Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Yong Xiang Wayne Foo Date

Physical Activity, Sedentary Time, and Intrinsic Subtypes of Breast Cancer

By

Yong Xiang Wayne Foo
Master of Public Health

Epidemiology

_____ [Chair's signature]
Lauren E. McCullough, PhD, MSPH
Committee Chair

_____ [Member's signature]
Mia M. Gaudet, PhD
Committee Member

Physical Activity, Sedentary Time, and Intrinsic Subtypes of Breast Cancer

By

Yong Xiang Wayne Foo

B.S., Nanyang Technological University, 2015

Thesis Committee Chair: Lauren E. McCullough, PhD, MSPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

Abstract

Physical Activity, Sedentary Time, and Intrinsic Subtypes of Breast Cancer
By Yong Xiang Wayne Foo

Previous epidemiologic studies have accumulated evidence that physical activity is associated with reductions in breast cancer risk. Given the heterogeneous nature of breast cancer, is it possible that associations with physical activity may vary by molecular subtype. The relation between sedentary sitting time and breast cancer subtypes is also unclear. Nine-hundred invasive breast cancer cases with molecular data between 1992 and 2011 were identified among 71,057 women in the American Cancer Society Prevention Study II Nutrition Cohort. Joint Cox regression was used to estimate multivariable-adjusted relative risks (RR) of breast cancer subtypes in relation to total recreational physical activity and sedentary sitting time. While the observed associations were not statistically significant ($p>0.05$), the strongest associations were observed among women with basal-like breast cancer. The most active women (>17.5 MET-hours/week) had a 31% reduction in risk of breast cancer when compared to the least active women [>0-8.75 MET-hours/week; 95% confidence interval (CI), 0.39-1.20; P_{trend}=0.14]. Additionally, women who reported sitting at least 3 hours per week experienced 32% increased risk of basal-like breast cancer than the less sedentary (<3 hours per week; 95% CI, 0.83-2.19). Given the poorer prognosis associated with basal-like tumors, the evidence of a modest association between basal-like breast cancer, physical activity and sedentary time may be of public health interest, especially among premenopausal African-American women who are more susceptible to basal-like breast cancer.

Physical Activity, Sedentary Time, and Intrinsic Subtypes of Breast Cancer

By

Yong Xiang Wayne Foo

B.S., Nanyang Technological University, 2015

Thesis Committee Chair: Lauren E. McCullough, PhD, MSPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

Acknowledgements

I would like to thank Dr. Lauren McCullough, my thesis advisor, for her constant guidance and patience throughout the entire course of completing this thesis. Her deadlines which seemed challenging to meet at first pushed me to tackle this project whole fully, and I am glad that I could finish this with ample time to spare. She not only went the extra mile to provide me with useful and beneficial feedback, but her knowledge and insight allowed me to approach my thesis from a way a professional epidemiologist would. I would also like to thank Lauren for giving me the opportunity to help her with her research project. The project has enriched me to learn and be exposed to topics relevant to my thesis.

I would also like to thank Dr. Mia Gaudet, my field supervisor at the American Cancer Society. Mia's advice pushed me to think of ways and approaches from a professional workplace standpoint rather than using the traditional classroom approach. Her flexibility has allowed me to juggle and balance between my work and thesis commitments at work. I would also like to thank the epidemiological research department at the American Cancer Society for their work in collecting, cleaning, and maintaining the Cancer Prevention Study II data which made my cleaning work a whole lot easier to do. I would also like to thank Maret Maliniak, the analyst at the American Cancer Society for her precious feedback and effort in checking my cleaning and analysis code for my thesis. Other mentions include Tareyton Clayton for organizing my thesis presentation at the American Cancer Society, Brian Carter for his expert analytical advice on SAS, Alpa Patel for sharing her knowledge on her expertise of physical activity data, and Susan Gapstur for approving this project at ACS and giving me valuable feedback.

And finally, I would like to thank my family for all their love and support: my parents, for whom without their sacrifice I would never dream of living in this amazing country and city, enrolling in this wonderful program at Emory. To Rebecca, thank you for your endless support and understanding. Your character that fights through all adversities will be one that I will always look up to and model after.

Table of Contents

List of Tables

CHAPTER I: BACKGROUND/LITERATURE REVIEW

1.1 Incidence of Breast Cancer

Breast cancer is the most commonly diagnosed cancer in women worldwide, with an estimated 1.7 million cases in 2012 and accounting for 25% of all cancer cases among females (1). Among U.S. women, breast cancer is the most frequently diagnosed cancer where approximately 252,710 U.S. women were expected to develop the disease in 2017 (2). Incidence rates of breast cancer increased sharply in the 1980s due to the increased widespread uptake in mammography screening which led to earlier detection of pre-existing cancer (3). Age-specific incidence rates began to decrease from 1999 to 2003, and the start of this trend reflected a plateau and saturation in mammography screening (4). A sharp drop of nearly 7% in breast cancer rates from 2002 to 2003 were mostly due to a decrease in the use of menopausal hormones that was sparked after the 2002 publication of Women's Health Initiative clinical trial results (5, 6). Breast cancer incidence rates stabilized since 1985 among women aged 50 years and younger, and since 2004 among women aged 50 years and older (7).

1.2 Intrinsic Subtypes of Breast Cancer

Current classification of breast cancer tumors has been further refined based on gene expression analyses to identify molecularly distinct breast cancer subtypes including luminal A, luminal B, HER2-enriched and triple negative breast cancer (TNBC) (8, 9). Using immunohistochemical staining for estrogen, progesterone, and HER2 receptors, surrogate definitions of these subtypes can be obtained (Luminal A: ER+ and/or PR+,

HER2-; Luminal B: ER+ and/or PR+, HER2+; HER2-enriched: ER-/PR-/HER2+; TNBC: ER-/PR-/HER2-), and fluorescent in situ hybridization can be performed to confirm HER2 status. More recently, breast tumor classification considers markers of tumor proliferation and aggressiveness, such as histologic grade or Ki-67 status to differentiate the classification of luminal A and luminal B tumors (9, 10).

Prognosis and responsiveness to treatment has been shown differ among breast cancer subtypes, with a poor prognosis being associated with HER2-enhancing and basal-like tumors compared to luminal A and B. Poorer prognosis among HER2-enhancing and basal-like tumors could be explained by a higher likelihood of relapse in patients who did not achieve a complete pathologic response (11). While luminal A is the most common breast cancer subtype, premenopausal African American women have a high prevalence of basal-like breast cancer. However, there appear to be no racial differences in the aggressiveness of basal-like breast cancer between African American women and whites (12).

1.3 Risk Factors of Breast Cancer

1.3.1 Reproductive factors

Reproductive factors, such as age of menarche, has been consistently associated with both pre- and postmenopausal breast cancer. For every additional year that menarche is delayed, there is a reduction of 5-20% in the risk of breast cancer (13, 14). Late age at menopause is also a well-established risk factor, increasing the risk of breast cancer by an average of 3% for every one year that menopause is delayed (15). On average, nulliparous women have a higher risk of breast cancer, with relative risks

ranging from 1.2 to 1.7 (14). The literature around lactation and breast cancer has been mixed, a substantial number of studies show risk reductions among women who breast-fed for a longer period of time (16). A pooled analysis of 47 studies found a significant risk reduction of about 4% for every 1 year of breast feeding (17).

In a systematic review that examined breast cancer risk factors among intrinsic subtypes, older age at menarche, greater parity, younger age at first birth was associated with a lower risk of luminal A cancer, while older age at menopause was associated with a higher risk of luminal A breast cancer. Older age at menarche was consistently associated with a lower risk of TNBC, and a probable association with a higher risk of TNBC for older age at menopause has been shown (10).

1.3.2 Hormonal factors

Hormonal factors such as endogenous hormones have been associated with breast cancer. Women with the highest quintile of estrogen and androgen levels have a two- to three-fold increase in the risk of breast cancer compared to the lowest quintile (18). In a pooled analysis from 7 studies, there is an average increase of 18% in the odds of breast cancer with doubling concentrations of estradiol and testosterone among premenopausal women (19). In addition, a pooled analysis of 9 epidemiologic studies that found a significant increase in the risk of breast cancer with increasing concentrations of all the sex hormones they examined. These hormones include: total estradiol, free estradiol, non-sex hormone-binding globulin (SHBG)-bound estradiol (which comprises of free and albumin-bound estradiol), estrone, estrone sulfate, androstenedione, dehydroepiandrosterone, dehydroepiandrosterone sulfate, and testosterone (20).

Studies that examined the association between oral contraceptives and breast cancer have been inconsistent. A pooled analysis in 1996 showed an increased risk among current and recent users by 24% (21). A recent systematic review found an 8% increase in breast cancer risk among current users, with a higher risk associated with more recent use of oral contraceptives (22). Use of oral contraceptives has consistently shown an increased risk of TNBC, while the results are inconsistent for luminal A breast cancers (10).

Hormonal replacement therapy (HRT) has been associated with an increased risk of breast cancer. A pooled analysis of 51 studies showed an increased risk of 35% for women who used HRT for 5 years or longer, compared to never use (23). The Women's Health Initiative trial in 2002 which examined the effect of combined estrogen plus progestin therapy on breast cancer showed a 24% increase in breast cancer risk when compared to the placebo group (5). Among intrinsic subtypes, HRT use is strongly associated with an increased risk of luminal A breast cancer, with reported hazard ratios as high as 2.07 comparing fifteen years of HRT use to never use. There is some suggestion that HRT use increases the risk of luminal B breast cancer (10).

1.3.3 Anthropometric factors

Height has been consistently associated with an increase in breast cancer risk, with a large European prospective study showing a 11% increase in risk with increasing height (24). Similarly, another European prospective study found a 26% increase in breast cancer risk for ER+PR+ breast cancer and 8% increase for ER-PR- breast cancer for women in the highest tertile for standing height (25).

The association between body mass index (BMI) and breast cancer varies by menopausal status. Independent studies have shown modest reductions or null associations between BMI and premenopausal breast cancer risk. Conversely, among postmenopausal women a consistent positive relationship between BMI and breast cancer risk has been reported. A meta-analysis of cohort studies observed a 15% reduction in breast cancer risk for every 5 kg/m^2 increase in BMI among premenopausal women (26). On the other hand, there is a 12% increase in postmenopausal breast cancer risk for every 5 kg/m^2 increase in BMI (27). Higher BMI is associated with a decreased risk of luminal A breast cancer and an increased risk of TNBC for among pre-menopausal women (10, 28). Among post-menopausal women, there appears to be a null relationship between BMI and HER2-enhancing breast cancer. Post-menopausal BMI is associated with an increased risk among luminal (receptor-positive) breast cancers (29, 30), and one study has shown that increased BMI is associated with an increased risk in triple negative breast cancer among post-menopausal women (31).

1.3.4 Physical activity

Previous epidemiologic studies have accumulated evidence that physical activity is associated with reductions in breast cancer risk ranging from 10-25% when comparing the most active women to the women in the lowest recreational physical activity category (32-35). Physical activity can be categorized into four different settings: recreational, occupational, household and transportation. A systematic review showed that all four types of physical activity are associated with a reduction in breast cancer risk, with a 21% risk reduction from recreational and household activity, 18% risk reduction from

transportation (walking and cycling) and a 13% risk reduction from occupational activity (36). Vigorous intensity activity was associated with a slightly greater risk reduction (18%) compared to moderate intensity activity (15%). Engaging in physical activity for 2-3 hours per week is associated with a 7% risk reduction, while 6.5 hours of weekly physical activity is associated with a 28% risk reduction of breast cancer. In addition, activity performed during menopause or after the age of 50 provided a greater risk reduction (17%) than activity done in the early adulthood or 20s (8%) (36). When examining the effect by menopause status, physical activity decreases the risk of breast cancer for both premenopausal and postmenopausal, but the average risk reduction is slightly greater among postmenopausal women (31%) than premenopausal women (27%) (36).

Because of the accumulating evidence in the recreational physical activity-breast cancer relationship, as well as the benefits of moderate-to-vigorous intensity associated with it, recreational physical activity is one of the few well-established risk factors that are easily modifiable compared to the other types of activity. This makes recreational physical activity an important domain to investigate in its association with breast cancer.

1.3.5 Sitting Time

Although limited, there is growing research on sitting time as a risk factor for breast cancer. A prospective cohort study using the Cancer Prevention Study II Nutrition Cohort found that longer leisure-time sitting time ($\geq$6 hours vs. <3 hours per day) was associated with an increased risk of invasive breast cancer by 10% among women, after adjusting for physical activity, BMI, and other risk factors (37). A prospective cohort

study of African American women showed an increased risk of breast cancer by 38% with longer sitting time (≥10 hours vs. <5 hours per day) (38), while a population-based study on Chinese women presented a 23% increased risk in breast cancer with longer time (≥4 hours vs. <1.2 hours per day) (39).

1.4 Physical Activity and Breast Cancer: Mechanisms

1.4.1 Sex hormones

Elevated levels of endogenous estrogens and androgens and lower levels of circulating sex hormone binding globulin (SHBG) are associated with increased risks of breast cancer (20). Estradiol is the most bioactive form of estrogen, and postmenopausally, estrone is the source of most circulating estradiol (40). Endogenous estrogens play a convincing role in the etiology of breast cancer due to their ability to inhibit apoptosis and acting as mitogens in the breast through estrogen receptor binding (41). Physical activity can alter menstrual characteristics by delaying the onset of menarche, prolonging menstrual cycles and reducing the number of ovulatory cycles, which potential reduces a woman's cumulative lifetime exposure to endogenous estrogen levels (42, 43). Among menopausal women, adipose tissue is the major source of estrogen, and higher levels of endogenous estrogens are observed in obese postmenopausal women (44).

A systematic review of nine exercise-only randomized controlled trials (RCTs) in postmenopausal women showed consistent results across all nine studies, where there was a decrease in sex hormone levels and an increase in SHBG levels in exercise groups, with <10% change in magnitude. The evidence of adiposity change as a mediating role

in the causal pathway between exercise and decreased estrogen levels was also supported, as exercise groups who lost more body fat in general had lower estrogen levels (41).

Endogenous levels of steroid hormones have been associated with postmenopausal breast cancer, especially ER+/PR+ cancer(20, 45). Given that physical activity could reduce the risk of postmenopausal breast cancer through hormonal pathways, it is possible that physical activity could reduce the risk of luminal A and B, which are hormone receptor-positive breast cancers, via estrogenic mechanisms.

1.4.2 Adipokines

Adipokines are secreted by adipose tissue, which is known to be an active endocrine organ. Adipokines such as leptin, TNF-α, and IL-6, can act as mitogens in the breast and inhibit apoptosis. This allow them to act directly in the mediation of breast cancer growth through the influence on tumor cell migration and invasion. Adipokines can also act indirectly by promotion of insulin resistance and increasing estogen bioactivity (46). Adiponectin, on the other hand, is an adipokine that is anti-inflammatory and promotes insulin sensitivity. Adiponectin is associated with lower levels of obesity. A nested case-control study within Campaign Against Cancer and Heart Disease (CLUE II) found a two-fold increase in odds of breast cancer for increased levels of leptin and a 60% increase in odds for decreased levels of adiponectin (47). Findings from RCTs suggest that exercise, together with sufficient weight loss can decrease levels of circulating leptin (41).

It is reported that in TNBC cell lines, leptin directly increased activity of IGF-1 receptor, which reciprocally increased activity of the leptin receptor to upregulate

epidermal growth factor receptor (EGFR) to promote proliferation and migration of TNBC cells (48). A study in mouse models showed that 92% of TNBC tumors expressed leptin receptor and 86% expressed leptin (49). The suggestive association between leptin and TNBC cells indicate a probable mechanism that physical activity may reduce the risk of TNBC through the reduction of leptin levels.

Studies have shown that adiponectin have an antiproliferative effect on cell growth in both ER+ and ER- cell lines, although data suggest that the dominant mechanisms in ER+ and ER- cell lines are different (50). While further research is required to elucidate the dominant mechanisms in different breast cancer subtypes, changes in adiponectin levels via physical activity may reduce the risk of various subtypes of breast cancer.

1.4.3 Insulin resistance

On top of being an established predictor for diabetes, insulin resistance may provide etiologic importance in the risk of breast cancer. History of diabetes, which is associated with elevated insulin levels is associated with risk of breast cancer (51). Increased insulin levels from insulin resistance can lead to reduced blood levels of insulin-like growth factor binding proteins (IGFBP) 1 and 2. IGFBP promote mitosis and inhibit apoptosis in breast cancer. In addition, elevated insulin levels chronically promote estrogen bioactivity as well as the promotion of adipokines related to breast cancer (52).

Insulin also inhibits SHBG production and increases the levels of IGF-I in blood, which results in the increase of mitogenic activity (Singh et al., 1990). Although estrogen increases IGF-I receptors in ER+ cell lines, TNBC cell lines expressed IGF-I receptors

that were at similar levels to ER+ cell lines (53). This suggests that IGF-I increased proliferation of cancer cell lines regardless of subtypes of breast cancer.

1.4.4 Other mechanisms: immune functioning, oxidative stress, and epigenetics

Physical activity might slow down carcinogenesis through moderation of innate and adaptive immune systems. This, in turn, enhances immunosurveillance and tumor suppression capacity of the immune system (41). Studies on the acute effects of exercise on immune function showed beneficial effects with moderate-intensity, but detrimental effects with high-intensity exercise (54). While some studies have demonstrated that long-term exercise alters the function of circulating immune cells, there is a lack of evidence from randomized-clinical trials (41).

Oxidative stress may play a central role in breast carcinogenesis and in causal pathways linked to obesity (55). Telomere attrition is an important etiologic pathway through which oxidative stress may affect the risk of breast cancer (56). Exercising regularly enhances antioxidant and oxidative damage repairing enzyme capacity, thus reducing oxidative damage (57). However, although epidemiologic evidence supporting that regular exercise reduces oxidative damage via telomere attrition is suggestive, it is limited (41).

DNA methylation is a key epigenetic mechanism in the regulation of gene expression and chromosomal stability, with accumulating epidemiologic evidence of global DNA hypomethylation and increased breast cancer risk (58, 59) .Observational studies have showed positive associations between physical activity and prevalent repetitive sequences (LINE-1) methylation, which is a surrogate of global methylation

(60). Among white middle-aged women with a family history of breast cancer, higher self-reported physical activity was associated with a 33% increase of LINE-1 methylation (61). In another study, more active cancer-free adults, measured by an accelerometer, had higher LINE-1 methylation (62). Luminometric methylation assay (LUMA), which was used to assess global methylation, was associated with a greater than two-fold increased risk of postmenopausal women among women who were more physically active (63).

1.5 Physical Activity and Intrinsic Subtypes of Breast Cancer

Given the heterogeneous nature of breast cancer, it is possible that associations with physical activity may vary by molecular subtype. A meta-analysis study that examined 9 studies looking at associations by estrogen receptor (ER) and progesterone receptor (PR) status reported stronger inverse associations between physical activity and risk of ER-/PR- (RR=0.80, 95% CI: 0.73-0.87) compared to ER+/PR+ tumors (RR=0.92, 95% CI: 0.87-0.98) (64). This suggests that hormonal pathways are not the only mechanism through which physical activity reduces the risk of breast cancer.

There are five epidemiologic studies (two prospective cohort and three population-based case-control studies) to date that examined the association of physical activity in relation to breast cancer subtypes classified by ER, PR and HER2 status and they produced mixed results (65-69). Two studies (one prospective cohort and one population-based case-control) showed a statistically significant inverse association (RRs ranged from 0.52 to 0.73) with physical activity and risk of triple negative breast cancer (TNBC) (65, 69). Among the four studies that looked at physical activity with HER2-enriched breast cancer, one study found a statistically significant inverse association

(RR=0.53, 95% CI: 0.31-0.92) (69), the other two studies found no signification association (RRs ranged from 0.88 to 1.28) (65, 66), while one study found no effect modification by HER2 status (HER2-: RR=0.84 ; HER2+: RR=0.74) (67). There was a statistically significant inverse association with physical activity and luminal A breast cancer in all the four studies that investigated this association (RRs ranged from 0.57 to 0.88) (65-67, 69). In the same four studies, one study found an association between physical activity and risk of luminal B breast cancer (67), while the rest of the studies revealed a null association (65, 66, 69).

No study has evaluated the association between sitting time and breast cancer intrinsic subtypes.

Previous studies that examined the association between physical activity and intrinsic subtypes of breast cancer only looked at ER, PR and HER2 status to classify the breast cancer subtypes. TNBC can be further defined by using basal markers, whereby triple negative cancers with cytokeratin (CK) 5/6+ and/or EGFR+ are classified as basal-like cancers, while triple negative cancers with negative results for CK5/6 and EGFR are classified as “unclassified” cancers. While basal-like breast cancer make up of about 50-75% of TNBC and the two terms have been used interchangeably, they are not biologically synonymous (70, 71). There is still biological heterogeneity within TNBC (72).

In addition, previous studies did not validate HER2 expression using flurorescence in situ hybridization (FISH) after immunohistochemical testing. This could result in a potential misclassification of luminal B breast cancer.

1.6 Classification of Intrinsic Subtypes of Breast Cancer

Intrinsic subtypes of breast cancer in this study are derived using the assessment of biomarkers from tumor blocks to provide a more accurate classification of the subtypes. Only two of the studies (both population-based case-control) assessed the biomarkers from representative tumor blocks (66, 69), while the rest of the studies derived the biomarker data from medical reports (65, 67, 68).

Given the close correlation between proliferation rate and histologic grade, histologic grade is used as a surrogate for proliferation rate to further differentiate between the classification of luminal A and B cancers (73). This is done to be more aligned with the recently proposed classification scheme (9). In addition, given the information of the biomarker data on CK5/6+ and EGFR+, TNBC can be classified into basal-like and unclassified breast cancers.

1.7 Significance of thesis

Triple-negative breast cancer (ER-/PR-/HER2-) is a high-risk breast cancer that have poor prognosis due to the lack of specific treatment to target specific proteins. Immunohistochemical analyses and DNA microarray have previously identified 80-90% of triple negative breast tumors to be basal-like (74). This study will be the first to provide data on the association between physical activity and basal-like breast cancer. Given the poor prognosis of basal-like breast cancer, it is of significance to examine physical activity as a potential protective risk factor.

While there is accumulating evidence of an etiological role of physical activity in breast cancer overall, investigation into the role physical activity play across intrinsic subtypes will allow for better prevention interventions based on the demographic

distributions of intrinsic subtypes, such as targeting pre-menopausal African-American women who are more likely to get TNBC (75, 76).

1.8 Aims of thesis

Aim 1: To estimate the associations of recreational physical activity with intrinsic subtypes of breast cancer in U.S. women.

Hypothesis: Increased recreational physical activity will decrease the risk of breast cancer throughout all the intrinsic subtypes, with the luminal A and basal-like cancers associated with the greatest decrease in risk.

Aim 2: To estimate the association of leisure-time sitting with intrinsic subtypes of breast cancer in U.S. women.

Hypothesis: Longer leisure-time sitting hours will increase the risk of breast cancer throughout all intrinsic subtypes.

CHAPTER II: PHYSICAL ACTIVITY, SEDENTARY TIME, AND INTRINSIC SUBTYPES OF BREAST CANCER

2.1 Abstract

Previous epidemiologic studies have accumulated evidence that physical activity is associated with reductions in breast cancer risk. Given the heterogeneous nature of breast cancer, is it possible that associations with physical activity may vary by molecular subtype. The relation between sedentary sitting time and breast cancer subtypes is also unclear. Nine-hundred invasive breast cancer cases with molecular data between 1992 and 2011 were identified among 71,057 women in the American Cancer Society Prevention Study II Nutrition Cohort. Joint Cox regression was used to estimate multivariable-adjusted relative risks (RR) of breast cancer subtypes in relation to total recreational physical activity and sedentary sitting time. While the observed associations were not statistically significant ($p>0.05$), the strongest associations were observed among women with basal-like breast cancer. The most active women (>17.5 MET-hours/week) had a 31% reduction in risk of breast cancer when compared to the least active women [>0-8.75 MET-hours/week; 95% confidence interval (CI), 0.39-1.20; $P_{trend}=0.14$]. Additionally, women who reported sitting at least 3 hours per week experienced 32% increased risk of basal-like breast cancer than the less sedentary (<3 hours per week; 95% CI, 0.83-2.19). Given the poorer prognosis associated with basal-like tumors, the evidence of a modest association between basal-like breast cancer, physical activity and sedentary time may be of public health interest, especially among premenopausal African-American women who are more susceptible to basal-like breast cancer.

2.2 Introduction

Breast cancer is the most commonly diagnosed cancer in women worldwide. Among U.S. women, breast cancer is the most frequently diagnosed cancer where approximately 252,710 U.S. women were expected to develop the disease in 2017 (2). Previous epidemiologic studies have accumulated evidence that physical activity is associated with reductions in breast cancer risk ranging from 10-25% when comparing the most active women to the women in the lowest recreational physical activity category (34, 35).

Given the heterogeneous nature of breast cancer, it is possible that associations with physical activity may vary by molecular subtype. Current classification of breast cancer tumors has been further refined based on gene expression analyses to identify molecularly distinct breast cancer subtypes including luminal A, luminal B, HER2-enhancing and triple negative breast cancers (TNBC) (8, 9). More recently, breast tumor classification considers markers of tumor proliferation and aggressiveness, such as histologic grade or Ki-67 status to differentiate the classification of luminal A and luminal B tumors (9, 10). TNBC can also be further classified into basal-like or "unclassified" based on markers for cytokeratin 5/6 (CK 5/6) and epidermal growth factor (EGFR) (73). Due to the multifactorial etiology of breast cancer, it important to understand the biologic pathways through which physical activity contributes to breast cancer risk. Uncovering the association of physical activity within intrinsic subtypes of breast cancer may aid in providing mechanistic insight to the inverse association.

Five epidemiologic studies to date have examined the association between physical activity in relation to breast cancer classified by ER, PR and HER2 status (65-

69), which has primarily been ascertained by medical records. Overall, there was an inverse association between physical activity and luminal A breast cancer and potential evidence for risk reduction for TNBC. No association was found for luminal B cancer, while results found in the association between physical activity and HER2-enriched cancer were mixed.

No studies of physical activity and breast cancer subtype have included markers for basal-like tumors. Moreover, only two studies have assessed tumor biomarkers themselves (66, 69).To better understand physical activity and intrinsic subtypes of breast cancer, associations were examined in the Cancer Prevention Study-II (CPS-II) Breast Tissue Repository which has annotated breast tissue specimens. This study will be the first to provide data on basal-like breast cancer.

While there is strong evidence of an etiological role of physical activity in breast cancer overall, investigation into the role physical activity play across intrinsic subtypes will allow for better prevention interventions that could target specific subtypes of breast cancer.

2.3 Methods

Study cohort and follow up

Analysis of women were taken from the 97,783 women in the CPS-II Nutrition cohort, a prospective study established by the American Cancer Society in 1992 as a subgroup of the CPS-II Baseline Cohort initiated in 1982. The purpose of the CPS-II Nutrition cohort was to investigate the role of diet and other lifestyle factors and exposures on cancer incidence, mortality, and survival (77). The Nutrition Cohort

participants were aged 50-74 years at enrollment in 1992 and they completed a baseline questionnaire on demographic, medical, behavioral, environmental, and occupational factors. Follow-up questionnaires were mailed to the participants in 2-year intervals starting in 1997 to update on their exposure information as well as confirming newly diagnosed cancer outcomes.

Women who were ineligible for analysis included 2,701 women due to lost to follow up, 13,501 women who reported prevalent cancer (except non-melanoma skin cancer) at baseline, 7,255 women who reported inactivity (0 MET-hours/week) and 3,205 women who had missing data on recreational physical activity and sitting time. An additional 53 women with an unverified breast cancer diagnosis and 11 women who had a bad diagnosis date were excluded. After the exclusions, there were 71,057 women in the final cohort for analysis.

Case ascertainment and tissue specimen collection

In 2011, the CPS-II Breast Tissue Repository was initiated through the collection of breast tissue from the CPS-II cohort. Eligible cases had: 1) reported a diagnosis of breast cancer on a routine CPS-II Nutrition Cohort follow-up questionnaire; 2) subsequently provided written consent to obtain medical records, and; 3) a confirmation of the diagnosis from a review of medical records. Participants were sent consent forms for tumor acquisition and release. Tissue specimens were requested from the hospitals for consenting or deceased participants. All tissue blocks and original diagnostic slides were requested from the stored cancer and normal blocks. As an alternative to the tissue block request, unstained slides from a representative tumor block and a normal block were

requested. Long-term storage of the tissue specimens is at the ACS National Home Office, Epidemiology Research Program in Atlanta, GA.

Of the 4,403 breast cancer incident cases [International Classification of Diseases for Oncology (ICD-O) topography code C50] that have been confirmed through medical records and are eligible for the tissue repository, 1,720 tissue samples from hospitals were received. 69% of the cases sent at least one tissue block and the remaining cases sent only unstained slides.

Immunohistochemical analysis and classification of molecular subtypes

Tissue specimens were sent to the Pathology Laboratory at the Mayo Clinic Bioservices Laboratory who processed the specimens and performed immunohistochemical staining for ER, PR, HER2, cytokeratin 5/6 (CK5/6), and epidermal growth factor receptor (EGFR). The Cyogenetics Core performed HER2-flurorescence in situ hybridization (FISH) to confirm HER2/neu among cases with HER2/neu IHC stains scored as 2+ or greater. As of April 2016, 1,715 breast cancer cases with tissue have been processed through Mayo. A single pathologist read all tissue markers.

Cases that were ER-positive and/or PR-positive and HER-2 negative with histological grade 1 and 2 were classified as luminal A cancers; cases that were either i) ER-positive and/or PR-positive and HER2-positive or ii) ER-positive and/or PR-positive and HER2-negative with histologic grade 3 were classified as luminal B cancers; cases that were ER-negative, PR-negative and HER2-positive were classified as HER2 enriched cancers; and cases that were ER-negative, PR-negative, HER2-negative but

positive for CK5/6 and/or EGFR were categorized as basal-like. Cases that lack expression for all 5 markers were considered as “unclassified” (73). After exclusion, 900 cases had complete data on ER, PR, HER2, CK5/6 and EGFR and could be classified into one of these five molecular subtypes. Among these cases, there were 729 cases of luminal A, 47 luminal B, 29 HER2 enriched, 72 basal-like and 23 “unclassified” breast cancers.

Assessment of physical activity and sitting time

Data collection of weekly recreational activities at enrollment has been described previously (78, 79). A summary estimate of the metabolic equivalent (MET) hours per week from baseline data will be used to examine the association between physical activity and intrinsic subtypes of breast cancer. Using the Ainsworth compendium for physical activities, metabolic equivalent (MET) values of each activity was used to calculate the summary MET-hours/week for each participant by multiplying the MET score for each activity by the number of hours spent per week (80). Categorization of MET-hours/week was classified into three categories (>0-8.75, >8.75-17.5, >17.5 MET-hours/week), with the lowest MET-hours/week category as the referent category. 8.75 MET-hours/week is equivalent to 30 minutes of moderate activity (3.5 METS) for 5 days (2.5 hours/week), which is a widely-used recommendation for physical activity (81-83). 17.5 MET-hours/week represent twice the physical activity recommendation of 8.75 MET-hours/week. Women who reported being inactive (N=7,255) were excluded from the analysis due to the possibility that their inactivity may be due to underlying health conditions that may be related to breast cancer risk.

Leisure-time sitting (assessed using time spent watching TV, reading, etc.) was categorized as 0-<3 (referent) and ≥3 hours per day.

Covariate Assessment

Covariates that were considered were age, race, education, BMI, adult weight change from age 18, smoking status, alcohol intake, family history of breast cancer, HRT use, number of live births by age at first live birth, age at menarche, age at menopause, and history of cysts/breast lumps. Assessment of potential confounders were performed using the directed acyclic graph (DAG) approach (Appendix B). A covariate is deemed as a potential confounder if the covariate is associated with both the exposure (physical activity) and the outcome (breast cancer), and is not a descendant of the exposure and outcome.

After assessment, covariates included in the multivariate analyses are age, BMI (<18.5, 18.5-<22.5, 22.5-<25.0, 25.0-<30.0, ≥30 kg/m^2), education (≤high school graduate, some college, ≥college graduate), alcohol intake (non-drinker, <1drink/day, 1+drink/day), and age at menopause (<45, 45-54, 55+ years). Because BMI could potentially mediate the association between physical activity and breast cancer risk, analyses were also conducted without BMI (Appendix D).

Statistical Analyses

Age-adjusted chi square tests were performed to determine the participants' baseline characteristics by recreational physical activity categories. Joint Cox proportional hazards regression was used to calculate RRs between each covariate and

the risk of breast cancer stratified by cases with and without complete molecular data. *P* values for statistical significance of heterogeneity of associations between cases with and without molecular data were derived using the chi-squared heterogeneity statistic.

Joint Cox proportional hazards modeling was used to calculate age-adjusted and multivariable-adjusted hazard ratios for approximation of relative risks (RRs) and 95% confidence intervals to estimate the associations of recreational physical activity and leisure time spent sitting with the intrinsic subtypes of breast cancer (84). Follow-up time for each participant was calculated as person-years from the date of the 1992/3 baseline enrollment to the date of: i) diagnosis of breast cancer; ii) death; iii) the last cancer-free questionnaire when self-reported breast cancer on a subsequent questionnaire was not verified; iv) the last completed questionnaire if no subsequent questionnaire was completed or returned; v) date of end of follow-up, June 30, 2011. The multivariable-adjusted model included age, BMI, education, alcohol intake, and age at menopause. Trends tests of RR in relation to MET-hours/week were performed by fitting the median MET value to each category.

All analyses were conducted with SAS version 9.4 (SAS Institute, Cary, USA). All statistical tests were two sided, and p-values < 0.05 were considered statistically significant. Emory University's Institutional Review Board approved this study.

2.4 Results

A total of 71,057 women in the analytic cohort were followed for an average of 14.5 years. The median MET expenditure among women who reported being active was 9.5 MET-hours/week, which was approximately equivalent to three hours of moderately

paced walking per week. Physically active women were more likely to be leaner, more likely to be more educated, more likely to drink alcohol, and less likely to have gained weight since age 18 (Table 1).

Compared to women without complete molecular data (N=4,336), women with complete molecular data (N=900) had similar risk association for risk factors such as race, BMI, adult weight change since age 18, smoking status, HRT use, number of live births by age at first live birth, age at menarche, age at menopause, and history of cysts/breast lumps (P-heterogeneity ≤ 0.05) (Table 2).

No statistically significant association between total recreational physical activity and all intrinsic breast cancer subtypes were observed (all $P_{trend} \geq 0.14$) (Table 3). Among the five intrinsic subtypes, physical activity appeared to be most strongly associated with basal-like breast cancer (RR=0.69; 95% CI, 0.39-1.18; $P_{trend} = 0.14$). Albeit the lack of statistical significance, there is an indication for a modest risk reduction for luminal A (RR=0.91; 95% CI, 0.76-1.09; $P_{trend} = 0.33$) (Table 3). Compared to the age-adjusted analyses, adjusting for age, BMI, education, alcohol intake, smoking status and age at menopause in the multivariate analyses improved the inverse association estimates between total recreational physical activity and decreased risk of breast cancer for all subtypes across all categories of physical activity (Table 3).

While no statistically significant associations between sitting time and breast cancer subtypes were observed (Table 3), the greatest effect observed among all subtypes was for basal-like breast cancer (RR=1.32; 95% CI, 0.83-2.19). Multivariate analysis did not alter the observed age-adjusted estimates.

2.5 Discussion

The results from the study do not support an association between physical activity and breast cancer across intrinsic subtypes luminal A, luminal B, HER2-enriched, basal-like and "unclassified" breast cancer. However, there was evidence of a modest risk reduction for luminal A and sizeable reductions for basal-like breast cancer.

The results of five previous studies (two prospective cohort studies and three population-based case-control studies) that examined the association between physical activity and intrinsic breast cancer subtypes (classified by ER, PR and HER2 status) differed across subtypes (65-69). All four studies that investigated the association between physical activity and luminal A breast cancer found a statistically significant inverse association. Although the findings for luminal A breast cancer were not statistically significant, there is an indication of a modest risk reduction. Luminal A breast cancer is the most common breast cancer subtype with an estimated 70% incidence of breast cancer cases in the U.S. (85). Several established risk factors for breast cancer have known associations with the luminal A subtype earlier age at menarche, later age at menopause, and HRT which are likely to influence risk via increasing the number of lifetime hormonal cycles (10, 73, 86).

Given the evidence of hormonal factors affecting the risk of luminal A breast cancer and the hormone-receptor positive characteristic of luminal A breast cancer, a possible mechanism for physical activity reducing the risk of luminal A breast cancer is through hormonal pathways. Physical activity can alter menstrual characteristics to reduce a woman's cumulative lifetime exposure to endogenous estrogen levels, such as delaying the onset of menarche, prolonging menstrual cycles and reducing the number of

ovulatory cycles (42, 43). Estrogen can activate ER through genomic and non-genomic pathways, leading to breast cancer cell proliferation (87). Likewise, estrogen stimulation may up-regulate PR expression (88). There is also possible evidence that physical activity may decrease plasma levels of progesterone (89).

Despite previous epidemiologic evidence and strong biologic plausibility for physical activity as a modifiable risk factor for luminal A breast cancer, the findings did not find a significant association. This could be due to the limited categorization of physical activity in our study. Comparison of a higher MET-hours/week category to the referent group may have elicited a significant inverse association between physical activity and luminal A breast cancer.

The findings for Luminal B revealed no association with physical activity. Findings from previous studies were mixed, with one study showing a significant inverse association (67), while the rest showed no significant association between physical activity and luminal B breast cancer (65, 66, 69). The associations between established risk factors for breast cancer and luminal B breast cancer are generally not well understood. A systematic review showed that family history of breast cancer and lifetime duration of breastfeeding were the only risk factors associated with luminal B, although associations with age at menarche were suggestive (10). Similarly, the Nurses' Health Studies showed no evidence of reproductive risk factors associated with luminal B breast cancer (86).

While luminal B is a hormone-receptor positive breast cancer, they are resistant to hormone therapy and have a distinct molecular phenotype from luminal A breast cancer, with luminal B having several similar molecular features with ER- molecular subtypes

(90). Although they express estrogen receptors, luminal B breast cancers do not exhibit a similar expression of estrogen-regulated genes like luminal A cancers. Luminal B breast cancer may therefore rely on other pathways for growth. Thus, a possible explanation that established breast cancer risk factors such as physical activity and reproductive factors, which lead to the alteration of endogenous estrogen levels, are not associated with luminal B breast cancer is because they do not reduce risk through hormonal pathways.

The findings revealed no association between physical activity and HER2-enriched breast cancer. Results from previous studies (one prospective cohort study, three population-based case-control studies) that examined the effect of physical activity by HER2 status are mixed (65-67, 69). One population-based case-control-study showed that physical activity above the median level was associated with a risk reduction for HER2-enriched cancer (69), while another study showed a reduced risk for HER2- but not HER2-enriched breast cancer (66). In another population-based control of post-menopausal women, breast cancer risk did not vary by HER2 status (67).

The associated risk factors with HER2-enrinched breast cancer remain largely unknown, and more research is needed to understand the risk profile of HER2-enriched cancer (10). Evidence have shown that HER2 amplification occurs early in human breast tumorigenesis and HER2 amplified breast cancers consist of a unique molecular portrait that does not change during progression of disease to metastasis (91). HER2 amplification, through the disruption of normal cell control mechanisms, increases stem/progenitor cell populations which may lead to tumor invasion or metastasis (92). While there is no established explanation as to why there is a heterogeneity in the association between physical activity and breast cancer by HER2 status, it is possible that

there may not be a mechanism through which physical activity produces a protective effect if normal cell mechanisms are disrupted or if there is an increase in the stem/progenitor cell population through overexpression of HER2 (66).

Given that the classification of HER2 status in previous studies were assessed by immunohistochemical testing without validation by FISH analysis, there could be potential misclassification of HER2 status, which could contribute to the mixed findings between physical activity and HER2-enriched breast cancer.

While the results from the study did not reveal any significant association between physical activity and basal-like cancer, they show an indication of a modest inverse association. Although no studies further classified TNBC into basal-like and unclassified breast cancers, there are four studies to date that looked at the association between physical activity and TNBC (65, 66, 68, 69). Of these studies, two studies (one prospective cohort and one population-based case-control study) reported statistically significant association between physical activity and TNBC (65, 69). The Women's Health Initiative Cohort study found a modest but not significantly lower risk of TNBC among physically active women (68). TNBC is the second most common subtype after luminal A breast cancer (85). Given that majority of TNBC are basal-like breast cancers, the observed effect estimates in the study parallel most previous studies. Basal-like breast cancer, which is defined via gene expression microarray analysis, represents approximately 10-25% of all cases and consists of about 50-75% of all TNBC (72, 93). The basal-like molecular subtype is associated with aggressiveness, early sign of metastasis, limited effective targeted therapies, and poor prognosis (94). Moreover, pre-menopausal and African-American women are more likely to develop basal-like breast

cancer (75, 95). Thus, it is of importance to uncover the epidemiologic and mechanistic links between physical activity and basal-like breast cancer to develop effective preventive strategies for this poor prognostic subtype.

The mechanisms linking physical activity to basal-like breast cancer are unresolved, although there are promising hypotheses. *BRCA1* mutation has been associated with basal-like breast cancer (72, 93, 96, 97), and studies have revealed physical activity-associated risk reductions among *BRCA1* mutation carriers (98-100). A study investigating the effect of prepubertal physical activity on the expression of *BRCA1* in rat mammary glands showed that exercise before puberty up-regulates *BRCA1*, which could in turn be associated with reduced breast cancer risk (101). Thus, it is plausible that physical activity reduces the risk of basal-like breast cancer among women with BRCA1 mutation carriers.

The findings do not support an association between physical activity and unclassified breast cancer. No study to date have examined the effect of physical activity on the risk of unclassified breast cancer. Studies that examined the association of other risk factors with unclassified breast cancer provided inconsistent and inconclusive results (28, 73, 95). More research must be done to determine the relationship between physical activity and unclassified breast cancer.

The data provide the first evidence of lack of association between sedentary time and risk of breast cancer across subtypes, albeit observing a modest increased risk of basal-like cancer with increased sitting time. Although limited, there are previous studies that examined sedentary time as a risk factor of breast cancer and have showed that longer leisure-time sitting time was associated with an increased risk of breast cancer

(37-39). However, given the heterogeneity in the categorization of sitting hours across these studies, it is possible that the chosen two categories of sitting time (≥3 hours v. <3 hours per day) may not elicit the association that would have shown from analysis of a different categorization of sitting time. The response categories in the CPS-II questionnaire as well as the insufficient sample size of cases to compare between higher and lower hours of sitting time limited the choice of categorization of sitting time.

The study had some limitations. One major limitation was the lack of sample size, especially for all intrinsic subtypes except for luminal A breast cancer. Only a subset of the breast cancer cases in the CPS-II Nutrition Cohort were eligible due to their availability of tumor specimens that had available data on ER, PR, HER2, CK5/6 and EGFR markers. The small sample sizes for luminal B, HER2-enriched, basal-like and unclassified breast cancers may have led to imprecise results. However, the findings such as the indication of a modest risk reduction for basal-like breast cancer, and the null association for luminal B breast cancer are largely consistent with previous studies, suggesting the replicability of the results.

Another limitation of this study was that baseline reported physical activity was assessed as the main exposure classification of physical activity. There was no information on physical activity during adolescence and adulthood, which may affect the multistage progression of invasive breast cancer. In addition, information of physical activity during the follow-up period were not updated. However, the rationale for assessing the baseline reported physical activity as the main exposure was because physical activity levels reported at baseline were likely to be aligned with those women who have been exercising consistently over their lifetime, or even women who began

exercising recently. Previous assessment of physical activity levels in the CPS-II Cohort of women showed no differences in risk of participants when comparing physical activity measures at age 40 and in 1982 (78). The average age of the CPS-II Cohort women in 1982 was about 53 years old. It is very likely that risk assessment using physical activity levels at 1992 baseline would be similar.

Given that the CPS-II Nutrition Cohort is a subset of the original CPS-II cohort originally recruited in 1982, another limitation of the study was that participants were generally healthier than the general population because they volunteered to participate in 1992. Moreover, they represent a select population who are predominantly White, middle-aged, or elderly, and well educated. While these differences are unlikely to affect internal validity, the results may not be generalizable to populations who possess different demographic characteristics.

Another limitation of our study is that the assessment of physical activity may not reflect total physical activity especially in individuals who are physically active in their occupations. However, given the demographics of the cohort, any contribution of occupational activity is likely to be negligible. Moreover, most of the participants were, homemakers. In addition, another limitation was the inclusion of both post-menopausal and pre-/peri-menopausal women in the analysis, which presents difficulty in generalizing the results to either women who are post-menopausal or pre-/peri-menopausal. Firstly, the number of pre-/peri-menopausal women was substantially lower than the number of post-menopausal women in the cohort, and the lack of sample size limited the ability to address the association between physical activity and intrinsic subtypes of breast cancer among pre-/peri-menopausal women. Secondly, results between

post-menopausal women and the inclusion of both post-menopausal and pre-/peri-menopausal women were similar (Appendix C). Thus, given the similarity of results, it was decided to include post-menopausal and pre-/peri-menopausal women into one cohort to increase the case numbers for a higher statistical power.

There are few major strengths of this study, such as the prospective cohort design, comprehensive physical activity assessment that covered various recreational physical activities and frequency, and uniform assessment of biomarkers from tumor tissue specimens. Intrinsic subtypes of breast cancer were categorized through the assessment of biomarkers from tumor tissue specimens rather than relying on information from medical records. FISH analysis was also used to validate the expression of HER2 using immunohistochemical assays. The FISH method, although more expensive and time-consuming, is the preferred approach for determining HER2 status, and it is recommended that IHC results of weakly positive cases (2+) are to be confirmed with FISH (102, 103).

In conclusion, while no significant association was found between physical activity and each intrinsic subtype of breast cancer, there is an indication of a possible inverse association for both basal-like and luminal A breast cancer. The heterogeneity of results among breast cancer subtypes emphasizes the need for more research on subtype-specific breast cancer etiology and its association with physical activity. Given the poorer prognosis associated with basal-like breast cancer and the plausibility of physical activity as a potential risk modifier, future research should be directed in understanding the mechanism through which physical activity acts on the risk of basal-like breast cancer.

2.6 REFERENCES

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65(2):87-108.
2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67(1):7-30.
3. Holford TR, Cronin KA, Mariotto AB, et al. Changing patterns in breast cancer incidence trends. *J Natl Cancer Inst Monogr* 2006(36):19-25.
4. Jemal A, Ward E, Thun MJ. Recent trends in breast cancer incidence rates by age and tumor characteristics among U.S. women. *Breast Cancer Res* 2007;9(3):R28.
5. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288(3):321-33.
6. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* 2007;356(16):1670-4.
7. DeSantis C, Ma J, Bryan L, et al. Breast cancer statistics, 2013. *CA Cancer J Clin* 2014;64(1):52-62.
8. Sorlie T, Wang Y, Xiao C, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics* 2006;7:127.
9. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 2013;24(9):2206-23.
10. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. *Biochim Biophys Acta* 2015;1856(1):73-85.
11. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007;13(8):2329-34.
12. O'Brien KM, Cole SR, Tse CK, et al. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* 2010;16(24):6100-10.
13. Hsieh CC, Trichopoulos D, Katsouyanni K, et al. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. *Int J Cancer* 1990;46(5):796-800.
14. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. *Epidemiol Rev* 1993;15(1):36-47.
15. Collaborative Group on Hormonal Factors in Breast C. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol* 2012;13(11):1141-51.
16. Lipworth L, Bailey LR, Trichopoulos D. History of breast-feeding in relation to breast cancer risk: a review of the epidemiologic literature. *J Natl Cancer Inst* 2000;92(4):302-12.
17. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological

studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet* 2002;360(9328):187-95.
18. Hankinson SE, Eliassen AH. Endogenous estrogen, testosterone and progesterone levels in relation to breast cancer risk. *J Steroid Biochem Mol Biol* 2007;106(1-5):24-30.
19. Endogenous H, Breast Cancer Collaborative G, Key TJ, et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol* 2013;14(10):1009-19.
20. Key T, Appleby P, Barnes I, et al. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002;94(8):606-16.
21. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996;347(9017):1713-27.
22. Gierisch JM, Coeytaux RR, Urrutia RP, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. *Cancer Epidemiol Biomarkers Prev* 2013;22(11):1931-43.
23. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* 1997;350(9084):1047-59.
24. Wiren S, Haggstrom C, Ulmer H, et al. Pooled cohort study on height and risk of cancer and cancer death. *Cancer Causes Control* 2014;25(2):151-9.
25. Ritte R, Lukanova A, Tjonneland A, et al. Height, age at menarche and risk of hormone receptor-positive and -negative breast cancer: a cohort study. *Int J Cancer* 2013;132(11):2619-29.
26. Wiseman M. The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *Proc Nutr Soc* 2008;67(3):253-6.
27. Renehan AG, Tyson M, Egger M, et al. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371(9612):569-78.
28. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev* 2007;16(3):439-43.
29. Munsell MF, Sprague BL, Berry DA, et al. Body mass index and breast cancer risk according to postmenopausal estrogen-progestin use and hormone receptor status. *Epidemiol Rev* 2014;36:114-36.
30. Horn J, Alsaker MD, Opdahl S, et al. Anthropometric factors and risk of molecular breast cancer subtypes among postmenopausal Norwegian women. *Int J Cancer* 2014;135(11):2678-86.
31. Chen L, Cook LS, Tang MT, et al. Body mass index and risk of luminal, HER2-overexpressing, and triple negative breast cancer. *Breast Cancer Res Treat* 2016;157(3):545-54.

32. Friedenreich CM, Cust AE. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *Br J Sports Med* 2008;42(8):636-47.
33. Lynch BM, Dunstan DW, Healy GN, et al. Objectively measured physical activity and sedentary time of breast cancer survivors, and associations with adiposity: findings from NHANES (2003-2006). *Cancer Causes Control* 2010;21(2):283-8.
34. Friedenreich CM. The role of physical activity in breast cancer etiology. *Semin Oncol* 2010;37(3):297-302.
35. Moore SC, Lee IM, Weiderpass E, et al. Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. *JAMA Intern Med* 2016;176(6):816-25.
36. Friedenreich CM. Physical activity and breast cancer: review of the epidemiologic evidence and biologic mechanisms. *Recent Results Cancer Res* 2011;188:125-39.
37. Patel AV, Hildebrand JS, Campbell PT, et al. Leisure-Time Spent Sitting and Site-Specific Cancer Incidence in a Large U.S. Cohort. *Cancer Epidemiol Biomarkers Prev* 2015;24(9):1350-9.
38. Nomura SJ, Dash C, Rosenberg L, et al. Sedentary time and breast cancer incidence in African American women. *Cancer Causes Control* 2016;27(10):1239-52.
39. Pronk A, Ji BT, Shu XO, et al. Physical activity and breast cancer risk in Chinese women. *Br J Cancer* 2011;105(9):1443-50.
40. Roberts KD, Rochefort JG, Bleau G, et al. Plasma estrone sulfate levels in postmenopausal women. *Steroids* 1980;35(2):179-87.
41. Neilson HK, Conroy SM, Friedenreich CM. The Influence of Energetic Factors on Biomarkers of Postmenopausal Breast Cancer Risk. *Curr Nutr Rep* 2014;3:22-34.
42. McTiernan A, Ulrich C, Slate S, et al. Physical activity and cancer etiology: associations and mechanisms. *Cancer Causes Control* 1998;9(5):487-509.
43. Hoffman-Goetz L, Apter D, Demark-Wahnefried W, et al. Possible mechanisms mediating an association between physical activity and breast cancer. *Cancer* 1998;83(3 Suppl):621-8.
44. Cleary MP, Grossmann ME. Minireview: Obesity and breast cancer: the estrogen connection. *Endocrinology* 2009;150(6):2537-42.
45. Missmer SA, Eliassen AH, Barbieri RL, et al. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J Natl Cancer Inst* 2004;96(24):1856-65.
46. Rose DP, Vona-Davis L. Biochemical and molecular mechanisms for the association between obesity, chronic inflammation, and breast cancer. *Biofactors* 2014;40(1):1-12.
47. Gross AL, Newschaffer CJ, Hoffman-Bolton J, et al. Adipocytokines, inflammation, and breast cancer risk in postmenopausal women: a prospective study. *Cancer Epidemiol Biomarkers Prev* 2013;22(7):1319-24.
48. Saxena NK, Taliaferro-Smith L, Knight BB, et al. Bidirectional crosstalk between leptin and insulin-like growth factor-I signaling promotes invasion and migration of breast cancer cells via transactivation of epidermal growth factor receptor. *Cancer Res* 2008;68(23):9712-22.

49. Otvos L, Jr., Kovalszky I, Riolfi M, et al. Efficacy of a leptin receptor antagonist peptide in a mouse model of triple-negative breast cancer. *Eur J Cancer* 2011;47(10):1578-84.
50. Korner A, Pazaitou-Panayiotou K, Kelesidis T, et al. Total and high-molecular-weight adiponectin in breast cancer: in vitro and in vivo studies. *J Clin Endocrinol Metab* 2007;92(3):1041-8.
51. Wu AH, Yu MC, Tseng CC, et al. Diabetes and risk of breast cancer in Asian-American women. *Carcinogenesis* 2007;28(7):1561-6.
52. Rose DP, Vona-Davis L. The cellular and molecular mechanisms by which insulin influences breast cancer risk and progression. *Endocr Relat Cancer* 2012;19(6):R225-41.
53. Davison Z, de Blacquiere GE, Westley BR, et al. Insulin-like growth factor-dependent proliferation and survival of triple-negative breast cancer cells: implications for therapy. *Neoplasia* 2011;13(6):504-15.
54. Walsh NP, Gleeson M, Shephard RJ, et al. Position statement. Part one: Immune function and exercise. *Exerc Immunol Rev* 2011;17:6-63.
55. Vera-Ramirez L, Sanchez-Rovira P, Ramirez-Tortosa MC, et al. Free radicals in breast carcinogenesis, breast cancer progression and cancer stem cells. Biological bases to develop oxidative-based therapies. *Crit Rev Oncol Hematol* 2011;80(3):347-68.
56. Prescott J, Wentzensen IM, Savage SA, et al. Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutat Res* 2012;730(1-2):75-84.
57. Radak Z, Chung HY, Goto S. Systemic adaptation to oxidative challenge induced by regular exercise. *Free Radic Biol Med* 2008;44(2):153-9.
58. Delgado-Cruzata L, Wu HC, Perrin M, et al. Global DNA methylation levels in white blood cell DNA from sisters discordant for breast cancer from the New York site of the Breast Cancer Family Registry. *Epigenetics* 2012;7(8):868-74.
59. Deroo LA, Bolick SC, Xu Z, et al. Global DNA methylation and one-carbon metabolism gene polymorphisms and the risk of breast cancer in the Sister Study. *Carcinogenesis* 2014;35(2):333-8.
60. Yang AS, Estecio MR, Doshi K, et al. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 2004;32(3):e38.
61. White AJ, Sandler DP, Bolick SC, et al. Recreational and household physical activity at different time points and DNA global methylation. *Eur J Cancer* 2013;49(9):2199-206.
62. Zhang FF, Cardarelli R, Carroll J, et al. Physical activity and global genomic DNA methylation in a cancer-free population. *Epigenetics* 2011;6(3):293-9.
63. McCullough LE, Chen J, White AJ, et al. Global DNA Methylation, Measured by the Luminometric Methylation Assay (LUMA), Associates with Postmenopausal Breast Cancer in Non-Obese and Physically Active Women. *J Cancer* 2015;6(6):548-54.
64. Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. *Breast Cancer Res Treat* 2013;137(3):869-82.

65. Ma H, Xu X, Clague J, et al. Recreational physical activity and risk of triple negative breast cancer in the California Teachers Study. *Breast Cancer Res* 2016;18(1):62.
66. Ma H, Xu X, Ursin G, et al. Reduced risk of breast cancer associated with recreational physical activity varies by HER2 status. *Cancer Med* 2015;4(7):1122-35.
67. Schmidt ME, Steindorf K, Mutschelknauss E, et al. Physical activity and postmenopausal breast cancer: effect modification by breast cancer subtypes and effective periods in life. *Cancer Epidemiol Biomarkers Prev* 2008;17(12):3402-10.
68. Phipps AI, Chlebowski RT, Prentice R, et al. Body size, physical activity, and risk of triple-negative and estrogen receptor-positive breast cancer. *Cancer Epidemiol Biomarkers Prev* 2011;20(3):454-63.
69. Trivers KF, Lund MJ, Porter PL, et al. The epidemiology of triple-negative breast cancer, including race. *Cancer Causes Control* 2009;20(7):1071-82.
70. Badve S, Dabbs DJ, Schnitt SJ, et al. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod Pathol* 2011;24(2):157-67.
71. Alluri P, Newman LA. Basal-like and triple-negative breast cancers: searching for positives among many negatives. *Surg Oncol Clin N Am* 2014;23(3):567-77.
72. Perou CM. Molecular stratification of triple-negative breast cancers. *Oncologist* 2011;16 Suppl 1:61-70.
73. Tamimi RM, Colditz GA, Hazra A, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. *Breast Cancer Res Treat* 2012;131(1):159-67.
74. Rakha EA, El-Sayed ME, Green AR, et al. Prognostic markers in triple-negative breast cancer. *Cancer* 2007;109(1):25-32.
75. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295(21):2492-502.
76. Lund MJ, Trivers KF, Porter PL, et al. Race and triple negative threats to breast cancer survival: a population-based study in Atlanta, GA. *Breast Cancer Res Treat* 2009;113(2):357-70.
77. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* 2002;94(2):500-11.
78. Patel AV, Callel EE, Bernstein L, et al. Recreational physical activity and risk of postmenopausal breast cancer in a large cohort of US women. *Cancer Causes Control* 2003;14(6):519-29.
79. Hildebrand JS, Gapstur SM, Campbell PT, et al. Recreational physical activity and leisure-time sitting in relation to postmenopausal breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2013;22(10):1906-12.
80. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011;43(8):1575-81.
81. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and

maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc* 2011;43(7):1334-59.

82. Kushi LH, Doyle C, McCullough M, et al. American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J Clin* 2012;62(1):30-67.
83. Physical activity guidelines for Americans. *Okla Nurse* 2008;53(4):25.
84. Xue X, Kim MY, Gaudet MM, et al. A comparison of the polytomous logistic regression and joint cox proportional hazards models for evaluating multiple disease subtypes in prospective cohort studies. *Cancer Epidemiol Biomarkers Prev* 2013;22(2):275-85.
85. Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst* 2014;106(5).
86. Sisti JS, Collins LC, Beck AH, et al. Reproductive risk factors in relation to molecular subtypes of breast cancer: Results from the nurses' health studies. *Int J Cancer* 2016;138(10):2346-56.
87. Yamaguchi Y. Microenvironmental regulation of estrogen signals in breast cancer. *Breast Cancer* 2007;14(2):175-81.
88. Fazzari A, Catalano MG, Comba A, et al. The control of progesterone receptor expression in MCF-7 breast cancer cells: effects of estradiol and sex hormone-binding globulin (SHBG). *Mol Cell Endocrinol* 2001;172(1-2):31-6.
89. Montagnani CF, Arena B, Maffulli N. Estradiol and progesterone during exercise in healthy untrained women. *Med Sci Sports Exerc* 1992;24(7):764-8.
90. Creighton CJ. The molecular profile of luminal B breast cancer. *Biologics* 2012;6:289-97.
91. Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* 2007;26(45):6469-87.
92. Korkaya H, Paulson A, Iovino F, et al. HER2 regulates the mammary stem/progenitor cell population driving tumorigenesis and invasion. *Oncogene* 2008;27(47):6120-30.
93. Schneider BP, Winer EP, Foulkes WD, et al. Triple-negative breast cancer: risk factors to potential targets. *Clin Cancer Res* 2008;14(24):8010-8.
94. Anders CK, Carey LA. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clin Breast Cancer* 2009;9 Suppl 2:S73-81.
95. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat* 2008;109(1):123-39.
96. Prat A, Cruz C, Hoadley KA, et al. Molecular features of the basal-like breast cancer subtype based on BRCA1 mutation status. *Breast Cancer Res Treat* 2014;147(1):185-91.
97. Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. *Oncogene* 2006;25(43):5846-53.
98. van Erkelens A, Derks L, Sie AS, et al. Lifestyle Risk Factors for Breast Cancer in BRCA1/2-Mutation Carriers Around Childbearing Age. *J Genet Couns* 2016.

99. Pijpe A, Manders P, Brohet RM, et al. Physical activity and the risk of breast cancer in BRCA1/2 mutation carriers. *Breast Cancer Res Treat* 2010;120(1):235-44.
100. Pettapiece-Phillips R, Narod SA, Kotsopoulos J. The role of body size and physical activity on the risk of breast cancer in BRCA mutation carriers. *Cancer Causes Control* 2015;26(3):333-44.
101. Wang M, Yu B, Westerlind K, et al. Prepubertal physical activity up-regulates estrogen receptor beta, BRCA1 and p53 mRNA expression in the rat mammary gland. *Breast Cancer Res Treat* 2009;115(1):213-20.
102. Jacobs TW, Gown AM, Yaziji H, et al. Comparison of fluorescence in situ hybridization and immunohistochemistry for the evaluation of HER-2/neu in breast cancer. *J Clin Oncol* 1999;17(7):1974-82.
103. Ross JS. Point: Fluorescence in situ hybridization is the preferred approach over immunohistochemistry for determining HER2 status. *Clin Chem* 2011;57(7):980-2.

2.7 TABLES

Table 1. CPS-II Nutrition Cohort women's baseline characteristics by recreational physical activity MET expenditure per week

Characteristic	Recreational physical activity MET expenditure (MET-h/week)		
	>0-8.75 (n=35,381)	>8.75-17.5 (n=16,582)	>17.5 (n=19,094)
Median MET-h/week	3.5	14.0	24.5
Age at baseline			
<50	2.4	1.8	2.2
50-59	35.2	32.5	33.4
60-69	47.9	50.4	50.1
70+	14.5	15.3	14.3
	Age-adjusted percentage[a]		
Race			
White	97.3	97.5	97.4
Black	1.5	1.3	1.4
Other	1.0	1.0	1.0
Missing	0.2	0.2	0.2
Education			
≤High school graduate	37.8	34.5	32.7
Some college	31.4	31.5	31.8
≥College grad	30.1	33.3	34.9
Missing	0.7	0.7	0.6
BMI (kg/m^2)			
<18.5	1.5	1.8	2.1
18.5-<22.5	22.9	27.4	30.9
22.5-<25.0	24.0	26.1	27.1
25.0-<30.0	32.4	30.9	28.3
≥30.0	17.4	12.5	10.3
Missing	1.8	1.2	1.3
Adult weight change, from age 18 (lbs.)			
>5 loss	4.4	5.5	6.7
≤5 loss-≤5 gain	8.9	11.0	13.3
>5-15 gain	14.8	17.8	19.5
>15-25 gain	17.4	18.6	18.7
>25-35 gain	16.0	15.7	15.2
>35 gain	36.5	29.7	24.9
Missing	2.0	1.8	1.6

Table 1. Continued…

Characteristic	Recreational physical activity MET expenditure (MET-h/week)		
	>0-8.75	>8.75-17.5	>17.5
Smoking status			
Never smoker	55.5	52.8	50.4
Former smoker	35.6	38.5	40.9
Current smoker	8.1	7.8	7.9
Missing	0.8	0.9	0.8
Alcohol intake			
Non-drinker	47.9	42.9	40.4
<1 drink/day	37.2	40.4	41.0
1+ drink/day	11.0	13.2	14.7
Missing	3.9	3.5	3.9
Family history of breast cancer			
No	86.2	86.6	86.3
Yes	13.8	13.4	13.7
HRT use			
Never	42.6	41.7	41.9
Former	17.0	16.9	16.9
Current	27.6	28.0	28.0
Unknown	12.8	13.4	13.2
Number of live births by age at first live birth			
Nulliparous	7.6	7.2	7.4
<25, 1-2 live births	16.0	15.6	16.2
25-29, 1-2 live births	11.9	12.2	12.1
30+, 1-2 live births	5.5	5.4	5.2
<25, 3+ live births	39.8	39.8	39.4
25-29, 3+ live births	14.5	15.3	15.1
30+, 3+ live births	2.3	2.1	2.1
Missing	2.3	2.4	2.5
Age at menarche (years)			
<12	19.4	19.3	19.6
12	25.3	25.7	24.8
13	29.7	29.7	29.0
>13	24.1	23.8	24.8
Missing	1.5	1.5	1.7

Table 1. Continued…

Characteristic	Recreational physical activity MET expenditure (MET-h/week)		
	>0-8.75	>8.75-17.5	>17.5
Age at menopause			
Not menopausal	5.5	5.4	5.5
<45	22.3	22.0	22.3
45-54	61.4	61.3	60.8
55+	8.4	8.9	9.1
Missing	2.4	2.4	2.3
History of cysts/breast lumps			
No	79.9	80.0	79.7
Yes	20.1	19.9	20.2
Missing	0.0	0.1	0.1

[a]Adjusted to the age distribution of the CPS-II Nutrition Cohort women

Table 2. Overall risk association of risk factors with invasive breast cancer stratified by cases with or without complete molecular data

	Cases with complete molecular data (N=900)		Cases with incomplete molecular data (N=4,336)		*P*-Heterogeneity
	N	HR (95% CI)	N	HR (95% CI)	
Age					0.001
<50	19	1.00 (Referent)	71	1.00 (Referent)	
50-59	391	1.38 (0.87-2.19)	1,506	1.40 (1.10-1.77)	
60-69	406	1.13 (0.72-1.79)	2,188	1.52 (1.20-1.93)	
70+	84	1.04 (0.63-1.71)	571	1.56 (1.22-2.00)	
Race					0.06
White	882	1.00 (Referent)	4,215	1.00 (Referent)	
Black	6	0.48 (0.22-1.08)	70	1.15 (0.91-1.46)	
Other	8	0.89 (0.44-1.79)	47	1.09 (0.82-1.46)	
Missing	4		4		
Education					<0.0001
≤High school graduate	245	1.00 (Referent)	1,441	1.00 (Referent)	
Some college	258	1.16 (0.98-1.39)	1,339	1.04 (0.96-1.12)	
≥College grad	393	1.67 (1.42-1.95)	1,525	1.13 (1.05-1.21)	
Missing	4		31		
BMI (kg/m^2)					0.36
<18.5	6	0.42 (0.19-0.93)	59	0.86 (0.66-1.11)	
18.5-<22.5	242	1.00 (Referent)	1,113	1.00 (Referent)	
22.5-<25.0	234	0.98 (0.83-1.16)	1,075	0.98 (0.91-1.07)	
25.0-<30.0	265	0.91 (0.77-1.08)	1,318	0.99 (0.92-1.07)	
≥30.0	137	1.07 (0.87-1.31)	704	1.18 (1.07-1.29)	
Missing	16		67		
Adult weight change, from age 18 (lbs.)					0.36
>5 loss	39	1.01 (0.68-1.48)	191	0.97 (0.81-1.15)	
≤5 loss-≤5 gain	80	1.00 (Referent)	413	1.00 (Referent)	
>5-15 gain	159	1.22 (0.94-1.60)	686	1.04 (0.92-1.17)	
>15-25 gain	168	1.22 (0.93-1.59)	737	1.04 (0.92-1.18)	
>25-35 gain	147	1.21 (0.93-1.59)	665	1.07 (0.95-1.21)	
>35 gain	293	1.25 (0.97-1.60)	1,573	1.29 (1.16-1.43)	
Missing	15		71		
Smoking status					0.61
Never smoker	483	1.00 (Referent)	2,219	1.00 (Referent)	
Former smoker	352	1.07 (0.93-1.23)	1,759	1.15 (1.08-1.23)	
Current smoker	61	0.99 (0.76-1.29)	320	1.07 (0.95-1.20)	
Missing	4		38		

Table 2. Continued…

	Cases with complete molecular data (N=900)		Cases with incomplete molecular data (N=4,336)		*P*-Heterogeneity
	N	HR (95% CI)	N	HR (95% CI)	
Alcohol intake					0.001
Non-drinker	346	1.00 (Referent)	1,849	1.00 (Referent)	
<1 drink/day	365	1.14 (0.98-1.32)	1,727	1.03 (0.97-1.10)	
1+ drink/day	172	1.75 (1.46-2.10)	621	1.19 (1.09-1.31)	
Missing	17		139		
Family history of breast cancer					0.04
No	752	1.00 (Referent)	3,492	1.00 (Referent)	
Yes	148	1.28 (1.07-1.53)	844	1.56 (1.45-1.68)	
HRT use					0.36
Never	355	1.00 (Referent)	1,676	1.00 (Referent)	
Former	118	0.87 (0.71-1.07)	640	0.98 (0.89-1.07)	
Current	306	1.22 (1.05-1.42)	1,337	1.16 (1.08-1.25)	
Unknown	121		683		
Number of live births by age at first live birth					0.39
Nulliparous	70	1.00 (Referent)	355	1.00 (Referent)	
<25, 1-2 live births	149	0.93 (0.70-1.24)	671	0.85 (0.75-0.97)	
25-29, 1-2 live births	97	0.82 (0.60-1.11)	574	0.98 (0.86-1.12)	
30+, 1-2 live births	67	1.31 (0.94-1.83)	277	1.07 (0.92-1.25)	
<25, 3+ live births	322	0.79 (0.61-1.02)	1,568	0.78 (0.70-0.88)	
25-29, 3+ live births	152	1.04 (0.79-1.38)	693	0.95 (0.84-1.08)	
30+, 3+ live births	22	1.08 (0.67-1.74)	103	0.99 (0.80-1.23)	
Missing	21		95		
Age at menarche (years)					0.82
<12	187	1.00 (Referent)	848	1.00 (Referent)	
12	226	0.93 (0.77-1.13)	1,110	1.01 (0.92-1.10)	
13	277	0.97 (0.81-1.17)	1,316	1.02 (0.93-1.11)	
>13	200	0.86 (0.71-1.05)	1,007	0.95 (0.87-1.04)	
Missing	10		55		
Age at menopause					0.07
<45	153	1.00 (Referent)	867	1.00 (Referent)	
45-54	558	1.33 (1.11-1.59)	2,710	1.14 (1.05-1.23)	
55+	100	1.75 (1.36-2.25)	426	1.29 (1.15-1.45)	
Not menopausal	66	1.51 (1.14-2.03)	246	1.06 (0.92-1.22)	
Missing	23		87		
History of cysts/breast lumps					0.57
No	673	1.00 (Referent)	3,203	1.00 (Referent)	
Yes	227	1.33 (1.14-1.54)	1,128	1.39 (1.30-1.49)	
Missing	0		5		

Table 3. Relative risk of invasive breast cancer by molecular subtypes per recreational physical activity and leisure-time sitting, CPS-II Nutrition Cohort of Women

		Luminal A			Luminal B		
	Person-years	Cases	RR (95% CI) [a]	RR (95% CI)[b]	Cases	RR (95% CI)[a]	RR (95% CI)[b]
MET-h/week total recreational physical activity							
>0-8.75	509,231	367	1.00 (Referent)	1.00 (Referent)	23	1.00 (Referent)	1.00 (Referent)
>8.75-17.5	240,457	174	1.02 (0.85-1.22)	1.00 (0.83-1.19)	11	1.04 (0.51-2.14)	1.02 (0.50-2.09)
>17.5	278,954	188	0.94 (0.79-1.12)	0.91 (0.76-1.09)	13	1.04 (0.53-2.06)	1.01 (0.51-2.00)
				P_{trend}=0.25[b]			P_{trend}=0.31[b]
Sitting h/day[c]							
<3	513,721	376	1.00 (Referent)	1.00 (Referent)	22	1.00 (Referent)	1.00 (Referent)
≥3	514,920	353	1.00 (0.86-1.16)	1.00 (0.86-1.16)	25	1.17 (0.62-2.02)	1.17 (0.68-2.03)

Table 3. Continued…

		HER2-enriched			Basal-like		
	Person-years	Cases	RR (95% CI)[a]	RR (95% CI)[b]	Cases	RR (95% CI)[a]	RR (95% CI)[b]
MET-h/week total recreational physical activity							
>0-8.75	509,231	18	1.00 (Referent)	1.00 (Referent)	43	1.00 (Referent)	1.00 (Referent)
>8.75-17.5	240,357	5	0.60 (0.23-1.61)	0.59 (0.22-1.57)	12	0.59 (0.31-1.11)	0.58 (0.30-1.09)
>17.5	278,954	6	0.62 (0.25-1.56)	0.60 (0.24-1.51)	17	0.71 (0.41-1.24)	0.69 (0.39-1.20)
				P_{trend}=0.83[b]			P_{trend}=0.14[b]
Sitting h/day[c]							
<3	513,721	14	1.00 (Referent)	1.00 (Referent)	31	1.00 (Referent)	1.00 (Referent)
≥3	514,920	15	1.20 (0.58-2.49)	1.21 (0.58-2.49)	41	1.33 (0.83-2.19)	1.32 (0.83-2.19)

Table 3. Continued…

		Unclassified		
	Person-years	Cases	RR (95% CI)[a]	RR (95% CI)[b]
MET-h/week total recreational physical activity				
>0-8.75	509,231	10	1.00 (Referent)	1.00 (Referent)
>8.75-17.5	240,357	8	1.68 (0.66-4.26)	1.65 (0.65-4.18)
>17.5	278,954	5	0.90 (0.31-2.63)	0.87 (0.30-2.56)
				P_{trend}=0.44[b]
Sitting h/day[c]				
<3	513,721	11	1.00 (Referent)	1.00 (Referent)
≥3	514,920	12	1.08 (0.48-2.39)	1.08 (0.49-2.40)

[a]Adjusted for age
[b]Adjusted for age, BMI, education, alcohol intake, and age at menopause
[c]Also adjusted for MET expenditure from total recreational physical activity

CHAPTER III: SUMMARY, PUBLIC HEALTH IMPLICATIONS, POSSIBLE FUTURE DIRECTIONS

The goal of this thesis was to improve the understanding of the relationship between recreational physical activity and breast cancer risk across intrinsic subtypes of breast cancer which are defined by gene expression profiling. Although no statistically significant association was identified in this thesis, the findings from this study still provide a valuation contribution to the field of breast cancer epidemiology.

The indication of evidence for modest risk reductions for luminal A and basal-like breast cancer in their associations with physical activity are largely aligned with the results of previous epidemiology studies. It is widely suggested that traditional breast cancer risk factors are associated with luminal A breast cancer (10, 73, 86). Physical activity, which is well-supported by epidemiologic evidence as a modifiable risk factor, could reduce the risk of luminal A breast cancer through hormonal pathways in the same way that hormonal risk factors, such as age at menarche, age at menopause and HRT use act on the risk of luminal A breast cancer.

A large part of the significance of this thesis is to shed light on the existing knowledge in the relationship between physical activity and basal-like breast cancer. While there are studies that examined the association between physical activity and TNBC, this is the first study to specifically look at basal-like breast cancer and its relationship with physical activity. Basal-like breast cancer is associated with poorer prognosis and it has the highest prevalence among premenopausal women (75, 95). In the findings, basal-like breast cancer was most strongly associated with physical activity among the five intrinsic subtypes, although results were not statistically significant. Interestingly, among the four previous studies related to physical activity and TNBC, two

studies found significant inverse associations, while one study revealed a modest risk reduction for TNBC.

The results of the findings related to basal-like breast cancer highlight the need for further research to examine the relationship between physical activity and basal-like breast cancer. The findings show substantial potential for physical activity as a modifiable risk factor to reduce the risk of basal-like cancer. Given the hormone-receptor negative characteristics of basal-like breast cancer, additional research is needed to identify the major mechanisms through which physical activity may potentially act to reduce breast cancer risk for basal-like tumors. Understanding these mechanisms could further identify additional potential modifiable risk factors and improve risk reduction strategies that could be used to specifically target high-risk demographics for basal-like cancer, such as young African-American women.

The findings from examining sedentary siting time and intrinsic subtypes of breast cancer showed no statistically significant association, but the results likely indicate a potential for an increased risk for breast cancer risk from increased sitting time, especially for basal-like breast cancer. This area of research related to sedentary sitting time and breast cancer is relatively new. Given the different categorizations of sitting time hours across limited previous studies, one possible reason for the null findings could be due to our chosen category of sitting time. The chosen comparison of sitting time categories ($\geq$3 hours vs. <3 hours per day) were limited by the combination of the response categories given in the CPS-II 1992 questionnaire as well as the lack of representation of cases in the higher sitting time categories. Future research related to

sedentary sitting time and breast cancer should encompass a meaningful categorization of sitting hours that could be standardized across studies.

Except for luminal A breast cancer, lack of power likely played a role in the imprecise findings of this study. Compared to previous prospective cohort studies, this study had the lowest number of invasive breast cancer cases (N=907), with about 81% of the cases (N=757) comprising of luminal A breast cancer cases. However, despite the shortcomings of the lack of power, the classification of intrinsic subtype breast cancer cases in the study was derived from immunohistochemical staining of tumor specimens at a pathology laboratory, as compared to previous prospective cohort studies that derived their cases based on medical records [1,4]. This contributes to the strength of this thesis, reducing the possibility of outcome misclassification.

In summary, the finding from this thesis revealed the potential indication for risk reductions for luminal A and basal-like breast cancer through the effects of physical activity. These findings support previous investigations which show established breast cancer risk factors, including physical activity, are associated with luminal A breast cancer. The findings related to basal-like breast cancer provides optimism that a modifiable lifestyle factor such as physical activity and sedentary sitting time could potentially reduce the risk of basal-like breast cancer. This optimism should fuel the need for future research in understanding the different mechanisms that differentially impact subtype-specific breast cancer risk. This could help inform future recommendations for breast cancer risk reduction based on known epidemiologic and demographic characteristics.

4 Appendices

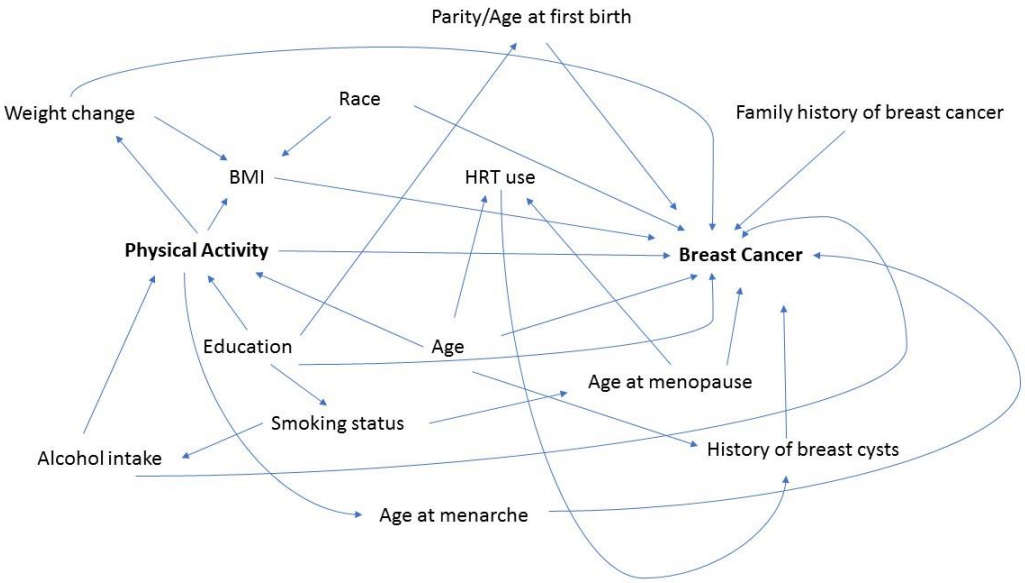
4.1 Appendix A

Appendix A. Summary of previous studies examining effect of physical activity on intrinsic subtypes of breast cancer.

Study Name	**Study Type**	**Population**	**Luminal A** RR (95% CI)	**Luminal B** RR (95% CI)	**HER2-enriched** RR (95% CI)	**Triple Negative** RR (95% CI)	**Source of biomarker data**
Ma et al., 2016	Prospective Cohort	California Teachers Study Cases: 4,827 Cohort: 108,907 women	Baseline PA: 0.87 (0.75-1.00) P_{trend}=0.05 Long-term PA: 0.88 (0.77-1.00) P_{trend}=0.01	Baseline PA: 0.82 (0.57-1.18) P_{trend}=0.14 Long Term PA: 1.16 (0.76-1.77) P_{trend}=0.75	Baseline PA: 0.70 (0.42-1.18) P_{trend}=0.41 Long Term PA: 1.08 (0.57-2.03) P_{trend}=0.97	Baseline PA: 0.72 (0.50-1.02) P_{trend}=0.07 Long Term PA: 0.73 (0.49-1.09) P_{trend}=0.11	Medical Reports
Phipps et al., 2011	Prospective Cohort	Women's Health Initiative Cases: 3,116 Cohort: 155,725 postmenopausal women	-	-	-	Moderate and low intensity PA: 0.75 (0.54-1.04) P_{trend}=0.07	Medical Reports
Ma et al., 2015	Population Case-Control	Women's CARE Study Cases: 1,195 Controls: 2,012	Average PA: 0.70 (0.52-0.94) P_{trend}=0.02	Average PA: 1.08 (0.58-2.02) P_{trend}=0.84	Average PA: 0.88 (0.44-1.74) P_{trend}=0.94	Average PA: 0.92 (0.64-1.31) P_{trend}=0.52	Tissue Repository
Schmidt et al., 2008	Population Case-Control	MARIE Study Cases: 3,414 Controls: 6,569	Average PA: 0.76 (0.63-0.92) P_{trend}=0.0069	Average PA: 0.70 (0.46-1.05) P_{trend}=0.0161	-	-	Medical Reports
Trivers et al., 2009	Population Case-Control	Atlanta arm of multi-center case-control study Cases: 476 Controls: 913	Baseline PA: 0.57 (0.45-0.71) (<median vs. median+)	Baseline PA: 0.89 (0.53-1.50) (<median vs. median+)	Baseline PA: 0.53 (0.31-0.92) (<median vs. median+)	Baseline PA: 0.73 (0.55-0.98) (<median vs. median+)	Tissue Repository

4.2 Appendix B

Directed acyclic graph (DAG) of the association between physical activity (exposure) and breast cancer (outcome)



4.3 Appendix C

Appendix C. Relative risk of invasive breast cancer by molecular subtypes per recreational physical activity and leisure-time sitting, CPS-II Nutrition Cohort of postmenopausal women only

		Luminal A			Luminal B		
	Person-years	Cases	RR (95% CI)[a]	RR (95% CI)[b]	Cases	RR (95% CI)[a]	RR (95% CI)[b]
MET-h/week total recreational physical activity							
>0-8.75	475,059	342	1.00 (Referent)	1.00 (Referent)	22	1.00 (Referent)	1.00 (Referent)
>8.75-17.5	226,972	157	0.97 (0.80-1.17)	0.95 (0.79-1.15)	11	1.08 (0.52-2.22)	1.05 (0.51-2.17)
>17.5	261,664	171	0.91 (0.76-1.09)	0.88 (0.73-1.06)	12	1.00 (0.49-2.02)	0.97 (0.48-1.95)
				P_{trend}=0.32[b]			P_{trend}=0.36[b]
Sitting h/day[c]							
<3	472,222	338	1.00 (Referent)	1.00 (Referent)	21	1.00 (Referent)	1.00 (Referent)
≥3	490,472	332	1.00 (0.86-1.17)	1.01 (0.86-1.17)	24	1.14 (0.66-2.00)	1.15 (0.66-2.01)

Appendix C. Continued…

		HER2-enriched			Basal-like		
	Person-years	Cases	RR (95% CI)[a]	RR (95% CI)[b]	Cases	RR (95% CI)[a]	RR (95% CI)[b]
MET-h/week total recreational physical activity							
>0-8.75	475,059	16	1.00 (Referent)	1.00 (Referent)	40	1.00 (Referent)	1.00 (Referent)
>8.75-17.5	226,972	5	0.66 (0.24-1.79)	0.64 (0.24-1.75)	11	0.57 (0.29-1.11)	0.56 (0.29-1.09)
>17.5	261,664	6	0.68 (0.27-1.75)	0.66 (0.26-1.67)	16	0.71 (0.40-1.27)	0.69 (0.39-1.23)
				P_{trend}=0.98[b]			P_{trend}=0.15[b]
Sitting h/day[c]							
<3	472,222	13	1.00 (Referent)	1.00 (Referent)	29	1.00 (Referent)	1.00 (Referent)
≥3	490,472	14	1.16 (0.55-2.45)	1.16 (0.55-2.46)	38	1.27 (0.78-2.05)	1.27 (0.78-2.06)

Appendix C. Continued…

		Unclassified		
	Person-years	Cases	RR (95% CI)[a]	RR (95% CI)[b]
MET-h/week total recreational physical activity				
>0-8.75	475,059	10	1.00 (Referent)	1.00 (Referent)
>8.75-17.5	226,972	8	1.67 (0.66-4.24)	1.64 (0.65-4.16)
>17.5	261,664	5	0.90 (0.31-2.63)	0.87 (0.30-2.56)
				P_{trend}=0.44[b]
Sitting h/day[c]				
<3	472,222	11	1.00 (Referent)	1.00 (Referent)
≥3	490,472	13	1.08 (0.49-2.40)	1.08 (0.49-2.40)

[a]Adjusted for age
[b]Adjusted for age, BMI, education, alcohol intake, and age at menopause
[c]Also adjusted for MET expenditure from total recreational physical activity

4.4 Appendix D

Appendix D. Relative risk of invasive breast cancer by molecular subtypes per recreational physical activity and leisure-time sitting, CPS-II Nutrition Cohort Women (multivariate analysis without BMI)

		Luminal A		Luminal B	
	Person-years	Cases	RR (95% CI)[a]	Cases	RR (95% CI)[a]
MET-h/week total recreational physical activity					
>0-8.75	509,231	367	1.00 (Referent)	23	1.00 (Referent)
>8.75-17.5	240,457	174	0.98 (0.82-1.18)	11	1.01 (0.49-2.07)
>17.5	278,954	188	0.89 (0.75-1.07)	13	0.99 (0.50-1.97)
			P_{trend}=0.25[a]		P_{trend}=0.31[a]
Sitting h/day[b]					
<3	513,721	376	1.00 (Referent)	22	1.00 (Referent)
≥3	514,920	353	1.02 (0.88-1.18)	25	1.19 (0.69-2.06)

Appendix D. Continued…

		HER2-enriched		Basal-like	
	Person-years	Cases	RR (95% CI)[a]	Cases	RR (95% CI)[a]
MET-h/week total recreational physical activity					
>0-8.75	509,231	18	1.00 (Referent)	43	1.00 (Referent)
>8.75-17.5	240,457	5	0.58 (0.22-1.55)	12	0.57 (0.30-1.08)
>17.5	278,954	6	0.59 (0.24-1.49)	17	0.68 (0.39-1.18)
			P_{trend}=0.83[b]		P_{trend}=0.12[b]
Sitting h/day[b]					
<3	513,721	14	1.00 (Referent)	31	1.00 (Referent)
≥3	514,920	15	1.23 (0.58-2.49)	41	1.34 (0.84-2.15)

Appendix D. Continued…

		Unclassified	
	Person-years	Cases	RR (95% CI)[a]
MET-h/week total recreational physical activity			
>0-8.75	509,231	10	1.00 (Referent)
>8.75-17.5	240,457	8	1.63 (0.64-4.13)
>17.5	278,954	5	0.86 (0.29-2.51)
			P_{trend}=0.44[a]
Sitting h/day[b]			
<3	513,721	11	1.00 (Referent)
≥3	514,920	12	1.09 (0.49-2.44)

[a]Adjusted for age, education, alcohol intake, and age at menopause
[b]Also adjusted for MET expenditure from total recreational physical activity