the pre-clinical detectable phase for breast cancer for both women using HT and non-users. While there are estimates of this distribution, these estimates likely include a mixed group of women who are both non-users and current users of HT. Since exogenous estrogen leads to increased proliferation of breast tissue, it is possible that the duration of the pre-clinical detectable phase is shortened for women who use HT. For this reason, it was determined that this method could not be employed. Instead, a sensitivity analysis was used to assess the potential impact of misclassification of new cancers as false-negative results. Even under extreme circumstances where more false-negative results were misclassified in the HT cohort than amongst non-users, the difference in sensitivity between the HT cohorts remained and was statistically significant.

Cancers diagnosed following a normal screening result were ascertained through periodic merging of the Screen Test database with the Alberta Cancer Registry. Therefore, the ascertainment of the false-negative results was dependent on the completeness of the registry. If there was under-ascertainment of interval cancers then the sensitivity of screening mammography may have been overestimated. If there were a differing degree of under-ascertainment in the two HT cohorts then it is possible that a

difference between the cohorts could be an artifact of an incomplete registry. However, this explanation of the results found in this study is unlikely. The North American Association of Central Cancer Registries adjusted estimate of the completeness of the Alberta Cancer Registry is 94.7% with 98.2% of breast cancers being microscopically confirmed (78). Given that the cancer registry was so complete it is unlikely that enough cancers would have been missed so as to impact the study results.

## B. Specificity

The specificity of screening mammography was significantly less for current users of HT compared to non-users.

### 1. Comparison to current evidence

The difference in specificity with current use of HT is similar to the results of other studies that have assessed this association (6-8). Laya et al calculated an adjusted estimate of specificity for the HT cohort by using the mean value of significant covariates for each cohort in their logistic regression equation (6). The magnitude of the difference in specificity between the cohorts when this method was employed was 4%. In the current study the magnitude of the difference in the specificity of screening mammography between the HT cohorts at the mean age of the study population (60 years) for non-Aboriginal women with no history of breast biopsy was 1.3% for women with a previous Screen Test mammogram and 2.6% for women with no previous Screen

Test mammogram. This difference was smaller than that found be Laya et al. This is likely due to the fact that the abnormality rate was considerably lower in the current study compared to that in the program used by Laya et al. Thurfjell et al found a difference between the HT cohorts of 1% which was significant (7). However, their estimate was not adjusted for potentially confounding which could potentially explain the smaller size and, as described previously, the type of mammography used may have varied between the HT cohorts. Seradour et al also found a smaller difference in specificity between the HT cohorts (8). One reason for the smaller magnitude of this difference could have been the low abnormality rate in the screening program (4%). Potential explanations could be that the response of breast in the population in this study could have a different response to HT, which could lead to a smaller decrease in specificity with use of HT. Another explanation for this smaller magnitude could be that it is due to an artifact resulting from a bias such as confounding. This study did adjust for age and first versus subsequent screening visit but there may have been residual confounding by other factors that were not adjusted for.

## 2. Potential explanations for reduced specificity

The reduced specificity that was seen with current use of HT could be the result of several factors. Increased radiographic breast density has been associated with use of HT (4, 5, 28-32). Increased radiographic breast density could lead to an increased number of

abnormal readings. This study did find that women who were currently using HT were less likely to have a fatty radiographic appearance of the breast on mammography than non-users. The magnitude of the difference in the proportion of women fatty radiographic breast density between the HT cohorts increased with increasing age. This increase resulted because the proportion of women with fatty radiographic breast density amongst women in the HT cohort remained constant with increasing age while the proportion increased with age amongst non-users. The differences in specificity mirrored these differences in breast density. For women with no previous Screen Test mammogram and no previous breast biopsy, the specificity was approximately 91% for women of all ages who were currently using HT while it increased from 91% in women aged 50 years to 95% in women aged 69 years amongst women who were not using HT. The differences in specificity were not significant for younger women. Younger women are likely closer to menopause and it is possible that the involutional process by which ductal tissue is replaced with adipose tissue may not be complete and as a result there is no difference between the specificity amongst current and non-users of HT.

Another potential explanation of the change in specificity could be that the threshold for a radiologist calling a mammogram abnormal is lower for women using HT, potentially due to the perceived possibility of an increased risk for breast cancer amongst these women. However, if this were the case one would expect that decreasing the

threshold for a mammogram to be abnormal would result in an increase in sensitivity along with a decreased specificity. The sensitivity of screening mammography was also decreased with current use of HT, which suggests that the explanation of a different abnormal threshold level for women who use HT is less likely.

There were a number of other interesting findings in regards to specificity. The difference between the cohorts was greater for a first screening visit compared to a subsequent screening visit. The smaller difference between the HT cohorts at the subsequent screening visit may be the result of a larger proportion of women in this group having a high quality mammogram available for comparison. All women who had a subsequent screening visit had previous mammographic films of comparable quality that were used for comparison. Women being screened at Screen Test for the first time may not have previously had a mammogram. If these women had had a previous mammogram there may be no films available for comparison or the quality of these films may have been variable.

Another hypothesis for the smaller difference between the HT cohorts with women who had been screened previously through the Screen Test program was that perhaps these women had been using HT at their previous appointment and there were fewer new changes and subsequently fewer false-positive results. However, a post hoc analysis assessing the effect of use of HT at the previous appointment on specificity did

not confirm this hypothesis. Of the women currently using HT who had been screened previously at Screen Test and had no history of a previous breast biopsy, 79% had reported hormone use at their previous visit. When stratified by age, the specificity was higher for women who reported no use of HT at her previous appointment compared to women who reported past use.

Another finding of interest was that the specificity was higher for women who reported being of Aboriginal descent compared to that of women who reported non-Aboriginal ethnic origins and there was no significant difference in specificity between the HT cohorts for women who were of Aboriginal descent. This may indicate that there are differences in the response of the breast tissue of Aboriginal women to exogenous estrogen. Another potential explanation could be that the effect of HT on breast density is less pronounced in women who are obese. An increased weight, when adjusted for height, is associated with a lower radiographic breast density (42, 44, 47). The prevalence of obesity is high in Aboriginal populations (75). A high prevalence of obese women among women of Aboriginal descent could potentially be the reason for the higher specificity in this group. There was no information in the Screen Test database on weight or height so these hypotheses could not be pursued.

## C. Positive predictive value

The effect of HT on the positive predictive value of screening mammography has not been previously assessed. This study found that positive predictive value was significantly reduced in current users of HT compared to non-users and that the amount of the reduction increased with increasing age. Positive predictive value is determined by both the specificity of screening mammography and the prevalence of breast cancer. The prevalence of breast cancer increased with age and there was no difference in the prevalence of breast cancer between the HT cohorts. When the increase in breast cancer prevalence with age is considered the differences in positive predictive value. as with specificity. reflect the differences in radiographic breast density between the HT cohorts. Amongst non-users both the specificity of screening mammography and the prevalence of breast cancer increase with age while in current users of HT specificity remains constant while the prevalence of breast cancer increases.

## D. Likelihood ratios and post-mammography breast cancer risk

This study attempted to put the differences in the performance of screening mammography between the HT cohorts into a form that was useful for clinicians by calculating likelihood ratios for screening mammography for current and non-users of HT. These likelihood ratios may be employed by clinicians to understand the implications of the decrease in sensitivity and specificity of mammography on what a particular

mammography result means in terms of breast cancer risk. Likelihood ratios can be employed to determine the post-screening mammography risk of breast cancer given a particular test result and a pre-screening risk of breast cancer. Women using HT who had an abnormal screening result had a lower likelihood ratio positive than non-users and, therefore, a lower post-mammography risk of breast cancer than non-users. Women using HT who had a normal screening result had a higher likelihood ratio negative than non-users and therefore a higher risk of breast cancer. Since likelihood ratios are determined using sensitivity and specificity the potential explanation for the differences in sensitivity and specificity between current and non-users of HT also apply to the differences that were found with likelihood ratios.

## III. Strengths and weaknesses of study

### A. Strengths of study

This study had many strengths including its size, its generalizability, the quality of the screening program in which it was conducted and the wealth of information in the Screen Test program database.

#### 1. Size

The study included more than 37, 000 women and was larger than any study conducted to date that assessed the effect of HT on the performance measures of

screening mammography in the context of a North American, organized breast cancer screening program which utilized 2-view mammography. There were 37, 590 women in the study who did not have breast cancer and 2, 349 women who had abnormal screening results. Therefore, the power to detect a difference in specificity or positive predictive value between the cohorts was extremely high. There was also a high statistical power to determine if there was in fact confounding or effect modification by the potentially confounding factors.

The study had 327 breast cancer cases which was greater than the previous two studies that had attempted to assess this association which had only 63 and 142 breast cancers (6, 7). Therefore, the study had increased statistical power to detect a difference in sensitivity between the HT cohorts. Additionally, it was possible to attempt to adjust the sensitivity estimate for factors that could potentially confound the association.

**2. Population based screening program**

The study was conducted using the population based screening program in Alberta and the potential that the results are generalizable to the women who utilize programmatic breast cancer screening that utilizes two-view mammography is high.

### 3. Organized screening program

The study was conducted in a screening program that meets the guidelines for an organized screening program, most importantly the abnormality rate. The abnormality rate for the program was seven percent for current users of HT and six percent for non-users. One of the previous studies that found a decrease in the specificity of screening mammography was conducted in a program that had an abnormality rate of 16% (6). This study shows that the decrease in specificity with current use of HT, although smaller, holds even when the abnormality rate is within the guidelines for an organized breast cancer-screening program.

### 4. First and subsequent screening visits

The study employed information from first and subsequent screening visits. Therefore, it was possible to determine if differences in the performance of screening mammography between current and non-users of HT varied depending on whether it was a woman's first or subsequent visit.

### 5. Information on factors which could potentially confound the association

Study participants completed a detailed questionnaire at the time of screening which provided information on nearly all of the factors which could potentially confound the association between current use of HT and a decrease in the effectiveness of screening

mammography. This information was then used in order to adjust the estimates for potentially confounding factors.

## B. Limitations of the study

The limitations of this study stem from the fact that secondary data - data not collected for the intent of addressing the study hypothesis - was employed. The Screen Test database did not contain detailed information on HT use and there was only crude information on radiographic breast density.

### 1. Limited details on HT use

#### (a) Duration and timing of use

The information on HT status was limited to only information on current use or non-use of HT. There was no information on former use of HT. There is a suggestion that the effect of HT are limited to current use. One study found that the effect of HT on the breast is transient (35). Additionally, the study by Laya et al assessing the effect of HT on the effectiveness of mammography found that the effects of HT on the specificity of mammography was limited to current users and there was no effect for former users (6). This effect could not be explored in this study due to the lack of information on past use of HT.

There was no information on compliance to therapy. If women who reported current use of HT were non-compliant then the association between HT and a decrease in sensitivity, specificity and positive predictive value may be underestimated. Additionally, there was no information on the date when HT was commenced. As a result, the group of women who reported current use of HT could have included women who had used HT for several years and women who had only been on the therapy for weeks or days. For women who had been on HT for only a short time it is possible that their breast tissue may not have yet responded to the same degree as women who had used the therapy for a longer period of time. If this was the case, then the magnitude of the association between HT and a decrease in the effectiveness of mammography may have been underestimated.

**(b) Type of preparation**

There is evidence that the degree of increase in breast density is dependent on the type of HT preparation used. Greendale et al have shown that women using combined estrogen-progesterone preparations have a greater increase in breast density than women who are using only estrogen preparations (32). The Screen Test database contains no information on the type of preparation. Thus, this study was unable to assess the impact of type of HT preparation on the performance of mammography.

### 2. Measurement of radiographic breast density

In this study the radiographic breast density was classified using a subjective measurement. Radiographic breast density was not a study outcome and, therefore, its misclassification does not impact the validity of the study results. However, the validity of the measurement of breast density is still a concern as it has been used as an explanation for the study results. If misclassification of breast density took place but occurred independently of HT status then the association between HT status and breast density would be underestimated. However, if misclassification was dependent on HT status then the association between current use of HT and increased breast density could be overestimated. The latter is impossible to rule out as the radiologists who read the mammograms do have access to information on current HT use. However, this information is located on the self-administered questionnaire that is separate from the mammography report where radiologists classify breast density.

## IV. Threats to validity of results

As with any epidemiological study there are potential threats to the internal validity of the results obtained from this study. The potential for misclassification of HT exposure and cancer status and the potential misclassification of new cancers as false-negative results were potential limitations of the study. The study was also limited by not having information on known confounding factors.

## A. Misclassification bias

There is the potential that HT exposure could have been misclassified. HT exposure status was determined from a self-administered questionnaire. However, a study examining characteristics of HT users found that over 90% of women who reported use of HT and filling their prescription at a particular pharmacy did have evidence of a prescription for HT being dispensed in the pharmacy's database (76). The information on HT status was reported in advance of the mammography so there is no reason to suspect that misclassification of HT status would depend on screening outcome (true-positive result, false-positive result, etc.). Additionally, a sensitivity analysis quantifying the impact of non-differential misclassification of HT status on the estimates of sensitivity and specificity showed that with increasing misclassification the true difference between the cohorts would be underestimated.

## B. Potential for residual confounding

The Screen Test database did not contain information on alcohol intake and anthropomorphic measurements such as weight and height. These factors have been identified to be associated with both use of HT and breast parenchymal density. Alcohol has been found to be only weakly correlated with breast density (52). Given that the effect of alcohol on breast density is likely small in relation to use of HT it is unlikely to be a possible explanation for the study findings. The lack of information on body mass

index is a concern. Use of HT is associated with a leaner body mass (BMI<27) (11, 15, 17, 18, 55, 58) and, when adjusted for height, a lower weight was also associated with increased radiographic breast density (42, 44, 47). However, a high quality study assessing the impact of use of HT on radiographic breast density found no difference in the relative risk of increased radiographic breast density with current use of HT when age was the only covariate compared to a model with age, height, weight, parity, menstrual status and previous use of oral contraceptives (4). Additionally, the one previous study that assessed the impact of HT on the specificity of screening mammography which adjusted for potentially confounding factors did not find that body mass index was a predictor of specificity (6).

It is also possible that other confounding factors of the association between use of HT and screening mammography outcomes exist but have not yet been identified.

## C. Random error

The study was large and, therefore, had high power to assess differences in the performance of screening mammography between the two HT cohorts. The one instance where power became an issue was adjusting the sensitivity estimate for potentially confounding factors. However, when all of the potentially confounding factors were included in the logistic regression model, regardless of whether or not they were significatn, the effect of HT was not changed.

## V. Implications of results

Reductions in the sensitivity, specificity, positive and negative predictive values of screening mammography for women using HT are important. Even small reductions in specificity result in increased costs and potentially increased morbidity from the screening program. The majority of women who are screened through the Screen Test Program are non-Aboriginal women who have had no previous Screen Test mammogram and no previous biopsy. The specificity of screening mammography amongst women in this group who were aged 60 years is reduced from 94% in non-users to 91% in current users of HT. If the prevalence of breast cancer is 5 per 1000 women screened then for every 10. 000 women using HT who are screened an additional 299 false-positive results will occur compared to 10. 000 women screened who are not using HT.

At a minimum these women have additional views with some going on to biopsy. A biopsy following an abnormal result is not uncommon. In the period from 1995 to 1997 in the Screen Test Program there were 1.266 abnormal mammograms which resulted in 4.331 diagnostic procedures of which nearly 12% were biopsies (69). There is the potential for morbidity with biopsy so biopsies associated with no diagnosis of cancer are concerning. Additionally, diagnostic procedures associated with false-positive outcomes result in increased financial costs to the health care system. Finally, false-positive results are associated with increased morbidity through psychological stress and anxiety (79, 80).

With a reduction in sensitivity, as was found to occur with current use of HT, fewer existing breast cancers are diagnosed at screening than in non-users of HT. As a result these cancers may present as interval cancers or they may go undetected during the interval between screening mammograms and be diagnosed at a subsequent screening visit. The concern with both of these scenarios is that women using HT may have a breast cancer for a longer period at the point at which it is diagnosed, as an interval cancer or by subsequent mammographic screening. Thus the potential exists that breast cancers diagnosed in women who take HT may be of a later stage and a larger size, characteristics that are associated with a poorer prognosis, and screening mammography may be less able to reduce mortality from breast cancer in these women.

The reduction in positive predictive values means that for every one-hundred women who do not use HT who have an abnormal screening result fourteen will have a subsequent diagnosis of breast cancer compared to only eight women who are current users of HT. While the difference between the cohorts with respect to negative predictive value is extremely small it is important because the majority of women who are screened have a normal screening result. For every 10, 000 current users of HT who have as normal screening result 20 will have a subsequent diagnosis of breast cancer. However, only seven of 10, 000 non-users with normal screening results will have a subsequent diagnosis of breast cancer.

Changes in the performance of screening mammography for current users of HT mean that abnormal or normal results are associated with different breast cancer risks than for women who are not using HT. The likelihood ratio positive that has been determined in this study provides a useful means for clinicians to determine the post-mammography risk of breast cancer following an abnormal for a women given her HT status, age, and whether or not she is having a first of subsequent screen. For instance, the likelihood ratio positive for a woman who is aged between 60 to 64 years and has had a previous mammogram is about 39 for women who do not use HT and 10 for women who are currently using HT. If the risk of breast cancer prior to screening mammography is 6 per 1000 in this group of women then the risk of breast cancer following an abnormal screening mammogram is 188 per 1000 for non-users but is only 56 out of 1000 current users.

The likelihood ratio negative was found to be significantly higher for women who are current users compared to non-users of HT meaning that women who use HT are more likely to have breast cancer following a normal result. However, this difference is extremely small and may not be as informative in a clinical setting. For example, for the same group of women described above the likelihood ratio negative for women not using HT is 0.05 and 0.40 for women currently using HT. Assuming the breast cancer risk prior to mammography is 6 per 1000 then the risk of breast cancer following a normal

screening mammogram is 0.3 per 1000 for women not using HT compared to 2.4 per 1000 for women who are currently using HT.

## VI. Recommendations based on findings

The current use of HT has been found to be associated with a small decrease in specificity and a large decrease in sensitivity. Based on these findings, it is intuitive to suggest that practices surrounding the mammographic screening of women who use HT should be assessed and potentially modified in order to ensure the maximum benefit from screening. However, while the effectiveness of a screening program is intimately associated with these performance indicators, this study provides no direct evidence that mammographic screening as it is currently performed is less effective in reducing mortality from breast cancer in women who use HT. Furthermore, there is no evidence to suggest that modifying screening practices would improve the effectiveness of screening.

It is important to consider potential recommendations and to understand the implications of their implementation. One recommendation could be that radiologists should ensure that they are aware of whether or not a women is currently using HT and take this into consideration when reading a mammogram. However, there is the potential that this practice may result in a further decrease in specificity resulting in more false-positive results with an unknown impact on sensitivity.

In Alberta, the current recommendation for the early detection of breast cancer is mammographic screening for women aged 50 to 69 years every two years coupled with yearly clinical breast exam by a trained health care practitioner and regular breast self examination by the woman (81). Given that even in a one-year follow-up period 24% of breast cancers diagnosed in women who were current users of HT were diagnosed as interval cancers, it is tempting to suggest that the screening interval for these women should be compressed to no more than yearly. However, there is no evidence that a change in practice such as this would increase the number of breast cancers detected. Furthermore, given that the specificity of screening mammography is reduced with current use of HT, the shorter screening interval may succeed only in increasing the number of false-positive screening results. As has been mentioned previously, these false-positive results increase the cost and morbidity associated with mammographic screening. Additionally, women who use HT, especially for long durations, are potentially at an increased risk for breast cancer (22) and the effect of additional radiation exposure through more frequent mammographic screening in this group is not known. It may also be tempting to encourage women using HT to have a clinical breast examination performed more frequently than is recommended. However, there is no evidence that clinical breast examination reduces mortality from breast cancer (82) and more frequent visits to a primary health provider for clinical breast examination may result only in increased costs to the health care system.

There is a suggestion that changes in mammographic density may regress only two weeks after cessation of HT (35). Discontinuation hormone therapy for a period of time preceding mammography may improve its effectiveness. However, there could be negative effects associated with the discontinuation of HT. Additionally, there is no evidence that this is effective in improving the performance characteristics of screening mammography so making a recommendation in this regard would be premature.

Women should be informed about the limitations of screening mammography. However, the manner in which this information is communicated should be carefully considered. Currently, women who participate in Screen Test and receive a normal screening result are informed that 10% of breast cancers are not found by mammography and that they should practice regular breast self examination and have a clinical breast examination yearly. While it may be appealing to modify this message on the limitations of mammography so that women who use HT are informed that nearly 25% of cancers are not detected mammographically, there are potential adverse consequences of doing so. Women may, as a result of this information, choose not to undergo screening mammography. Another consequence may be that women may cease use of HT. Despite the negative consequences of use of HT there are proven benefits of its use (83). Physicians should add the potential effect of the current use of HT on the performance of screening mammography to the issues that they discuss with women who are using or are

considering using hormone therapy. Women should be informed that the performance of screening mammography is poorer if they are using HT but a simple statement of 25% of cancers may not be detected by mammography may not be the most prudent means of presenting this information.

## VII. Areas for further research

There are many opportunities for further research in the area of use of HT and its effects on screening mammography. Further research that assesses the impact of use of HT on the performance of screening mammography should be carried out. These studies would serve two purposes. One would be to attempt to replicate the results that were found in this study in other organized population based screening programs. The second would be to address the limitations of this, and other previous studies, particularly in regards to the type of HT preparation used and the dosing employed. The ability to conduct this sort of analysis is dependent on having a large number of breast cancer cases so that there is sufficient power to be able to ascertain differences between HT cohorts with respect to sensitivity. It would take many years for one screening program to amass a sufficient number of breast cancer cases in order to have sufficient statistical power to conduct this sort of analysis. Therefore, it is essential that collaborative research amongst organized breast cancer screening programs take place in order to be truly able to further research in this area.

Perhaps the most important area of further study that should be pursued is to assess whether or not the reductions in the performance measures for screening mammography that have been observed do lead to a decreased effectiveness of screening mammography for women who are current users of HT. It is unlikely that it would be feasible to assess this association through the use of a randomized controlled trial. A trial of this sort would require that large numbers of women would have to be randomized to use or non-use of HT and would require long term follow-up. This could potentially be addressed a cohort study but this would also require long-term follow-up and particular attention would have to be paid to the impact of potential biases.

It is also important to assess if there are means by which the performance of screening mammography for women who use HT could be improved. This could entail assessing whether or not cessation of HT for a period preceding mammography results in an improvement in the sensitivity and specificity of screening mammography. Finally, it would be useful to assess whether or not the decreased specificity that was observed with current use of HT results in an increased cost of diagnostic procedures following screening mammography for women who are currently using HT.

## VIII. Conclusions

This study found current use of HT is associated with a decrease in the sensitivity, specificity and positive predictive value of screening mammography. The increase in the

radiographic breast density that has been found to be associated with use of HT (4, 5, 28-32) is likely the explanation for these decreases. Further research should to be undertaken in order to elucidate whether or not the poorer performance of screening mammography results in a decrease in the effectiveness of screening mammography for women who are current users of HT.

## 5. References

1. Bird R, Wallace T, Yankaska B. Analysis of cancers missed at screening mammography. Radiology 1992;184:613-617.

2. Kerlikowske K, Grady D, Barclay J, Sickles E, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammogram. JAMA 1996;276(1):33-38.

3. Ma L, Fishell E, Wright B, Hanna W, Allan S, Boyd N. Case-control study of factors associated with failure to detect breast cancer by mammography. J Natl Cancer Inst 1992;84:781-785.

4. Persson I, Thurfjell E, Holmberg L. Effect of estrogen and estrogen-progestin replacement regimens on mammographic breast parenchymal density. J Clin Oncol 1997;15(10):3201-7.

5. Stomper PC, DJ DS, DiNitto PA, Arredondo MA. Analysis of parenchymal density on mammograms in 1353 women 25-79 years old. AJR Am J of Roentgenol 1996;167(5):1261-5.

6. Laya MB, Larson EB, Taplin SH, White E. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. J Natl Cancer Inst 1996;88(10):643-9.

7. Thurfjell EL, Holmberg LH, Persson IR. Screening mammography: sensitivity and specificity in relation to hormone replacement therapy. Radiology 1997;203(2):339-41.

8. Seradour B, Esteve J, Heid P, Jacquemier J. Hormone replacement therapy and screening mammography: analysis of the results in the Bouches du Rhone programme. J Med Screen 1999;6:99-102.

9. Litherland JC, Stallard S, Hole D, Cordiner C. The effect of hormone replacement therapy on the sensitivity of screening mammograms. Clin Radiol 1999;54:285-288.

10. Kaufert PL. The menopausal transition; the use of estrogen. Can J Public Health 1986;Supp 77:86-91.

11. Brett KM, Madans JH. Use of postmenopausal hormone replacement therapy: estimates from a nationally representative cohort study. Am J Epidemiol 1997;145(6):536-45.

12. Nagata C, Matsushita Y, Shimizu H. Prevalence of hormone replacement therapy and users' characteristics: a community survey in Japan. Maturitas 1996;25:201-207.

13. Moorhead T, Hannaford P, Warskyj M. Prevalence and characteristics associated with use of hormone replacement therapy in Britain. Br J Obstet Gynaecol 1997;104:290-297.

14. MacLennan AH, MacLennan A, Wilson D. The prevalence of oestrogen replacement therapy in South Australia. Maturitas 1993;16(3):175-83.

15. Hammar M, Brynhildsen J, Dabrosin L, Frisk J, Lindgren R, Nedstrand E. Hormone replacement therapy and previous use of oral contraceptives among Swedish women. Maturitas 1996;25:193-199.

16. Stadberg E, Mattsson L, Milsom I. The prevalence and severity of climacteric symptoms and the use of different treatment regimens in a Swedish population. Acta Obstet Gynecol Scand 1997;76:442-448.

17. Derby CA, Hume AL, Barbour MM, McPhillips JB, Lasater TM, Carleton RA. Correlates of postmenopausal estrogen use and trends through the 1980s in two southeastern New England communities. Am J Epidemiol 1993:137(10):1125-35.

18. Johannes CB, Crawford SL, Posner JG, McKinlay SM. Longitudinal patterns and correlates of hormone replacement therapy use in middle-aged women. Am J Epidemiol 1994;140(5):439-52.

19. Topo P, Klaukka T, Hemminki E, Uutela A. Use of hormone replacement therapy in 1976-89 by 45-64 year old Finnish women. J Epidemiol Community Health 1991;45:277-280.

20. Bergkvist L, Persson I. Hormone replacement therapy and breast cancer. Pharmacoepidemiology 1996;15(5):360-370.

21. Brinton LA. Hormone replacement therapy and risk for breast cancer. Endocrinol Metab Clin North Am 1997;26(2):361-378.

22. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 075 women with breast cancer and 108 411 without breast cancer. Lancet 1997;350:1047-1059.

23. Peck D, Lowman R. Estrogen and the postmenopausal breast. JAMA 1978;240:1733-35.

24. Berkowitz J, Gatewood O, Goldblum L, Gayler B. Hormonal replacement therapy: mammographic manifestations. Radiology 1990;174:199-201.

25. Cyrlak D, Wong CH. Mammographic changes in postmenopausal women undergoing hormonal replacement therapy. AJR Am J Roentgenol 1993;161(6):1177-83.

26. Laya MB, Gallagher JC, Schreiman JS, Larson EB, Watson P, Weinstein L. Effect of postmenopausal hormonal replacement therapy on mammographic density and parenchymal pattern. Radiology 1995;196(2):433-7.

27. Stomper PC, Van Voorhis BJ, Ravnikar VA, Meyer JE. Mammographic changes associated with postmenopausal hormone replacement therapy: a longitudinal study. Radiology 1990;174(2):487-90.

28. Kaufman Z, Garstin W, Hayes R, Mitchell M, Baum M. The mammographic parenchymal patterns of women on hormonal replacement therapy. Clin Radiol 1991;43:389-392.

29. Marugg RC, van der Mooren MJ, Hendricks JHCL, Rolland R, Ruijs SHJ. Mammographic changes in post menopausal women on hormonal replacement therapy. Eur Radiol 1997;7:749-755.

30. McNicholas MM, Heneghan JP, Milner MH, Tunney T, Hourihane JB, MacErlaine DP. Pain and increased mammographic density in women receiving hormone replacement therapy: a prospective study. AJR Am J Roentgenol 1994;163(2):311-5.

31. Leung W, Goldberg F, Zee B, Sterns E. Mammographic density in women on postmenopausal hormone replacement therapy. Surgery 1997;122(4):669-73.

32. Greendale GA, Reboussin BA, Sie A, et al. Effects of estrogen and estrogen-progestin on mammographic parenchymal density. Ann Intern Med 1999;130:262-269.

33. Wolfe JN. Breast patterns as an index of risk for developing breast cancer. AJR Am J Roentgenol 1976;126(6):1130-7.

34. Wolfe JN. Breast parenchymal patterns and their changes with age. Radiology 1976;121(3 Pt. 1):545-52.

35. Harvey JA, Pinkerton JV, Herman CR. Short-term cessation of hormone replacement therapy and improvement of mammographic specificity. J Natl Cancer Inst 1997;89(21):1623-5.

36. Erel CT, Seyisoglu H, Senturk ML, et al. Mammographic changes in women on hormonal replacement therapy. Maturitas 1996;25(1):51-7.

37. Holland R, Hendricks J, Mrevunac M. Mammographically occult breast cancer. Cancer 1982;52:1810-1819.

38. Fajardo LL, Hillman BJ, Frey C. Correlation between breast parenchymal patterns and mammographers' certainty of diagnosis. Invest Radiol 1988;23:505-508.

39. van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? Eur J Epidemiol 1998;14(4):315-20.

40. Junqueira LC, Carneiro J, Kelley RO. Basic Histology. (7 ed.) Norwalk, Connecticuit: Appleton and Lange, 1992.

41. Hutson SW, Cowen PN, Bird CC. Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer. J Clin Pathol 1985;38:281-287.

42. Bartow SA, Pathak DR, Mettler FA, Key CR, Pike MC. Breast mammographic pattern: a concatenation of confounding and breast cancer risk factors. Am J Epidemiol 1995;142(8):813-9.

43. Bergkvist L, Tabar L, Bergstrom R, Adami HO. Epidemiologic determinants of the mammographic parenchymal pattern. A population-based study within a mammographic screening program. Am J Epidemiol 1987;126(6):1075-81.

44. Brisson J, Morrison A, Khalid N. Mammographic parenchymal features and breast cancer in the breast cancer detection demonstration project. J Natl Cancer Inst 1988;80:1534-1540.

45. Buchanan J, Weisberg B, Sandoz J, Gray L, Bland K. Selected prognostic variables for mammographic parenchymal patterns. Cancer 1980;47:2135-2137.

46. Whitehouse GH, Leinster SJ. The variation of breast parenchymal patterns with age. Br J Radiol 1985;58(688):315-8.

47. Gravelle IH, Bultrode JC, Bulbrook RD. The relation between radiological patterns of the breast and body weight and height. Br J Radiol 1982;55:23-25.

48. Leinster SJ, Walsh PV, Whitehouse GH, al-Sumidaie AM. Factors associated with mammographic parenchymal patterns. Clin Radiol 1988;39(3):252-6.

49. White E, Velentgas P, Mandelson MT, et al. Variation in mammographic breast density by time in menstrual cycle among women aged 40-49 years. J Natl Cancer Inst 1998;90(12):906-10.

50. Wolfe JN, Albert S, Belle S, Salane M. Familial influences on breast parenchymal patterns. Cancer 1980;46(11):2433-7.

51. Brenner RJ, Pfaff JM. Mammographic changes after excisional breast biopsy for benign disease. AJR Am J Roentgenol 1996;167(4):1047-52.

52. Herrinton LJ, Saftlas AF, Stanford JL, Brinton LA, Wolfe JN. Do alcohol intake and mammographic densities interact in regard to the risk of breast cancer? Cancer 1993;71:3029-35.

53. Turnbull AE, Kapera L, Cohen ME. Mammographic parenchymal patterns in Asian and Caucasian women attending for screening. Clin Radiol 1993;48(1):38-40.

54. Gravelle IH, Bulbrook RD, Wang DY, et al. A comparison of mammographic parenchymal patterns in premenopausal Japanese and British women. Breast Cancer Research and Treatment 1991;18:S93-S95.

55. Parazzini F, La Vecchia C, Negri E, Bianchi C, Fedele L. Determinants of estrogen replacement therapy use in northern Italy. Rev Epidem et Sante Publ 1993;41:53-58.

56. Groeneveld FPMJ, Bareman FP, Barentsen R, Dokter HJ, Drogendijk AC, Hoes AW. Determinants of first prescription of hormone replacement therapy: a follow-up study among 1689 women aged 45-60 years. Maturitas 1995;20:81-89.

57. Limouzin-Lamothe MA, Mairon N, Joyce CR, Le Gal M. Quality of life after the menopause: influence of hormonal replacement therapy. Am J Obstet Gynecol 1994;170(2):618-24.

58. Cauley JA, Cummings SR, Black DM, Mascioli SR, Seeley DG. Prevalence and determinants of estrogen replacement therapy in elderly women. Am J Obstet Gynecol 1990;163(5 Pt 1):1438-44.

59. Ghali WA, Freund KM, Boss RD, Ryan CA, Moskowitz MA. Menopausal hormone therapy: physician awareness of patient attitudes. Am J Med 1997;103(1):3-10.

60. Ringa V, Ledesert B, Breart G. Determinants of hormone replacement therapy among postmenopausal women enrolled in the French GAZEL cohort. Osteoporos Int 1994;4(1):16-20.

61. Fletcher SW, Black W, Harris R, Rimer B, Shapiro S. Report of the international workshop on screening for breast cancer. J Natl Cancer Inst 1993;85:1644-56.

62. Kerlikowske K, Grady D, Rubin S, Sandrock C, Ernster VL. Efficacy of screening mammography: a meta-analysis. JAMA 1995;273(2):149-154.

63. Rosenberg RD, Hunt WC, Williamson MR, et al. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183, 134 screening mammograms in Alburquerque, New Mexico. Radiology 1998;209:511--518.

64. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 1992;30(1):187-210.

65. Roubidoux MA, Wilson TE, Orange RJ, Fitzgerald JT, Helvie MA, Packer SA. Breast cancer in women who undergo screening mammography: relationship of hormone replacement therapy to stage and detection method. Radiology 1998;208(3):725-8.

66. Litherland JC, Evans AJ, Wilson AR. The effect of hormone replacement therapy on recall rate in the National Health Service Breast Screening Programme. Clin Radiol 1997;52(4):276-9.

67. Going J, Anderson T, Battersby. Proliferative and secretory activity in human breast during natural and artificial menstrual cycles. Am J Pathol 1988;130:193-204.

68. Bernstein L, Ross R. Endogenous hormones and breast cancer risk. Epidemiol Rev 1993;15:48-65.

69. Screen Test: Alberta Program for the Early Detection of Breast Cancer. 1995/97 Biennial Report. Calgary, AB: Alberta Cancer Board, 1998.

70. Statistics Canada. Individuals File (Flat ASCII File) (public use microdata files: 1996 Census of Population). 1999.

71. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: Van Norstrand Reinhold, 1982.

72. Coughlin SS, Trock B, Criqui MH, Pickle LW, Browner D, Tefft MC. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. J Clin Epidemiol 1992;45(1):1-7.

73. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley and Sons, 1989.

74. Rothman KJ, Greenland S. Modern epidemiology. (2nd ed.) Philadelphia: Lippincott-Raven, 1998.

75. MacMillan HL, MacMillan AB, Offord DR, Dingle JL. Aboriginal Health. CMAJ 1996;155(11):1569-1578.

76. Buist DSM, LaCroix A, Newton KM, Keenan NL. Are long-term hormone replacement therapy users different from short-term and never users? Am J Epidemiol 1999;149(3):275-81.

77. Day NE. Estimating the sensitivity of a screening test. J Epidemiol Community Health 1985;39:364-366.

78. Chen VW, Wu XC, Andrews PA, eds. Cancer in North America, 1991-1995. Vol 1. Louisiana: North American Association of Central Cancer Registries, 1999.

79. Rimer BK, Bluman LG. The psychosocial consequences of mammography. Journal of the National Cancer Institute. Monographs 1997(22):131-8.

80. Scaf-Klomp W, Sanderman R, van de Weil HBM, Otter R, van den Heuvel WJA. Distressed or relieved? Psychological side effects of breast cancer screening in the Netherlands. J Epidemiol Community Health 1997;51:705-710.

81. Early Detection of Breast Cancer Working Group. Guideline for the early detection of breast cancer. Edmonton: Alberta Medical Association, 1999.

82. Baines CJ, Miller AB. Mammography versus clinical examination of the breasts. J Natl Cancer Inst Monog 1997(22):125-9.

83. Grady D, Rubin S, Petitti D, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. Ann Intern Med 1992;117(12):1016-37.

# 6. Appendix

*Screen Test*
*Alberta Program for the Early Detection of Breast Cancer*

1

ID#: Date:

First Visit Questionnaire, v4.5

Please check off (✔) the box beside all that is correct. Please add or correct any information where needed.

Full birth (maiden) name:

First Middle Last (Maiden) Last

☐ Current Name:
☐ Date of birth:
☐ Mailing Address:
☐ Phone number:
☐ Family physician name and address: Dr.
☐ Alberta Health Care Insurance Number:
☐ I do not have any breast lumps or bloody discharge.
☐ I have never had breast cancer.

**Please fill out the rest of this questionnaire by checking off (✔) the box beside your single best choice, or by filling out the blank line.**

**First we would like to ask a few questions about breast examination and mammography.**

1. How many years ago was your last physical examination of the breasts by a physician or trained health professional?
☐ Less than 1 ☐ 1 to 2 ☐ More than 2 ☐ Never had one

2. Who taught you how to do breast self-examination? Check one only.
☐ Physician ☐ Other health professional in private session ☐ A class
☐ Printed material ☐ Other ☐ Never been taught

3. How many times a year do you do breast self-examination?
☐ 0 ☐ 1-3 ☐ 4-8 ☐ 9-15 ☐ more than 15

4. Do you feel confident that you are doing breast self-examination correctly?
☐ No ☐ Yes ☐ I do not do breast self-examination

5. How many years ago was your last mammogram?
☐ less than 1 ☐ between 1-2 ☐ between 2-3 ☐ between 3-4
☐ between 4-5 ☐ 5 or more ☐ Never had one

➔ Please turn to page 2 to continue

**You may know that a history of breast cancer runs in some families. We would like to ask a few questions about your family history now. Please remember that we are only talking about BLOOD RELATIVES, not step-families, in-laws, or adopted families.**

6. How many sisters do or did you have? (deceased or living, please don't include half-sisters here) _____

7. For each of your SISTERS please tell us more about her in the following table:

| | [illegible] | [illegible] | [illegible] |
|---|---|---|---|
| Sister 1 | | | ☐No ☐Yes \| |
| Sister 2 | | | ☐No ☐Yes \| |
| Sister 3 | | | ☐No ☐Yes \| |
| Sister 4 | | | ☐No ☐Yes \| |
| Sister 5 | | | ☐No ☐Yes \| |
| Sister 6 | | | ☐No ☐Yes \| |

* If you have more than six sisters, please ask us for more paper.

8. Is your mother alive?
☐No → If no, how old was she when she died? _____
☐Yes → If yes, how old is she now? _____
☐Don't know

9. Does or did your mother have breast cancer?
☐No
☐Yes → If yes, how old was she when she found out? _____
☐Don't know

10. Which of your other female relatives have been diagnosed with breast cancer, including daughters, half-sisters, nieces, cousins, aunts or grandmothers? (For example, 1 grandmother and 2 aunts) _____ ☐None

**Now we would like to ask about your past medical history, as it affects your breast health.**

11. How many breast biopsies have you had? _____

12. Are you currently taking estrogen (hormones for menopause)?
☐No ☐Yes

13. Have you ever used birth control pills (oral contraceptives)?
☐No ☐Yes

14. How old were you when you had your first menstrual period?
☐10 or younger ☐11 ☐12 ☐13 ☐14 ☐15 ☐16 or older

15. Have you had a hysterectomy (womb removed)?
☐ No → Please go to question 16.
☐ Yes → Please go to question 17.
☐ Don't Know → Please go to question 17.

16. Have you had menopause ("change of life")?
☐ No
☐ Yes → If yes, how old were you when you had your last menstrual period? _____
☐ Don't know

17. Have you ever been pregnant?
☐ No
☐ Yes → if yes,
a) How many total pregnancies have you had? (including miscarriages and stillbirths) _____
b) How many children have you had? (including stillbirths) _____
c) What was your age when your first child was born? _____

**Finally we would like to ask a few questions about yourself. Screen Test's goal is to decrease breast cancer by 30% within 15 years. This means that we must screen at least 80% of all women in Alberta. To do so, we need to know which groups of women are coming in to be screened, and which groups of women we may be missing. To find this out, we need to ask a few additional questions about yourself.**

18. Which **province** or territory were you born in? (Specify country if you were born outside of Canada.) _____

19. What would you say is your **main** ancestral ethnic group? Check one only.

| | |
|---|---|
| ☐ British | ☐ E and SE Asian |
| ☐ Western European | ☐ Southern European |
| ☐ Eastern European | ☐ South Asian |
| ☐ French | ☐ African |
| ☐ Northern European | ☐ Other, please specify |
| ☐ Aboriginal | _____ |

20. What was your **highest** level of education completed? Check one only.
☐ Grade 9 or less ☐ Some high school ☐ High school diploma
☐ Any college/some university ☐ University degree

21. If you have been employed outside of the home, what was your **main** occupation? Check one only.
☐ Farming/Processing ☐ Sales/Services ☐ Clerical ☐ Management
☐ Professional ☐ Other, specify _____ ☐ Not applicable

22. What is your **current** employment status? Check one only.
☐ Homemaker ☐ Retired/student ☐ Unemployed ☐ Employed

- **Thank you for taking time to fill out this questionnaire.**
- **Please ensure that you have read the back page before returning this questionnaire to the receptionist.**

*Screen Test*
*Alberta Program for the Early Detection of Breast Cancer*



**Consent Letter**

**Authorization for Participation and Follow-up**

I understand that as a participant in the Alberta Program for the Early Detection of Breast Cancer I will have a mammogram (X-ray of the breasts) done today. It will be interpreted by a radiologist with special training in reading mammograms. I give permission for the program to provide the results of the radiologist's report to the doctor I have named. Also, if any further tests are required as a result of my visit, I give permission for the program to obtain the results of these from my doctor, or any other doctor she/he may refer to.

Screening site _____ Date _____

Signature _____

→ Please turn over to continue.