LIFESTYLE, HORMONES, AND BREAST CANCER

HANNAH OH

A Dissertation Submitted to the Faculty of
The Harvard School of Public Health
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Science
in the Departments of Epidemiology and Nutrition
Harvard University
Boston, Massachusetts

March 2015
Dissertation Advisor: Dr. A. Heather Eliassen

Lifestyle, hormones, and breast cancer

ABSTRACT

Breast cancer is a leading cause of cancer and the second leading cause of cancer death among women in the US. Although many risk factors for breast cancer are known, few are modifiable and little is known about ways to prevent its incidence.

Early-life body size is inversely associated with both premenopausal and postmenopausal breast cancer risk, suggesting an excess risk in lean girls. In a prospective analysis within the Nurses' Health Study (NHS) II, Chapter 1 examines whether adolescent physical activity mitigates the excess risk of breast cancer associated with early-life body leanness. Lean girls were at higher risk of breast cancer, regardless of the level of adolescent physical activity; however, the association was slightly, though not significantly, attenuated among the most active girls.

Breast cancer is hormone-related cancer; estrogen metabolites (EM) are both estrogenic and genotoxic, suggesting factors that alter the pattern of estrogen metabolism may contribute to breast carcinogenesis. With the application of advanced technology that measures 15 different individual estrogens and EM in urine, Chapter 2 examines the associations of dietary fiber and macronutrients intake with detailed estrogen metabolism in a cross-sectional analysis within the NHSII. Few significant associations were identified: a positive association between total fiber intake and 4-methoxyestradiol, an inverse association between total fiber intake and 17-epiestriol, and inverse associations for polyunsaturated and trans-fat intakes with 17-epiestriol.
The tissue-specific responsiveness to potentially carcinogenic hormones, estrogen and progesterone, is partially regulated by the tissue expression of receptors that bind these hormones. Using benign breast biopsy samples collected in a nested case-control study within the NHS and NHSII, Chapter 3 assesses estrogen receptor (ER), progesterone receptor (PR), and proliferative marker Ki67 expression in normal breast tissue in relation to subsequent breast cancer risk. In this case-control analysis, PR expression in normal breast tissue was significantly positively associated with breast cancer risk in premenopausal women. ER and Ki67 expression was not significantly associated with breast cancer risk; however, our power was limited.

Results of this dissertation help elucidate the underlying biologic mechanisms of breast cancer and enhance our understanding of the link between risk factors and breast cancer risk.
TABLE OF CONTENTS

INTRODUCTION 1
   References 4

CHAPTER 1: The interaction between early-life body size and physical activity on risk of breast cancer
   Abstract 8
   Introduction 9
   Materials and methods 10
   Results 15
   Discussion 17
   Tables 22
   Figures 28
   References 30

CHAPTER 2: Dietary macronutrient and fiber intake and patterns of estrogen metabolism in premenopausal women
   Abstract 37
   Introduction 39
   Materials and methods 41
   Results 46
   Discussion 48
   Tables 52
   References 69

CHAPTER 3: Expression of estrogen receptor, progesterone receptor, and Ki67 in normal breast tissue and subsequent risk of breast cancer
   Abstract 75
   Introduction 77
   Materials and methods 78
   Results 83
   Discussion 85
   Tables 90
   References 94
CHAPTER 1

Figure 1.1. Hazard ratios of breast cancer according to self-reported adolescent physical activity (average between ages 12-17 years) and body size at three different age periods (ages 5-10, 10-20, and 20 years) in the Nurses’ Health Study II, 1997-2011
LIST OF TABLES

CHAPTER 1

**Table 1.1.** Baseline characteristics of the study population in 1997 by average childhood (ages 5-10 years) self-reported body size in the Nurses’ Health Study II (N=74,723) 22

**Table 1.2.** Hazard ratios of breast cancer according to self-reported early-life body size in the Nurses’ Health Study II, 1997-2011 23

**Table 1.3.** Hazard ratios of breast cancer according to self-reported early-life body size, stratified by adolescent physical activity (average between ages 12-17 years) in the Nurses’ Health Study II, 1997-2011 26

CHAPTER 2

**Table 2.1.** Characteristics of the study population by average dietary fiber and total fat intake at urine collection in the Nurses’ Health Study II 52

**Table 2.2.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of total fiber in 1995 and 1999: Nurses’ Health Study II (n = 598) 53

**Table 2.3.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of total fat in 1995 and 1999: Nurses’ Health Study II (n = 598) 55

**Table 2.4.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of carbohydrates in 1995 and 1999: Nurses’ Health Study II (n = 598) 57

**Table 2.5.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of total protein in 1995 and 1999: Nurses’ Health Study II (n = 598) 59

**Supplementary table 2.S1.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of different fiber from different sources in 1995 and 1999: Nurses’ Health Study II (n = 598) 61

**Supplementary table 2.S2.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of different fat subtypes in 1995 and 1999: Nurses’ Health Study II (n = 598) 63

**Supplementary table 2.S3.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average glycemic load in 1995 and 1999: Nurses’ Health Study II (n = 598) 65

**Supplementary table 2.S4.** Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of protein from different sources in 1995 and 1999: Nurses’ Health Study II (n = 598) 67
CHAPTER 3

Table 3.1. Characteristics of the study population at benign breast biopsy by breast cancer case-control status in the Nurses’ Health Study and the Nurses’ Health Study II

Table 3.2. Odds ratios (95% confidence interval) of developing subsequent breast cancer according to tertiles of mean percentage of ER, PR, Ki67 expression in normal breast tissue in the Nurses’ Health Study and the Nurses’ Health Study II

Table 3.3. Odds ratios (95% confidence interval) of developing subsequent breast cancer according to ER, PR, Ki67 expression in normal breast tissue and benign lesion type in the Nurses’ Health Study and the Nurses’ Health Study II

Supplementary table 3.S1. Spearman correlation among tissue markers in normal breast tissue in the Nurses’ Health Study and the Nurses’ Health Study II

Supplementary table 3.S2. Spearman correlation among tissue markers in normal breast tissue vs. tumor tissue among cases in the Nurses’ Health Study and the Nurses’ Health Study II
ACKNOWLEDGEMENTS

Over the past four years, I have received support and encouragement from a great number of individuals. I am very grateful for their contributions to this dissertation. First, I would like to thank God for leading and guiding me through this journey. I could do all this through Him who gives me strength (Philippians 4:13).

I would like to express my deepest gratitude to my advisor/mentor Dr. A. Heather Eliassen for her time, patience, support, mentoring and guidance. She helped me with all her heart and effort to make sure my doctoral training is good and preparing me well for the next step of my career. Her guidance has made this a thoughtful and rewarding journey.

I also wish to thank the members of my dissertation and oral exam committee—Drs. A. Heather Eliassen, Rulla M. Tamimi, Stephanie A. Smith-Warner, Molin Wang—for their thoughtful comments and support over the past two years. Without their crucial contribution, this dissertation would not have been completed. I truly enjoyed and learned much during our discussion at committee meetings. I also would like to give extra thanks to Heather, Rulla, and Stephanie for bolstering my postdoctoral applications with their letters and constantly encouraging me during the process.

I also acknowledge Drs. Meir J. Stampfer and Walter C. Willett, and the National Institute of Health for supporting my doctoral studies with the T32 training grant (tuition and stipend support).
I would like to thank the participants and staff of the Nurses’ Health Study (NHS) and the NHSII for their valuable contributions as well as state cancer registries for their help. Their participation made this study possible.

I would like to extend my appreciation to my colleagues and friends (written qualifying exam study group)—Seoyoung Kim, Dae Hyun Kim, NaNa Keum, Mi-Hye Lee—for their friendship, emotional support, and thoughtful advice over the past four years.

Finally, I give special thanks to my family (parents, brother, in-laws, and my dear husband) for their love, support, and patience. Without them, this effort would have been worth nothing. I especially thank my husband, Gidong Sim, for his unfailing love and understanding. He made the past two years of my life full of joy and happiness.

Hannah Oh
INTRODUCTION

Breast cancer is a leading cause of cancer and the second leading cause of cancer death among women in the US [1]. The risk of breast cancer increases with age and the majority of cases occur among women of age 50 years or older [2]. Many risk factors for breast cancer are well-studied and support that breast cancer is a hormone-related cancer: a) reproductive factors such as age at menarche, parity, and age at first birth are established risk factors for breast cancer [3-5], b) exogenous estrogens (e.g., oral contraceptives [6], postmenopausal hormone therapy [3, 7]) are positively associated with breast cancer risk, c) adult obesity [8] increases postmenopausal breast cancer risk as adipose tissue produces estrogens via aromatase activity in postmenopausal women, and d) tamoxifen [9], an estrogen receptor antagonist in breast tissue, can treat and prevent hormone-dependent breast cancer. However, among the known risk factors for breast cancer, few are modifiable (i.e., body fatness, physical activity, diet, oral contraceptives use, postmenopausal hormone therapy use). Modifiable risk factors, including body fatness, physical activity, and diet, are the focus of Chapters 1 and 2.

While many identified breast cancer risk factors are related to later-life exposure, breast tissues are particularly susceptible to carcinogens during the early-life period, especially between menarche and the first childbirth [5, 10, 11]. Starting with menarche, girls are exposed to ovarian hormones (e.g., estrogen) that promote the replication of breast cells and the expansion of initiated mutations [12]. After the first pregnancy, terminal differentiation of the mammary gland takes place; the differentiated cells are less likely to transform into tumor cells [12]. The longer the duration of this window of
susceptibility the higher the risk of breast cancer [10, 13]. Although evidence is inconsistent, adolescent factors such as diet, alcohol intake, and smoking behavior are suggestively associated with the risk of breast cancer [14-16]. Childhood to adolescent body size and physical activity are also associated with breast cancer risk in later life [17-21]. Thus, lifestyle modifications in early life may contribute importantly to the prevention of breast cancer. We investigated early-life body size and physical activity in Chapter 1.

Several types of hormones (e.g., estrogen, progesterone, and growth factors) that regulate cell proliferation and apoptosis are involved in breast carcinogenesis [22, 23]. Hormone levels measured in biospecimens (e.g., plasma, urine, fecal) may be used to elucidate the underlying biology of breast cancer and may enhance our understanding of the link between risk factors and breast cancer risk. Although the tissue-level exposure to key hormones may be the most relevant measure of exposure, we used several related biomarkers to identify patterns associated with breast cancer risk or with breast cancer risk factors: estrogens metabolites measured in urine (as a measure of excreted estrogens) in Chapter 2, and expression of hormone receptors in normal breast tissue (which may indicate tissue-specific responsiveness to these hormones) in Chapter 3.

Overall, this dissertation focuses on lifestyle factors, including body fatness, physical activity, and diet, that are likely modifiable, and their relationships with breast cancer risk or with hormone-related biomarkers, as well as the relation between hormone
receptor expression and breast cancer risk. These analyses were conducted within the Nurses’ Health Study (NHS) and NHSII cohorts.

**Chapter 1** investigates the risk of breast cancer related to two important early-life modifiable lifestyle factors: physical activity and body fatness. To assess whether adolescent physical activity modifies the excess breast cancer risk associated with early-life body leanness, we assessed the interaction between early-life physical activity and body size on risk of breast cancer. **Chapter 2** examines whether diet, particularly macronutrients and fiber intake, is associated with patterns of estrogen metabolism using urine samples of premenopausal women. Finally, **Chapter 3** assesses tissue-level expression of hormone receptors (estrogen and progesterone receptor) and proliferative marker Ki67 in relation to breast cancer risk using normal breast tissue in benign biopsy samples.
REFERENCES


CHAPTER 1

The interaction between early-life body size and physical activity on risk of breast cancer

Hannah Oh\textsuperscript{1, 2, 3}
Caroline E. Boeke\textsuperscript{1, 2}
Rulla M. Tamimi\textsuperscript{1, 2}
Stephanie A. Smith-Warner\textsuperscript{1, 3}
Molin Wang\textsuperscript{1, 2, 4}
Walter C. Willett\textsuperscript{1, 2, 3}
A. Heather Eliassen\textsuperscript{1, 2}

\textsuperscript{1} Department of Epidemiology, Harvard School of Public Health
\textsuperscript{2} Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School
\textsuperscript{3} Department of Nutrition, Harvard School of Public Health
\textsuperscript{4} Department of Biostatistics, Harvard School of Public Health
ABSTRACT

While early-life body leanness is associated with increased breast cancer risk, early-life physical activity may protect against breast cancer. We examined whether the excess risk among lean girls is modified by their levels of prior, concurrent, or future physical activity. We conducted an analysis among 74,723 women in the Nurses’ Health Study II (follow-up 1997-2011). Participants recalled their body size at ages 5, 10, and 20 years in 1989 using a 9-level pictogram (level 1: most lean). In 1997, they reported adolescent levels of physical activity (ages 12-13 and 14-17 years). Cox proportional hazards models estimated the overall association of body size with breast cancer risk and assessed interactions of adolescent physical activity with body size at three different age periods (5-10, 10-20, and 20 years), adjusting for early-life and adult risk factors for breast cancer. Regardless of levels of adolescent physical activity, early-life body leanness (level 1-2 vs. 4.5+) was significantly associated with higher breast cancer risk. The association was slightly attenuated among those who were active (60+ MET-hr/wk) during adolescence compared to those who were inactive (<30 MET-hr/wk) (body size at ages 5-10 years: hazard ratio=1.37, 95% confidence interval=1.04-1.81 vs. 1.66, 1.29-2.12), but the interaction was not significant (p=0.72). The results were similar for body size at three different age periods. Being lean during early life is a risk factor for breast cancer among both inactive and active girls. Adolescent physical activity did not significantly modify the association, although some interaction cannot be excluded.
INTRODUCTION

Evidence consistently supports that early-life body fatness is associated with decreased breast cancer risk in both premenopausal [1-5] and postmenopausal [1, 2, 4-7] women. In our prior analysis in the Nurses' Health Study (NHS) and NHSII cohorts we observed that childhood and adolescent body fatness was significantly associated with a 20-50% decreased breast cancer risk regardless of menopausal status [4]. The association did not change after adjustment for adult body mass index (BMI) [4] suggesting an independent, lifelong protection from breast cancer. Although most studies were conducted among Caucasian women [3-6, 8, 9], the inverse association was confirmed in African American [2, 10], Hispanic [11], and Asian women [1, 7]. This finding contrasts with adult obesity, which increases the risk of postmenopausal breast cancer [12]. The hypothesized underlying mechanisms are increased anovulation [13, 14], altered circulating levels of hormones (e.g., estrogen, IGF-1) [15, 16], altered development of breast tissue (e.g., breast density) [17, 18], and decelerated adolescent growth [19] in heavier girls.

While early-life body leanness may be a risk factor for breast cancer, some evidence suggests a modest inverse association between early-life physical activity and later breast cancer risk [20]. In a recent analysis in the NHSII, we observed that physical activity (highest vs. lowest quintiles) at ages 14-17 years was associated with a 15% lower risk of premenopausal breast cancer [21]. Thus, we hypothesized that an excess risk associated with early-life body leanness may be attenuated among women who engaged in high levels of activity during their adolescence. An interaction between these two factors is biologically plausible because they share potential mechanisms.
such as change in IGF-1 and sex hormones levels [15, 16, 22], age at menarche [23-25], and menstrual cycle patterns [23, 26]. Furthermore, high levels of physical activity may help maintain a lean body weight, which we hypothesized may not increase the risk among girls who became lean as a result of physical activity. In contrast, being lean despite being inactive may result from a carcinogenic hormonal profile (e.g., high androgen).

To our knowledge, only one study to date has examined the potential interaction between early-life body size and physical activity on breast cancer risk (p-interaction=0.02) [27]. However, this study was a retrospective, case-control study where participants recalled their body size and activity levels after their diagnosis of breast cancer. In a prospective analysis within the NHSII cohort, we considered three different questions when assessing the interactions between body size and physical activity on the risk of breast cancer, with respect to the timing of each. First, we assessed an interaction between body size and ‘future’ physical activity to address the question of whether a girl with a given body size could alter her risk profile by engaging in high levels of activity immediately afterward. Next, we assessed an interaction between ‘concurrent’ body size and physical activity. Finally, we assessed physical activity ‘prior’ to body size to address the question of risk among lean girls who may have been lean as a result of physical activity.

MATERIALS AND METHODS

Study population
The NHSII is an ongoing cohort study that began in 1989, including 116,430 female registered nurses aged 25-42 years. An initial self-administered questionnaire was mailed to collect information on health behaviors, lifestyle and reproductive factors, and medical histories. Biennial follow-up questionnaires updated information on known and suspected risk factors for breast cancer, as well as new disease diagnoses. Response rates are >90% based on cumulative person-time of follow-up.

The follow-up of this analysis began in 1997 when adolescent physical activity was assessed. We excluded women who died or had a diagnosis of cancer, other than nonmelanoma skin cancer, prior to 1997, or did not report their early-life body size or physical activity.

This study was approved by the Committee on the Use of Human Subjects in Research at Brigham and Women’s Hospital (Boston, Massachusetts). Completion of the self-administered questionnaire was considered to imply informed consent.

**Assessment of body size**

In 1989, participants were asked to recall their body size at ages 5, 10, and 20 years using a 9-level pictogram [28]. According to a validation study conducted within the Third Harvard Growth study, a Boston-area longitudinal study of school-children initiated in 1922, recalled body size using this pictogram by older women at ages 71-76 years (n=100) has a good correlation with their BMI measured while they were at ages 5 (Pearson r=0.60), 10 (r=0.65), and 20 years (r=0.66) [29]. We averaged body size at ages 5 and 10 years and ages 10 and 20 years to obtain estimates of childhood and adolescent body size, respectively. Higher levels (>5) of body size were collapsed
because there were fewer women in those levels. For stratification analyses, the levels with a similar risk (i.e., level 1 and 2) were collapsed to increase power.

Assessment of physical activity

In 1997, participants were asked to recall their average hours per week spent on strenuous (e.g. running, swimming laps) and moderate (e.g. hiking, casual cycling) recreational activities, and walking (to and from school or work) during grades 7-8 (ages 12-13) and 9-12 (ages 14-17). Within a subsample of 160 NHSII participants, the 4-year reproducibility of recalled activities using this questionnaire was good (r=0.64 for total activity) [30].

We calculated participants’ metabolic equivalent (MET)-hour/week of total physical activity by multiplying the number of hours per week of each activity with its corresponding average MET values (strenuous = 7 METs; moderate = 4.5 METs; walking=3 METs) [31] and then summing the values from all activities. We averaged total activity (sum of recreational activities and walking) at ages 12-13 and 14-17 years to obtain an estimate of adolescent activity. Because we previously observed the strongest association with breast cancer by activity at ages 14-17 years [21], we conducted a sensitivity analysis using only physical activity at ages 14-17 years.

Ascertainment of breast cancer cases

Breast cancer cases diagnosed through May 31, 2011 were identified by self-administered biennial questionnaire and the National Death Index, and confirmed by reviewing medical records. Information on tumor characteristics was abstracted from the
pathology reports. The analysis included both premenopausal and postmenopausal cases and both invasive and in situ cases because the association with early-life body size did not vary by these characteristics [4]. A total of 2,641 cases were identified during 1997-2011 follow-up (622 in situ and 1,769 invasive cases confirmed by medical records; 1,683 estrogen receptor-positive and 381 estrogen receptor-negative tumors). Among those with medical records, 99% of reported cases were confirmed; therefore, we included 250 breast cancer cases for whom the medical records were missing and only verbally confirmed by telephone.

**Covariate assessment**

Height, adolescent alcohol intake, adolescent smoking status, age at menarche, menstrual length and pattern at ages 18-22 years, and BMI at age 18 were reported once via the 1989 baseline questionnaire. Age, parity/age at first birth, personal history of benign breast disease, oral contraceptive use, smoking, menopausal status, and current BMI were updated biennially and family history of breast cancer was updated every 4 years since 1989. Birth weight was assessed in 1991 and adolescent physical activity in 1997. Adult alcohol intake was reported in 1991 and updated every 4 years. Adult physical activity was reported in 1989, 1991, 1997, 2001, 2005 and 2009.

**Statistical analysis**

Each woman contributed person-time from the return date of her questionnaire in 1997 until a diagnosis of cancer except nonmelanoma skin cancer, death, or the end of follow-up on June 1, 2011, whichever came earlier. Cox proportional hazards models
were used to estimate hazard ratios (HR) and their 95% confidence intervals (CI) for categories of body size. We performed tests for trend by modeling body size as a continuous variable (level 1-9).

Multivariate-adjusted models included both early-life and adult risk factors for breast cancer. Body size at different age periods during early-life were not simultaneously adjusted in the model because they were strongly correlated with each other ($r = 0.5$-$0.8$). Covariates that were reported more than once during the follow up were treated as time-varying covariates. Because age at menarche, menstrual pattern at ages 18-22 years, BMI at age 18 (for body size at ages 5-10 years only), and current BMI could be intermediate variables through which early-life body size is associated with breast cancer, we adjusted for them only in secondary analyses.

To assess whether the association varied by the level of adolescent physical activity, we stratified the participants by approximate tertiles (inactive, moderately active, active: <30, 30-59.9, 60+ MET-hr/wk). To assess the joint effect of body size and physical activity, we cross-classified by both body size and physical activity and estimated the HR using the most overweight and inactive (<30 MET-hr/wk) group as a reference. The same analyses were performed on: 1) childhood body size and adolescent (‘future’) physical activity, 2) adolescent body size and adolescent (‘concurrent’) physical activity, and 3) body size at age 20 years and adolescent (‘prior’) physical activity. Likelihood ratio tests compared models with and without interaction terms to determine significance of the interactions. We repeated the analyses after restricting to premenopausal women (1,364 cases) because adolescent activity is suggestively more strongly associated with the risk of premenopausal breast cancer.
For these analyses, women contributed person-time only while they were premenopausal.

Lastly, we used a competing risks Cox proportional hazards model [32, 33], using a data augmentation method [34], to assess whether the interaction differed by estrogen receptor (ER) status of tumor. A separate observation was created for each subject for each type of tumor (ER+/-) and stratified the data on event type, allowing for estimation of separate associations with each type of tumor. Likelihood ratio tests compared models that assumed different associations for each type of tumor with models that assumed the same association. All statistical tests were two-sided with 5% type I error and performed using SAS software version 9.

RESULTS

The mean age at the beginning of follow-up was 42 years for the 74,723 women who were included in our analysis. Compared with women who were overweight during childhood, those who were lean were more likely to be older at menarche, engage in higher levels of physical activity, and have lower BMI later in life (Table 1.1). These women were also less likely to have weighed 3.9 kg or more at birth and have smoked during adolescence and adulthood, and more likely to have a personal history of benign breast disease.

Body size at ages 5 to 20 years were all significantly inversely associated with breast cancer risk (Table 1.2). Women who were lean (level 1 vs. 4.5+) during their childhood had a 55% higher risk compared to the women who were overweight, adjusting for adolescent and adult risk factors for breast cancer (HR=1.55, 95%
CI=1.31-1.83; p-trend<0.0001). The estimates were similar to those from age-adjusted models. The association became slightly stronger when menstrual characteristics were added to the model; the change was primarily driven by age at menarche. Adjustment for current adult BMI only slightly attenuated the association (HR=1.48, 95% CI=1.24-1.75) and adjustment for BMI at age 18 somewhat attenuated the association (HR=1.38, 95% CI=1.15-1.65).

Within each stratum of adolescent activity, early-life body leanness was significantly associated with higher breast cancer risk (Table 1.3). Multivariate-adjusted HR for women who were lean (level 1-2 vs. 4.5+) during their childhood were 1.66 (95% CI=1.29-2.12) among inactive women, 1.59 (95% CI=1.22-2.08) among moderately active women, and 1.37 (95% CI=1.04-1.81) among active women. Similar patterns were observed for each time comparison of body size and physical activity, with slightly attenuated positive associations for body leanness among women who engaged in higher levels of activity during adolescence. However, the interactions were not statistically significant.

When we cross-classified body size and physical activity, the risk in women who were lean and active in early-life was only suggestively lower compared to women who were lean and inactive (future activity: HR=1.55, 95% CI=1.15-2.09 vs. HR=1.66, 95% CI=1.30-2.13; concurrent activity: HR=1.61, 95% CI=1.21-2.15 vs. HR=1.72, 95% CI=1.36-2.17; prior activity: HR=1.69, 95% CI=1.24-2.30 vs. HR=1.73, 95% CI=1.34-2.25) and their confidence intervals overlapped considerably (Figure 1.1). The results did not differ by ER status of tumors (p-heterogeneity_{future activity}=0.47, p-heterogeneity_{concurrent activity}=0.40, p-heterogeneity_{prior activity}=0.58) and were consistent
when we restricted the analyses to premenopausal women or using activity (tertile/quintile) at ages 14-17 years only.

**DISCUSSION**

In this large prospective study, body leanness during childhood to early adulthood was associated with a significantly higher breast cancer risk, regardless of the level of adolescent physical activity. The association was independent of early-life and adult risk factors for breast cancer and adult BMI. Although the associations were slightly attenuated among women who were active compared to those who were inactive, adolescent physical activity did not significantly modify the association. The results did not vary by ER status of tumors.

Our findings of a positive association between early-life body leanness and breast cancer risk are consistent with those from previous studies [1-4, 8, 9, 11] in which lean girls had a 10-70% increased risk compared with overweight girls. The magnitude of the association may vary among studies due to variations in the range of BMI, sample size, length of follow-up, and population characteristics (e.g., ethnicity, age). For example, in our study of primarily Caucasian women, the corresponding median BMI at age 18 (the most proximate age available, r=0.69) for women in the lowest and highest somatotype groups at age 20 were 18 and 26 kg/m², respectively. In a Danish study (3,340 cases), a more narrow contrast of BMI at age 14 (median 16 vs. 22 kg/m²) was associated with a 20% higher risk [8]. Another study in African American women (1,062 cases) reported a 32-47% higher risk in women with BMI <20 vs. ≥25 kg/m² at age 18 [2]. Most studies, including ours, did not include women who were
extremely lean (such as those in the settings of extreme caloric restriction) during early life; little is known about the relationship in extremely lean girls although some studies observed reduced risk in women who experienced severe famine [35] or anorexia nervosa [36] during adolescence.

While the biology underlying the association between early-life body size and breast cancer risk is not clear, multiple mechanisms, including both estrogen and non-estrogen pathways, are likely to be involved as the association is significant for both ER+ and ER- tumors [4]. First, overweight girls are more likely to experience irregular menstrual cycles [23] and anovulation [13, 14], despite having earlier menarche [23]. However, these factors are unlikely to explain the strong association of early-life body size because adjustment for menstrual pattern between ages 14-22 years did not alter our results. Second, although biomarker studies have generally not observed an association between childhood body size and the levels of adolescent [37] and adult [16] endogenous sex hormones, we cannot exclude the possibility that the breast tissue-specific response to sex hormones (e.g., estrogen sensitivity) may vary by early-life body size. Third, early-life body leanness may contribute to a lifelong set point of IGF-1. In the NHS and NHSII, we observed 14% higher adult levels of plasma IGF-1 among women who were lean (vs. overweight) at young ages [15], and adult IGF-1 levels have been associated with a higher risk of breast cancer [38]. However, genetically-determined IGF-1 levels do not seem to play a major role because no association was observed between childhood body size and SNPs in the IGF-1 and IGFBP-1/-3 gene regions [39]. Lastly, early-life body size has been inversely associated with mammographic density [17, 18] and the risk of benign breast disease [40, 41].
Given the strong association between mammographic density and breast cancer, breast density may play a role in the association between early-life body size and breast cancer risk; however, adjustment for mammographic density did not substantially alter the association in our prior results [42]. Evidence from mathematical models [43] and epidemiological studies [44] supports that breast tissues are particularly susceptible to carcinogens during early life, especially between menarche and the first childbirth [45]. For instance, overweight girls may have earlier differentiation and altered development of breast tissue structures [46] that decreases susceptibility to breast cancer throughout their lives; differentiated cells become less likely to proliferate or transform into tumor cells [43, 45].

A lack of significant interaction in our study suggests that the modest association of adolescent physical activity with breast cancer risk may not be strong enough to counteract the robust association of early-life body size. While early-life physical activity and body size share several potential mechanisms, it is possible that early-life body size acts upon breast carcinogenesis in a way that is not modifiable by physical activity. For example, early-life body size is inversely associated with mammographic density [17, 18] while physical activity is not [47]. If early-life body leanness increases risk primarily through higher breast density, engaging in high levels of activity during early life may not modify the association between early-life body size and breast cancer risk. However, we observed a slightly attenuated association of early-life body size among women who were physically active during adolescence. Our results contrast with those from the only prior study to date, which has suggested a significantly stronger association of BMI at age 18 among postmenopausal women who engaged in higher levels of activity at ages
14-22 year [27]. Possible recall bias in their retrospective study may explain the contrast, as suggested by their stronger inverse association of adolescent physical activity (≥364 vs. 0 times/year: OR=0.55, 95% CI=0.39-0.78) and weaker positive association of early-life body leanness (BMI <18.6 vs. >21.8 kg/m²: OR=1.09, 95% CI=0.97-1.22) compared to ours.

Given some limitations of our study, we cannot exclude the possibility of some interaction between early-life body size and physical activity. Because participants recalled their early-life body size and adolescent physical activity when they were ages 25-50 years, our exposure data are inevitably imperfect. Physical activity is particularly difficult to measure accurately using questionnaires as there are many components including intensity, frequency, and timing that may be differentially related to breast cancer risk. Our physical activity data also did not include non-recreational activity such as household and occupational activities. Given we focused on adolescence, levels of these activities were likely minimal in our study population; however, failure to meet this assumption would result in measurement error in physical activity data. In our analysis where participants recalled body size and physical activity prior to breast cancer diagnosis, the resulting measurement error is most likely independent of outcome (i.e., non-differential) and may have attenuated the association and interaction. For early-life body size, the association was similar in other studies that used measured BMI [8, 9] supporting the validity of our recalled body size data. Despite these limitations, this study included a long follow-up and adjustment for both early-life and adult risk factors for breast cancer.
In summary, while early-life body leanness was significantly associated with a higher risk of breast cancer, adolescent physical activity did not significantly modify the association, although some interaction cannot be excluded. Further study is required to elucidate the mechanisms underlying this association and to identify strategies to mitigate the excess risk of breast cancer in lean girls. Despite the higher risk of breast cancer observed among lean girls, lifetime obesity and physical inactivity increase the risk of other types of cancer and various other chronic diseases (e.g., cardiovascular, endocrine, pulmonary, renal, gastrointestinal, musculoskeletal) [48]. Further, adult body fatness increases postmenopausal breast cancer risk, and girls who are overweight are likely to remain overweight as adults [49]. Therefore, public health recommendations should still support early life maintenance of a healthy weight and physical activity.
Table 1.1. Baseline characteristics of the study population in 1997 by average childhood (ages 5-10 years) self-reported body size in the Nurses’ Health Study II (N=74,723)a

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1.5-2</th>
<th>2.5-3</th>
<th>3.5-4</th>
<th>4.5+ c</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>12,609</td>
<td>22,846</td>
<td>18,960</td>
<td>12,759</td>
<td>7,549</td>
</tr>
<tr>
<td>Age (years)</td>
<td>42.5 (4.8)</td>
<td>41.8 (4.7)</td>
<td>42.0 (4.6)</td>
<td>42.3 (4.6)</td>
<td>42.5 (4.6)</td>
</tr>
<tr>
<td>Height (inches)</td>
<td>64.9 (2.6)</td>
<td>64.9 (2.6)</td>
<td>64.8 (2.6)</td>
<td>64.9 (2.6)</td>
<td>65.1 (2.6)</td>
</tr>
<tr>
<td>Birth weight &gt;=3.9 kg, %</td>
<td>10.0</td>
<td>11.5</td>
<td>13.9</td>
<td>16.2</td>
<td>18.7</td>
</tr>
<tr>
<td>Age at menarche (years)</td>
<td>12.8 (1.4)</td>
<td>12.6 (1.4)</td>
<td>12.3 (1.4)</td>
<td>12.1 (1.4)</td>
<td>12.0 (1.4)</td>
</tr>
</tbody>
</table>

Menstrual cycle pattern at ages 18-22 years
- Irregular or no periods, % 24.5 23.3 23.9 24.4 26.5
- Time to menstrual regularity since menarche 23.3 21.8 21.7 22.4 24.8

BMI at age 18 (kg/m²) 19.4 (2.1) 20.2 (2.3) 21.5 (2.8) 22.7 (3.3) 24.4 (4.7)
Total physical activity at ages 12-17 years (MET-hr/wk) 56.3 (37.9) 54.5 (35.9) 51.9 (35.2) 48.4 (33.9) 46.0 (33.8)
Alcohol drinkers at ages 15-17 years, % 22.8 23.0 23.0 23.6 23.2
Alcohol intake at ages 15-17 years among drinkers (g/day) 4.6 (6.3) 4.2(5.4) 4.3(5.4) 4.7(6.5) 4.9(7.5)
Total caloric intake during high school, kcal/d 2465 (580) 2488 (566) 2493 (557) 2486 (573) 2505 (561)
Smoking at ages 10-20 years, % 31.0 30.1 31.4 34.1 38.0
Current BMI (kg/m²) 23.8 (4.2) 24.6 (4.8) 26.4 (5.9) 28.1 (6.7) 30.0 (7.7)
Current total physical activity (MET-hr/wk) 19.1 (24.7) 18.7 (22.3) 18.2 (22.2) 18.3 (22.5) 17.5 (21.8)
Current alcohol intake (g/day) 3.5 (6.6) 3.5 (6.4) 3.5 (6.6) 3.5 (6.9) 3.3 (6.7)
Current smoking status - Never, % 66.7 68.0 66.6 63.8 59.6
- Past, % 22.9 23.4 24.4 25.4 27.2
- Current, % 10.3 8.6 9.0 10.8 13.2
Ever use of oral contraceptives - Never, % 11.7 12.9 14.0 14.2 15.6
- Past users, % 79.1 78.4 77.9 77.5 76.6
- Current users, % 9.1 8.7 8.1 8.3 7.7
Parous, % 80.7 82.4 81.1 78.9 73.1
Parity among parous women 2.2 (0.9) 2.3 (0.9) 2.3 (0.9) 2.2 (0.9) 2.2 (0.9)
Age at first birth among parous women (years) 26.3 (4.6) 26.5 (4.6) 26.5 (4.6) 26.5 (4.6) 26.5 (4.7)
Premenopausal women, % 88.2 89.2 89.7 89.4 88.7
First-degree family history of breast cancer, % 9.6 9.7 9.3 9.3 9.5
Personal history of benign breast disease, % 46.9 45.5 44.7 43.4 43.3

aValues are standardized to the age distribution of the study population. Abbreviations: SD: standard deviation; kg: kilograms; g: grams; m: meters; MET-hr/wk: metabolic equivalent of task-hour/week; kcal/d: kilocalorie per day.

bParticipants recalled their body size at ages 5 and 10 years using a 9-level pictogram (level 1: most lean; level 9: most overweight). We averaged body size at ages 5 and 10 years to obtain an estimate of childhood body size.

cHigher levels (4.5 to 9) were collapsed because there were fewer women in those levels.
<table>
<thead>
<tr>
<th>Average body size (level)</th>
<th>Hazard Ratio (95% Confidence Interval)</th>
<th>Per 1-unit decrease in body size</th>
<th>p-trend'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5 - 2</td>
<td>2.5 - 3</td>
<td>3.5 - 4</td>
</tr>
<tr>
<td>AGES 5-10 YEARS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of cases</td>
<td>503</td>
<td>895</td>
<td>643</td>
</tr>
<tr>
<td>Person-years</td>
<td>168,232</td>
<td>305,585</td>
<td>253,570</td>
</tr>
<tr>
<td>Age-adjusted</td>
<td>1.53</td>
<td>1.57</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>(1.30, 1.81)</td>
<td>(1.34, 1.83)</td>
<td>(1.15, 1.58)</td>
</tr>
<tr>
<td>MV^a -adjusted</td>
<td>1.55</td>
<td>1.56</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>(1.31, 1.83)</td>
<td>(1.33, 1.82)</td>
<td>(1.15, 1.58)</td>
</tr>
<tr>
<td>MV^b without physical activity^c</td>
<td>1.53</td>
<td>1.54</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>(1.30, 1.81)</td>
<td>(1.32, 1.80)</td>
<td>(1.14, 1.57)</td>
</tr>
<tr>
<td>MV^b +</td>
<td>1.63</td>
<td>1.61</td>
<td>1.38</td>
</tr>
<tr>
<td>Menstrual characteristics^d</td>
<td>(1.38, 1.93)</td>
<td>(1.38, 1.89)</td>
<td>(1.17, 1.62)</td>
</tr>
<tr>
<td>MV^b +</td>
<td>1.48</td>
<td>1.49</td>
<td>1.31</td>
</tr>
<tr>
<td>Current BMI^b</td>
<td>(1.24, 1.75)</td>
<td>(1.27, 1.75)</td>
<td>(1.11, 1.54)</td>
</tr>
<tr>
<td>MV^b +</td>
<td>1.38</td>
<td>1.41</td>
<td>1.26</td>
</tr>
<tr>
<td>BMI at age 18</td>
<td>(1.15, 1.65)</td>
<td>(1.19, 1.67)</td>
<td>(1.07, 1.49)</td>
</tr>
<tr>
<td>AGES 10-20 YEARS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of cases</td>
<td>110</td>
<td>768</td>
<td>935</td>
</tr>
<tr>
<td>Person-years</td>
<td>33,311</td>
<td>253,995</td>
<td>348,090</td>
</tr>
<tr>
<td>Age-adjusted</td>
<td>1.64</td>
<td>1.63</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>(1.31, 2.06)</td>
<td>(1.41, 1.88)</td>
<td>(1.29, 1.72)</td>
</tr>
<tr>
<td>MV^a –adjusted</td>
<td>1.64</td>
<td>1.63</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>(1.30, 2.06)</td>
<td>(1.40, 1.89)</td>
<td>(1.29, 1.72)</td>
</tr>
<tr>
<td>MV^b without physical activity^c</td>
<td>1.62</td>
<td>1.60</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>(1.29, 2.03)</td>
<td>(1.38, 1.86)</td>
<td>(1.28, 1.70)</td>
</tr>
</tbody>
</table>
### Table 1.2 (Continued)

<table>
<thead>
<tr>
<th>Average body size (level)(^a)</th>
<th>Hazard Ratio (95% Confidence Interval)</th>
<th>Per 1-unit decrease in body size</th>
<th>p-trend(^f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.74 (1.38, 2.19)</td>
<td>1.53 (1.20, 1.62)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>1.5 - 2</td>
<td>1.70 (1.47, 1.98)</td>
<td>1.53 (1.20, 1.62)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>2.5 - 3</td>
<td>1.53 (1.32, 1.77)</td>
<td>1.40 (1.17, 1.58)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>3.5 - 4</td>
<td>1.40 (1.20, 1.62)</td>
<td>1.00 (1.11, 1.20)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>4.5+</td>
<td>1.00 (1.11, 1.20)</td>
<td>1.12 (1.08, 1.17)</td>
<td>1.00 (Ref)</td>
</tr>
</tbody>
</table>

### Body size (level)\(^a\)

<table>
<thead>
<tr>
<th>Body size (level)(^a)</th>
<th>Hazard Ratio (95% Confidence Interval)</th>
<th>Per 1-unit decrease in body size</th>
<th>p-trend(^f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.54 (1.23, 1.90)</td>
<td>1.32 (1.12, 1.55)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>2</td>
<td>1.56 (1.33, 1.82)</td>
<td>1.40 (1.21, 1.64)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>3</td>
<td>1.44 (1.15, 1.74)</td>
<td>1.34 (1.15, 1.57)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>4</td>
<td>1.47 (1.25, 1.74)</td>
<td>1.30 (1.10, 1.53)</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td>5+</td>
<td>1.44 (1.15, 1.81)</td>
<td>1.30 (1.10, 1.53)</td>
<td>1.00 (Ref)</td>
</tr>
</tbody>
</table>

### AGE 20 YEARS

<table>
<thead>
<tr>
<th>No. of cases</th>
<th>135</th>
<th>778</th>
<th>996</th>
<th>526</th>
<th>206</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person-years</td>
<td>42,612</td>
<td>258,124</td>
<td>378,017</td>
<td>210,564</td>
<td>109,065</td>
</tr>
</tbody>
</table>

| Age-adjusted  | 1.53 (1.23, 1.90) | 1.39 (1.20, 1.62) | 1.32 (1.12, 1.55) | 1.00 (Ref) | 1.10 <0.0001 |
| MV\(^b\) –adjusted | 1.54 (1.34, 1.82) | 1.40 (1.21, 1.64) | 1.33 (1.13, 1.57) | 1.00 (Ref) | 1.10 <0.0001 |
| MV\(^b\) without physical activity\(^c\) | 1.52 (1.22, 1.89) | 1.39 (1.19, 1.62) | 1.33 (1.13, 1.56) | 1.00 (Ref) | 1.10 <0.0001 |
| MV\(^b\) + | 1.61 (1.29, 2.00) | 1.43 (1.23, 1.66) | 1.34 (1.14, 1.58) | 1.00 (Ref) | 1.11 <0.0001 |
| Menstrual characteristics\(^d\) | 1.61 (1.29, 2.00) | 1.43 (1.23, 1.66) | 1.34 (1.14, 1.58) | 1.00 (Ref) | 1.11 <0.0001 |
| MV\(^b\) + current BMI\(^e\) | 1.44 (1.15, 1.81) | 1.34 (1.15, 1.57) | 1.30 (1.10, 1.53) | 1.00 (Ref) | 1.09 <0.0001 |

\(^a\) Participants recalled their body size at ages 5, 10, and 20 years using a 9-level pictogram (level 1: most lean; level 9: most overweight). Higher levels (4.5+ or 5+) were collapsed because there were fewer women in those levels.

\(^b\) Multivariate (MV)-adjusted model includes age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent

\(^c\) Multivariate (MV)-adjusted model includes age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent

\(^d\) Multivariate (MV)-adjusted model includes age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent

\(^e\) Multivariate (MV)-adjusted model includes age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent

\(^f\) Multivariate (MV)-adjusted model includes age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent
Table 1.2 (Continued)

smoking/current with no adolescent smoking), adult alcohol intake (g/d, continuous), family history of breast cancer (yes/no), personal history of benign breast disease (yes/no), adult physical activity (MET-hr/wk, continuous), menopausal status/postmenopausal hormone use (premenopausal/postmenopausal never user/postmenopausal past user/postmenopausal current user/dubious), and age at menopause (year, continuous).

Adolescent physical activity (average MET-hr/wk at ages 12-17 years, continuous) and adult physical activity (MET-hr/wk, continuous).

Menstrual characteristics includes age at menarche (<12/12/13/14/15+ years), menstrual pattern at ages 18-22 years (regular/irregular or no periods), time to menstrual regularity since menarche (<1/1-4/5+ years or never).

Current adult BMI (kg/m$^2$, continuous)

p-trend was estimated using a Wald test for a continuous variable of body size (level 1-9).
Table 1.3. Hazard ratios of breast cancer according to self-reported early-life body size, stratified by adolescent physical activity (average between ages 12-17 years) in the Nurses' Health Study II, 1997-2011

<table>
<thead>
<tr>
<th>Average body size (level)</th>
<th>1-2</th>
<th>2.5-3</th>
<th>3.5-4</th>
<th>4.5+</th>
<th>Per 1-unit decrease in body size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stratified by ‘future’ physical activity:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average childhood body size (ages 5-10 years) and average adolescent physical activity (ages 12-17 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All women</td>
<td>Cases</td>
<td>1,398</td>
<td>643</td>
<td>404</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>HR</td>
<td>1.55</td>
<td>1.35</td>
<td>1.24</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(1.33, 1.81)</td>
<td>(1.15, 1.58)</td>
<td>(1.04, 1.47)</td>
<td>(1.07, 1.14)</td>
</tr>
<tr>
<td>Inactive women:</td>
<td>Cases</td>
<td>431</td>
<td>216</td>
<td>145</td>
<td>76</td>
</tr>
<tr>
<td>&lt;30 MET-hr/wk</td>
<td>HR</td>
<td>1.66</td>
<td>1.45</td>
<td>1.22</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(1.29, 2.12)</td>
<td>(1.11, 1.89)</td>
<td>(0.92, 1.62)</td>
<td>(1.06, 1.19)</td>
</tr>
<tr>
<td>Moderately active women:</td>
<td>Cases</td>
<td>479</td>
<td>229</td>
<td>149</td>
<td>63</td>
</tr>
<tr>
<td>30-59.9 MET-hr/wk</td>
<td>HR</td>
<td>1.59</td>
<td>1.40</td>
<td>1.34</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(1.22, 2.08)</td>
<td>(1.06, 1.86)</td>
<td>(1.00, 1.81)</td>
<td>(1.05, 1.18)</td>
</tr>
<tr>
<td>Active women:</td>
<td>Cases</td>
<td>488</td>
<td>198</td>
<td>110</td>
<td>57</td>
</tr>
<tr>
<td>60+ MET-hr/wk</td>
<td>HR</td>
<td>1.37</td>
<td>1.18</td>
<td>1.09</td>
<td>1.00 (Ref)</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(1.04, 1.81)</td>
<td>(0.88, 1.59)</td>
<td>(0.79, 1.51)</td>
<td>(1.02, 1.15)</td>
</tr>
</tbody>
</table>

**p-interaction** = 0.72

| **Stratified by ‘concurrent’ physical activity:** |       |       |       |       |
| Average adolescent body size (ages 10-20 years) and average adolescent physical activity (ages 12-17 years) |       |       |       |       |
| All women                 | Cases | 878   | 935   | 590   | 238 |
|                           | HR    | 1.63  | 1.49  | 1.38  | 1.00 (Ref) |
|                           | 95% CI| (1.41, 1.88) | (1.29, 1.72) | (1.19, 1.61) | (1.09, 1.18) |
| Inactive women:           | Cases | 264   | 294   | 215   | 95 |
| <30 MET-hr/wk             | HR    | 1.74  | 1.67  | 1.52  | 1.00 (Ref) |
|                           | 95% CI| (1.37, 2.21) | (1.32, 2.11) | (1.19, 1.94) | (1.09, 1.23) |
| Moderately active women:  | Cases | 314   | 323   | 202   | 81 |
| 30-59.9 MET-hr/wk         | HR    | 1.59  | 1.40  | 1.26  | 1.00 (Ref) |
|                           | 95% CI| (1.24, 2.04) | (1.09, 1.79) | (0.97, 1.63) | (1.07, 1.21) |
| Active women:             | Cases | 300   | 318   | 173   | 62 |
| 60+ MET-hr/wk             | HR    | 1.52  | 1.41  | 1.38  | 1.00 (Ref) |
|                           | 95% CI| (1.15, 2.00) | (1.07, 1.86) | (1.03, 1.85) | (1.03, 1.18) |

**p-interaction** = 0.72
Table 1.3 (Continued)

<table>
<thead>
<tr>
<th>Body size (level)</th>
<th>1-2</th>
<th>3</th>
<th>4</th>
<th>5+</th>
<th>Per 1-unit decrease in body size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multivariate hazard ratio (95% confidence interval)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stratified by ‘prior’ physical activity:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Body size at age 20 years and average adolescent physical activity (ages 12-17 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All women</td>
<td>Cases</td>
<td>913</td>
<td>996</td>
<td>526</td>
<td>206</td>
</tr>
<tr>
<td>HR</td>
<td>1.56</td>
<td>1.40</td>
<td>1.33</td>
<td><strong>1.00</strong></td>
<td>1.10</td>
</tr>
<tr>
<td>95% CI</td>
<td>(1.34, 1.82)</td>
<td>(1.21, 1.64)</td>
<td>(1.13, 1.57)</td>
<td><em>(Ref)</em></td>
<td>(1.07, 1.14)</td>
</tr>
<tr>
<td>Inactive women:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30 MET-hr/wk</td>
<td>Cases</td>
<td>271</td>
<td>330</td>
<td>193</td>
<td>74</td>
</tr>
<tr>
<td>HR</td>
<td>1.73</td>
<td>1.72</td>
<td>1.52</td>
<td><strong>1.00</strong></td>
<td>1.12</td>
</tr>
<tr>
<td>95% CI</td>
<td>(1.33, 2.25)</td>
<td>(1.33, 2.22)</td>
<td>(1.16, 1.99)</td>
<td><em>(Ref)</em></td>
<td>(1.06, 1.19)</td>
</tr>
<tr>
<td>Moderately active women:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-59.9 MET-hr/wk</td>
<td>Cases</td>
<td>331</td>
<td>337</td>
<td>176</td>
<td>76</td>
</tr>
<tr>
<td>HR</td>
<td>1.39</td>
<td>1.17</td>
<td>1.11</td>
<td><strong>1.00</strong></td>
<td>1.10</td>
</tr>
<tr>
<td>95% CI</td>
<td>(1.08, 1.79)</td>
<td>(0.91, 1.51)</td>
<td>(0.85, 1.46)</td>
<td><em>(Ref)</em></td>
<td>(1.04, 1.17)</td>
</tr>
<tr>
<td>Active women:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60+ MET-hr/wk</td>
<td>Cases</td>
<td>311</td>
<td>329</td>
<td>157</td>
<td>56</td>
</tr>
<tr>
<td>HR</td>
<td>1.48</td>
<td>1.32</td>
<td>1.37</td>
<td><strong>1.00</strong></td>
<td>1.08</td>
</tr>
<tr>
<td>95% CI</td>
<td>(1.11, 1.98)</td>
<td>(0.99, 1.76)</td>
<td>(1.00, 1.86)</td>
<td><em>(Ref)</em></td>
<td>(1.01, 1.15)</td>
</tr>
</tbody>
</table>
| **p-interaction**<sup>c</sup> = 0.64

<sup>a</sup> Participants recalled their body size at ages 5, 10, and 20 years using a 9-level pictogram (level 1: most lean; level 9: most overweight). Higher levels (4.5+ or 5+) were collapsed because there were fewer women in those levels.

<sup>b</sup> Adjusted for age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent smoking/current with no adolescent smoking), adult alcohol intake (g/d, continuous), family history of breast cancer (yes/no), personal history of benign breast disease (yes/no), adult physical activity (MET-hr/wk, continuous), menopausal status/postmenopausal hormone use (premenopausal/postmenopausal never user/postmenopausal past user/postmenopausal current user/dubious), and age at menopause (year, continuous).

<sup>c</sup> p-interaction was estimated using a likelihood ratio test comparing models with and without an interaction term (between a continuous body size level 1-9 and a continuous physical activity MET-hr/wk).
Figure 1.1. Hazard ratios\textsuperscript{a} of breast cancer according to self-reported adolescent physical activity (average between ages 12-17 years) and body size\textsuperscript{b} at three different age periods (ages 5-10, 10-20, and 20 years) in the Nurses’ Health Study II, 1997-2011 (reference = body size 4.5+ or 5+ level, physical activity <30 MET-hr/wk)
Figure 1.1 (Continued)

*Statistically significant (p<0.05)

*Adjusted for age (years, continuous), average alcohol intake between ages 15-18 years (g/day, continuous), average physical activity between ages 12-17 years (MET-hr/wk, continuous), birthweight (<5.5/5.5-6.9/7-8.4/8.5+), height (inches, continuous), parity/age at first birth (nulliparous/1-2 at < 25 years/1-2 at 25-29 years/1-2 at 30+ years/3+ at <25 years/3+ at 25+ years), ever oral contraceptive use (never/past/current), smoking status (never/past with adolescent smoking/past with no adolescent smoking/current with adolescent smoking/current with no adolescent smoking), adult alcohol intake (g/d, continuous), family history of breast cancer (yes/no), personal history of benign breast disease (yes/no), adult physical activity (MET-hr/wk, continuous), menopausal status/postmenopausal hormone use (premenopausal/postmenopausal never user/postmenopausal past user/postmenopausal current user/dubious), and age at menopause (year, continuous).

b Participants recalled their body size at ages 5, 10, and 20 years using a 9-level pictogram (level 1: most lean; level 9: most overweight). Higher levels (4.5+ or 5+) were collapsed because there were fewer women in those levels.
REFERENCES


CHAPTER 2
Dietary macronutrient and fiber intake and patterns of estrogen metabolism in
premenopausal women

Hannah Oh¹, ², ³
Stephanie A. Smith-Warner¹, ³
Rulla M. Tamimi¹, ²
Molin Wang¹, ², ⁴
Xia Xu⁵
Susan E. Hankinson¹, ², ⁶
Barbara J. Fuhrman⁷, ⁸
Regina G. Ziegler*⁸
A. Heather Eliassen*¹, ²

¹Department of Epidemiology, Harvard School of Public Health, Boston, MA
²Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical
School, Boston, MA
³Department of Nutrition, Harvard School of Public Health, Boston, MA
⁴Department of Biostatistics, Harvard School of Public Health, Boston, MA
⁵Cancer Research Technology Program, Leidos Biomedical Research, Inc., Frederick National
Laboratory for Cancer Research, Frederick, MD
⁶Division of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA
⁷Department of Epidemiology, Fay W. Boozman College of Public Health, University of
Arkansas for Medical Sciences, Little Rock, AR
⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD

*Co-last authors
ABSTRACT

Background: Patterns of estrogen synthesis, metabolism, and excretion affect the bioavailability of potentially carcinogenic estrogen and estrogen metabolites; thus, factors that alter the pattern of estrogen metabolism may be important. Dietary intake of fiber and macronutrients may be associated with patterns of estrogen metabolism.

Methods: We conducted a cross-sectional analysis among 598 premenopausal women who participated in a reproducibility study (n=109) or served as controls in a nested case-control study of breast cancer (n=489) within the Nurses' Health Study II. Dietary intakes of fiber, fat, carbohydrate, and protein were assessed via semi-quantitative food frequency questionnaires in 1995 and 1999. Mid-luteal urine samples were collected between 1996-1999 and estrogen metabolites (EM) were quantified using a high-performance liquid chromatography-tandem mass spectrometry. Linear mixed models were used to estimate creatinine-adjusted geometric means for individual EM and their pathway groups across categories of dietary intake while controlling for total energy intake and potential confounders.

Results: Higher total dietary fiber intake (>25 vs. ≤ 15 g/day) was associated with significantly higher levels of 4-methoxyestradiol (55% difference, p-difference=0.01, p-trend=0.004) and lower levels of 17-epiestriol (-27% difference, p-difference=0.03, p-trend=0.03), but was not associated with any other EM. The associations did not vary by fiber intake from different sources. Total fat intake was not significantly associated with any individual EM but was suggestively positively associated with 17-epiestriol. Polyunsaturated and trans fat intakes were significantly positively associated with 17-
epiestriol. No patterns of metabolism were observed with either carbohydrate or protein intake.

**Conclusion:** Fiber and macronutrient intakes were not significantly associated with patterns of estrogen metabolism.
INTRODUCTION

Higher exposure to endogenous estrogens and estrogen metabolites (EM) are hypothesized to increase the risk of several hormone-related cancers (e.g., breast, ovarian, endometrial cancer) [1, 2]. Cumulative evidence suggests genotoxic (i.e., DNA damage) and estrogenic (i.e., increased proliferation) effects of EM may be important in the etiology of such cancers [1, 3]. Parent estrogens (estrone and estradiol) are irreversibly hydroxylated at three different carbon positions of the steroid ring, resulting in EM of three different pathways (i.e., 2-, 4-, 16-hydroxylation pathways). The degree of genotoxicity and estrogenicity of the EM vary by their pathways of metabolism and their methylation status [1, 4]. For example, 4-hydroxy metabolites were suggested to be more estrogenic than 2-hydroxy metabolites because 2-hydroxy metabolites have a faster rate of dissociation from estrogen receptors than 4-hydroxy metabolites [5]. When parent estrogens are hydroxylated at the 2- or 4-carbon positions, they become catechols that can induce DNA damage directly by forming quinone DNA adducts and indirectly via redox cycling [3, 5-7]. However, methylation by catechol-o-methyl transferase (COMT) stabilizes these metabolites so they do not undergo further redox cycling [3]. Some studies have shown that 16-pathway EM may be more carcinogenic than 2-pathway EM [8, 9] but such a difference has not been observed in other studies [8, 10-12].

Patterns of estrogen synthesis, metabolism, and excretion altogether affect the bioavailability of estrogens and EM. Metabolism of parent estrogens determines the production of EM in the body; conjugation of estrogens, allowing water-solubility, increases the body’s ability to excrete the metabolites. Therefore, factors that affect the
pattern of estrogen metabolism may be important. We previously reported that several risk factors for breast cancer (e.g., irregular menstrual cycles, menstrual cycle length, physical activity, height, adult BMI, smoking, and analgesic use) were associated with urinary EM in premenopausal women [13-17]. Dietary intake of fiber and macronutrients also may be associated with the patterns of estrogen metabolism.

Although we previously did not find an association of dietary fiber, fat, and carbohydrate intake with the levels of premenopausal plasma sex hormones [18], other studies suggest that a low-fat, high-fiber diet may reduce endogenous estrogen levels by increasing their fecal excretion [19-21]. The associations are somewhat inconsistent for urinary estrogen excretion [19, 21-23]. Aubertin-Leheudre et al. reported lower follicular levels of urinary 2-hydroxyestradiol, 16α-hydroxyestrone, and 2-hydroxyestrone/4-hydroxyestrone ratio among twenty-two premenopausal women on a low-fat, high-fiber diet compared to another twenty-two women on a high fat, low-fiber diet [23]. However, little is known about the association between diet and estrogen metabolism beyond these metabolites. With a recent advance in technology, we were able to measure 15 different individual estrogens and EM in urine using high-performance liquid chromatography-tandem mass spectrometry (LC/MS-MS) [24]. No study to date, to our knowledge, has examined the association of each macronutrient and fiber intake with detailed estrogen metabolism using a large sample size. In a cross-sectional analysis within the Nurses’ Health Study II (NHSII) cohort, we examined the relationship of dietary fiber, fat, carbohydrate and protein intake with the luteal levels of 15 urinary EM in premenopausal women.
MATERIALS AND METHODS

Study population

The NHSII is an ongoing cohort study that began in 1989, including 116,430 female registered nurses aged 25-42 years. An initial mailed, self-administered questionnaire collected information on participants’ health behaviors, lifestyle factors, reproductive factors, and medical histories. Biennial follow-up questionnaires assessed updated information on a variety of known and suspected risk factors for chronic diseases, as well as new disease diagnoses. Response rates are >90% based on cumulative person-time of follow-up.

Between 1996 and 1999, 29,611 healthy women provided blood and urine samples. Among them, 18,521 women were premenopausal, had not used oral contraceptives, had not been pregnant or breastfed during the past 6 months, and provided samples timed within their menstrual cycles (early follicular blood and mid-luteal blood and urine samples). Details of the blood and urine collection have been published previously [10, 25]. Urine samples were collected 7-9 days before the anticipated start of the next menstrual cycle and shipped overnight to our laboratory with an ice pack. To allow accurate counting of luteal day of collection, participants were asked to return postcards indicating the start date of their next menstrual cycle. Approximately, 93% of samples were received within 26 hours of collection. Samples have been stored in liquid nitrogen freezers since collection.

Among the 18,521 premenopausal women who provided timed, luteal urine samples, a random sample of 412 women were invited to participate in the Within Person Stability Study and asked to provide a second and a third set of samples over a 2-3 year period. A nested case-control study was also conducted among women who
gave timed samples. Two controls were selected for each case and were matched on age, menopausal status at diagnosis, month and year of collection, ethnicity, and luteal day. Our analyses included participants from the reproducibility study [26] (first urine sample; N=109) and controls from the nested case-control study (N=489) [10]. Women were excluded if they had missing urinary creatinine measures (n=9) or had total energy intake <600 or >3500 kcal/d based on the 1995 and 1999 Food Frequency Questionnaires (FFQs) (n=5). A total of 598 women were included in the analyses.

This investigation was approved by the Institutional Review Board of the Brigham and Women’s Hospital.

**Exposure assessment**

Women were asked to recall their previous-year average intake of 133 food items via semi-quantitative FFQs every four years since 1991. Women reported their frequency of each food item ranging from never to more than 6 times per day. Using the frequency information and specified portion size on the FFQs, average daily nutrient intake (e.g., dietary fiber, fat, carbohydrates, protein) of each participant was calculated based on the USDA database [27]. The FFQs were validated in randomly selected women from the NHS cohort in 1980-1981 and 1986-1987 [28, 29]. When women were asked to record their intake via two one-week diet records throughout the year in 1986 (the same period as their FFQ’s), their average energy-adjusted carbohydrate, total fat, and protein intake from diet records were fairly well correlated with the estimates from the FFQs (r=0.64, 0.57, 0.50, respectively) [29].
Average energy-adjusted intake of fiber (total as well as fiber from fruits, cereals, vegetables, legumes) in grams per day and average percentage of energy from dietary fat (total, saturated, polyunsaturated, monounsaturated, trans), carbohydrates (total, glycemic load), and protein (total, animal, dairy, vegetable) from the 1995 and 1999 FFQs were used in the analyses to reduce measurement error and to represent the period closest to the date of urine collection. For each nutrient, we categorized participants into four groups using the near-quartile whole number cutpoints.

Covariate assessment

Covariate information was collected by the questionnaire assessed at urine collection in 1996-1999 or the biennial questionnaires in 1997-2001. Information on first morning urine (yes/no), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), anovulatory cycle (luteal plasma progesterone < 400 ng/dl; yes/no), age at urine collection (years; continuous), and BMI at urine collection (kg/m^2; continuous) were reported on the questionnaire returned with the blood and urine samples. Height (inches, continuous) was assessed via the baseline questionnaire in 1989. Alcohol consumption (average of 1995 and 1999 FFQs; <3 drinks/month, 3 drinks/month to 2 drinks/wk, 3-5 drinks/wk, >5 drinks/wk), physical activity (average MET-hr/wk in 1997 and 2001; continuous), menstrual cycle length (<26, 26-31, 31+ days) and pattern (extremely regular/very regular/regular/usually or always irregular), parity/age at first birth (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), total energy intake (average kcal/d in 1995 and 1999 FFQs; continuous), and smoking (yes/no) were
reported on biennial questionnaires. Age at menarche was not included in our multivariate models because we previously reported no association between age at menarche and urinary estrogen metabolites in premenopausal women [14].

**Laboratory assays**

Urine concentrations of EM were assayed at the Laboratory of Proteomics and Analytical Technology (Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD) using LC/MS-MS, which allowed concurrent measurement of 15 individual estrogens and metabolites with high sensitivity, specificity and reproducibility [24, 30]. LC-MS/MS was performed with a TSQ Quantum-AM triple quadruple mass spectrometer coupled with a Surveyor high performance liquid chromatography system (Thermo, San Jose, CA), both of which were controlled by Xcalibur software (Thermo). Details of the assay were published previously [10, 24, 31]. Masked quality control samples were placed in each batch to assess laboratory variability. Coefficients of variation (CV) were <7% for all EM except for 4-methoxyestrone (17%) and 4-methoxyestradiol (15%). Our reproducibility study of urinary EM showed good reproducibility over 2-3 years (r=0.52-0.72) [26].

Plasma progesterone, used to identify ovulatory cycles, was measured at Quest Diagnostics-Nichols Institute (San Juan Capistrano, CA) and the Royal Marsden Hospital (London, U.K.) with CVs <17%. Urinary creatinine was measured in 3 different laboratories (Endocrine Core Laboratory at Emory University, Atlanta, GA; Boston Children’s Hospital, Boston, MA; Brigham and Women’s Hospital, Boston, MA) with overall CVs <9.2%.
All individual urinary EM were adjusted for creatinine, resulting in units of pmol/mg creatinine, because the raw EM levels may vary simply due to the variations in urine concentration. Analyses were performed on 15 individual EM, as well as on metabolic pathways, pathway ratios, catechol-to-methylated catechol ratios, parent estrogens-to-total EM ratio, and pathway-to-parent estrogens ratios.

**Statistical analyses**

Each individual EM was log-transformed to improve normality. Outlying values (up to 16 values) were identified and excluded using the extreme Studentized deviate many-outlier procedure [32]. To account for correlations among the controls within the matched sets, linear mixed models, with robust variance, were used to estimate multivariate-adjusted geometric means (and 95% confidence intervals) of EM by categories of dietary fiber, fat, carbohydrate, and protein intake. Because urinary EM levels were modeled on the natural logarithm scale, the regression coefficients for exposure were exponentiated for presentation in the original scale. We performed a test for trend by including exposures (dietary intake of each nutrient) in the model as a continuous variable using the median of each category. Because we adjusted for total energy and alcohol intake in the model, the interpretation of regression coefficients for macronutrients, for example fat, is the difference in EM when replacing a percentage of energy from fat by the equal amount of energy from other macronutrients (carbohydrate and protein).

We tested for interactions between each nutrient and other lifestyle factors: BMI (≥25 vs. <25 kg/m²) and physical activity (≥ vs. < median 15.7 MET-hr/wk) using
likelihood ratio tests. For sensitivity analyses, we repeated the same analyses after restricting to women who provided samples during an ovulatory cycle (luteal plasma progesterone > 400 ng/dl) (N=538), women who did not become postmenopausal within 4 years after urine collection (N=524), women within 3-10 luteal days (N=534), and women with first morning urine (N=468), separately. We found similar estimates from models using different ways of adjustment for creatinine (creatinine-standardized EM vs. adjustment for creatinine as a covariate in the model); thus, only the creatinine-standardized results are presented. All statistical tests were two-sided with 5 % type I error. Analyses were conducted with SAS version 9 (SAS Institute).

RESULTS

The mean age at urine collection was 43 years for 598 women who were included in our analysis. Eighty percent of study participants provided a first morning urine sample; 90% of study participants provided a sample during an ovulatory cycle. Compared to women with a lower level of total dietary fiber intake, women with a higher level of fiber intake were older and engaged in higher levels of physical activity at urine collection (Table 2.1). Women with higher fiber intake were also less likely to be parous, current smokers, and have irregular menstrual cycles. These women were also more likely to have used oral contraceptives in the past. Compared to women with lower energy intake from total fat, those with higher intake had higher BMI at urine collection, were younger at first birth, had lower levels of physical activity, and were more likely to be current smokers and have irregular menstrual cycles.
Compared to women who reported ≤ 15 g/day energy-adjusted total fiber intake, women with >25 g/d total fiber intake had significantly higher levels of 4-methoxyestradiol (0.06 vs. 0.04 pmol/mg creatinine, p-difference=0.01, p-trend=0.004), significantly lower levels of 17-epiestriol (1.24 vs. 1.69 pmol/mg creatinine, p-difference=0.03, p-trend=0.03) and estriol (25.9 vs. 30.2 pmol/mg creatinine, p-difference=0.15, p-trend=0.03), and suggestively higher levels of estrone and 2-catechols (Table 2.2). Following these associations, total fiber intake was positively associated with the EM ratios that included 16-pathway EM in the denominator (e.g., catechols/16-pathway, 2-pathway/16-pathway) and inversely associated with the ratios that included 2-catechols in the denominator (e.g., 4-catechols/2-catechols). We observed a similar pattern of association with fiber intake from different sources, although the associations with 2-catechols was statistically significant for fiber intake from cereals and the association with estrone was statistically significant for fiber intake from vegetables and cereals (Supplementary table 2.S1). Only fiber intake from fruits was significantly inversely associated with 17-epiestriol.

We did not observe significant associations between total dietary fat intake and EM, although there was a suggestively positive association with 16-pathway EM, 17-epiestriol (>35 vs. ≤ 25 % energy: 1.77 vs. 1.44 pmol/mg creatinine, p-difference=0.14, p-trend=0.06) (Table 2.3). When we examined specific types of fat, we observed significantly positive associations of 17-epiestriol with polyunsaturated fat (PUFA) (>6 vs. ≤ 4.5 % energy: 37% difference, p-difference=0.01, p-trend=0.01) and trans-fat intakes (>1.75 vs. ≤ 1.15 % energy: 36% difference, p-difference=0.01, p-trend=0.01) (Supplementary table 2.S2), that were slightly stronger than the association observed
with total fat intake. Higher trans-fat intake also was significantly associated with lower levels of 4-hydroxyestrone (-24% difference, p-difference=0.01, p-trend=0.03) and a significantly higher level of 16α-hydroxyestrone (31% difference, p-difference=0.01, p-trend=0.01). Saturated fat and monounsaturated fat (MUFA) intake did not show any significant associations with EM.

We observed no clear associations between dietary carbohydrate intake and urinary EM levels (Table 2.4). There were no differences in associations by glycemic load (Supplementary table 2.S3). Protein intake (total, as well as intake from animal, dairy, and vegetable sources) also was not associated with EM (Table 2.5, Supplementary table 2.S4).

The associations of fiber, fat, carbohydrate, and protein intake with EM did not vary by BMI and physical activity (data not shown). Results were unchanged in sensitivity analyses restricting to women who provided samples during an ovulatory cycle, women who did not become postmenopausal within 4 years after urine collection, women within 3-10 luteal days, and women with first morning urine (data not shown).

**DISCUSSION**

In this cross-sectional study, we found few consistent associations of macronutrient and dietary fiber intakes with urinary EM levels among premenopausal women. A high-fiber diet was positively associated with 4-methoxyestradiol and inversely associated with 17-epiestriol. Total fat, carbohydrate, and protein intake were not significantly associated with any EM, although significant associations were
observed for a few specific subtypes of fat. Results were similar by BMI and physical activity.

To date, there have been only a few small studies of diet and EM in premenopausal women. Earlier studies compared proxy groups such as vegetarians vs. omnivores (n=20) [19] and Caucasian vs. Asian immigrants (n=24) [21], with an assumption that their usual diets represented low-fat, high-fiber diets and high-fat, low-fiber diets, respectively; significantly lower premenopausal urinary excretion of parent estrogens (-60% difference) in Asian women [21] and estriol (-37% difference) in vegetarians were observed compared to Caucasian women and omnivores, respectively [19]. In a recent study that examined follicular levels of urinary parent estrogens and 12 EM (n=44), a low-fat, high-fiber diet (< 21% energy from total fat and > 21g/day total fiber intake) was significantly associated with lower levels of estriol (-27% difference), 2-hydroxyestradiol (-41% difference), and 2-hydroxyestrone/4-hydroxyestrenone ratio (-29% difference) compared to a high-fat, low-fiber diet [23]. Although these studies were very small and unable to separate the differences in EM levels due to high fiber intake from those due to low fat intake, our results are consistent in that we observed significantly lower levels of estriol and 17-epiestriol among women with a higher fiber intake and suggestively lower levels of these 2 EM among women with a lower fat intake. Interestingly, urinary 17-epiestriol was the only urinary EM positively associated with the risk of breast cancer in this cohort [10].

To our knowledge, this study is the first to examine the association of each macronutrient and fiber with 15 individual luteal EM in premenopausal women. We previously reported no association between diet and plasma follicular and luteal sex
hormones in premenopausal women [18]; we also observed few significant associations in the current study of luteal urinary EM although plasma estrogens and urinary EM are only modestly correlated (r=0.26-0.33) [26].

The biological mechanisms for potential associations between dietary factors and urinary EM are not well understood. Fiber has the ability to bind enteral carcinogens and circulating estrogens [33], thereby facilitating elimination of estrogens from the body. Components of cruciferous vegetables may induce hepatic CYP1A1 enzymatic activity that is involved in 2-hydroxylation of parent estrogens [34, 35]. Dietary fat and carbohydrate intake that influence blood cholesterol and insulin levels may contribute to sex steroid hormones synthesis and balance [36, 37]. However, we only found significant associations with fiber intake.

We acknowledge several limitations of this study. First, with our cross-sectional analysis, we cannot determine the temporal relationship between the exposure and the outcome. Second, we used only one time urine collection per participant; however, our reproducibility study showed urinary EM, except methylated 4-catechols which have the lowest absolute concentrations, have high intraclass correlation coefficients (ICC range 0.4-0.7) over three years indicating that a single measurement adequately represents exposures over time. Furthermore, with multiple comparisons, positive findings in our analyses could be due to chance. None of our results remained significant after a Bonferroni correction (corrected alpha=0.003), although the Bonferroni method likely is too conservative in our study given the correlated outcomes (r= -0.002-0.85). We have interpreted our results with caution and this study can serve as a hypothesis-generating analysis that can provide a basis for future study. Lastly, the majority of our study
population was white women in the US, where fat intake is generally high and fiber intake is generally low (e.g., average 33% of energy intake from total fat in the US vs. 20% in Japan) [38, 39]. A small difference in EM between the highest vs. lowest intake groups may be due to insufficient variations in dietary intake (e.g., total fiber: >25 vs. ≤ 15 g/d; total fat: >35 vs. ≤ 25 % energy) in our study population and may not translate into a clinically relevant difference in risk. Additionally, the study results may not be generalizable to other populations with lower dietary intake of fat or higher intake of fiber.

Despite the limitations, the proposed study has several strengths. High performance LC/MS-MS assay allowed comprehensive evaluations of 15 individual EM and their pathways. By using timed luteal samples, the study allowed more accurate estimates of association in premenopausal women because levels of hormones vary widely over the menstrual cycle. A careful adjustment for total energy and potential confounders assessed at urine collection allowed an isocaloric interpretation of the results and increased the validity of the associations.

In this comprehensive analysis of dietary intake and urinary EM, we did not observe significant associations between macronutrients and urinary patterns of estrogen metabolism. We observed suggested associations between dietary fiber and 17-epiestriol, but further study is required to confirm these results.
Table 2.1. Characteristics of the study population by average dietary fiber and total fat intake at urine collection in the Nurses’ Health Study II

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Average energy-adjusted dietary fiber intake (g/d)</th>
<th>Mean (SD) or Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 15</td>
<td>15.1-20</td>
</tr>
<tr>
<td>N</td>
<td>109</td>
<td>224</td>
</tr>
<tr>
<td>Age, years</td>
<td>41.9 (4.2)</td>
<td>43.3 (3.6)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>24.6 (4.2)</td>
<td>25.1 (4.5)</td>
</tr>
<tr>
<td>Height, inches</td>
<td>64.8 (1.8)</td>
<td>65.2 (2.5)</td>
</tr>
<tr>
<td>Age at menarche, years</td>
<td>12.7 (1.0)</td>
<td>12.8 (1.4)</td>
</tr>
<tr>
<td>Parous, %</td>
<td>85.4</td>
<td>83.9</td>
</tr>
<tr>
<td>Parity (among parous women)</td>
<td>2.4 (0.6)</td>
<td>2.5 (1.0)</td>
</tr>
<tr>
<td>Age at first birth, years</td>
<td>25.8 (3.3)</td>
<td>26.6 (4.1)</td>
</tr>
<tr>
<td>Physical activity, MET-hr/week</td>
<td>15.8 (14.2)</td>
<td>19.7 (15.5)</td>
</tr>
<tr>
<td>Alcohol drinker, %</td>
<td>61.1</td>
<td>77.5</td>
</tr>
<tr>
<td>Alcohol intake (among drinkers), g/d</td>
<td>4.4 (4.3)</td>
<td>5.8 (7.3)</td>
</tr>
<tr>
<td>Total energy intake, kcal/d</td>
<td>1719 (336)</td>
<td>1900 (417)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage of energy intake from total fat (%)</th>
<th>≤ 25</th>
<th>25.1-30</th>
<th>30.1-35</th>
<th>&gt; 35</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>132</td>
<td>201</td>
<td>165</td>
<td>100</td>
</tr>
<tr>
<td>Age, years</td>
<td>43.4 (3.8)</td>
<td>42.7 (3.8)</td>
<td>42.9 (3.8)</td>
<td>42.5 (3.9)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>24.9 (3.7)</td>
<td>24.2 (3.6)</td>
<td>25.1 (4.7)</td>
<td>26.4 (4.7)</td>
</tr>
<tr>
<td>Height, inches</td>
<td>64.8 (1.9)</td>
<td>65.3 (2.1)</td>
<td>65.3 (2.4)</td>
<td>65.1 (2.1)</td>
</tr>
<tr>
<td>Age at menarche, years</td>
<td>12.3 (0.9)</td>
<td>12.6 (1.3)</td>
<td>12.7 (1.3)</td>
<td>12.7 (1.1)</td>
</tr>
<tr>
<td>Parous, %</td>
<td>78.2</td>
<td>78.3</td>
<td>89.4</td>
<td>81.9</td>
</tr>
<tr>
<td>Parity (among parous women)</td>
<td>2.3 (0.6)</td>
<td>2.4 (1.0)</td>
<td>2.4 (0.9)</td>
<td>2.2 (0.6)</td>
</tr>
<tr>
<td>Age at first birth, years</td>
<td>27.0 (3.8)</td>
<td>27.0 (3.9)</td>
<td>26.2 (3.7)</td>
<td>25.7 (3.7)</td>
</tr>
<tr>
<td>Physical activity, MET-hr/week</td>
<td>24.5 (19.6)</td>
<td>23.4 (18.6)</td>
<td>19.1 (13.7)</td>
<td>17.7 (16.6)</td>
</tr>
<tr>
<td>Alcohol drinker, %</td>
<td>75.1</td>
<td>75.4</td>
<td>72.8</td>
<td>69.2</td>
</tr>
<tr>
<td>Alcohol intake (among drinkers), g/d</td>
<td>5.1 (4.4)</td>
<td>6.4 (8.1)</td>
<td>4.9 (5.0)</td>
<td>3.5 (3.0)</td>
</tr>
<tr>
<td>Total energy intake, kcal/d</td>
<td>1807 (368)</td>
<td>1898 (397)</td>
<td>1954 (418)</td>
<td>1670 (347)</td>
</tr>
<tr>
<td>Past oral contraceptive use, %</td>
<td>83.2</td>
<td>84.2</td>
<td>90.1</td>
<td>79.9</td>
</tr>
<tr>
<td>Current smoker, %</td>
<td>1.7</td>
<td>7.2</td>
<td>6.7</td>
<td>11.9</td>
</tr>
<tr>
<td>Menstrual cycle length 26-31 days, %</td>
<td>73.7</td>
<td>68.5</td>
<td>71.9</td>
<td>67.1</td>
</tr>
<tr>
<td>Irregular menstrual cycles, %</td>
<td>1.4</td>
<td>4.1</td>
<td>6.0</td>
<td>6.1</td>
</tr>
<tr>
<td>First morning urine, %</td>
<td>80.6</td>
<td>82.6</td>
<td>80.0</td>
<td>79.0</td>
</tr>
<tr>
<td>Luteal day outside 3-10 days before next period, %</td>
<td>12.7</td>
<td>7.7</td>
<td>13.4</td>
<td>6.7</td>
</tr>
<tr>
<td>Anovulatory cycle, %</td>
<td>10.2</td>
<td>7.5</td>
<td>10.3</td>
<td>7.2</td>
</tr>
</tbody>
</table>

Values are means (SD) or percentages and are standardized to the age distribution of the study population.

a Average of 1997 and 2001 activities.

b Average of 1995 and 1999 intakes.

c Usually or always irregular menstrual cycles: more than 7 days deviation from expected.
Table 2.2. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of total fiber in 1995 and 1999: Nurses’ Health Study II (n = 598)\(^a\)

<table>
<thead>
<tr>
<th>Energy-adjusted intake of fiber (g/d)</th>
<th>≤ 15</th>
<th>15.1-20</th>
<th>20.1-25</th>
<th>&gt;25</th>
<th>% diff(^b)</th>
<th>p-diff(^c)</th>
<th>p-trend(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>109</td>
<td>224</td>
<td>166</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total estrogens and estrogen metabolites</td>
<td>190</td>
<td>197</td>
<td>207</td>
<td>195</td>
<td>2.4</td>
<td>0.74</td>
<td>0.61</td>
</tr>
<tr>
<td>Parent estrogens</td>
<td>38.0</td>
<td>38.4</td>
<td>43.1</td>
<td>41.3</td>
<td>8.7</td>
<td>0.33</td>
<td>0.12</td>
</tr>
<tr>
<td>Estrone</td>
<td>25.0</td>
<td>26.5</td>
<td>28.7</td>
<td>28.1</td>
<td>12.7</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td>Estradiol</td>
<td>13.1</td>
<td>13.0</td>
<td>14.0</td>
<td>13.0</td>
<td>-0.5</td>
<td>0.95</td>
<td>0.78</td>
</tr>
<tr>
<td>Catechols</td>
<td>59.3</td>
<td>58.8</td>
<td>63.1</td>
<td>66.1</td>
<td>11.4</td>
<td>0.33</td>
<td>0.18</td>
</tr>
<tr>
<td>2-Catechols</td>
<td>51.7</td>
<td>50.3</td>
<td>55.4</td>
<td>58.3</td>
<td>12.7</td>
<td>0.30</td>
<td>0.12</td>
</tr>
<tr>
<td>2-Hydroxyestrone</td>
<td>45.9</td>
<td>44.4</td>
<td>49.0</td>
<td>52.1</td>
<td>13.5</td>
<td>0.28</td>
<td>0.11</td>
</tr>
<tr>
<td>2-Hydroxyestradiol</td>
<td>5.17</td>
<td>5.05</td>
<td>5.82</td>
<td>5.77</td>
<td>11.7</td>
<td>0.33</td>
<td>0.11</td>
</tr>
<tr>
<td>4-Catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxysterone</td>
<td>5.81</td>
<td>5.68</td>
<td>5.92</td>
<td>6.20</td>
<td>6.7</td>
<td>0.64</td>
<td>0.52</td>
</tr>
<tr>
<td>Methyalted catechols</td>
<td>10.3</td>
<td>9.21</td>
<td>10.8</td>
<td>10.2</td>
<td>-0.7</td>
<td>0.95</td>
<td>0.39</td>
</tr>
<tr>
<td>Methyalted 2-catechols</td>
<td>9.98</td>
<td>8.85</td>
<td>10.5</td>
<td>9.96</td>
<td>-0.2</td>
<td>0.99</td>
<td>0.33</td>
</tr>
<tr>
<td>2-Methoxyestrone</td>
<td>7.75</td>
<td>7.07</td>
<td>8.13</td>
<td>7.80</td>
<td>0.6</td>
<td>0.96</td>
<td>0.41</td>
</tr>
<tr>
<td>2-Methoxyestradiol</td>
<td>0.67</td>
<td>0.69</td>
<td>0.73</td>
<td>0.74</td>
<td>11.1</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>2-Hydroxyestrone-3-methyl ether</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyalted 4-catechols</td>
<td>1.20</td>
<td>1.06</td>
<td>1.23</td>
<td>1.09</td>
<td>-8.8</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>4-Methoxyestrone</td>
<td>0.18</td>
<td>0.19</td>
<td>0.20</td>
<td>0.20</td>
<td>10.0</td>
<td>0.51</td>
<td>0.33</td>
</tr>
<tr>
<td>4-Methoxyestradiol</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>-2.5</td>
<td>0.87</td>
<td>0.93</td>
</tr>
<tr>
<td>2-Hydroxylation pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxylation pathway</td>
<td>63.0</td>
<td>60.5</td>
<td>68.2</td>
<td>69.7</td>
<td>10.5</td>
<td>0.35</td>
<td>0.12</td>
</tr>
<tr>
<td>16-Hydroxylation pathway</td>
<td>65.2</td>
<td>71.4</td>
<td>69.2</td>
<td>59.5</td>
<td>-8.8</td>
<td>0.32</td>
<td>0.17</td>
</tr>
<tr>
<td>16α-Hydroxyestrone</td>
<td>11.6</td>
<td>11.9</td>
<td>12.8</td>
<td>9.97</td>
<td>-13.7</td>
<td>0.23</td>
<td>0.30</td>
</tr>
<tr>
<td>Estriol</td>
<td>30.2</td>
<td>32.9</td>
<td>29.1</td>
<td>25.9</td>
<td>-14.5</td>
<td>0.15</td>
<td>0.03</td>
</tr>
<tr>
<td>17-Epistriol</td>
<td>1.69</td>
<td>1.58</td>
<td>1.54</td>
<td>1.24</td>
<td>-27.1</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>16-Ketoestradiol</td>
<td>13.0</td>
<td>14.5</td>
<td>14.5</td>
<td>13.1</td>
<td>0.8</td>
<td>0.92</td>
<td>0.77</td>
</tr>
<tr>
<td>16-Epistriol</td>
<td>6.10</td>
<td>6.69</td>
<td>6.43</td>
<td>5.81</td>
<td>-4.7</td>
<td>0.56</td>
<td>0.29</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td>0.12</td>
<td>0.12</td>
<td>0.10</td>
<td>0.09</td>
<td>-24.5</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>2-Catechols/16-pathway</td>
<td>0.78</td>
<td>0.70</td>
<td>0.82</td>
<td>0.97</td>
<td>23.1</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Catechols/16-pathway</td>
<td>0.87</td>
<td>0.82</td>
<td>0.94</td>
<td>1.09</td>
<td>26.0</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.09</td>
<td>-18.2</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>0.93</td>
<td>0.85</td>
<td>1.02</td>
<td>1.16</td>
<td>24.3</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>4.2</td>
<td>0.78</td>
<td>0.63</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>1.04</td>
<td>0.98</td>
<td>1.14</td>
<td>1.29</td>
<td>23.5</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>2-pathway/4,16-pathway</td>
<td>0.83</td>
<td>0.74</td>
<td>0.88</td>
<td>0.99</td>
<td>19.2</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>2-Catechols/methyalted 2-catechols</td>
<td>5.06</td>
<td>5.54</td>
<td>5.45</td>
<td>5.80</td>
<td>14.5</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>4-Catechols/methyalted 4-catechols</td>
<td>31.5</td>
<td>30.1</td>
<td>29.2</td>
<td>27.3</td>
<td>-13.5</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>Catechols/methyalted catechols</td>
<td>5.59</td>
<td>6.22</td>
<td>6.06</td>
<td>6.39</td>
<td>14.2</td>
<td>0.10</td>
<td>0.23</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
<td>0.28</td>
<td>8.4</td>
<td>0.19</td>
<td>0.09</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>1.59</td>
<td>1.52</td>
<td>1.61</td>
<td>1.68</td>
<td>5.2</td>
<td>0.55</td>
<td>0.26</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>0.17</td>
<td>0.16</td>
<td>0.14</td>
<td>0.15</td>
<td>-11.5</td>
<td>0.31</td>
<td>0.19</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>1.73</td>
<td>1.79</td>
<td>1.59</td>
<td>1.43</td>
<td>-17.2</td>
<td>0.03</td>
<td>0.003</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>3.90</td>
<td>3.69</td>
<td>3.89</td>
<td>5.30</td>
<td>36.1</td>
<td>0.04</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 2.2 (Continued)

a Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m$^2$; continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usually or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)

b % diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

c p-difference was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

d p-trend was estimated using a Wald test for a continuous variable of category-specific median EM levels (pmol/mg creatinine).
Table 2.3. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of total fat in 1995 and 1999: Nurses’ Health Study II (n = 598)°

<table>
<thead>
<tr>
<th></th>
<th>% Energy from total fat</th>
<th>% diffb</th>
<th>p-diffc</th>
<th>p-trendd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 25        25.1-30 30.1-35 &gt; 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total estrogens and estrogen metabolites</td>
<td>132        201    165    100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent estrogens</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrone</td>
<td>40.8       40.0    37.5    43.2</td>
<td>5.8</td>
<td>0.48</td>
<td>0.80</td>
</tr>
<tr>
<td>Estradiol</td>
<td>27.6       26.6    26.0    28.4</td>
<td>3.0</td>
<td>0.71</td>
<td>0.91</td>
</tr>
<tr>
<td>Catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Hydroxyestrone</td>
<td>54.5       51.6    52.3    55.1</td>
<td>1.2</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>2-Hydroxyestradiol</td>
<td>5.77       5.22    5.20    5.46</td>
<td>-5.4</td>
<td>0.60</td>
<td>0.57</td>
</tr>
<tr>
<td>Methylated catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyalted 2-catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Methoxyestrone</td>
<td>7.98       7.52    7.18    7.89</td>
<td>-1.2</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>2-Methoxyestradiol</td>
<td>0.72       0.69    0.70    0.74</td>
<td>2.9</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>Methylated 4-catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Methoxyestrone</td>
<td>10.4       9.92    9.53    10.1</td>
<td>-2.6</td>
<td>0.78</td>
<td>0.63</td>
</tr>
<tr>
<td>4-Methoxyestradiol</td>
<td>10.1       9.55    9.24    9.79</td>
<td>-3.4</td>
<td>0.73</td>
<td>0.61</td>
</tr>
<tr>
<td>4-Hydroxyestrone</td>
<td>6.14       5.76    5.62    6.14</td>
<td>0.1</td>
<td>0.99</td>
<td>0.92</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td>0.11       0.11    0.11    0.12</td>
<td>-9.9</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>2-Catechols/16-pathway</td>
<td>0.85       0.77    0.73    0.77</td>
<td>-9.2</td>
<td>0.44</td>
<td>0.36</td>
</tr>
<tr>
<td>Catechols/16-pathway</td>
<td>0.98       0.89    0.82    0.89</td>
<td>-9.1</td>
<td>0.44</td>
<td>0.30</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>0.10       0.10    0.10    0.10</td>
<td>8.6</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>1.05       0.93    0.86    0.93</td>
<td>-11.0</td>
<td>0.34</td>
<td>0.23</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>0.10       0.09    0.09    0.10</td>
<td>-2.2</td>
<td>0.87</td>
<td>0.71</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>1.18       1.06    0.97    1.06</td>
<td>-10.7</td>
<td>0.34</td>
<td>0.22</td>
</tr>
<tr>
<td>2-pathway/16-pathway</td>
<td>0.91       0.81    0.77    0.82</td>
<td>-9.4</td>
<td>0.38</td>
<td>0.32</td>
</tr>
<tr>
<td>2-Catechols/methylated 2-catechols</td>
<td>5.56       5.45    5.39    5.50</td>
<td>-1.1</td>
<td>0.89</td>
<td>0.85</td>
</tr>
<tr>
<td>4-Catechols/methylated 4-catechols</td>
<td>32.1       29.9    27.6    28.3</td>
<td>-11.7</td>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td>0.11       0.11    0.11    0.12</td>
<td>9.9</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>2-Catechols/16-pathway</td>
<td>0.85       0.77    0.73    0.77</td>
<td>-9.2</td>
<td>0.44</td>
<td>0.36</td>
</tr>
<tr>
<td>Catechols/16-pathway</td>
<td>0.98       0.89    0.82    0.89</td>
<td>-9.1</td>
<td>0.44</td>
<td>0.30</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>0.10       0.10    0.10    0.10</td>
<td>8.6</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>1.05       0.93    0.86    0.93</td>
<td>-11.0</td>
<td>0.34</td>
<td>0.23</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>0.10       0.09    0.09    0.10</td>
<td>-2.2</td>
<td>0.87</td>
<td>0.71</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>1.18       1.06    0.97    1.06</td>
<td>-10.7</td>
<td>0.34</td>
<td>0.22</td>
</tr>
<tr>
<td>2-pathway/16-pathway</td>
<td>0.91       0.81    0.77    0.82</td>
<td>-9.4</td>
<td>0.38</td>
<td>0.32</td>
</tr>
<tr>
<td>2-Catechols/methylated 2-catechols</td>
<td>5.56       5.45    5.39    5.50</td>
<td>-1.1</td>
<td>0.89</td>
<td>0.85</td>
</tr>
<tr>
<td>4-Catechols/methylated 4-catechols</td>
<td>32.1       29.9    27.6    28.3</td>
<td>-11.7</td>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>6.23       6.10    6.04    6.09</td>
<td>-2.3</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>1.63       1.55    1.57    1.50</td>
<td>-8.1</td>
<td>0.28</td>
<td>0.34</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>0.16       0.15    0.16    0.16</td>
<td>0.4</td>
<td>0.97</td>
<td>0.87</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>1.53       1.66    1.83    1.60</td>
<td>4.7</td>
<td>0.58</td>
<td>0.31</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>4.49       3.99    3.61    3.98</td>
<td>-11.5</td>
<td>0.37</td>
<td>0.24</td>
</tr>
</tbody>
</table>
Table 2.3 (Continued)

*a* Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m\(^2\); continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usually or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous).

*b* % diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

*c* p-difference was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

*d* p-trend was estimated using a Wald test for a continuous variable of category-specific median EM levels (pmol/mg creatinine).
Table 2.4. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of carbohydrates in 1995 and 1999: Nurses’ Health Study II (n = 598)\(^a\)

<table>
<thead>
<tr>
<th>Pathway/parent estrogens</th>
<th>% Energy from carbohydrates ≤ 45</th>
<th>45.1-50</th>
<th>50.1-55</th>
<th>&gt; 55</th>
<th>% diff(^b)</th>
<th>p-diff(^c)</th>
<th>p-trend(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>109</td>
<td>135</td>
<td>195</td>
<td>159</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total estrogens and estrogen metabolites</td>
<td>210</td>
<td>188</td>
<td>198</td>
<td>195</td>
<td>-6.9</td>
<td>0.27</td>
<td>0.49</td>
</tr>
<tr>
<td>Parent estrogens</td>
<td>Estrone</td>
<td>27.6</td>
<td>25.3</td>
<td>27.4</td>
<td>27.0</td>
<td>-2.3</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Estradiol</td>
<td>14.6</td>
<td>12.5</td>
<td>13.3</td>
<td>13.3</td>
<td>-8.9</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Catechols</td>
<td>40.8</td>
<td>37.7</td>
<td>40.0</td>
<td>40.3</td>
<td>-1.4</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>C2-Catechols</td>
<td>55.0</td>
<td>50.2</td>
<td>54.3</td>
<td>51.9</td>
<td>-5.6</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>2-Hydroxyestrone</td>
<td>48.5</td>
<td>44.3</td>
<td>48.2</td>
<td>45.9</td>
<td>-5.3</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>2-Hydroxyestradiol</td>
<td>5.61</td>
<td>5.20</td>
<td>5.37</td>
<td>5.36</td>
<td>-4.4</td>
<td>0.66</td>
</tr>
<tr>
<td>Methylated catechols</td>
<td>Estradiol</td>
<td>27.6</td>
<td>25.3</td>
<td>27.4</td>
<td>27.0</td>
<td>-2.3</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>C2-Methylated catechols</td>
<td>55.0</td>
<td>50.2</td>
<td>54.3</td>
<td>51.9</td>
<td>-5.6</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>2-Methoxyestrone</td>
<td>7.79</td>
<td>6.98</td>
<td>7.67</td>
<td>7.77</td>
<td>-0.3</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>2-Methoxyestradiol</td>
<td>0.76</td>
<td>0.70</td>
<td>0.66</td>
<td>0.72</td>
<td>-5.7</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>4-Hydroxyestrone</td>
<td>6.42</td>
<td>5.18</td>
<td>6.22</td>
<td>5.85</td>
<td>-8.8</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Methylated 4-catechols</td>
<td>10.1</td>
<td>9.38</td>
<td>10.2</td>
<td>10.0</td>
<td>-0.4</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Methylated 2-catechols</td>
<td>9.69</td>
<td>9.11</td>
<td>9.82</td>
<td>9.77</td>
<td>0.8</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>2-Methoxyestrone</td>
<td>7.79</td>
<td>6.98</td>
<td>7.67</td>
<td>7.77</td>
<td>-0.3</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>2-Methoxyestradiol</td>
<td>0.76</td>
<td>0.70</td>
<td>0.66</td>
<td>0.72</td>
<td>-5.7</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>2-Hydroxyestrone-3-methyl ether</td>
<td>1.05</td>
<td>1.16</td>
<td>1.12</td>
<td>1.19</td>
<td>12.4</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Methylated 4-catechols</td>
<td>0.22</td>
<td>0.17</td>
<td>0.21</td>
<td>0.18</td>
<td>-19.3</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Methylated 2-catechols</td>
<td>0.14</td>
<td>0.10</td>
<td>0.13</td>
<td>0.11</td>
<td>-22.3</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>2-Methoxyestrone</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>1.0</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>2-Methoxyestradiol</td>
<td>0.76</td>
<td>0.70</td>
<td>0.66</td>
<td>0.72</td>
<td>-5.7</td>
<td>0.55</td>
</tr>
<tr>
<td>4-Hydroxylation pathway</td>
<td>66.1</td>
<td>60.4</td>
<td>65.8</td>
<td>63.2</td>
<td>-4.3</td>
<td>0.65</td>
<td>0.91</td>
</tr>
<tr>
<td>16-Hydroxylation pathway</td>
<td>6.84</td>
<td>5.59</td>
<td>6.54</td>
<td>6.10</td>
<td>-10.7</td>
<td>0.33</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>16α-Hydroxyestrone</td>
<td>11.5</td>
<td>12.3</td>
<td>11.9</td>
<td>11.1</td>
<td>2.7</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Estriol</td>
<td>31.9</td>
<td>31.8</td>
<td>30.4</td>
<td>27.7</td>
<td>-13.3</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>17-Epiestradiol</td>
<td>1.71</td>
<td>1.63</td>
<td>1.42</td>
<td>1.51</td>
<td>-11.6</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>16-Ketoestradiol</td>
<td>14.0</td>
<td>14.2</td>
<td>14.1</td>
<td>13.2</td>
<td>-5.9</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>16-Epiestradiol</td>
<td>6.88</td>
<td>6.48</td>
<td>6.18</td>
<td>6.12</td>
<td>-11.0</td>
<td>0.12</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td>4-Catechols/2-catechols</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.11</td>
<td>-5.5</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>2-Catechols/16-pathway</td>
<td>0.79</td>
<td>0.69</td>
<td>0.81</td>
<td>0.82</td>
<td>4.5</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Catechols/16-pathway</td>
<td>0.91</td>
<td>0.76</td>
<td>0.94</td>
<td>0.95</td>
<td>4.3</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>4-Pathway/2-pathway</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>-8.9</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>2-Pathway/16-pathway</td>
<td>0.95</td>
<td>0.80</td>
<td>0.99</td>
<td>1.00</td>
<td>5.7</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>4-pathway/16-pathway</td>
<td>0.10</td>
<td>0.08</td>
<td>0.10</td>
<td>0.10</td>
<td>-4.0</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2,4-pathway/16-pathway</td>
<td>1.08</td>
<td>0.91</td>
<td>1.12</td>
<td>1.13</td>
<td>5.2</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>2-pathway/4,16-pathway</td>
<td>0.83</td>
<td>0.73</td>
<td>0.85</td>
<td>0.87</td>
<td>4.9</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>2-Catechols/methylated 2-catechols</td>
<td>5.57</td>
<td>5.16</td>
<td>5.68</td>
<td>5.29</td>
<td>-4.9</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>4-Catechols/methylated 4-catechols</td>
<td>26.5</td>
<td>30.8</td>
<td>30.2</td>
<td>31.3</td>
<td>18.2</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Catechols/methylated catechols</td>
<td>6.16</td>
<td>5.77</td>
<td>6.28</td>
<td>5.98</td>
<td>-2.8</td>
<td>0.69</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>0.27</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
<td>2.1</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>1.55</td>
<td>1.48</td>
<td>1.65</td>
<td>1.55</td>
<td>0</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>0.16</td>
<td>0.15</td>
<td>0.16</td>
<td>0.15</td>
<td>-8.8</td>
<td>0.39</td>
<td>0.60</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>1.65</td>
<td>1.83</td>
<td>1.67</td>
<td>1.54</td>
<td>-6.4</td>
<td>0.41</td>
<td>0.20</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>4.27</td>
<td>3.46</td>
<td>4.08</td>
<td>4.16</td>
<td>-2.5</td>
<td>0.85</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Table 2.4 (Continued)

\(^a\) Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m\(^2\); continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usually or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)

\(^b\) % diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

\(^c\) p-difference was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

\(^d\) p-trend was estimated using a Wald test for a continuous variable of category-specific median EM levels (pmol/mg creatinine).
Table 2.5. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of total protein in 1995 and 1999: Nurses’ Health Study II (n = 598)

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>% Energy from protein ≤ 15</th>
<th>15.1-18</th>
<th>18.1-20</th>
<th>&gt; 20</th>
<th>diff&lt;sup&gt;b&lt;/sup&gt;</th>
<th>p-diff&lt;sup&gt;c&lt;/sup&gt;</th>
<th>P-trend&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total estrogens and estrogen metabolites</strong></td>
<td>72</td>
<td>216</td>
<td>179</td>
<td>131</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parent estrogens</strong></td>
<td>199</td>
<td>193</td>
<td>195</td>
<td>205</td>
<td>2.9</td>
<td>0.70</td>
<td>0.56</td>
</tr>
<tr>
<td>Estrone</td>
<td>41.4</td>
<td>40.1</td>
<td>39.1</td>
<td>40.5</td>
<td>-2.2</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Estradiol</td>
<td>27.5</td>
<td>26.9</td>
<td>26.1</td>
<td>28.2</td>
<td>2.7</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Catechols</strong></td>
<td>14.2</td>
<td>13.4</td>
<td>13.0</td>
<td>13.5</td>
<td>-4.7</td>
<td>0.56</td>
<td>0.55</td>
</tr>
<tr>
<td>2-Catechols</td>
<td>56.0</td>
<td>62.5</td>
<td>59.0</td>
<td>65.1</td>
<td>16.2</td>
<td>0.19</td>
<td>0.32</td>
</tr>
<tr>
<td>2-Hydroxyestrone</td>
<td>47.8</td>
<td>54.4</td>
<td>51.3</td>
<td>56.3</td>
<td>17.8</td>
<td>0.17</td>
<td>0.31</td>
</tr>
<tr>
<td>2-Hydroxyestradiol</td>
<td>42.0</td>
<td>48.2</td>
<td>45.5</td>
<td>49.7</td>
<td>18.3</td>
<td>0.17</td>
<td>0.31</td>
</tr>
<tr>
<td>4-Catechols</td>
<td>5.05</td>
<td>5.32</td>
<td>5.24</td>
<td>5.90</td>
<td>16.7</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>4-Hydroxyestrone</td>
<td>6.01</td>
<td>5.91</td>
<td>5.47</td>
<td>6.46</td>
<td>7.6</td>
<td>0.61</td>
<td>0.68</td>
</tr>
<tr>
<td>Methylated catechols</td>
<td>9.95</td>
<td>10.0</td>
<td>9.62</td>
<td>10.4</td>
<td>4.0</td>
<td>0.72</td>
<td>0.80</td>
</tr>
<tr>
<td>Methylated 2-catechols</td>
<td>9.73</td>
<td>9.74</td>
<td>9.24</td>
<td>10.1</td>
<td>3.4</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td>2-Methoxyestrone</td>
<td>7.49</td>
<td>7.73</td>
<td>7.30</td>
<td>7.70</td>
<td>2.8</td>
<td>0.80</td>
<td>0.99</td>
</tr>
<tr>
<td>2-Methoxyestradiol</td>
<td>0.67</td>
<td>0.70</td>
<td>0.70</td>
<td>0.76</td>
<td>13.9</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>2-Hydroxyestrone-3-methyl ether</td>
<td>1.19</td>
<td>1.17</td>
<td>1.06</td>
<td>1.20</td>
<td>0.7</td>
<td>0.95</td>
<td>0.85</td>
</tr>
<tr>
<td>Methylated 4-catechols</td>
<td>0.17</td>
<td>0.19</td>
<td>0.20</td>
<td>0.19</td>
<td>13.7</td>
<td>0.38</td>
<td>0.43</td>
</tr>
<tr>
<td>4-Methoxyestrone</td>
<td>0.10</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>16.5</td>
<td>0.35</td>
<td>0.47</td>
</tr>
<tr>
<td>4-Methoxyestradiol</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>-11.1</td>
<td>0.52</td>
<td>0.57</td>
</tr>
<tr>
<td>2-Hydroxylation pathway</td>
<td>59.0</td>
<td>65.9</td>
<td>61.6</td>
<td>68.1</td>
<td>15.4</td>
<td>0.20</td>
<td>0.36</td>
</tr>
<tr>
<td>4-Hydroxylation pathway</td>
<td>5.97</td>
<td>6.22</td>
<td>6.00</td>
<td>6.63</td>
<td>11.0</td>
<td>0.44</td>
<td>0.50</td>
</tr>
<tr>
<td>16-Hydroxylation pathway</td>
<td>71.5</td>
<td>64.9</td>
<td>68.4</td>
<td>68.0</td>
<td>-4.9</td>
<td>0.60</td>
<td>0.90</td>
</tr>
<tr>
<td>16α-Hydroxyestrone</td>
<td>12.6</td>
<td>12.3</td>
<td>11.3</td>
<td>10.8</td>
<td>-14.4</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>Estriol</td>
<td>34.1</td>
<td>28.2</td>
<td>29.7</td>
<td>32.6</td>
<td>-4.2</td>
<td>0.70</td>
<td>0.82</td>
</tr>
<tr>
<td>17-Epiestriol</td>
<td>1.68</td>
<td>1.62</td>
<td>1.48</td>
<td>1.44</td>
<td>-13.9</td>
<td>0.32</td>
<td>0.22</td>
</tr>
<tr>
<td>16-Ketoestradiol</td>
<td>15.1</td>
<td>13.4</td>
<td>13.9</td>
<td>14.2</td>
<td>-6.0</td>
<td>0.50</td>
<td>0.82</td>
</tr>
<tr>
<td>16-Epiestriol</td>
<td>6.31</td>
<td>6.26</td>
<td>6.19</td>
<td>6.74</td>
<td>6.8</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>Ratios of metabolic pathways</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td>0.12</td>
<td>0.10</td>
<td>0.11</td>
<td>0.11</td>
<td>-4.6</td>
<td>0.73</td>
<td>0.93</td>
</tr>
<tr>
<td>2-Catechols/16-pathway</td>
<td>0.66</td>
<td>0.81</td>
<td>0.77</td>
<td>0.83</td>
<td>25.4</td>
<td>0.11</td>
<td>0.23</td>
</tr>
<tr>
<td>Catechols/16-pathway</td>
<td>0.78</td>
<td>0.93</td>
<td>0.86</td>
<td>0.93</td>
<td>19.0</td>
<td>0.21</td>
<td>0.43</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>1.6</td>
<td>0.89</td>
<td>0.68</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>0.82</td>
<td>0.99</td>
<td>0.91</td>
<td>0.97</td>
<td>18.2</td>
<td>0.22</td>
<td>0.46</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>17.2</td>
<td>0.30</td>
<td>0.41</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>0.94</td>
<td>1.11</td>
<td>1.03</td>
<td>1.11</td>
<td>18.4</td>
<td>0.20</td>
<td>0.41</td>
</tr>
<tr>
<td>2-pathway/4,16-pathway</td>
<td>0.71</td>
<td>0.85</td>
<td>0.81</td>
<td>0.85</td>
<td>20.3</td>
<td>0.14</td>
<td>0.30</td>
</tr>
<tr>
<td>2-Catechols/methylated 2-catechols</td>
<td>4.96</td>
<td>5.52</td>
<td>5.44</td>
<td>5.71</td>
<td>15.2</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>4-Catechols/methylated 4-catechols</td>
<td>33.0</td>
<td>29.4</td>
<td>26.9</td>
<td>31.8</td>
<td>-3.5</td>
<td>0.86</td>
<td>0.83</td>
</tr>
<tr>
<td>Catechols/methylated catechols</td>
<td>5.64</td>
<td>6.15</td>
<td>6.01</td>
<td>6.37</td>
<td>13.0</td>
<td>0.15</td>
<td>0.24</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>0.27</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
<td>-1.3</td>
<td>0.84</td>
<td>0.91</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>1.42</td>
<td>1.63</td>
<td>1.55</td>
<td>1.57</td>
<td>10.5</td>
<td>0.25</td>
<td>0.57</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
<td>0.16</td>
<td>10.3</td>
<td>0.43</td>
<td>0.49</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>1.73</td>
<td>1.65</td>
<td>1.70</td>
<td>1.61</td>
<td>-6.8</td>
<td>0.45</td>
<td>0.56</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>3.30</td>
<td>3.83</td>
<td>4.09</td>
<td>4.63</td>
<td>40.3</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Table 2.5 (Continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **a** Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m²; continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usually or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)  
| **b** % diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.  
| **c** p-*difference was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.  
| **d** p-*trend was estimated using a Wald test for a continuous variable of category-specific median EM levels (pmol/mg creatinine).  

---
| Pathway/2 | Catechols/16 | Catechols/2 | Hydroxylation pathway | Methylated 2 | 4-Catechols | 2-Hydroxyestradiol | 4-Hydroxyestrone | Methyalted catechols | Methyalted 2-catechols | 2-Methoxyestrone | 2-Methoxyestradiol | 2-Hydroxyestrone-3-methyl ether | 16-Hydroxylation pathway | 4-Hydroxylation pathway | 16α-Hydroxyestrone | Estradiol | 17-Epiestradiol | 16-Ketoestradiol | 16-Epiestradiol | Ratios of metabolic pathways |
|----------|-------------|-------------|-----------------------|--------------|-------------|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|----------------------------|-----------------|-----------------|------------------|------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Total estrogens and estrogen metabolites | 5.3 | 0.39 | 0.44 | 10.0 | 0.10 | 0.13 | 7.6 | 0.22 | 0.18 | 6.6 | 0.29 | 0.39 | 17.0 | 0.03 | 0.04 | 16.7 | 0.02 | 0.03 | 14.0 | 0.07 | 0.03 | 9.3 | 0.21 | 0.15 | 14.3 | 0.06 | 0.11 | 16.0 | 0.03 | 0.03 | 15.4 | 0.05 | 0.04 | 13.9 | 0.07 | 0.06 | 5.5 | 0.42 | 0.39 | 5.1 | 0.45 | 0.58 | 6.8 | 0.33 | 0.25 | 15.0 | 0.10 | 0.17 | 16.0 | 0.03 | 0.03 | 15.4 | 0.05 | 0.04 | 13.9 | 0.07 | 0.06 | 12.9 | 0.18 | 0.21 | 24.3 | 0.01 | 0.02 | 13.4 | 0.17 | 0.28 | 17.2 | 0.08 | 0.16 | 14.8 | 0.14 | 0.16 | 30.3 | 0.004 | 0.01 | 15.7 | 0.12 | 0.25 | 16.9 | 0.09 | 0.17 | 14.8 | 0.15 | 0.17 | 32.0 | 0.003 | 0.01 | 16.1 | 0.12 | 0.25 | 17.7 | 0.09 | 0.17 | 20.4 | 0.05 | 0.06 | 23.4 | 0.02 | 0.03 | 20.3 | 0.05 | 0.07 | 16.9 | 0.09 | 0.12 |
| Fiber from fruit<sup>e</sup> (> 4 vs. ≤ 2 g/day) | 11.7 | 0.32 | 0.29 | 2.1 | 0.85 | 0.95 | 10.3 | 0.39 | 0.20 | 13.6 | 0.26 | 0.21 | 9.4 | 0.29 | 0.26 | 11.9 | 0.17 | 0.22 | 8.6 | 0.33 | 0.37 | 8.6 | 0.33 | 0.25 | 8.4 | 0.36 | 0.31 | 12.4 | 0.17 | 0.21 | 9.6 | 0.30 | 0.33 | 9.2 | 0.31 | 0.24 | 7.3 | 0.43 | 0.39 | 15.1 | 0.10 | 0.10 | 6.8 | 0.47 | 0.50 | 9.3 | 0.31 | 0.27 | 19.6 | 0.05 | 0.04 | 7.0 | 0.44 | 0.53 | 17.9 | 0.07 | 0.05 | 12.9 | 0.18 | 0.26 | 10.6 | 0.25 | 0.30 | 7.0 | 0.43 | 0.56 | 5.3 | 0.56 | 0.69 | 1.4 | 0.87 | 0.83 | 29.9 | 0.03 | 0.06 | 2.6 | 0.82 | 0.92 | 15.6 | 0.22 | 0.14 | 3.0 | 0.81 | 0.89 | 16.0 | 0.25 | 0.39 | -7.2 | 0.56 | 0.49 | 8.2 | 0.55 | 0.44 | 5.2 | 0.70 | 0.83 | 52.2 | 0.003 | 0.004 | 24.0 | 0.13 | 0.13 | 32.3 | 0.06 | 0.02 | 7.4 | 0.63 | 0.51 | 15.7 | 0.10 | 0.12 | 25.8 | 0.01 | 0.01 | 16.1 | 0.09 | 0.19 | 16.1 | 0.09 | 0.15 | 8.7 | 0.42 | 0.53 | 0.4 | 0.97 | 0.92 | -1.8 | 0.86 | 0.93 | 4.1 | 0.71 | 0.65 | -7.2 | 0.32 | 0.39 | -3.2 | 0.66 | 0.72 | -5.4 | 0.47 | 0.71 | -3.0 | 0.70 | 0.62 | -12.4 | 0.19 | 0.35 | 5.9 | 0.56 | 0.64 | -7.0 | 0.47 | 0.73 | 7.8 | 0.46 | 0.80 | -12.2 | 0.14 | 0.07 | -8.3 | 0.32 | 0.44 | -17.4 | 0.04 | 0.07 | -12.7 | 0.13 | 0.18 | -27.2 | 0.01 | 0.01 | 4.0 | 0.74 | 0.99 | -19.7 | 0.07 | 0.05 | -2.8 | 0.82 | 0.62 | -3.2 | 0.65 | 0.79 | 5.8 | 0.42 | 0.45 | 0.2 | 0.98 | 0.93 | 1.4 | 0.84 | 0.78 | 0.5 | 0.94 | 0.98 | -7.4 | 0.26 | 0.11 | 0.2 | 0.98 | 0.93 | 1.4 | 0.84 | 0.78 | 0.5 | 0.94 | 0.98 | -14.5 | 0.13 | 0.10 | -22.5 | 0.02 | 0.03 | -21.9 | 0.02 | 0.06 | -16.4 | 0.10 | 0.14 | 30.3 | 0.02 | 0.03 | 33.8 | 0.01 | 0.01 | 25.7 | 0.04 | 0.14 | 17.7 | 0.14 | 0.20 | 31.1 | 0.01 | 0.02 | 27.5 | 0.02 | 0.02 | 23.4 | 0.06 | 0.10 | 19.0 | 0.11 | 0.12 | -11.3 | 0.17 | 0.13 | -21.4 | 0.01 | 0.01 | -15.0 | 0.07 | 0.18 | -11.4 | 0.18 | 0.19 |

Supplementary table 2.S1. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of different fiber from different sources in 1995 and 1999: Nurses’ Health Study II (n = 598)<sup>a</sup>
### Supplementary table 2.S1 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Fiber from fruit( &gt; 4 \text{ vs.} \leq 2 \text{ g/day})</th>
<th>Fiber from cereals( &gt; 8 \text{ vs.} \leq 5 \text{ g/day})</th>
<th>Fiber from vegetables( &gt;12 \text{ vs.} \leq 7 \text{ g/day})</th>
<th>Fiber from legumes( &gt; 2 \text{ vs.} \leq 0.75 \text{ g/day})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff (b)</td>
<td>(p_{\text{diff}}^{c})</td>
<td>(p_{\text{trend}}^{d})</td>
<td>% diff (b)</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>34.6</td>
<td>0.01</td>
<td>0.01</td>
<td>28.7</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>14.8</td>
<td>0.25</td>
<td>0.32</td>
<td>-1.2</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>32.1</td>
<td>0.01</td>
<td>0.01</td>
<td>23.7</td>
</tr>
<tr>
<td>2-pathway/4,16-pathway</td>
<td>29.4</td>
<td>0.01</td>
<td>0.01</td>
<td>30.8</td>
</tr>
</tbody>
</table>
| 2-Catechols/methylated 2-
  catechols                  | 10.2          | 0.14          | 0.11          | 17.7          | 0.01          | 0.01          | 8.9           | 0.21          | 0.25          | 8.1           | 0.24          | 0.38          |
| 4-Catechols/methylated 4-
  catechols                  | -10.3         | 0.49          | 0.67          | 4.3           | 0.79          | 0.82          | -11.0         | 0.47          | 0.56          | -0.8          | 0.96          | 0.96          |
| Catechols/methylated catechols | 8.3           | 0.22          | 0.18          | 13.0          | 0.06          | 0.04          | 7.2           | 0.30          | 0.27          | 9.3           | 0.18          | 0.31          |
| Parent estrogens/EM       | 10.8          | 0.04          | 0.07          | 0.8           | 0.87          | 0.92          | 13.7          | 0.01          | 0.02          | 4.6           | 0.37          | 0.22          |
| 2-Pathway/parent estrogens | 6.2           | 0.38          | 0.34          | 14.8          | 0.04          | 0.04          | 1.9           | 0.79          | 0.99          | 3.9           | 0.58          | 0.61          |
| 4-Pathway/parent estrogens | -6.2          | 0.51          | 0.50          | -11.0         | 0.22          | 0.21          | -12.5         | 0.17          | 0.26          | -6.2          | 0.51          | 0.47          |
| 16-Pathway/parent estrogens | -21.0         | 0.001         | 0.003         | -11.5         | 0.09          | 0.10          | -20.2         | 0.002         | 0.005         | -11.9         | 0.08          | 0.06          |
| 2-Hydroxyestrone/16a-
  hydroxyestrone            | 38.2          | 0.01          | 0.02          | 27.5          | 0.04          | 0.03          | 29.3          | 0.04          | 0.14          | 8.4           | 0.51          | 0.37          |

\(a\) Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m\(^2\); continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usually or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)

\(b\) \% diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

\(c\) \(p_{\text{diff}}^{c}\) was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

\(d\) \(p_{\text{trend}}^{d}\) was estimated using a Wald test for a continuous variable of quartile-specific median EM levels (pmol/mg creatinine).

\(e\) The highest and the lowest intake groups were categorized based on quartiles of the nutrient intake.
Supplementary table 2.S2. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of different fat subtypes in 1995 and 1999: Nurses’ Health Study II (n = 598)\(^{h}\)

<table>
<thead>
<tr>
<th></th>
<th>Saturated fat(^{a}) (&gt;12 vs. ≤8 % energy)</th>
<th>Monounsaturated fat(^{c}) (&gt;13 vs. ≤10 % energy)</th>
<th>Polyunsaturated fat(^{e}) (&gt;6 vs. ≤4.5 % energy)</th>
<th>Trans-fat(^{d}) (&gt;1.75 vs. ≤1.15% energy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff(^{b})</td>
<td>p(_{diff})^{c}</td>
<td>p(_{trend})^{d}</td>
<td>% diff(^{b})</td>
</tr>
<tr>
<td>Total estrogens and estrogen metabolites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent estrogens</td>
<td>-2.0 0.76 0.83</td>
<td>4.8 0.42 0.58</td>
<td>0.5 0.94 0.72</td>
<td>-2.2 0.71 0.91</td>
</tr>
<tr>
<td>Estrone</td>
<td>-9.9 0.18 0.12</td>
<td>5.6 0.42 0.64</td>
<td>5.7 0.41 0.40</td>
<td>-9.8 0.14 0.11</td>
</tr>
<tr>
<td>Estradiol</td>
<td>2.5 0.73 0.98</td>
<td>9.4 0.16 0.19</td>
<td>4.7 0.46 0.41</td>
<td>-0.7 0.92 0.80</td>
</tr>
<tr>
<td>Catechools</td>
<td>-11.2 0.23 0.49</td>
<td>4.0 0.65 0.69</td>
<td>-2.4 0.78 0.84</td>
<td>-10.2 0.22 0.34</td>
</tr>
<tr>
<td>2-Catecholes</td>
<td>-12.1 0.21 0.45</td>
<td>5.8 0.53 0.54</td>
<td>0.5 0.96 0.61</td>
<td>-9.1 0.30 0.40</td>
</tr>
<tr>
<td>2-Hydroxyestrone</td>
<td>-13.1 0.18 0.42</td>
<td>6.0 0.52 0.53</td>
<td>1.3 0.89 0.57</td>
<td>-8.5 0.34 0.45</td>
</tr>
<tr>
<td>2-Hydroxyestradiol</td>
<td>-11.5 0.24 0.28</td>
<td>0.03 1.00 0.88</td>
<td>-5.2 0.55 0.93</td>
<td>-15.8 0.07 0.10</td>
</tr>
<tr>
<td>4-Catecholes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxyestrone</td>
<td>-7.8 0.51 0.62</td>
<td>-6.5 0.53 0.49</td>
<td>-14.6 0.14 0.28</td>
<td>-24.2 0.01 0.03</td>
</tr>
<tr>
<td>Methyalted catecholes</td>
<td>-12.3 0.16 0.17</td>
<td>3.4 0.68 0.78</td>
<td>1.4 0.86 0.76</td>
<td>-7.9 0.33 0.35</td>
</tr>
<tr>
<td>Methyalted 2-catecholes</td>
<td>-12.8 0.15 0.16</td>
<td>3.3 0.70 0.81</td>
<td>1.1 0.89 0.75</td>
<td>-8.6 0.30 0.32</td>
</tr>
<tr>
<td>2-Methoxyestrone</td>
<td>-14.0 0.12 0.15</td>
<td>6.0 0.50 0.62</td>
<td>1.7 0.85 0.63</td>
<td>-7.5 0.38 0.41</td>
</tr>
<tr>
<td>2-Methoxyestradiol</td>
<td>-7.7 0.42 0.65</td>
<td>4.9 0.58 0.63</td>
<td>1.5 0.78 0.79</td>
<td>-7.9 0.35 0.42</td>
</tr>
<tr>
<td>2-Hydroxyestrone-3-methyl ether</td>
<td>-8.2 0.38 0.27</td>
<td>-4.8 0.57 0.49</td>
<td>-5.3 0.52 0.32</td>
<td>-2.5 0.77 0.85</td>
</tr>
<tr>
<td>Methyalted 4-catecholes</td>
<td>2.2 0.87 0.99</td>
<td>19.0 0.12 0.09</td>
<td>16.9 0.17 0.46</td>
<td>-0.4 0.97 0.97</td>
</tr>
<tr>
<td>4-Methoxyestrone</td>
<td>11.8 0.43 0.61</td>
<td>19.2 0.16 0.11</td>
<td>7.7 0.55 0.93</td>
<td>4.9 0.71 0.80</td>
</tr>
<tr>
<td>4-Methoxyestradiol</td>
<td>-14.7 0.32 0.24</td>
<td>6.7 0.64 0.66</td>
<td>33.2 0.04 0.14</td>
<td>-6.0 0.67 0.92</td>
</tr>
<tr>
<td>2-Hydroxylation pathway</td>
<td>-13.3 0.14 0.31</td>
<td>3.7 0.67 0.68</td>
<td>0.1 0.99 0.69</td>
<td>-10.4 0.21 0.28</td>
</tr>
<tr>
<td>4-Hydroxylation pathway</td>
<td>5.2 0.65 0.76</td>
<td>-2.6 0.80 0.71</td>
<td>-12.6 0.19 0.29</td>
<td>-18.9 0.05 0.09</td>
</tr>
<tr>
<td>16-Hydroxylation pathway</td>
<td>7.9 0.37 0.38</td>
<td>9.2 0.24 0.29</td>
<td>4.4 0.56 0.58</td>
<td>17.7 0.03 0.02</td>
</tr>
<tr>
<td>16β-Hydroxyestrone</td>
<td>19.6 0.11 0.08</td>
<td>12.7 0.22 0.24</td>
<td>13.9 0.18 0.20</td>
<td>31.3 0.01 0.01</td>
</tr>
<tr>
<td>Estriol</td>
<td>7.2 0.48 0.47</td>
<td>10.0 0.28 0.34</td>
<td>0.8 0.93 0.90</td>
<td>16.9 0.08 0.05</td>
</tr>
<tr>
<td>17-Epiestriol</td>
<td>17.5 0.23 0.31</td>
<td>21.8 0.10 0.06</td>
<td>37.0 0.01 0.01</td>
<td>36.1 0.01 0.01</td>
</tr>
<tr>
<td>16-Ketoestradiol</td>
<td>3.3 0.68 0.81</td>
<td>5.1 0.47 0.57</td>
<td>7.6 0.28 0.38</td>
<td>10.6 0.16 0.10</td>
</tr>
<tr>
<td>16-Epiestriol</td>
<td>9.7 0.22 0.26</td>
<td>9.1 0.19 0.30</td>
<td>5.6 0.40 0.63</td>
<td>9.3 0.19 0.15</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catecholes/2-catecholes</td>
<td>10.6 0.41 0.38</td>
<td>-8.5 0.40 0.36</td>
<td>-15.9 0.10 0.11</td>
<td>-1.9 0.86 0.92</td>
</tr>
<tr>
<td>2-Catecholes/16-pathway</td>
<td>-21.6 0.04 0.11</td>
<td>-3.0 0.78 0.79</td>
<td>-5.9 0.56 0.98</td>
<td>-24.5 0.01 0.01</td>
</tr>
<tr>
<td>Catecholes/16-pathway</td>
<td>-21.6 0.04 0.09</td>
<td>-3.8 0.71 0.66</td>
<td>-9.6 0.33 0.66</td>
<td>-28.6 0.002 0.002</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>14.0 0.18 0.17</td>
<td>-4.8 0.57 0.58</td>
<td>-12.7 0.11 0.11</td>
<td>-2.8 0.75 0.76</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>-23.6 0.02 0.05</td>
<td>-4.2 0.68 0.65</td>
<td>-7.6 0.44 0.77</td>
<td>-28.3 0.002 0.002</td>
</tr>
</tbody>
</table>
### Supplementary table 2.S2 (Continued)

<table>
<thead>
<tr>
<th>Method/Pathway</th>
<th>Saturated fat&lt;sup&gt;a&lt;/sup&gt; (% &gt;12 vs. ≤8% energy)</th>
<th>Monounsaturated fat&lt;sup&gt;a&lt;/sup&gt; (% &gt;13 vs. ≤10% energy)</th>
<th>Polyunsaturated fat&lt;sup&gt;a&lt;/sup&gt; (% &gt;6 vs. % ≤4.5 energy)</th>
<th>Trans-fat&lt;sup&gt;a&lt;/sup&gt; (% &gt;1.75 vs. ≤1.15% energy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff&lt;sup&gt;b&lt;/sup&gt;</td>
<td>p&lt;sub&gt;diff&lt;/sub&gt;&lt;sup&gt;c&lt;/sup&gt;</td>
<td>p&lt;sub&gt;trend&lt;/sub&gt;&lt;sup&gt;d&lt;/sup&gt;</td>
<td>% diff&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>-12.4</td>
<td>0.32</td>
<td>0.48</td>
<td>-6.7</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>-22.6</td>
<td>0.03</td>
<td>0.06</td>
<td>-5.2</td>
</tr>
<tr>
<td>2-pathway/4,16-pathway</td>
<td>-21.8</td>
<td>0.02</td>
<td>0.05</td>
<td>-3.2</td>
</tr>
<tr>
<td>2-Catechols/methylated 2-catechols</td>
<td>-6.5</td>
<td>0.36</td>
<td>0.53</td>
<td>-0.8</td>
</tr>
<tr>
<td>4-Catechols/methylated 4-catechols</td>
<td>-10.3</td>
<td>0.54</td>
<td>0.66</td>
<td>-22.7</td>
</tr>
<tr>
<td>Catechols/methylated catechols</td>
<td>-6.3</td>
<td>0.37</td>
<td>0.51</td>
<td>-2.6</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>-2.7</td>
<td>0.62</td>
<td>0.22</td>
<td>3.2</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>-11.5</td>
<td>0.11</td>
<td>0.40</td>
<td>-2.9</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>1.3</td>
<td>0.91</td>
<td>0.50</td>
<td>-5.4</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>16.3</td>
<td>0.06</td>
<td>0.03</td>
<td>2.1</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>-29.1</td>
<td>0.01</td>
<td>0.02</td>
<td>-4.5</td>
</tr>
</tbody>
</table>

<sup>a</sup>Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m<sup>2</sup>; continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usualy or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)

<sup>b</sup>% diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

<sup>c</sup>p<sub>diff</sub> was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

<sup>d</sup>p<sub>trend</sub> was estimated using a Wald test for a continuous variable of quartile-specific median EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

<sup>e</sup>The highest and the lowest intake groups were categorized based on quartiles of the nutrient intake.
### Supplementary table 2.S3. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average glycemic load in 1995 and 1999: Nurses’ Health Study II (n = 598)\(^a\)

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Glycemic load(^b)</th>
<th>% diff(^b)</th>
<th>p&lt;sub&gt;diff&lt;/sub&gt;(^c)</th>
<th>p&lt;sub&gt;trend&lt;/sub&gt;(^d)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤100</td>
<td>100.1-130</td>
<td>130.1-155</td>
<td>&gt;155</td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total estrogens and estrogen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>metabolites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent estrogens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrone</td>
<td>37.4</td>
<td>38.1</td>
<td>41.1</td>
<td>42.5</td>
<td>13.6</td>
</tr>
<tr>
<td>Estradiol</td>
<td>26.5</td>
<td>26.1</td>
<td>27.0</td>
<td>27.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Catechols</td>
<td>63.1</td>
<td>59.1</td>
<td>58.7</td>
<td>62.3</td>
<td>-1.3</td>
</tr>
<tr>
<td>2-Catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Hydroxyestrone</td>
<td>54.7</td>
<td>51.6</td>
<td>50.9</td>
<td>53.7</td>
<td>-1.8</td>
</tr>
<tr>
<td>2-Hydroxyestriadiol</td>
<td>48.3</td>
<td>45.7</td>
<td>45.2</td>
<td>47.7</td>
<td>-1.4</td>
</tr>
<tr>
<td>4-Catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxyestrone</td>
<td>5.98</td>
<td>5.46</td>
<td>6.14</td>
<td>6.01</td>
<td>0.5</td>
</tr>
<tr>
<td>Methylated catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylated 2-catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Methoxyestrone</td>
<td>9.51</td>
<td>9.73</td>
<td>9.91</td>
<td>10.8</td>
<td>13.1</td>
</tr>
<tr>
<td>2-Methoxyestriadiol</td>
<td>9.23</td>
<td>9.40</td>
<td>9.61</td>
<td>10.3</td>
<td>11.6</td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxyestrone/16α-hydroxyestrone</td>
<td>64.5</td>
<td>62.5</td>
<td>62.7</td>
<td>65.9</td>
<td>2.0</td>
</tr>
<tr>
<td>16-Hydroxylation pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16α-Hydroxyestrone</td>
<td>69.8</td>
<td>67.3</td>
<td>65.0</td>
<td>68.0</td>
<td>-2.5</td>
</tr>
<tr>
<td>Estriol</td>
<td>11.6</td>
<td>12.1</td>
<td>11.0</td>
<td>12.3</td>
<td>5.8</td>
</tr>
<tr>
<td>17-Epiestradiol</td>
<td>32.8</td>
<td>31.2</td>
<td>28.0</td>
<td>28.7</td>
<td>-12.3</td>
</tr>
<tr>
<td>16-Ketoestradiol</td>
<td>1.79</td>
<td>1.57</td>
<td>1.25</td>
<td>1.58</td>
<td>-11.7</td>
</tr>
<tr>
<td>16-Epiestradiol</td>
<td>14.3</td>
<td>14.0</td>
<td>13.0</td>
<td>14.2</td>
<td>-1.0</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td>0.12</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>-8.2</td>
</tr>
<tr>
<td>2-Catechols/16α-pathway</td>
<td>0.77</td>
<td>0.76</td>
<td>0.80</td>
<td>0.80</td>
<td>4.1</td>
</tr>
<tr>
<td>Catechols/16α-pathway</td>
<td>0.88</td>
<td>0.85</td>
<td>0.90</td>
<td>0.92</td>
<td>4.0</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td>0.09</td>
<td>-14.9</td>
</tr>
<tr>
<td>2-Pathway/16α-pathway</td>
<td>0.90</td>
<td>0.90</td>
<td>0.97</td>
<td>0.98</td>
<td>9.6</td>
</tr>
<tr>
<td>4-pathway/16α-pathway</td>
<td>0.10</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td>-6.4</td>
</tr>
<tr>
<td>2,4-pathway/16α-pathway</td>
<td>1.03</td>
<td>1.02</td>
<td>1.09</td>
<td>1.10</td>
<td>7.0</td>
</tr>
<tr>
<td>2-pathway/4,16-pathway</td>
<td>0.79</td>
<td>0.80</td>
<td>0.85</td>
<td>0.85</td>
<td>8.1</td>
</tr>
<tr>
<td>2-Catechols/methylated 2-catechols</td>
<td>5.60</td>
<td>5.21</td>
<td>5.52</td>
<td>5.42</td>
<td>-3.2</td>
</tr>
<tr>
<td>4-Catechols/methylated 4-catechols</td>
<td>30.5</td>
<td>30.4</td>
<td>28.2</td>
<td>29.1</td>
<td>-4.4</td>
</tr>
<tr>
<td>Catechols/methylated catechols</td>
<td>6.27</td>
<td>5.79</td>
<td>6.14</td>
<td>6.03</td>
<td>-3.8</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>0.25</td>
<td>0.26</td>
<td>0.28</td>
<td>0.26</td>
<td>4.5</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>1.57</td>
<td>1.56</td>
<td>1.51</td>
<td>1.62</td>
<td>2.6</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>0.17</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>-11.8</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>1.78</td>
<td>1.73</td>
<td>1.54</td>
<td>1.62</td>
<td>-9.0</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>4.08</td>
<td>3.78</td>
<td>4.19</td>
<td>3.91</td>
<td>-4.1</td>
</tr>
</tbody>
</table>
Supplementary table 2.S3 (Continued)

aAdjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m^2; continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usual or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)

b % diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

^P^diff was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

^d^P^trend was estimated using a Wald test for a continuous variable of quartile-specific median EM levels (pmol/mg creatinine).

The highest and the lowest intake groups were categorized based on quartiles of the nutrient intake.
Supplementary table 2.S4. Adjusted geometric means (pmol/mg creatinine) of urinary estrogen/estrogen metabolite levels by average dietary intake of protein from different sources in 1995 and 1999: Nurses’ Health Study II (n = 598)\(^a\)  

<table>
<thead>
<tr>
<th></th>
<th>Non-dairy animal protein(^e) (% &gt;10 vs. ≤6 % energy)</th>
<th>Dairy protein(^e) (% &gt; 5 vs. ≤ 2.2 % energy)</th>
<th>Vegetable protein(^e) (% &gt; 6.3 vs. ≤ 4.4 % energy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff(^b)</td>
<td>p(_{diff})^c</td>
<td>p(_{trend})^d</td>
</tr>
<tr>
<td>Total estrogens and estrogen metabolites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent estrogens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrone</td>
<td>-2.8</td>
<td>0.65</td>
<td>0.66</td>
</tr>
<tr>
<td>Estradiol</td>
<td>-3.9</td>
<td>0.58</td>
<td>0.77</td>
</tr>
<tr>
<td>Hydroxyestrone</td>
<td>0.7</td>
<td>0.92</td>
<td>0.82</td>
</tr>
<tr>
<td>Hydroxyestriadiol</td>
<td>1.4</td>
<td>0.84</td>
<td>0.51</td>
</tr>
<tr>
<td>Catechols</td>
<td>-3.2</td>
<td>0.72</td>
<td>0.48</td>
</tr>
<tr>
<td>2-Catechols</td>
<td>-1.5</td>
<td>0.88</td>
<td>0.64</td>
</tr>
<tr>
<td>2-Hydroxyestrone</td>
<td>-1.3</td>
<td>0.89</td>
<td>0.64</td>
</tr>
<tr>
<td>2-Hydroxyestriadiol</td>
<td>-1.7</td>
<td>0.86</td>
<td>0.72</td>
</tr>
<tr>
<td>4-Catechols</td>
<td>-15.5</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>4-Hydroxyestrone</td>
<td>-6.4</td>
<td>0.45</td>
<td>0.48</td>
</tr>
<tr>
<td>Methylated catechols</td>
<td>-6.9</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td>Methylated 2-catechols</td>
<td>-7.9</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>2-Methoxyestrone</td>
<td>-0.5</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>2-Methoxyestriadiol</td>
<td>-8.0</td>
<td>0.37</td>
<td>0.44</td>
</tr>
<tr>
<td>Methylated 4-catechols</td>
<td>-7.0</td>
<td>0.56</td>
<td>0.49</td>
</tr>
<tr>
<td>4-Methoxyestrone</td>
<td>-7.1</td>
<td>0.59</td>
<td>0.54</td>
</tr>
<tr>
<td>4-Methoxyestriadiol</td>
<td>-16.4</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>2-Hydroxylation pathway</td>
<td>-3.1</td>
<td>0.73</td>
<td>0.54</td>
</tr>
<tr>
<td>4-Hydroxylation pathway</td>
<td>-12.0</td>
<td>0.26</td>
<td>0.18</td>
</tr>
<tr>
<td>16-Hydroxylation pathway</td>
<td>-2.2</td>
<td>0.78</td>
<td>0.88</td>
</tr>
<tr>
<td>16α-Hydroxyestrone</td>
<td>-4.0</td>
<td>0.70</td>
<td>0.87</td>
</tr>
<tr>
<td>Estriol</td>
<td>-4.0</td>
<td>0.66</td>
<td>0.87</td>
</tr>
<tr>
<td>17-Epiestriol</td>
<td>4.6</td>
<td>0.73</td>
<td>0.67</td>
</tr>
<tr>
<td>16-Ketoestriadiol</td>
<td>-3.1</td>
<td>0.68</td>
<td>0.81</td>
</tr>
<tr>
<td>16-Epiestriol</td>
<td>0.9</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>Ratios of metabolic pathways</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Catechols/2-catechols</td>
<td>5.7</td>
<td>0.63</td>
<td>0.72</td>
</tr>
<tr>
<td>2-Catechols/16-pathway</td>
<td>-0.7</td>
<td>0.95</td>
<td>0.47</td>
</tr>
<tr>
<td>Catechols/16-pathway</td>
<td>-8.1</td>
<td>0.45</td>
<td>0.15</td>
</tr>
<tr>
<td>4-Pathway/2-pathway</td>
<td>0.4</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>2-Pathway/16-pathway</td>
<td>-7.6</td>
<td>0.47</td>
<td>0.18</td>
</tr>
<tr>
<td>4-pathway/16-pathway</td>
<td>-8.3</td>
<td>0.49</td>
<td>0.21</td>
</tr>
<tr>
<td>2,4-pathway/16-pathway</td>
<td>-8.2</td>
<td>0.43</td>
<td>0.14</td>
</tr>
</tbody>
</table>
Supplementary table 2.S4 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Non-dairy animal protein(^e) (&gt;10 vs. ≤6 % energy)</th>
<th>Dairy protein(^e) (&gt; 5 vs. ≤2.2 % energy)</th>
<th>Vegetable protein(^e) (&gt; 6.3 vs. ≤4.4 % energy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff(^b)</td>
<td>(p_{\text{diff}})^c</td>
<td>(p_{\text{trend}})^d</td>
</tr>
<tr>
<td>2-pathway/4,16-pathway</td>
<td>0.2</td>
<td>0.98</td>
<td>0.60</td>
</tr>
<tr>
<td>2-Catechols/methylated 2-catechols</td>
<td>-0.8</td>
<td>0.91</td>
<td>0.60</td>
</tr>
<tr>
<td>4-Catechols/methylated 4-catechols</td>
<td>-10.4</td>
<td>0.52</td>
<td>0.45</td>
</tr>
<tr>
<td>Catechols/methylated catechols</td>
<td>-2.2</td>
<td>0.75</td>
<td>0.42</td>
</tr>
<tr>
<td>Parent estrogens/EM</td>
<td>6.2</td>
<td>0.26</td>
<td>0.21</td>
</tr>
<tr>
<td>2-Pathway/parent estrogens</td>
<td>-7.7</td>
<td>0.26</td>
<td>0.09</td>
</tr>
<tr>
<td>4-Pathway/parent estrogens</td>
<td>-7.9</td>
<td>0.43</td>
<td>0.21</td>
</tr>
<tr>
<td>16-Pathway/parent estrogens</td>
<td>-0.7</td>
<td>0.93</td>
<td>0.77</td>
</tr>
<tr>
<td>2-Hydroxyestrone/16α-hydroxyestrone</td>
<td>1.3</td>
<td>0.92</td>
<td>0.67</td>
</tr>
</tbody>
</table>

\(^a\)Adjusted for age (years; continuous), laboratory batch, luteal day (3-5, 6-7, 8-9, 10-28 days before the next period), first morning urine (yes/no), BMI (kg/m\(^2\); continuous), alcohol consumption (percent energy; continuous), physical activity (MET-hr/wk; continuous), menstrual cycle length (<26, 26-31, 32+ days), menstrual cycle pattern (extremely regular/very regular/regular/usually or always irregular), current smoking (yes/no), age at first birth/parity (nulliparous/1-2 children at <25 years/1-2 children at 25-29 years/1-2 children at 30+ years/3+ children at <25 years/3+ children at 25+ years), height (inches; continuous), anovulatory cycle (yes/no), and total energy intake (kcal/d; continuous)

\(^b\) % diff indicates the percentage of difference in mean EM levels (pmol/mg creatinine) comparing the highest intake vs. the lowest intake groups.

\(^c\) \(p_{\text{diff}}\) was estimated using a Wald test for an indicator variable comparing mean EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

\(^d\) \(p_{\text{trend}}\) was estimated using a Wald test for a continuous variable comparing quartile-specific median EM levels (pmol/mg creatinine) of the highest intake vs. the lowest intake groups.

\(^e\) The highest and the lowest intake groups were categorized based on quartiles of the nutrient intake.
REFERENCES


22. Young, L.R., et al., Total dietary fat and omega-3 fatty acids have modest effects on urinary sex hormones in postmenopausal women. Nutrition & Metabolism, 2013. 10.


CHAPTER 3

Expression of estrogen receptor, progesterone receptor, and Ki67 in normal breast tissue and subsequent risk of breast cancer

Hannah Oh¹, ², ³
A. Heather Eliassen¹, ²
Molin Wang¹, ², ⁴
Stephanie A. Smith-Warner¹, ³
Andrew H. Beck⁵
Stuart J. Schnitt⁵
Laura C. Collins⁵
James L. Connolly⁵
Laleh Montaser-Kouhsari⁵
Rulla M. Tamimi¹, ²

¹ Department of Epidemiology, Harvard School of Public Health, Boston, MA
² Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA
³ Department of Nutrition, Harvard School of Public Health, Boston, MA
⁴ Department of Biostatistics, Harvard School of Public Health, Boston, MA
⁵ Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA
ABSTRACT

Background: Biological activity, including potential carcinogenic effects, of estrogen and progesterone in breast tissue is primarily mediated by their receptors in the tissue. Ki67 is a marker of cell cycle activation. We examined the associations of estrogen receptor (ER), progesterone receptor (PR), and Ki67 expression in normal breast tissue from benign biopsies with subsequent breast cancer risk.

Methods: We conducted an analysis among 385 women (90 cases, 295 controls) with benign breast disease (BBD) in a nested case-control study within the Nurses' Health Study (NHS) and the NHSII. Tissue microarrays (TMA) were constructed using cores obtained from benign biopsies containing normal terminal duct lobular units (TDLU). Immunohistochemical staining for ER, PR, and Ki67 was performed on sections cut from the TMAs. Staining results were interpreted by computational image analysis that scored the percentage of positively stained cells for each marker. Unconditional logistic regression models, adjusting for matching factors and benign lesion subtype, were used to estimate odds ratios (OR) for developing subsequent breast cancer by tertiles of marker expression.

Results: ER and Ki67 expression (highest vs. lowest tertiles) in normal breast TDLUs was not significantly associated with subsequent breast cancer risk (≥14.7 vs. <7.3 % ER-positive cells: OR=0.55, 95% CI=0.21-1.44, p-trend=0.85; ≥6.2 vs. < 2.4 % Ki67-positive cells: OR=1.75, 95% CI=0.87-3.50, p-trend=0.15). PR expression was suggestively positively associated with breast cancer risk (≥9 vs. <4%: OR=2.08, 95% CI=1.00-4.31, p-trend=0.06); the positive association was significant among women who were premenopausal at BBD biopsy (OR=3.55, 95% CI=1.28-9.87, p-trend=0.03).
**Conclusion:** PR expression in normal breast tissue was significantly positively associated with subsequent breast cancer risk in premenopausal women. Although we did not observe significant results with ER and Ki67, we cannot exclude associations given the limited power in this study. Further studies are required to confirm these results.
INTRODUCTION

Development of breast neoplasia involves hormones, including estrogens and progesterone, that regulate cell proliferation and apoptosis. Hormone-related factors such as reproductive factors (e.g., age at menarche, parity, age at first birth) [1-3], exogenous estrogens (e.g., oral contraceptives [4, 5], postmenopausal hormone therapy [1, 6, 7]), and postmenopausal obesity (adipose tissue serves as an endocrine organ that produces estrogens via aromatase activity) [1, 8] are among the most important risk factors for breast cancer. Furthermore, tamoxifen, an estrogen receptor antagonist in breast tissue, is also used to prevent and treat hormone receptor-positive breast cancer [9, 10]. Epidemiologic evidence consistently supports a positive association between plasma estrogens and breast cancer risk in postmenopausal women [11, 12], although the evidence is mixed in premenopausal women [12]. Evidence for a role of plasma progesterone in breast cancer risk is primarily from small studies and is somewhat inconsistent [12].

The tissue-specific responsiveness to these hormones is partially regulated by the tissue expression of receptors that bind them [13-15]. For example, when estrogen binds to the estrogen receptor (ER), it activates a transduction-signaling pathway that increases the expression of genes involved in cell division and DNA replication [13, 16]. Highly proliferating cells are at high risk of mutagenesis because there is less time available for DNA repair [16]. Mammary tissues with high levels of ER expression are more likely to bind estrogens and develop tumors by initiating mutations and stimulating proliferation of mutated mammary cells [13, 16]. Although progesterone and the
progesterone receptor (PR) are not well-studied, progesterone is believed to stimulate cell proliferation and contribute to breast carcinogenesis via similar mechanisms [15].

Ki67 is a cell proliferation marker, as it is present only during active phases of the cell cycle (G1, S, G2, and M) and its level peaks at mitosis [17, 18]. Although the exact role of Ki67 in breast tumorigenesis is poorly understood, Ki67 expression is positively correlated with several breast cancer risk factors including breast density and precancerous lesions [17, 19, 20]. Ki67 also may play a crucial role in cell division, as the disruption of Ki67 activity results in arrest of cell proliferation in vitro [21, 22].

In breast cancer, expression of ER, PR and Ki67 contribute to different tumor subtype classifications (e.g., ER+ and/or PR+, luminal A/B) and serve as predictive and prognostic factors [23]. However, little is known about ER, PR, and Ki67 expression in normal breast tissue. In a nested case-control study within the Nurses’ Health Study (NHS) and NHSII cohorts, we examined the associations of ER, PR, and Ki67 expression levels in normal terminal duct lobular units (TDLUs) with subsequent risk of breast cancer among women with a previous diagnosis of benign breast disease (BBD). We hypothesized that the women with higher levels of ER, PR, and Ki67 expression in their normal TDLUs may be at higher risk of subsequent breast cancer.

MATERIALS AND METHODS

Study design and population

This study is a case-control study nested within the subcohort of women who reported a diagnosis of biopsy-confirmed BBD in the NHS and NHSII cohorts. The NHS is an ongoing cohort study that began in 1976, including 121,700 female registered
nurses aged 30-55 years. The NHSII is an ongoing cohort study that began in 1989, including 116,430 female registered nurses aged 25-42 years. In both cohorts, participants completed initial mailed, self-administered questionnaires collected information on participants’ health behaviors, lifestyle factors, reproductive factors, and medical histories. Subsequent biennial follow-up questionnaires were used to assess updated information on a variety of known and suspected risk factors for chronic diseases, as well as newly diagnosed diseases (e.g., BBD, breast cancer) that were then confirmed via medical record review. Details of this nested case-control study and the BBD assessment have been previously described [24, 25]. Cumulative response rates for both cohorts are >90% and are similar among women regardless of their BBD diagnosis.

Cases were women with biopsy-confirmed BBD who reported a diagnosis of breast cancer following their BBD diagnosis. Cases of breast cancer were diagnosed during 1976-1998 for the NHS and 1989-1999 for the NHSII. When possible, 4 controls were selected for each case, matched on year of birth and year of benign breast biopsy, among women with biopsy-confirmed BBD who remained free of breast cancer at the time the matching case was diagnosed. We attempted to obtain BBD pathology records and archived biopsy specimens for all cases and controls from their hospital pathology departments; our ability to obtain biopsy blocks did not significantly differ by case and control status. To reduce potential reverse causation due to subclinical tissue change, women were excluded if they had evidence of in situ or invasive carcinoma at biopsy or reported a diagnosis of breast cancer within 6 months of their biopsy (n=24).
This investigation was approved by the Institutional Review Board of the Brigham and Women’s Hospital. Completion of the self-administered questionnaire was presumed to imply informed consent.

**Tissue microarray (TMA) construction and laboratory assays**

Two study pathologists (SJS, JLC) independently reviewed the Hematoxylin and eosin (H&E) slides from the biopsy blocks and completed detailed worksheets on the subtype of BBD lesion (i.e., non-proliferative, proliferative without atypia, atypical ductal hyperplasia, atypical lobular hyperplasia) [26] in a blinded manner. After collecting eligible participants’ archived formalin-fixed paraffin-embedded breast biopsy blocks, the corresponding H&E slides were re-reviewed by a single pathologist (JLC) to identify the areas of lesions and normal TDLUs and to circle the target areas for TMA coring. Six TMA blocks were constructed in the Dana Farber Harvard Cancer Center Tissue Microarray Core Facility, Boston, MA by obtaining 0.6-mm cores from the targeted area (up to 3 cores for normal TDLU) in each donor block and inserting them into the recipient TMA blocks. We previously evaluated our TMA construction methods and confirmed a high success rate (76%) of capturing normal TDLUs in these TMA blocks [27].

For each immunohistochemical stain (i.e., ER, PR, Ki67), a 5-µm paraffin section was cut from each TMA block and immunostained with its antibodies (ERα: RM-9101-S, Neomarkers, CA; PR: M3569, Dako Corporation, CA; Ki67: VP-RM04, Vector Laboratories, CA) after deparaffinizing the section in two 5-min changes of xylene and
rehydration through graded alcohols to distilled water. Appropriate positive and negative controls were included in all staining runs. (ER refers to ERα throughout the paper)

Immunostaining results of each core were interpreted using an automated computational image analysis system (Definiens Tissue Studio software, Munich, Germany). For each stain, we used the Tissue Studio software to define an intensity and size threshold for nucleus identification and to define an intensity threshold for nuclear stain positivity. Based on these criteria, each cell was classified as positive or negative. Based on the intensity of the nuclear stain, positive cells were further categorized into low-positive, medium-positive, or high-positive. For each woman, we estimated the mean percentage of stain-positive cells (at any intensity) across the cores, by weighting each core by its total cell count. Percent stain-positive cells were then categorized into tertiles (ER: <7.3, 7.3-14.6, ≥14.7 %; PR: <4, 4-8.9, ≥9 %; Ki67: <2.4, 2.5-6.1, ≥6.2 %). In order to ensure enough cells to estimate the percentage of ER/PR/Ki67 expression, we excluded women with less than 100 total cell count (sum of all the cores); less strict cutoff points (<30, <50 cells) were used in sensitivity analyses. A total of 295 controls and 90 cases with at least one core of normal TDLUs and evaluable ER/PR/Ki67 staining were included in our analyses.

Because intensity of the stains is influenced by both storage time (age of block) and variability in processing [28], as well as the magnitude of expression, we considered intensity in our secondary analyses only by calculating a score encompassing the percentage of stain-positive cells weighed by the intensity of each cell (i.e., intensity score = %low-positive + 2* %medium-positive + 3* %high-positive). To account for potential heterogeneity within the tissue, we also repeated the analyses
using the core with the maximum percentage of positive-stained cells. To assess heterogeneity across the cores, we calculated intraclass correlation coefficients (ICCs) by dividing the between-person variances by the sum of the within- and the between-person variances that were estimated using a linear mixed model.

**Statistical analysis**

To avoid losing data due to incomplete matched sets (tissue was not available for some cases and controls), unconditional logistic regression, adjusting for matching factors (age at biopsy, year at biopsy, time since biopsy), was performed to estimate odds ratios (OR) and 95% confidence intervals (CI) for breast cancer by tertiles of tissue marker expression. BBD subtype, an important potential confounder, was additionally adjusted for in the multivariate models. We performed a test for trend by including exposures (ER, PR, Ki67 expression) in the model as continuous variables (weighted mean percentage of stain-positive cells). We also assessed non-linearity of the association using cubic spline models with four computer-selected knots; we compared the model-fit of nonlinear vs. linear models using likelihood ratio tests, if linear models were significant. Both linear and non-linear models were assessed again after excluding potential outliers selected based on Quantile-Quantile plots (n=4 for ER, n=3 for PR, n=2 for Ki67). Since levels of ER, PR, and Ki67 expression may vary by menopausal status, we restricted our analyses to women who were premenopausal at BBD biopsy (ER: 25 cases and 87 controls, PR: 33 cases and 114 controls, Ki67: 44 cases and 161 controls) in sensitivity analyses. To assess whether the associations varied by BBD subtype (proliferative vs. non-proliferative), we also estimated the risk after cross-
classifying the study population by BBD subtype and tissue marker expression (above vs. below the median percentage of stain-positive cells). We used women with high marker expression (above median) and proliferative lesions as a common reference group because there were few cases in women with low marker expression (below median) and non-proliferative lesions. All statistical tests were two-sided with 5 % type I error. Analyses were conducted with SAS version 9 (SAS Institute).

RESULTS

The mean age at BBD biopsy was 45.4 years among 385 women (333 from the NHS and 52 from the NHSII) included in our analyses. Average time between the BBD biopsy and the breast cancer diagnosis was 9.0 years. Compared to controls, cases were more likely to be nulliparous, have proliferative lesions, have used oral contraceptives, have smoked, and have a family history of breast cancer (Table 3.1).

Among 174 women (42 cases and 132 controls) who had at least one core of normal TDLUs with evaluable ER staining, the percentage of ER-positive cells ranged from 0 to 44.6 % with a mean of 12.1 %, a standard deviation (SD) of 9.5 % and a median of 9.9 % (Table 3.1). Among 237 women (59 cases and 178 controls) who had at least one core of normal TDLUs with evaluable PR staining, the percentage of PR-positive cells ranged from 0 to 36.5 % with a mean of 8.4 % (SD=7.3 %, median=6.5 %). Among 283 women (66 cases and 217 controls) who had at least one core of normal TDLUs with evaluable Ki67 staining, the percentage of Ki67-positive cells ranged from 0 to 31.6 % with a mean of 6.1 % (SD=6.3 %, median=4.2 %). The ICCs across cores were 0.40 for ER, 0.33 for PR, and 0.24 for Ki67 expression. The
correlation between ER and PR expression was modest (Spearman r=0.35) but minimal between ER and Ki67 (r= -0.03), and between PR and Ki67 (r= -0.12) in normal breast tissue (Supplementary Table 3.S1). The correlations were weak when comparing marker expression in normal breast tissue with that in the subsequent breast tumor tissue among 27 breast cancer cases for whom Definiens-measured IHC results were available from their breast tumor TMAs (Spearman r=0.02 – 0.24) (Supplementary Table 3.S2).

ER expression in normal breast TDLUs was not significantly associated with the risk of subsequent breast cancer when comparing the highest vs. lowest tertiles (≥14.7 vs. <7.3 %: multivariate OR=0.55, 95% CI=0.21-1.44, p-trend=0.85) (Table 3.2); similar results were observed when restricted to premenopausal women (data not shown). PR expression was suggestively positively associated with breast cancer risk (≥9 vs. <4%: OR=2.08, 95% CI=1.00-4.31, p-trend=0.06); the positive association was significant when restricting the analysis to premenopausal women (OR=3.55, 95% CI=1.28-9.87, p-trend=0.03, p_{non-linear}=0.15) (data not shown). Ki67 expression was not significantly associated with the risk of subsequent breast cancer (≥6.2 vs. <2.4%: OR=1.75, 95% CI=0.87-3.50, p-trend=0.15); results were similar in the subgroup of premenopausal women (data not shown).

When cross-classified by BBD subtype, the risk was highest in women with high PR expression (above median: ≥5.6 %) and proliferative lesions; their risk was significantly higher compared to women with low PR expression (below median: <5.6 %) regardless of BBD subtype (low PR and non-proliferative lesion vs. high PR and proliferative lesion: RR=0.19, 95% CI=0.05-0.69; low PR and proliferative lesion vs. high
PR and proliferative lesion: RR=0.46, 95% CI=0.22-0.93) (Table 3.3). The risk was also significantly higher in women with high Ki67 expression (above median: ≥4.3 %) and proliferative lesion compared to women with low Ki67 expression (below median: <4.3 %) and non-proliferative lesions (low Ki67 and non-proliferative lesion vs. high Ki67 and proliferative lesion: RR=0.20, 95% CI=0.04-0.97).

When we mutually adjusted for PR expression in the model, given the association we observed between PR expression and breast cancer risk, ER expression was significantly inversely associated (OR=0.28, 95% CI=0.08-0.95) and Ki67 expression was significantly positively association with breast cancer risk (OR=2.88, 95% CI=1.07-7.74, p-trend=0.04) (data not shown).

Results did not change after excluding potential outliers (data not shown) or after additionally adjusting for other important breast cancer risk factors: age at menarche (continuous, years), height (continuous, inches), BMI at age 18 years (continuous, kg/m²), family history of breast cancer (yes/no), ever oral contraceptive use (yes/no) (data not shown). Results also were similar for analyses of intensity scores or when the core with maximum expression was analyzed (data not shown). Similar patterns were observed in sensitivity analyses using < 30 cell or < 50 cell cutoff points for exclusion criteria (data not shown).

**DISCUSSION**

In this study of normal breast TDLUs from benign biopsies, ER and Ki67 expression was not significantly associated with the risk of subsequent breast cancer. In contrast, PR expression in normal breast TDLUs was suggestively positively associated
with the risk of subsequent breast cancer; the positive association was significant among premenopausal women.

ER, PR, and Ki67 are often overexpressed in breast cancers; approximately 70% of invasive breast carcinomas express ER [29], 60% express PR [30], and some show high, but variable, levels of Ki67 expression [31], suggesting their potential roles in breast carcinogenesis. Some evidence suggests ER and PR expression increase breast cancer risk by mediating the biologic activity of estrogens and progesterone, which stimulate proliferation of surrounding cells via paracrine signaling (e.g., growth factors) [29]. In normal breast tissue of cancer-free women, much lower mean levels of expression have been observed compared to breast cancers [29, 31-35]: 3-14%, 12-30%, and 0.9-1% for ER, PR, and Ki67, respectively. With a larger sample size than these prior studies, we observed comparable levels of ER expression (mean 12%, median 10%) in normal breast TDLUs from cancer-free benign biopsies, although PR expression (mean 8%, median 7%) was somewhat lower and Ki67 (mean 6%, median 4%) was somewhat higher than previously reported, possibly due to variations in assay methods and population characteristics (e.g., menopausal status, age, BBD diagnosis and subtype).

Beyond assessing the distribution and the expression levels of these markers in different breast tissue types (e.g., normal, benign lesions, cancers), very few studies have examined their expression in normal breast tissue in relation to breast cancer risk. Lawson et al. compared cancer-free white Australian (n=74) vs. Japanese women (n=92) as proxy groups for women at high risk and women at low risk for breast cancer and found higher ER expression in normal breast tissue of Australian women (14 vs. 9
%, suggesting these tissue markers may explain the risk difference between the two groups. In an earlier hospital-based case-control study, Khan et al. reported women with a previous diagnosis (< 3 years) of breast cancer (n=51) were 6.5 times (OR=6.5, 95% CI=1.5-27.4) more likely to express >1% ER in their benign epithelium compared to women without a history of breast cancer (n=69) [37]; however, with their retrospective design, the difference in ER expression may be due to reverse causation (e.g., cancers increasing ER expression of surrounding normal breast tissue through a paracrine effect). In the same study, no significant association with PR expression was observed. In a prospective study of women with atypical hyperplasia (n=192), Santisteban et al. reported a 4-fold higher risk of developing breast cancer within 10 years among women with higher Ki67 expression (≥ 2 vs. <2%) in their benign biopsies showing atypical hyperplasia [35]. We previously observed expression of IGF1R, another hormone receptor associated with cell growth and proliferation [38, 39], in normal breast TDLUs in benign biopsies was significantly positively associated with subsequent breast cancer risk within the NHS and NHSII [25], further supporting the role of proliferation and tissue-specific receptor expression in breast carcinogenesis.

Some experimental studies also support positive associations of ER and PR expression with breast tumorigenesis; ER knock-out mice [40] and PR knock-out mice [41, 42] were resistant to developing tumors when introduced with oncogenes or carcinogen challenge. In contrast to findings from previous studies, we did not observe significant associations of ER and Ki67 expression with breast cancer risk.

We acknowledge several limitations of this study. Given the small sample size, our confidence intervals for the OR estimates were wide. Also, not all eligible women
were included in the analyses; women were excluded if they did not provide their biopsy specimens, did not have normal TDLUs or enough cells in their tissue blocks, or did not have evaluable staining of ER, PR, or Ki67. Hospitals generally did not store biopsy blocks for more than 10 years; thus, we were not able to obtain tissue blocks from women who were diagnosed with BBD a long time ago. However, our ability to obtain biopsy blocks did not significantly differ by breast cancer case and control status.

Another limitation of this study is the use of TMAs containing up to 3 cores per woman; although we have previously shown that 3 cores reasonably capture normal TDLUs from biopsy blocks [27], sampling variability in coring may result in measurement error. Our low ICCs (0.23-0.40) indicate large within-person variability among the cores, possibly due to true heterogeneity in tissues [43] and methodological artifacts.

Furthermore, our tissue samples among premenopausal women were not timed within menstrual cycles. Levels of ER and Ki67 expression in normal breast tissue vary depending on the phase of the cycle (follicular vs. ovulatory vs. luteal) [44-46]; PR levels do not [46]. Therefore, random variation in these exposures could have attenuated the results. Lastly, our analyses were conducted among a subgroup of women with a previous diagnosis of BBD. Little is known about how ER, PR, and Ki67 expression levels in normal breast TDLUs among this subgroup of women who had benign breast biopsies compare with those among healthy women who have never had a breast biopsy. Thus, our results may not be generalizable to women without BBD.

Despite these limitations, this is one of few prospective studies to examine the associations of ER, PR, and Ki67 expression levels in normal breast TDLUs with
subsequent breast cancer risk. Furthermore, by using an automated image analysis system, we reduced observer error and measurement error in exposure.

In summary, while ER and Ki67 expression in normal breast tissue at the time of benign biopsy was not significantly associated with subsequent breast cancer risk, levels of PR expression were positively associated with the risk of developing breast cancer after diagnosis of BBD, particularly among women premenopausal at the time of BBD biopsy. These findings contribute to our understanding of breast cancer biology and may suggest new targets for breast cancer risk assessment and prevention. However, further studies are required to confirm these results.
Table 3.1. Characteristics of the study population at benign breast biopsy by breast cancer case-control status in the Nurses' Health Study and the Nurses' Health Study II

<table>
<thead>
<tr>
<th>Stained for ER expression in normal breast TDLUs</th>
<th>Controls (n=295)</th>
<th>Cases (n=90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>132</td>
<td>42</td>
</tr>
<tr>
<td>Percentage of ER-positive cells, range</td>
<td>0 – 44.6</td>
<td>2.1 – 38.6</td>
</tr>
<tr>
<td>Percentage of ER-positive cells, mean (SD)</td>
<td>12.7 (9.7)</td>
<td>11.1 (9.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stained for PR expression in normal breast TDLUs</th>
<th>Controls (n=295)</th>
<th>Cases (n=90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>178</td>
<td>59</td>
</tr>
<tr>
<td>Percentage of PR-positive cell, range</td>
<td>0 – 32.0</td>
<td>0 – 36.5</td>
</tr>
<tr>
<td>Percentage of PR-positive cell, mean (SD)</td>
<td>7.9 (7.1)</td>
<td>10.9 (7.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stained for Ki67 expression in normal breast TDLUs</th>
<th>Controls (n=295)</th>
<th>Cases (n=90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>217</td>
<td>66</td>
</tr>
<tr>
<td>Percentage of Ki67-positive cell, range</td>
<td>0 – 31.6</td>
<td>0 – 27.7</td>
</tr>
<tr>
<td>Percentage of Ki67-positive cells, mean (SD)</td>
<td>5.8 (6.1)</td>
<td>7.4 (7.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean (SD) or Percentage</th>
<th>Controls (n=295)</th>
<th>Cases (n=90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>45.4 (8.9)</td>
<td>45.4 (9.5)</td>
</tr>
<tr>
<td>Type of benign lesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Non-proliferative, %</td>
<td>30.6</td>
<td>19.0</td>
</tr>
<tr>
<td>- Proliferative without atypia, %</td>
<td>55.0</td>
<td>51.6</td>
</tr>
<tr>
<td>- Proliferative with atypical hyperplasia, %</td>
<td>14.4</td>
<td>29.3</td>
</tr>
<tr>
<td>Height, inches</td>
<td>64.7(2.6)</td>
<td>64.3(2.2)</td>
</tr>
<tr>
<td>Average body size at ages 5-10 years(^a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Level 1 (most lean), %</td>
<td>34.6</td>
<td>33.6</td>
</tr>
<tr>
<td>- Level 1.5-2, %</td>
<td>28.6</td>
<td>43.0</td>
</tr>
<tr>
<td>- Level 2.5-3, %</td>
<td>16.2</td>
<td>7.3</td>
</tr>
<tr>
<td>- Level 3.5-4, %</td>
<td>11.6</td>
<td>13.2</td>
</tr>
<tr>
<td>- Level 4.5+ (most overweight), %</td>
<td>9.0</td>
<td>2.9</td>
</tr>
<tr>
<td>BMI at age 18 years, kg/m(^2)</td>
<td>21.1(2.9)</td>
<td>20.9(2.1)</td>
</tr>
<tr>
<td>BMI at biopsy, kg/m(^2)</td>
<td>23.1(5.6)</td>
<td>22.5(6.3)</td>
</tr>
<tr>
<td>Age at menarche, years</td>
<td>12.6(1.4)</td>
<td>12.4(1.2)</td>
</tr>
<tr>
<td>Parous, %</td>
<td>93.6</td>
<td>91.1</td>
</tr>
<tr>
<td>- Parity (among parous women)</td>
<td>3.1(1.5)</td>
<td>3.4(1.6)</td>
</tr>
<tr>
<td>- Age at first birth (among parous women)</td>
<td>24.7(3.4)</td>
<td>25.3(3.2)</td>
</tr>
<tr>
<td>Premenopausal, %</td>
<td>70.7</td>
<td>67.5</td>
</tr>
<tr>
<td>Age at menopause (among postmenopausal women)</td>
<td>46.5(7.1)</td>
<td>47.4(4.5)</td>
</tr>
<tr>
<td>Ever oral contraceptive use, %</td>
<td>43.8</td>
<td>60.1</td>
</tr>
<tr>
<td>Ever smoking, %</td>
<td>48.6</td>
<td>64.5</td>
</tr>
<tr>
<td>First-degree family history of breast cancer, %</td>
<td>17.9</td>
<td>24.8</td>
</tr>
<tr>
<td>Cumulative average lifetime alcohol consumption(^b), g/d</td>
<td>4.6(6.8)</td>
<td>4.3(5.1)</td>
</tr>
<tr>
<td>Cumulative average adult physical activity(^c), MET-hr/wk</td>
<td>16.8(19.3)</td>
<td>12.5(12.1)</td>
</tr>
</tbody>
</table>

Values are means (SD) or percentages and are standardized to the age distribution of the study population.

Value is not age adjusted

Abbreviations: BMI=body mass index, kg=kilogram, g=gram, MET-hr/wk=metabolic equivalent of task-hour/week

\(^a\)Participants recalled their body size at ages 5 and 10 years using a 9-level pictogram (level 1: most lean; level 9: most overweight). We averaged body size at ages 5 and 10 years to obtain an estimate of childhood body size.

\(^b\)Cumulative average of alcohol consumption starting at age 18 years to the years prior to benign biopsy.

\(^c\)Cumulative average of physical activity practiced during adulthood since the enrollment in the cohorts (1976 for the NHS and 1989 for the NHSII) to the years prior to benign biopsy. MET-hr/wk of total activity was estimated by multiplying the number of hr/wk of each activity with its corresponding average MET values (strenuous activity = 7 METs, moderate activity = 4.5 METs, walking = 3 METs) and summing the values from all activities [44].

90
Table 3.2. Odds ratios (95% confidence interval) of developing subsequent breast cancer according to tertiles of mean percentage of ER, PR, Ki67 expression in normal breast tissues in the Nurses’ Health Study and the Nurses’ Health Study II

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
<th>Model 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 2&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ER expression (N=174)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 7.3 %</td>
<td>17</td>
<td>43</td>
<td>1.0 (Ref)</td>
<td>1.0 (Ref)</td>
</tr>
<tr>
<td>7.3 – 14.6 %</td>
<td>15</td>
<td>46</td>
<td>0.71 (0.30, 1.70)</td>
<td>0.72 (0.30, 1.71)</td>
</tr>
<tr>
<td>≥ 14.7 %</td>
<td>10</td>
<td>43</td>
<td>0.54 (0.21, 1.41)</td>
<td>0.55 (0.21, 1.44)</td>
</tr>
<tr>
<td>Per 1% increase</td>
<td></td>
<td></td>
<td>1.00 (0.96, 1.04)</td>
<td>1.00 (0.96, 1.04)</td>
</tr>
<tr>
<td>p-trend</td>
<td></td>
<td></td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean PR expression (N=237)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 4.0 %</td>
<td>16</td>
<td>64</td>
<td>1.0 (Ref)</td>
<td>1.0 (Ref)</td>
</tr>
<tr>
<td>4.0 – 8.9 %</td>
<td>15</td>
<td>54</td>
<td>1.30 (0.58, 2.93)</td>
<td>1.31 (0.58, 2.98)</td>
</tr>
<tr>
<td>≥ 9.0 %</td>
<td>28</td>
<td>60</td>
<td>2.08 (1.01, 4.30)</td>
<td>2.08 (1.00, 4.31)</td>
</tr>
<tr>
<td>Per 1% increase</td>
<td></td>
<td></td>
<td>1.04 (1.00, 1.08)</td>
<td>1.04 (1.00, 1.08)</td>
</tr>
<tr>
<td>p-trend</td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Mean Ki67 expression (N=283)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 2.4 %</td>
<td>20</td>
<td>75</td>
<td>1.0 (Ref)</td>
<td>1.0 (Ref)</td>
</tr>
<tr>
<td>2.4 – 6.1 %</td>
<td>18</td>
<td>79</td>
<td>0.90 (0.43, 1.88)</td>
<td>0.93 (0.44, 1.96)</td>
</tr>
<tr>
<td>≥ 6.2 %</td>
<td>28</td>
<td>63</td>
<td>1.69 (0.85, 3.36)</td>
<td>1.75 (0.87, 3.50)</td>
</tr>
<tr>
<td>Per 1% increase</td>
<td></td>
<td></td>
<td>1.03 (0.99, 1.07)</td>
<td>1.03 (0.99, 1.08)</td>
</tr>
<tr>
<td>p-trend</td>
<td></td>
<td></td>
<td>0.19</td>
<td>0.15</td>
</tr>
</tbody>
</table>

<sup>a</sup>Model 1 (Matching factors only): age at diagnosis (continuous, years), calendar year of biopsy (<1970, 1970-1980, 1980+), time since biopsy (continuous, years)

<sup>b</sup>Model 2: Model 1 + proliferative benign lesion (yes/no)
Table 3.3. Odds ratios (95% confidence interval) of developing subsequent breast cancer according to ER, PR, Ki67 expression in normal breast tissues and benign lesion type in the Nurses’ Health Study and the Nurses’ Health Study II

<table>
<thead>
<tr>
<th>Tissue expression</th>
<th>Lesion type</th>
<th>Case</th>
<th>Control</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low ER (&lt;4.7 %)</td>
<td>Non-proliferative</td>
<td>5</td>
<td>17</td>
<td>1.19 (0.34, 4.13)</td>
</tr>
<tr>
<td></td>
<td>Proliferative</td>
<td>21</td>
<td>50</td>
<td>1.58 (0.70, 3.59)</td>
</tr>
<tr>
<td>High ER (4.7+ %)</td>
<td>Non-proliferative</td>
<td>3</td>
<td>17</td>
<td>0.69 (0.17, 2.82)</td>
</tr>
<tr>
<td></td>
<td>Proliferative</td>
<td>13</td>
<td>48</td>
<td>1.0 (Ref)</td>
</tr>
<tr>
<td>PR expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low PR (&lt;5.6 %)</td>
<td>Non-proliferative</td>
<td>3</td>
<td>28</td>
<td>0.19 (0.05, 0.69)</td>
</tr>
<tr>
<td></td>
<td>Proliferative</td>
<td>16</td>
<td>61</td>
<td>0.46 (0.22, 0.93)</td>
</tr>
<tr>
<td>High PR (5.6+ %)</td>
<td>Non-proliferative</td>
<td>7</td>
<td>25</td>
<td>0.53 (0.20, 1.39)</td>
</tr>
<tr>
<td></td>
<td>Proliferative</td>
<td>33</td>
<td>64</td>
<td>1.0 (Ref)</td>
</tr>
<tr>
<td>Ki67 expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Ki67 (&lt;4.3 %)</td>
<td>Non-proliferative</td>
<td>2</td>
<td>27</td>
<td>0.20 (0.04, 0.97)</td>
</tr>
<tr>
<td></td>
<td>Proliferative</td>
<td>30</td>
<td>85</td>
<td>0.99 (0.52, 1.87)</td>
</tr>
<tr>
<td>High Ki67 (4.3+ %)</td>
<td>Non-proliferative</td>
<td>8</td>
<td>31</td>
<td>0.71 (0.28, 1.80)</td>
</tr>
<tr>
<td></td>
<td>Proliferative</td>
<td>26</td>
<td>74</td>
<td>1.0 (Ref)</td>
</tr>
</tbody>
</table>

aER, PR, and Ki67 expressions in normal breast tissue were dichotomized around the median (above vs. below median percentage of stain-positive cells).
bOdds ratio and 95% confidence intervals were estimated using matching factors only model: age at diagnosis (continuous, years), calendar year of biopsy (<1970, 1970-1980, 1980+), time since biopsy (continuous, years)
### Supplementary table 3.S1. Spearman correlations among tissue markers in normal breast tissue in the Nurses’ Health Study and the Nurses’ Health Study II

<table>
<thead>
<tr>
<th></th>
<th>All women</th>
<th>Premenopausal women only</th>
<th>Postmenopausal women only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER</td>
<td>PR</td>
<td>Ki67</td>
</tr>
<tr>
<td>ER</td>
<td>1.0</td>
<td>0.35</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(n=149)</td>
<td>(n=110)</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>1.0</td>
<td>-0.12</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>(n=154)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ki67</td>
<td>1.0</td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Supplementary table 3.S2. Spearman correlations among tissue markers in normal breast tissue vs. breast tumor tissue among cases (n=27) in the Nurses’ Health Study and the Nurses’ Health Study II

<table>
<thead>
<tr>
<th>Normal tissue</th>
<th>ER</th>
<th>PR</th>
<th>Ki67</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>0.02</td>
<td>0.06</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(n=19)</td>
<td>(n=19)</td>
<td>(n=20)</td>
</tr>
<tr>
<td>PR</td>
<td>-0.24</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(n=24)</td>
<td>(n=21)</td>
<td>(n=27)</td>
</tr>
<tr>
<td>Ki67</td>
<td>-0.12</td>
<td>-0.18</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(n=25)</td>
<td>(n=21)</td>
<td>(n=26)</td>
</tr>
</tbody>
</table>
REFERENCES


