Lifestyle Risk Factors and Rise of Breast Cancer Overall and by Subtypes Defined by Hormone Receptor Status

Lifestyle Risk Factors and Risk of Breast Cancer Overall and By Subtypes Defined by Hormone Receptor Status

A dissertation presented

by

You Wu

to

The Department of Nutrition

Harvard T.H. Chan School of Public Health

and

Harvard Graduate School of Art and Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of Nutritional Epidemiology

Harvard University

Cambridge, Massachusetts

April 2020

Dissertation Advisor: Dr. Stephanie Smith-Warner You Wu

Lifestyle Risk Factors and Risk of Breast Cancer Overall and by Subtypes Defined By Hormone Receptor Status

Abstract

Worldwide, breast cancer is the most commonly diagnosed malignancy in women with over 2 million cases diagnosed each year. Breast cancer is a heterogeneous disease that could be classified into subtypes based on the molecular features, for example, by hormone receptor status. The majority of breast cancer cases are estrogen receptor (ER) positive, and less than 20% are ER-negative. These subtypes have different etiologies, clinical characteristics, and survival rates. The effects of risk factors may also differ by hormone receptor status.

For Chapter 1 through 3, I examined potential dietary factors in three pooled analyses of diet and risk of breast cancer in an international consortium of more than 20 prospective cohorts. Over 1 million women were included in the analyses, among whom around 40,000 incidence breast cancer cases were documented. In each chapter, a 2-stage approach was used for data analysis. In stage 1, Cox proportional hazard regression was used to get study-specific hazard ratios for breast cancer overall and the subtypes defined by ER status. In stage 2, the study-specific multivariable HRs were pooled using random-effects model.

Chapter 1 focused on dietary fiber as an example of nutrient. We found that dietary fiber intake was inversely associated with breast cancer risk. The association could mainly be attributed to fiber from fruits and vegetables, but not grains, and was modified by fat intake, where the

association became weaker and nonsignificant among people with higher fat intake. In subtype analysis, the association was stronger for ER- tumors than that for ER+, although difference by ER status reached statistical significance only for fiber from vegetables.

In Chapter 2, dairy products were examined as examples of food items. Individual dairy products showed null or very weak inverse associations with risk of overall breast cancer. Differences by ER status were suggested for yogurt and cottage/ricotta cheese where associations were observed for ER-negative tumors only. Dietary calcium intake was only weakly associated with breast cancer risk, and the effect estimates did not differ by ER status.

In Chapter 3, the focus was on red meat and other major protein sources, to explore the substitution effects of different food groups on risk of breast cancer. Total red meat, processed meat, and unprocessed meat intakes were not significantly associated with risk of breast cancer when holding other protein sources constant. However, inverse associations were observed when substituting red meat with an energy-equivalent amount of mature beans or dairy products. There were no significant substitution effect replacing total or unprocessed red meat with poultry, seafood, eggs, or nuts. The results were similar for ER-positive and ER-negative breast cancer.

To quantify the theoretical impact of interventions, population attributable risk (PAR%) helps set priority and guide personal decision. Wide range of PAR% of cancer incidence by modifiable risk factors has been reported, yet there is no consensus on what contributed to the variation.

Chapter 4 investigated the PAR% of breast cancer by a group of modifiable risk factors and examined its variation by choices of exposures and methods. Fruits and vegetable intake, physical activity, adult weight gain, and alcohol consumption were the exposures of interest in the analysis. Partial PAR% was calculated from three models: baseline only, simple updates of

repeated measures, or cumulative averages of repeated measures. For each model, two methods were applied - one based on the four factors individually, and the other based on comparison between an overall high- versus low-risk group. The models based on repeated measures yielded greater estimated PAR%. PAR% estimates by the low-risk method were higher than that based on each factor individually, but in similar pattern. Therefore, PAR% by modifiable risk factors in current literatures likely underestimated the preventable fraction, if relied on studies with baseline data only.

In conclusion, the first three chapters found modest inverse associations with risk of breast cancer for dietary fiber, especially that from fruits and vegetables; for yogurt and cottage cheese consumption and ER- tumors; and when substituting red meat by beans or dairy. The last chapter emphasized the importance of high-quality repeated measure data in PAR% calculation and called for cautious interpretation of PAR% in the current literature.

Table of Contents

Title Page ····· i

Copyright ····· ii

Abstract ····· iii

Table of Contents ····· vi

List of Tables ····· vii

List of Figures ····· x

Acknowledgements ····· xi

Chapter 1. A pooled analysis of dietary fiber intake and risk of breast cancer overall and for subtypes defined by estrogen receptor status ····· 1

Introduction ····· 3

Methods ····· 4

Results ····· 8

Discussion ····· 23

Conclusion ····· 28

References ····· 29

Chapter 2. Dairy foods, calcium and breast cancer risk: a pooled analysis of 21 cohort studies ····· 34

Introduction ····· 35

Methods ····· 36

Results ····· 39

Discussion ····· 52

Conclusion ····· 55

References ····· 57

Chapter 3. Total red meat, unprocessed red meat, and processed meat and risk of breast cancer - a pooled analysis of substitution with alternative protein sources ····· 64
Introduction ····· 66
Methods ····· 67
Results ····· 72
Discussion ····· 79
Conclusion ····· 84
References ····· 86

Chapter 4. Examining the influence of timing of measurement on assessing preventability of breast cancer by lifestyle risk factors ····· 88
Introduction ····· 90
Methods ····· 91
Results ····· 95
Discussion ····· 100
Conclusion ····· 103
References ····· 104

Supplemental Materials ····· 106

List of Tables

Table 1.1	Characteristics of the cohort studies included in the pooled analysis of dietary fiber intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer	9
Table 1.2	Distribution of intake of total dietary fiber and source-specific dietary fiber	11
Table 1.3	Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for total dietary fiber intake	13
Table 1.4	Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for source-specific dietary fiber intake	17
Table 1.5	Pooled HRs (95% CIs) for breast cancer overall and subtypes defined by ER status according to the combined effects of dietary fiber and alcohol	20
Table 1.6	Pooled multivariable HRs (95% CIs) per 10g/day of dietary fiber intake and risk of breast cancer by age at diagnosis, menopausal status, total fat intake, BMI, follow-up duration, and region.	21
Table 2.1	Characteristics of the cohort studies included in the pooled analysis of dietary fiber intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer	40
Table 2.2	Daily median intakes of dairy products and calcium by cohort study	42
Table 2.3	Pooled age and multivariable adjusted hazard ratios (HR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER) status according to intake of dairy products	44
Table 2.4	Pooled age and multivariable adjusted hazard ratios (HR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER) status according to calcium intake	49
Table 2.5	Pooled multivariable hazard adjusted ratios (MVHR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER) status stratified by participant characteristics	51

Table 3.1	Characteristics of the cohort studies included in the pooled analysis of red meat intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer	73
Table 3.2	Distribution of total red meat, processed meat, and unprocessed meat in the pooled analysis of red meat intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer	74
Table 3.3	Pooled multivariable hazard ratio (MVHR) of breast cancer overall and subtypes defined by ER status by major protein sources	76
Table 3.4	Substitution effect of alternative protein sources replacing total red meat[1] on risk of breast cancer by age at diagnosis, menopausal status, BMI, follow-up duration, and region	80
Table 4.1	Risk factors of breast cancer for postmenopausal women in the Nurses' Health Study	96
Table 4.2	Hazard ratio and 95% CI for breast cancer overall by WCRF cancer prevention recommendations using baseline, simple update, or cumulative average	97
Table 4.3	Prevalence, multivariable-adjusted hazard ratio, and partial PAR associated with high-risk groups of the risk factors	99
Table 4.4	Prevalence, multivariable-adjusted hazard ratio, and PAR associated with overall high-risk[1] profile	101
Table S1.1	Case status defined by PR or jointly by ER and PR and intake of dietary fiber from fruits, vegetables, and grains	107
Table S1.2a	Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for total dietary fiber intake	108
Table S1.2b	Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for sources of dietary fiber	109
Table S1.3	Pooled multivariable HRs (95% CIs) of total breast cancer for total dietary fiber intake adjusting for fruit and vegetable intake, vitamin C, and carotenoids	111

Table S1.4	Median Pearson correlations for dietary fiber, carotenoids, and vitamin C across all studies included in the breast cancer analyses in the Pooling Project of Prospective Studies of Diet and Cancer	114
Table S2.1	Number of cases of breast cancer subtypes defined by estrogen receptor (ER) and progesterone receptor (PR) status for cohort study included in the pooled analyses of dairy products and breast cancer risk	115
Table S2.2	Median Pearson correlations for dairy products and nutrients across all studies included in the breast cancer analyses in the Pooling Project of Prospective Studies of Diet and Cancer	116
Table S2.3	Pooled age and multivariable adjusted hazard ratios (HR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER) status according to calcium intake	117
Table S3.1	Median Pearson correlations[1] for food groups across all studies included in the breast cancer analyses in the Pooling Project of Prospective Studies of Diet and Cancer	118
Table S3.2	Sample conversions from grams to calories of common items in the food groups	119
Table S4.1	Multivariable-adjusted hazard ratio and 95% CI for breast cancer overall by WCRF cancer prevention recommendations using simple update data with lag	120

List of Figures

Figure 3.1 Substitution effect on risk of breast cancer overall and the subtypes defined by ER status when replacing red meat by alternative protein sources 78

Acknowledgements

My gratitude goes first and foremost to the participants, the fellow researchers, and the staff of all the cohort studies included in my thesis projects. The Pooling Project of Prospective Studies of Diet and Cancer (DCPP) would not have been possible without their commitment to improving human health.

I would like to thank my academic advisor, Dr. Stephanie Smith-Warner, for her extraordinary mentorship during my doctoral education. Learning from her devotion to rigorous research and attention to details, the past few years have shaped me into a reliable, responsible scientist. I would also like to thank my dissertation committee members, Dr. Walter Willett, Dr. Edward Giovannucci, and Dr. Molin Wang for their invaluable contribution to my research. The feedback, inspiration, and suggestions they gave me have taken this thesis to another level and made me confident about any new research question that's coming.

Thanks to my friends at Harvard Chan, and the peers in the PHS cohort. Our shared interests (in research and in life) brought us incredible friendship as well as a connected future. I would like to moreover acknowledge the administrative members of the Nutrition departments and the PHS office, for their amazing help to streamline everything at school.

I am grateful to my partner, Xiao Li, who was the first to encourage me to apply for the doctoral program and has since been the top supporter of my academic endeavors.

Lastly, I thank my parents for their unconditional love, care, and support. They raised me to be a person with kindness, positivity, strength, and integrity, so the world now has the best version of me.

Chapter 1

A pooled analysis of dietary fiber intake and risk of breast cancer overall and for subtypes defined by estrogen receptor status

ABSTRACT

Background: Dietary fiber intake may reduce breast cancer risk and attenuate the elevated risk of breast cancer with high alcohol consumption. Estrogen receptor negative (ER-) breast cancer is a more aggressive but less understood subtype than ER positive (ER+) subtype due to its low prevalence. As ER- breast cancer is less hormonally dependent, associations with diet may be more evident.

Method: This study included 24 prospective studies with over 1.5 million women among whom 56,844 (34,384 ER+; 7,828 ER-) breast cancer cases were diagnosed during follow-up. Using harmonized participant level data, study specific hazard ratios (HR) and 95% confidence intervals (CI) were estimated by Cox proportional hazards regression and pooled using a random-effects model.

Results: Results show that dietary fiber intake was inversely associated with breast cancer risk (pooled multivariable HR[MVHR] comparing the highest versus lowest quintile: 0.93, 95% CI 0.90-0.97). The association was stronger for ER- tumors (pooled MVHR comparing the highest versus lowest quintile = 0.87, 95% CI: 0.80-0.95) than that for ER+ (pooled MVHR = 0.95, 95% CI: 0.90-1.00), although difference by ER status did not reach statistical significance (p-value for common-effect for quintile 5 = 0.08). Secondary analyses show that the inverse association between dietary fiber and breast cancer (1) could mainly be attributed to fiber from fruits and vegetables, but not grains; (2) was modified by fat intake, where the association became weaker

and nonsignificant among people with higher fat intake (p-interaction = 0.028). Other potential factors, such as age of diagnosis, BMI, menopausal status, region, and follow-up years did not modify the inverse association between dietary fiber intake and breast cancer overall or the subtypes.

Conclusion: Higher intake of dietary fiber, especially fiber from vegetables and fruits, is associated with a modestly lower risk of breast cancer, particularly ER- breast cancer. These findings were consistently observed across subgroups defined by age, menopausal status and BMI, while we observed statistically significant effect modifications by alcohol consumption and fat intake. The associations for total breast cancer remained significant when we controlled for several potential bioactive constituents that are correlated with dietary fiber intake and are relevant to breast cancer risk, however the results for the tumor subtype were attenuated.

INTRODUCTION

Breast cancer, the leading cause of cancer death in women worldwide, is a heterogeneous disease in terms of etiology and clinical traits [1]. Estrogen receptor negative (ER-) breast cancer is a more aggressive but less understood subtype than ER positive (ER+) breast cancer due to its lower prevalence.

Dietary fiber is hypothesized to decrease breast cancer risk by indirectly decreasing estrogen binding and estrogen receptor transcription in breast cancer cells [2], decreasing intestinal reabsorption of estrogen and increases its fecal excretion, thus lowers the circulating estrogen level [3-5]. Dietary fiber may also influence risk of breast cancer through non-hormonal mechanisms. Butyrate, a product by fermentation of dietary fiber, has the potential to induce cell cycle arrest in G1 phase and apoptosis in a p53-independent manner, thus inhibit breast cancer cell growth [6,7]. Higher dietary fiber intake may also lead to reduced inflammation, with stronger association observed for ER- and PR- tumors[8].

Large cohort studies have also revealed an inverse association between dietary fiber intake and breast cancer, although not consistently across studies [9]. Further, few studies have reported results for breast cancer subtypes defined by hormone receptor status. As ER- breast cancer is less hormonally dependent than ER+ breast cancer, associations with diet may be more evident for ER- breast cancer. Taking advantage of the statistical power in the consortium, we examined the associations between dietary fiber intake and risk of breast cancer overall and by subtypes defined by ER and the combination of ER and PR status in 24 studies including over 1.5 million women participating in the Pooling Project of Prospective Studies of Diet and Cancer (DCPP). We also

conducted analyses by different sources of dietary fiber (fruits, vegetables, and grains), and tested whether the associations varied by other breast cancer risk factors.

METHODS

Study population. This study included 24 prospective cohort studies within the DCPP, a long-standing international consortium (Table 1). All participating studies met the following inclusion criteria: 1) ascertainment of at least 25 incident cases of ER negative breast cancer; 2) at least one publication on any diet and cancer analysis; 3) assessment of long-term dietary intake; 4) validation of the diet assessment method or a closely related instrument. All studies received approval by the institutional review board of their participating institutions.

Case ascertainment. Breast cancer was defined by International Classification of Diseases (ICD)-9 code 174.0 or ICD-10 code C50. Incident invasive breast cancer cases were identified in each study by self-report questionnaires with subsequent confirmation with medical record review, linkage with cancer registries, or both methods. Some studies additionally used mortality registries to ascertain cases. The follow-up rate generally exceeded 90% for the studies. Hormone receptor status was obtained from each study through cancer registries, pathology reports, medical records, or laboratory determinations. The cases with borderline ER/PR status were classified as positive for that receptor.

Dietary and non-dietary factors assessment. Dietary intake at baseline was assessed using a self-administered food frequency questionnaire (FFQ) or diet history asking about usual dietary intake generally in the past year. All studies inquired as to the frequency of consumption of food

and the usual amount of consumption. Each study estimated nutrient intakes by multiplying the frequency of consumption of each food item and their respective food composition data and then summing intakes across all food items to calculate overall daily intake of that nutrient. Nutrient intakes, including dietary fiber, were adjusted for total energy using the residual method. For the studies that assessed dietary fiber intake in their validation study, the correlation coefficients between intakes estimated from the FFQ or a closely related instrument, and the reference method, generally ranged from 0.4-0.7 for dietary fiber intake [10-26].Eighteen of the 24 studies classified dietary fiber intake into three major sources: fruits, vegetables, and grains. In secondary analyses, we compared dietary fiber intake from the three sources.

Information on non-dietary factors was collected using self-administered questionnaires at baseline. Studies collected demographics, age, height, weight, medical history, lifestyle, and menopausal status at baseline. Most studies had information on reproductive factors, exogenous hormone use, education attainment, physical activity, smoking status, and family history of breast cancer.

Statistical analyses. Within each study, participants were excluded based on study-specific exclusion criteria. We further excluded those who were diagnosed with cancer (other than non-melanoma skin cancer) before baseline, and who reported extreme energy intakes, i.e. outside of three standard deviations from study-specific $\log_e$-transformed mean energy intake. A more comprehensive dietary assessment was introduced in 1986 in the Nurses' Health Study, therefore we analyzed the Nurses' Health Study as two separate cohorts, [1980-1986, Nurses' Health Study (a), and 1986-2014, Nurses' Health Study (b)]. The Netherlands Cohort Study was analyzed as a

case-cohort study, since the dietary assessments were only processed for the cases and a random sample of the entire cohort.

We categorized total dietary fiber intake using both study-specific quintiles and groups defined by absolute intake cut points (<10, 10-<15, 15-<20, 20-<25, ≥25 g/day). Dietary fiber from different sources was categorized using study-specific quintiles and groups defined by absolute intake cut points (<3, 3-<6, 6-<9, 9-<12, ≥12 g/day).

We used a two-stage approach to calculate pooled hazard ratios. In stage 1, study-specific hazard ratios were estimated using Cox proportional hazard regression with 95% confidence intervals, for breast cancer overall and subtypes defined by ER or ER/PR status. We calculated person-years of follow-up from the date of baseline questionnaire return to the date of incident invasive breast cancer diagnosis, death, loss to follow-up, or end of follow-up, whichever came first. We adjusted all models for age at baseline and year of questionnaire return to control for age, calendar time, and time since entry to studies. In the multivariable analyses, we controlled for established and suspected confounders (as listed in table 2) directly in the models of studies with ≥ 200 cases of the outcome of interest. Otherwise, we adjusted for confounders using the propensity score method[27-29]. We handled missing data by creating missing indicator variables since the proportion of missing data was generally lower than 10%. In stage 2, the study-specific hazard ratios were pooled using a random-effects model, weighted by the sum of the inverse of the variance and the estimated between-studies variance components[30]. We tested for between-studies heterogeneity using the I^2 and Q statistic[31]. When we observed significant between-study heterogeneity in the results, we conducted meta-regression analyses to identify potential sources of heterogeneity of

study-level characteristics (median age, median follow-up duration of the population) were conducted. The p-value for the test for trend across categories was calculated by assigning the study-specific median value for each intake category to each individual and modelling the median as a continuous term. Nonlinearity was tested for models of total and source-specific dietary fiber intake and breast cancer overall and by subtypes, using restricted cubic spline terms selected by stepwise regression procedure [32,33]. For the associations in which the assumption of linearity held, we analyzed dietary fiber intake as a continuous variable.

Analyses for the three sources of dietary fiber were carried out individually and also when in the same model. We further adjusted for dietary (from foods only) intake of vitamin C, total fruit and vegetable, and carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lycopene, lutein) to investigate potential confounding by dietary factors with correlated sources of dietary fiber.

We evaluated whether the association between dietary fiber intake and breast cancer risk varied by: age at diagnosis (<55, 55-<65, ≥65 years); alcohol consumption (0, >0-<5, 5-<20, ≥20 g/day), BMI (<25, 25-<30 ≥30 kg/m^2); menopausal status at diagnosis (premenopausal, postmenopausal; estimated using a previously described algorithm based on age at diagnosis[34]); total fat intake (in tertiles); follow-up period (< 5, 5-<10, 10-<15, ≥15 years); and region (North America, others). The p-value for interaction was obtained by fitting the product term of the potential effect modifier and dietary fiber intake as a continuous interaction term in the model. Effect difference by region was evaluated using a mixed-effects meta-regression model. The test for statistical significance of different effect sizes by tumor subtype was conducted using the contrast test [35,36].

For studies that evaluated dietary fiber in their validation study, we corrected for the bias in the HRs (from continuous analysis) due to measurement error in dietary fiber intake, using a regression calibration method[37,38].

All statistical hypotheses were tested by calculating two-sided Wald 95% confidence intervals (CIs), and two-sided $p < 0.05$ was considered statistically significant. All analyses were conducted using SAS software versions 9.4 (SAS Institute, Inc., Cary, North Carolina).

The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

RESULTS

Across the 24 prospective cohort studies with a maximum of 6-21 years of follow-up among 1,545,757 women, 56,844 incident breast cancer cases were identified, including 34,384 ER+, 7,828 ER-, 27,306 PR+, and 12,217 PR- tumors (Table 1.1, Table S1.1). Median dietary fiber intake varied more than 2-fold across the studies (Table 1.2).

A weak inverse association was observed for total dietary fiber intake and risk of breast cancer overall (pooled age-adjusted HR comparing the highest versus lowest quintile = 0.89, 95% CI: 0.86-0.93; pooled multivariable HR [MVHR] = 0.93, 95% CI: 0.90-0.97, Table 1.3). There was little evidence of confounding, so we present only the results from the multivariable models for all

Table 1.1 Characteristics of the cohort studies included in the pooled analysis of dietary fiber intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer

Study	Baseline cohort size[2]	Baseline age range (yr)	Mean follow-up time (yr)	No. of cases[1]		
				Total	ER+	ER-
The NIH-AARP Diet and Health Study	200049	50-71	7.1	5972	2322	464
Breast Cancer Detection Demonstration Project Follow-up Study	42061	40-93	8.4	1305	793	166
Beta-Carotene and Retinol Efficacy Trial	6000	48-71	11.6	367	193	31
Campaign Against Cancer and Heart Disease	8279	18-93	14.9	288	198	50
Cancer Prevention Study II Nutrition Cohort	74137	40-87	9.3	2999	1835	323
California Teachers Study	100067	22-104	7.6	2696	1930	343
Canadian National Breast Screening Study	45185	40-59	16.3	1240	367	125
The European Prospective Investigation into Cancer and Nutrition	330292	19-98	14.1	10668	5514	1474
Iowa Women's Health Study	34584	52-71	15.9	1849	1329	238
The Japan Public Health Center-Based Study Cohort I	21609	40-59	13.7	289	111	69
Multiethnic Cohort Study	92435	45-78	10.1	3308	2169	543
Melbourne Collaborative Cohort Study	22456	31-75	13.2	799	493	171
Nurses' Health Study (part a)[3]	88618	34-67	6.3	1122	528	255
Nurses' Health Study (part b)	68394	40-67	24	5667	4272	907
Nurses' Health Study II	93778	26-46	21.3	2921	2170	576
The Netherlands Cohort Study[4]	62573	54-70	6.8	2013	700	183
New York University Women's Health Study	13257	31-70	15.7	919	392	121
The Hormones and Diet in the Etiology of Breast Cancer Study	9044	34-70	12	283	206	67
The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	28292	52-74	9.1	1090	858	137
Swedish Mammography Cohort	60950	38-76	15.2	2605	1605	384
The Vitamin D and Omega-3 trial	30158	50-76	9.1	1203	728	124
Women's Health Initiative	96034	49-81	16.2	4992	3997	694
Women's Health Study	38385	38-89	9.7	1177	937	187
Women's Lifestyle and Health Study	47514	32-50	14.9	1072	737	196
Total	1545757			56844	34384	7828

Table 1.1 (Continued)

1. ER, estrogen receptor; ER+, estrogen receptor positive; ER-, estrogen receptor negative.
2. Cohort size reflects the size after application of study-specific exclusion criteria and further exclusion of participants with energy intakes beyond 3 SDs of their loge-transformed study-specific mean energy intake and history of cancer diagnosis at baseline (except for nonmelanoma skin cancer).
3. The Nurses' Health Study were analyzed as two different cohorts [1980–1986, Nurses' Health Study (a); 1986–2006, Nurses' Health Study (b)]; The Nurses' Health Study (b) was not included in the total cohort size because these participants were included in the Nurses' Health Study (a).
4. The Netherlands Cohort Study was analyzed as a case-cohort study, and the above exclusions were not applied to its baseline cohort size.

Table 1.2 Distribution of intake of total dietary fiber and source-specific dietary fiber

Study Abbreviation	Total dietary fiber intake			Fiber from fruits			Fiber from vegetables			Fiber from grains		
	Median	10th Pctl	90th Pctl	Median	10th Pctl	90th Pctl	Median	10th Pctl	90th Pctl	Median	10th Pctl	90th Pctl
AARP	16.2	8.3	29.3	3.5	0.9	8.3	6.8	3.0	14.3	4.8	2.0	9.5
BCDDP	10.5	5.4	18.3	2.4	0.6	5.5	2.7	1.2	5.3	3.6	1.4	7.7
CARET	12.6	6.6	21.8	.	.	.	.	.	.	.	.	.
CLUE2	9.4	5.1	16.4	.	.	.	.	.	.	3.4	1.5	6.9
CPS2	11.0	5.9	18.0	2.5	0.8	5.7	3.6	1.7	6.9	3.7	1.6	7.1
CTS	13.3	7.4	22.4	.	.	.	.	.	.	.	.	.
CNBSS	16.7	10.0	26.5	3.9	1.1	8.1	7.4	4.0	13.7	3.8	1.8	6.8
EPIC	21.3	13.7	31.8	4.1	1.4	8.9	4.3	1.8	8.7	7.2	3.2	13.9
IWHS	19.0	11.0	30.9	4.5	1.5	9.1	7.5	3.9	12.9	4.9	2.1	9.8
JPHC1	9.5	6.0	12.8	.	.	.	.	.	.	.	.	.
MEC	16.1	7.6	33.1	4.9	1.2	13.9	7.0	3.0	17.3	6.0	2.4	14.1
MCCS	28.5	17.2	44.4	6.8	2.3	14.5	4.7	2.1	9.0	9.3	4.4	16.9
NHS 80	12.4	7.0	20.7	3.5	0.9	8.2	4.1	2.1	8.2	2.0	0.7	4.6
NHS 86	18.0	10.7	28.7	4.0	1.4	8.4	6.2	3.3	10.7	4.1	1.8	8.7
NHSII	17.4	10.1	28.2	2.7	0.9	6.1	6.0	2.8	11.9	5.3	2.7	9.7
NLCS	24.5	16.9	34.2	3.9	1.5	7.4	4.0	2.3	6.7	9.4	4.8	15.6
NYU	12.3	5.9	22.8	3.5	1.0	7.4	4.5	1.7	10.0	2.7	0.9	7.1
ORDET	19.0	12.4	26.9	5.2	1.8	9.3	3.2	1.6	6.1	8.2	4.4	12.7
PLCO	20.1	11.7	32.8	4.4	1.6	8.7	7.6	4.0	13.8	6.2	2.9	12.2
SMC	21.9	13.8	32.4	2.8	0.8	6.2	1.5	0.6	3.0	15.6	8.9	24.9
VITAL	14.7	7.9	25.5	2.4	0.6	6.8	6.2	3.7	11.4	4.4	1.8	8.7
WHI	15.4	8.3	25.5	.	.	.	.	.	.	.	.	.
WHS	17.7	10.0	29.2	3.6	1.1	7.7	5.9	2.9	11.5	4.0	2.0	7.6
WLHS	13.7	8.3	20.8	2.1	0.5	4.8	.	.	.	.	.	.

Table 1.2 (Continued)

* Abbreviations: BCDDP, Breast Cancer Detection Demonstration Project Follow-up Study; CARET, Beta-Carotene and Retinol Efficacy Trial; CLUE II, Campaign Against Cancer and Heart Disease; CPS II, Cancer Prevention Study II Nutrition Cohort; CTS, California Teachers Study; CNBSS, Canadian National Breast Screening Study; EPIC, The European Prospective Investigation into Cancer and Nutrition; IWHS, Iowa Women's Health Study; JPHC1, The Japan Public Health Center-Based Study Cohort I ; MEC, Multiethnic Cohort; MCCS, Melbourne Collaborative Cohort Study; NLCS, the Netherlands Cohort Study; NYU, New York University Women's Health Study; AARP, The NIH-AARP Diet and Health Study; NHSa, Nurses' Health Study (part a); NHSb, Nurses' Health Study (part b); NHS II, Nurses' Health Study II; ORDET, The Hormones and Diet in the Etiology of Breast Cancer Study; PLCO, The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SMC, Swedish Mammography Cohort; VITAL, the VITamin D and OmegA-3 TriaL; WHI, Women's Health Initiative; WHS, Women's Health Study; WLHS, Women's Lifestyle and Health Study.

Table 1.3 Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for total dietary fiber intake

Breast cancer		Quintile1	Quintile2	Quintile3	Quintile4	Quintile5	*p*-trend	*p*-het[1]	I^2	*p*-common effect [2]
					Pooled HR (95% CI)					
Overall	*No. of cases*	*11524*	*11332*	*11339*	*11401*	*11248*				
	Age-adjusted	1 ref	0.97 (0.95-1.00)	0.96 (0.94-0.99)	0.95 (0.92-0.98)	0.91 (0.88-0.94)	< 0.001	0.20	19.7%	
	Multivariable[5]	1 ref	0.98 (0.95-1.01)	0.97 (0.94-1.00)	0.96 (0.93-1.00)	0.93 (0.90-0.97)	< 0.001	0.04	37.8%	
ER+	*No. of cases*	*6791*	*6802*	*6889*	*6967*	*6935*				
	Multivariable[5]	1 ref	0.98 (0.94-1.02)	0.98 (0.94-1.03)	0.98 (0.93-1.03)	0.95 (0.90-1.00)	0.08	0.01	46.3%	
ER-	*No. of cases*	*1641*	*1555*	*1594*	*1569*	*1469*				0.08
	Multivariable[5]	1 ref	0.93 (0.86-1.01)	0.96 (0.88-1.04)	0.94 (0.86-1.03)	0.87 (0.80-0.95)	< 0.001	0.43	10.5%	

Table 1.3 (Continued)

		Continuous (per 10g/d)				
Breast cancer		Pooled HR (95% CI)		*p*-het[3]	I^2	*p*-common effect[4]
Overall	*No. of cases*	*56844*				
	Age-adjusted	0.94	(0.92-0.96)	0.37	7.2%	
	Multivariable[5]	0.96	(0.94-0.98)	0.17	22.1%	
ER+	*No. of cases*	*34384*				
	Multivariable[5]	0.97	(0.94-1.00)	0.02	42.9%	0.09
ER-	*No. of cases*	*7828*				
	Multivariable[5]	0.92	(0.88-0.97)	0.29	12.5%	

1. p for between-study heterogeneity for quintile 5.

2. p for common-effect by ER status for quintile 5.

3. p for between-study heterogeneity for continuous dietary fiber intake.

4. p for common-effect by ER status for continuous dietary fiber intake.

5. Multivariable models include race (White, African American, Hispanic, Asian, other), education (< high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m2), height(<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75m), alcohol (0, 0-<5, 5-<15, 15-<30 and ≥30 g/d), energy intake (continuous, kcal/d), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 except PLCO <10,10/11,12/13, ≥14 years), hormone replacement therapy use (never user, past user, current user), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25y), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

remaining analyses. When tumor subtypes defined by ER status were examined separately, no significant difference was detected between the two subtypes, but the association was only significant for ER- tumors (pooled MVHR comparing the highest to lowest quintile = 0.87, 95% CI: 0.80-0.95 for ER- tumors and 0.95, 95% CI: 0.90-1.00 for ER+ tumors, p-value, test for common-effects for quintile 5 = 0.08). Significant between-study heterogeneity was observed for breast cancer overall (p = 0.04) and for the ER+ subtype (p = 0.01). The study-specific HRs for ER+ tumors ranged from 0.61 to 1.48. Meta-regression analyses suggested that the heterogeneity observed for overall breast cancer might be attributable to different proportions of ER- tumors in each study (p = 0.05 for the highest quintile). The differences in the study populations' median age (p = 0.002 for the highest quintile, 0.005 for trend) and/or median follow-up time (p = 0.05 for the highest quintile, 0.04 for trend) contributed to the variation observed for ER+ tumors.

The results from analyses in which dietary fiber intake was modeled using common absolute intake cut points were comparable to those from the quintile analyses (Table S1.2a, S1.2b). In continuous analyses, an increment of 10g/day of dietary fiber intake was associated with 6% (95% CI: 4% – 8%) lower risk of total breast cancer. When we corrected for measurement error using the validation data, the pooled age- and energy-adjusted HR for a 10g/day increment of dietary fiber did not change much (pooled HR = 0.97, 95% CI 0.94 to 1.00; to 0.96, 95% CI: 0.91 to 1.02) for breast cancer overall. Excluding cases diagnosed within the first 5 years of follow-up to reduce possible bias by dietary changes due to prediagnostic symptoms did not substantially change the pooled results (data not shown). To address concerns that our results could be due to confounding by other nutrients that are correlated with dietary fiber, we further adjusted for dietary intake (from foods only) of vitamin C, α-carotene, β-carotene, β-cryptoxanthin, lycopene, and lutein in quintile

analyses (median Pearson correlation coefficients between intakes of these nutrients and dietary fiber ranged from 0.4 – 0.6 across studies, Table S1.4). There was no appreciable change in the associations observed for dietary fiber intake and risk of overall and ER+ breast cancer in models controlling for each of these dietary factors. The associations for ER- breast cancer, however, were attenuated and became nonsignificant when adjusted for the five major carotenoids or total fruit and vegetable intake; the pooled MVHRs ranged from 0.89-0.95 (Table S1.3).

We observed inverse associations for the three source-specific dietary fiber with breast cancer overall (pooled MVHR comparing the highest vs lowest quintile = 0.92 [95 % CI: 0.88-0.96] for fiber from fruits, 0.95 [95% CI: 0.91-0.99] for fiber from vegetables, and 0.96 [95% CI: 0.92-1.00] for fiber from grains). As observed for total dietary fiber intake, the strongest associations observed for dietary fiber intake from fruits and for dietary fiber intake from vegetables were observed for ER- breast cancer. Pooled MVHR comparing the highest vs lowest quintile of fiber from fruit = 0.87, 95% CI: 0.79-0.97 for ER-; and 0.93, 95% CI: 0.89-0.97 for ER+ tumors (p-value for common-effect by ER status for quintile 5 = 0.30). Association with fiber from vegetables, however, was only observed for ER- tumors (pooled MVHR = 0.87, 95% CI: 0.79-0.95, p-value for common-effect by ER status = 0.02). No significant associations were observed for dietary fiber intake from grains [for ER+/-] (Table 1.4). In a sensitivity analysis where we mutually adjusted intakes of (1) dietary fiber from fruits and vegetables and (2) total fruits and vegetables for the risk of ER- breast cancer, the pooled MVHR for total fruit and vegetable intake changed from 0.96 (95% CI 0.94-0.99) to 1.00 (95% CI 0.99-1.00), while the pooled MVHR for fiber from fruits and vegetables became non-significant while the point estimate remained unchanged (0.97, 95% 0.94-0.99 to 0.97, 95% CI 0.94-1.01).

Table 1.4 Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for source-specific dietary fiber intake

Breast cancer	Pooled HR (95% CI)					*p*-trend	*p*-het[1]	I^2	*p*-common effect [2]
	Quintile1	Quintile2	Quintile3	Quintile4	Quintile5				
				Fiber from fruit					
Overall	1 ref	0.98 (0.93-1.03)	0.95 (0.91-1.00)	0.96 (0.93-0.99)	0.92 (0.88-0.96)	< 0.001	0.03	43.6%	
ER+	1 ref	1.00 (0.95-1.05)	0.97 (0.92-1.02)	0.99 (0.95-1.03)	0.93 (0.89-0.97)	< 0.001	0.24	18.7%	
ER-	1 ref	0.97 (0.88-1.08)	0.95 (0.86-1.06)	0.95 (0.85-1.05)	0.87 (0.79-0.97)	0.01	0.14	28.4%	0.30
				Fiber from vegetable					
Overall	1 ref	0.99 (0.97-1.02)	0.99 (0.96-1.02)	0.98 (0.94-1.03)	0.95 (0.91-0.99)	0.02	0.03	44.1%	
ER+	1 ref	1.02 (0.98-1.07)	1.01 (0.96-1.06)	1.01 (0.95-1.07)	0.98 (0.91-1.04)	0.39	0.002	57.5%	
ER-	1 ref	0.95 (0.87-1.05)	0.92 (0.84-1.00)	0.88 (0.82-0.96)	0.86 (0.79-0.93)	0.004	0.53	0%	0.02
				Fiber from grains					
Overall	1 ref	0.98 (0.95-1.01)	0.99 (0.95-1.02)	1 (0.97-1.03)	0.96 (0.92-1.00)	0.03	0.17	25.7%	
ER+	1 ref	1.01 (0.97-1.05)	0.99 (0.94-1.04)	1.01 (0.97-1.05)	0.98 (0.93-1.02)	0.17	0.31	13.1%	
ER-	1 ref	0.96 (0.88-1.04)	1.01 (0.94-1.10)	1.03 (0.95-1.11)	0.97 (0.88-1.07)	0.85	0.23	19.4%	0.89

Table 1.4 (Continued)

Breast cancer	Continuous (per 3g/d)					Mutually adjusted (per 3g/d)			
	Pooled HR (95% CI)		p-het[3]	I^2	p-common effect[4]	Pooled HR[5] (95% CI)		p-het[3]	I^2
Overall	0.97	(0.95-0.98)	0.03	44.8%		0.97	(0.95-0.98)	0.03	45.3%
ER+	0.97	(0.95-0.99)	0.18	24.6%		0.97	(0.96-0.99)	0.29	14.7%
					0.43				
ER-	0.95	(0.92-0.99)	0.08	36.1%		0.96	(0.93-1.00)	0.09	34.4%
Overall	0.98	(0.97-1.00)	0.04	44.0%		0.99	(0.97-1.00)	0.03	44.3%
ER+	0.99	(0.97-1.01)	0.01	51.0%		0.99	(0.97-1.02)	0.01	50.1%
					0.07				
ER-	0.96	(0.93-0.99)	0.28	15.5%		0.96	(0.94-0.99)	0.37	8.1%
Overall	0.99	(0.98-1.00)	0.33	11.5%		0.99	(0.98-1.00)	0.54	0.0%
ER+	0.99	(0.97-1.00)	0.22	21.3%		0.99	(0.97-1.00)	0.25	18.8%
					0.69				
ER-	1.00	(0.97-1.03)	0.15	28.4%		0.99	(0.96-1.02)	0.21	22.2%

1. *p* for between-study heterogeneity for quintile 5;
2. *p* for common-effect by ER status for quintile 5;
3. *p* for between-study heterogeneity for continuous dietary fiber intake;
4. *p* for common-effect by ER status for continuous dietary fiber intake;
5. Pooled MVHRs estimated from a model including fiber from fruit and fiber from vegetables and fiber from grains.

Multivariable models include race (White, African American, Hispanic, Asian, other), education (< high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m2), height(<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75m), alcohol (0, 0-<5, 5-<15, 15-<30 and ≥30 g/d), energy intake (continuous, kcal/d), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 except PLCO <10,10/11,12/13, ≥14 years), hormone replacement therapy use (never user, past user, current user), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25y), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

The nonparametric regression analyses showed no evidence of nonlinearity for total or source-specific dietary fiber intake and risk of breast cancer overall or by ER subtype (all p-values for nonlinearity > 0.05). In continuous analyses, for every 10g/day increase in total dietary fiber intake, risk was 4% (95% CI: 2% - 6%) lower for total breast cancer, 3% (95% CI: 0% - 6%) lower for ER+ breast cancer, and 8% (95% CI: 3% - 12%) lower for ER- tumors (p-value, test for common effects = 0.09, Table 1.3).

We observed a statistically significant interaction between alcohol and dietary fiber intake and risk of overall breast cancer (p-value for interaction = 0.001, Table 1.5 and 1.6). High dietary fiber intake was associated with a weak lower risk of breast cancer in each alcohol consumption strata, but there was no significant difference across strata (p-for-interaction > 0.3). When we modeled alcohol consumption and dietary fiber intake jointly, the strongest association was observed for non-drinkers with the highest fiber intake (pooled MVHR compared to those with the lowest fiber intake and the highest alcohol consumption = 0.72, 95% CI: 0.68-0.76). Similar results were found for ER+ and ER- tumors (p-values for interaction ≤0.003) (Table 1.5). Significant effect modification was also seen for total fat intake. The association between dietary fiber intake and overall breast cancer risk became weaker and nonsignificant among women with higher fat intake (p-for-interaction = 0.03). A similar pattern was observed for the ER- subtype (p-for-interaction = 0.04 but not ER+ subtype (p-for-interaction = 0.70). Age of diagnosis, menopausal status at diagnosis, BMI, region, and follow-up years did not modify the inverse association between dietary fiber intake and risk of overall, ER+, or ER- breast cancer (Table 1.6).

Table 1.5 Pooled HRs (95% CI) for breast cancer overall and subtypes defined by ER status according to the combined effects of dietary fiber and alcohol

Total breast cancer		Total dietary fiber (g/day)			
		<15	15-20	20+	p-interaction
Alcohol (g/day)					
	Pooled HR	1	0.94	0.93	
20+	95% CI	Ref	(0.87-1.02)	(0.86-1.01)	
		n=2809	n=1659	n=954	
	Pooled HR	0.84	0.83	0.79	
5-20	95% CI	(0.80-0.88)	(0.79-0.87)	(0.75-0.84)	
		n=4698	n=4874	n=3986	
	Pooled HR	0.78	0.78	0.74	<0.001
0-5	95% CI	(0.74-0.81)	(0.75-0.82)	(0.71-0.78)	
		n=6277	n=7086	n=7107	
	Pooled HR	0.78	0.77	0.72	
0	95% CI	(0.74-0.82)	(0.74-0.81)	(0.68-0.76)	
		n=6063	n=5209	n=5094	
ER+ breast cancer					
		<15	15-20	20+	p-interaction
Alcohol (g/day)					
	Pooled HR	1	0.96	0.99	
20+	95% CI	Ref	(0.86-1.08)	(0.89-1.10)	
		n=1740	n=1045	n=576	
	Pooled HR	0.85	0.84	0.78	
5-20	95% CI	(0.80-0.90)	(0.79-0.89)	(0.73-0.84)	
		n=2934	n=3023	n=2335	
	Pooled HR	0.75	0.77	0.73	<0.001
0-5	95% CI	(0.71-0.80)	(0.73-0.82)	(0.69-0.78)	
		n=3848	n=4379	n=4183	
	Pooled HR	0.74	0.77	0.73	
0	95% CI	(0.68-0.80)	(0.72-0.82)	(0.68-0.78)	
		n=3592	n=3243	n=3064	
ER- breast cancer					
		<15	15-20	20+	p-interaction
Alcohol (g/day)					
	Pooled HR	1	0.97	1.09	
20+	95% CI	Ref	(0.81-1.17)	(0.85-1.40)	
		n=353	n=224	n=117	
	Pooled HR	0.91	0.84	0.88	
5-20	95% CI	(0.79-1.04)	(0.73-0.97)	(0.75-1.03)	
		n=663	n=648	n=533	
	Pooled HR	0.85	0.86	0.80	0.003
0-5	95% CI	(0.75-0.98)	(0.75-0.98)	(0.69-0.92)	
		n=931	n=1018	n=941	
	Pooled HR	0.78	0.79	0.72	
0	95% CI	(0.68-0.91)	(0.69-0.91)	(0.62-0.83)	
		n=885	n=720	n=666	

Table 1.6 Pooled multivariable HRs (95% CIs) per 10g/day of dietary fiber intake and risk of breast cancer by age at diagnosis, menopausal status, total fat intake, BMI, follow-up duration, and region

Total dietary fiber Increment: 10g/d	Breast cancer overall			ER+			ER-		
	No. of cases	MVHR (95% CI)	*p*-het	No. of cases	MVHR (95% CI)	*p*-het	No. of cases	MVHR (95% CI)	*p*-het
Continuous	56844	0.96(0.94-0.98)	0.17	34384	0.97(0.94-1.00)	0.02	7828	0.92(0.88-0.97)	0.29
Age at diagnosis									
<55	10960	0.93(0.90-0.97)	0.881	6177	0.91(0.86-0.96)	0.613	2055	0.94(0.86-1.03)	0.600
55- <65	19344	0.94(0.91-0.98)	0.250	11451	0.98(0.94-1.02)	0.361	2629	0.91(0.84-0.98)	0.494
≥65	26485	0.97(0.95-0.99)	0.576	16709	0.99(0.95-1.02)	0.201	2937	0.94(0.87-1.00)	0.467
p-interaction			0.206			0.177			0.786
Menopausal status									
Premenopausal	5726	0.95(0.88-1.01)	0.354	3247	0.91(0.84-0.99)	0.412	1190	0.97(0.86-1.10)	0.616
Postmenopausal	32735	0.97(0.95-1.00)	0.228	18847	0.99(0.95-1.03)	0.043	4027	0.94(0.88-1.00)	0.654
p-interaction			0.468			0.19			0.639
Alcohol consumption									
0	16347	0.95(0.92-0.98)	0.344	9876	0.97(0.91-1.02)	0.006	2228	0.91(0.84-0.99)	0.462
> 0 - 10	26838	0.96(0.94-0.99)	0.262	16248	0.97(0.93-1.01)	0.064	3773	0.92(0.87-0.99)	0.482
>10	12547	0.93(0.89-0.98)	0.230	7739	0.94(0.88-0.99)	0.312	1533	0.89(0.77-1.03)	0.173
p-interaction			0.317			0.354			0.934
Total fat intake									
Tertile 1	18637	0.95(0.92-0.99)	0.091	11410	0.96(0.92-1.00)	0.099	2430	0.91(0.85-0.98)	0.969
Tertile 2	19264	0.94(0.91-0.97)	0.888	11549	0.98(0.92-1.04)	0.035	2580	0.88(0.78-1.00)	0.065
Tertile 3	18943	1.00(0.96-1.04)	0.234	11425	0.99(0.93-1.04)	0.180	2584	1.03(0.92-1.16)	0.084
p-interaction			0.028			0.703			0.038

Table 1.6 (Continued)

Total dietary fiber Increment: 10g/d	Breast cancer overall			ER+			ER-		
	No. of cases	MVHR (95% CI)	*p*-het	No. of cases	MVHR (95% CI)	*p*-het	No. of cases	MVHR (95% CI)	*p*-het
BMI (kg/m2)									
<25	26949	0.95(0.92-0.98)	0.229	16362	0.96(0.92-1.00)	0.124	3801	0.94(0.88-1.00)	0.994
25- <30	15860	0.97(0.94-1.00)	0.491	9340	0.98(0.94-1.02)	0.600	2144	0.93(0.83-1.05)	0.059
≥30	8080	0.97(0.92-1.04)	0.058	4738	0.86(0.69-1.07)	0.130	966	0.86(0.69-1.07)	0.027
p-interaction			0.507			0.359			0.653
Follow-up years									
<5	23583	0.96(0.94-0.98)	0.634	12789	0.97(0.93-1.02)	0.184	3099	0.89(0.83-0.95)	0.779
5- <10	18093	0.96(0.93-0.99)	0.286	10843	0.99(0.94-1.05)	0.040	2611	0.92(0.85-1.00)	0.585
≥10	15148	0.95(0.91-0.99)	0.239	10024	0.93(0.88-0.98)	0.317	1994	1.02(0.91-1.14)	0.314
p-interaction			0.852			0.351			0.165
Region									
North America	37912	0.96(0.94-0.99)	0.054	25018	0.98(0.94-1.02)	0.006	5284	0.91(0.87-0.96)	0.530
Others	17729	0.94(0.91-0.97)	0.905	9366	0.93(0.89-0.98)	0.859	2544	0.97(0.84-1.12)	0.085
p-interaction			0.378			0.209			0.709

DISCUSSION

In this large pooled analysis of 24 prospective studies, we observed that dietary fiber intake was associated with a modestly lower risk of ER- breast cancer. Associations for overall and ER+ breast cancer were weak (and nonsignificant for ER+ tumors). The associations between total dietary fiber and risk of breast cancer overall and the ER- subtype were modified by total fat intake, where the weakest association was observed for the highest fat intake level. There was no evidence that alcohol consumption modifies the association, although we did see a 28% lower risk of breast cancer overall when comparing women with the highest fiber intake and the lowest alcohol consumption versus women with the lowest fiber intake and the highest alcohol consumption. These associations were also not modified by BMI, menopausal status at diagnosis, age at diagnosis, follow-up time, or region (North America vs. others). Of the three main food sources of dietary fiber, dietary fiber intake from fruits was inversely associated with risk of both ER+ and ER- breast cancer, higher intake of dietary fiber from vegetables was associated with a significantly lower risk of ER- breast cancer only, and dietary fiber intake from grains was not associated with either subtype.

The relation between dietary fiber intake and overall breast cancer risk has been investigated previously in several cohort studies with mostly weak to null results. A recently published meta-analysis of 20 prospective studies found a 8% (95% CI: 5% - 12%) lower risk of breast cancer comparing the highest versus the lowest intake category[39], and among the specific sources, only fruit fiber was significantly associated with a 7% (95% CI: 4% - 11%) lower risk of breast cancer. The results of the present study are in line with the meta-analysis as we observed an 8% (95% CI: 4% - 12%) lower risk of breast cancer overall comparing ≥25g/day versus <10g/day of fiber intake.

We pooled the latest data available from 24 high-quality cohort studies (among which 15 studies overlapped with the meta-analysis but had longer follow-up) to comprehensively assess the effect of total and source-specific dietary fiber on risk of breast cancer. Participant-level data were harmonized to reduce between-study heterogeneity and enabled us to evaluate dose-response relationship and effect modifications more accurately.

When dietary fiber intake was examined by specific food sources, our results showed that the effects of dietary fiber from fruits and from vegetables were similar. In a sensitivity analysis where we mutually adjusted (1) fiber from fruits and vegetables and (2) total fruit and vegetable intake to assess potential confounding, the pooled MVHR for total fruit and vegetable intake changed from significant to 1.00, while the pooled MVHR for fiber from fruits and vegetables became non-significant but were attenuated only slightly (data not shown). These results suggest that the associations we observed for dietary fiber from these two sources cannot be completely explained by other components in fruits or vegetables. It is also likely, though, that the associations were confounded by intake of other nutrients, such as carotenoids, in fiber-rich foods. Our prior pooled analysis of carotenoids intake and risk of breast cancer in a subset of the studies in the current analysis found that α-carotene, β-carotene, and lutein/zeaxanthin were significantly protective for ER- breast cancer [40]. In sensitivity analyses that adjusted for these possible confounders, there are evidence that the associations seen for dietary fiber were slightly attenuated for ER- tumors, since these nutrients are enriched in fruits and vegetables, thus have relatively high correlation with dietary fiber intake (Pearson correlations from 0.4 to 0.6). It is worth noting, though, that the point estimates for ER- tumors were still stronger than that for ER+ and overall breast cancer even after the attenuation. The public health message is nonetheless consistent – increase the consumption of

fruits and vegetables rich in dietary fiber and carotenoids. Fiber from grains was not associated with risk of breast cancer overall or the subtypes. Grains might be a good source of insoluble fiber, which forms bulk and accelerates excretion including that of estrogen, but the magnitude might be too low to exert protective effect on risk of breast cancer. On the other hand, vegetables and fruits are good sources of soluble fiber, which attracts water to form gel, controls blood glucose, and insulin-like growth factors that had been shown to positively relate to breast cancer risk.

Based on experimental studies, dietary fiber is hypothesized to decrease breast cancer risk through the indirect regulation of cancer cell invasion and migration. One plausible pathway involves butyrate produced from dietary fiber by colonic bacterial fermentation, which has been reported to decrease estrogen binding and estrogen receptor transcription in breast cancer cells [2], which would consequently reduce the stimulation by estrogen on the growth of the mammary epithelium and the differentiation of epithelial tissue [41] and thus lower the risk of breast cancer development. It has also been proposed that dietary fiber decreases intestinal reabsorption of estrogen and increases its fecal excretion, thus lowers the circulating estrogen level [3-5]. Human studies suggested that vegetarian women could excrete 2 to 3 times more estrogen in feces than do omnivores. The greater amount of the estrogen escaped reabsorption may partially explain the lower incidence of breast cancer in vegetarian women [42-44]. Estrogen could also affect breast cancer development via estrogen receptor independent mechanisms, supported by evidence where both exogenous and endogenous estradiol could accelerate mammary tumor formation in estrogen receptor α knock-out mice [45]. Dietary fiber may also influence risk of breast cancer through non-hormonal mechanisms. Butyrate, a product by fermentation of dietary fiber, has the potential to induce cell cycle arrest in G1 phase and apoptosis in a p53-independent manner, thus inhibit breast cancer cell

growth [6,7]. Higher dietary fiber intake may also lead to reduced inflammation, with stronger association observed for ER- and PR- tumors[8]. Mechanistically, the impact was indicated by lower serum CRP concentration and less IL cytokines infiltration, both of which are key inflammatory markers in tumor microenvironment, and has been associated with risk of breast cancer [46-48].

Dietary fiber may modulate the elevated risk of breast cancer by alcohol consumption, possibly by reducing the hydrolysis of conjugated estrogen, therefore counteracting the enhanced estrogen responsiveness due to alcohol consumption[49]. In our analysis, we saw a significant interaction of dietary fiber and alcohol. The inverse association between dietary fiber and risk of breast cancer overall and for ER+ and ER- tumors was more prominent among non-drinkers. It is worth noting that most of the effects could be due to alcohol. Although we saw a 28% risk reduction among women who were non-drinkers and had the highest fiber intake, we still see a 22% lower risk for non-drinkers with low fiber intake.

Our finding that total fat intake modifies the association between dietary fiber and breast cancer risk is supported by experimental studies. A cross-sectional study found that fat intake was positively associated with circulating estrogen, possibly via increased reabsorption of biliary estrogens [50]. They also observed significantly higher plasma estrone and estradiol concentrations in the high fat/low fiber group compared to the low fat/high fiber group. In the diet and androgens (DIANA) randomized trial that lasted for 4.5 months, the intervention group assigned to lower fat intake was shown to have reduced bioavailable sex hormones concentrations [51]. All of these supports the synergistic effect of low fat/high fiber combined with respect to lowering the circulating estrogen level, and subsequently risk reduction of breast cancer.

Our study has several strengths. The large sample size enabled us to examine associations separately for breast cancer subtypes defined by hormone receptor status with higher statistical power, and to evaluate whether these associations were modified by several breast cancer risk factors. Inclusion of studies with different dietary patterns enabled us to examine the exposure over a wide range of intake, thus making it less likely to miss an association as may occur in a single study. The prospective cohort study design minimized recall bias and selection bias of dietary intake as well as the relevant confounders, biases which are more likely in case-control studies[52,53]. Moreover, the exposures and covariates were harmonized on participant level, which enabled more comparable comparisons across studies and reduced potential sources of heterogeneity than in a meta-analysis.

Our study also has limitations. Measurement error in estimated dietary fiber intake can occur within each individual study. Meanwhile, misclassification might also happen across studies due to the varying methods of assessing dietary fiber intake and the covariates. After correcting for measurement error, the association between dietary fiber intake and risk of breast cancer changed only slightly. The study-specific quintile analyses ranked individuals according to their relative intake within each study, minimizing the influence of correlated measurement error; whereas the analyses using common absolute intake cut-off points of dietary fiber intake minimized the potential bias caused by between-studies heterogeneity in intake ranges. Both approaches yielded similar results, adding confidence to our conclusion. Similarly, although confounding variables were measured in varied ways across studies, we harmonized the coding of the covariates across studies to reduce heterogeneity and found that the age-adjusted results were similar to the multivariable results, suggesting that possible misclassification of the confounding variables was

not likely to strongly influence the observed associations. Although between-study heterogeneity is an inevitable concern for any pooled analysis, the p-values for heterogeneity were mostly nonsignificant in this study. Another limitation is that we only have a single measure of dietary fiber intake at study enrollment so we could not incorporate dietary changes earlier or later in life. We did, however, finely stratify by follow-up duration into 3 levels. No significant difference was observed for the varying lengths of follow-up. On the other hand, if there is a long latency period for dietary fiber to realize its protective effect, baseline exposure could be the most relevant. This points to the need for further investigation into the susceptible windows of exposure for different breast cancer subtypes.

CONCLUSION

This large pooled analysis of prospective cohort studies provides evidence that higher intake of dietary fiber, especially fiber from vegetables and fruits, is associated with a modestly lower risk of breast cancer, particularly ER- breast cancer. These findings were consistently observed across subgroups defined by age, menopausal status and BMI, while we observed statistically significant interaction between dietary fiber and alcohol with relation to risk of breast cancer, although the effect size was too small to be clinically meaningful. The associations for total breast cancer remained significant when we controlled for several potential bioactive constituents that are correlated with dietary fiber intake and are relevant to breast cancer risk, however the results for the tumor subtype were attenuated. Nonetheless, the study provides additional evidence supporting the U.S. Dietary Guidelines, which suggest consuming a variety of fruits and vegetables. Additional studies on dietary fiber and susceptible window of breast cancer development may be helpful in explaining the mechanisms underlying the association.

References

1. Cotterchio M, Kreiger N, Theis B, Sloan M, Bahl S. Hormonal factors and the risk of breast cancer according to estrogen- and progesterone-receptor subgroup. *Cancer Epidemiol Biomarkers Prev*. 2003;12(10):1053-1060.

2. deFazio A, Chiew YE, Donoghue C, Lee CS, Sutherland RL. Effect of sodium butyrate on estrogen receptor and epidermal growth factor receptor gene expression in human breast cancer cell lines. *J Biol Chem*. 1992;267(25):18008-18012.

3. Rose DP, Goldman M, Connolly JM, Strong LE. High-fiber diet reduces serum estrogen concentrations in premenopausal women. *Am J Clin Nutr*. 1991;54(3):520-525. doi:10.1093/ajcn/54.3.520.

4. Wu AH, Pike MC, Stram DO. Meta-analysis: dietary fat intake, serum estrogen levels, and the risk of breast cancer. *JNCI Journal of the National Cancer Institute*. 1999;91(6):529-534. doi:10.1093/jnci/91.6.529.

5. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk Factors for Breast Cancer According to Estrogen and Progesterone Receptor Status. *JNCI Journal of the National Cancer Institute*. 2004;96(3):218-228. doi:10.1093/jnci/djh025.

6. Janson W, Brandner G, Siegel J. Butyrate modulates DNA-damage-induced p53 response by induction of p53-independent differentiation and apoptosis. *ONCOGENE*. 1997;15(12):1395-1406. doi:10.1038/sj.onc.1201304.

7. Chopin V, Toillon R-A, Jouy N, Le Bourhis X. Sodium butyrate induces P53-independent, Fas-mediated apoptosis in MCF-7 human breast cancer cells. *BRITISH JOURNAL OF PHARMACOLOGY*. 2002;135(1):79-86. doi:10.1038/sj.bjp.0704456.

8. Harris HR, Willett WC, Vaidya RL, Michels KB. An Adolescent and Early Adulthood Dietary Pattern Associated with Inflammation and the Incidence of Breast Cancer. *Cancer Res*. 2017;77(5):1179-1187. doi:10.1158/0008-5472.CAN-16-2273.

9. Chen S, Chen Y, Ma S, et al. Dietary fibre intake and risk of breast cancer: A systematic review and meta-analysis of epidemiological studies. *Oncotarget*. 2016;7(49):80980-80989. doi:10.18632/oncotarget.13140.

10. Thompson FE, Kipnis V, Midthune D, et al. Performance of a food-frequency questionnaire in the US NIH-AARP (National Institutes of Health-American Association of Retired Persons) Diet and Health Study. *Public Health Nutrition*. 2008;11(2):183-195. doi:10.1017/S1368980007000419.

11. Mai V, Flood A, Peters U, Lacey JVJ, Schairer C, Schatzkin A. Dietary fibre and risk of colorectal cancer in the Breast Cancer Detection Demonstration Project (BCDDP) follow-up cohort. *Int J Epidemiol*. 2003;32(2):234-239. doi:10.1093/ije/dyg052.

12. Flagg EW, Coates RJ, Calle EE, Potischman N, Thun MJ. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort Food Frequency Questionnaire. *Epidemiology*. 2000;11(4):462-468. doi:10.1097/00001648-200007000-00017.

13. Horn-Ross PL, Lee VS, Collins CN, et al. Dietary assessment in the California Teachers Study: reproducibility and validity. *Cancer Causes & Control*. 2008;19(6):595-603. doi:10.1007/s10552-008-9124-y.

14. Jain MG, Harrison L, Howe GR, Miller AB. Evaluation of a self-administered dietary questionnaire for use in a cohort study. *Am J Clin Nutr*. 1982;36(5):931-935. doi:10.1093/ajcn/36.5.931.

15. Kaaks R. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol*. 1997;26(90001):26S–36. doi:10.1093/ije/26.suppl_1.S26.

16. Munger RG, Folsom AR, Kushi LH, Kaye SA, Sellers TA. Dietary Assessment of Older Iowa Women with a Food Frequency Questionnaire: Nutrient Intake, Reproducibility, and Comparison with 24-Hour Dietary Recall Interviews. *American Journal of Epidemiology*. 1992;136(2):192-200. doi:10.1093/oxfordjournals.aje.a116485.

17. Tsubono Y, Kobayashi M, Sasaki S, Tsugane S. Validity and reproducibility of a self-administered food frequency questionnaire used in the baseline survey of the JPHC Study Cohort I. *Journal of Epidemiology*. 2003;13(1 Suppl):S125-S133. doi:10.2188/jea.13.1sup_125.

18. Stram DO, Hankin JH, Wilkens LR, et al. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *American Journal of Epidemiology*. 2000;151(4):358-370. doi:10.1093/oxfordjournals.aje.a010214.

19. Bassett JK, English DR, Fahey MT, et al. Validity and calibration of the FFQ used in the Melbourne Collaborative Cohort Study. *Public Health Nutrition*. 2016;19(13):2357-2368. doi:10.1017/S1368980016000690.

20. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire. *American Journal of Epidemiology*. 1985;122(1):51-65. doi:10.1093/oxfordjournals.aje.a114086.

21. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *European journal of clinical nutrition*.

22. Riboli E, Toniolo P, Kaaks R, Shore RE, Casagrande C, Pasternack BS. Reproducibility of a food frequency questionnaire used in the New York University Women's Health Study: effect of self-selection by study subjects. *Eur J Clin Nutr*. 1997;51(7):437-442. doi:10.1038/sj.ejcn.1600422.

23. Kunzmann AT, Coleman HG, Huang W-Y, Kitahara CM, Cantwell MM, Berndt SI. Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Am J Clin Nutr*. 2015;102(4):881-890. doi:10.3945/ajcn.115.113282.

24. Subar AF, Thompson FE, Kipnis V, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study. *American Journal of Epidemiology*. 2001;154(12):1089-1099. doi:10.1093/aje/154.12.1089.

25. Suzuki R, Rylander-Rudqvist T, Ye W, Saji S, Adlercreutz H, Wolk A. Dietary fiber intake and risk of postmenopausal breast cancer defined by estrogen and progesterone receptor status--a prospective cohort study among Swedish women. *Int J Cancer*. 2008;122(2):403-412. doi:10.1002/ijc.23060.

26. Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. *Annals of Epidemiology*. 1999;9(3):178-187. doi:10.1016/s1047-2797(98)00055-6.

27. Joffe MM, Rosenbaum PR. Invited Commentary: Propensity Scores. *American Journal of Epidemiology*. 1999;150(4):327-333. doi:10.1093/oxfordjournals.aje.a010011.

28. Imai K, van Dyk DA. Causal Inference With General Treatment Regimes. *Journal of the American Statistical Association*. 2004;99(467):854-866. doi:10.1198/016214504000001187.

29. Cepeda MS. Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders. *American Journal of Epidemiology*. 2003;158(3):280-287. doi:10.1093/aje/kwg115.

30. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2.

31. Cochran WG. The Combination of Estimates from Different Experiments. *Biometrics*. 1954;10(1):101. doi:10.2307/3001666.

32. Stevens VL, McCullough ML, Sun J, Gapstur SM. Folate and other one-carbon metabolism–related nutrients and risk of postmenopausal breast cancer in the Cancer Prevention Study II Nutrition Cohort. *Am J Clin Nutr*. 2010;91(6):1708-1715. doi:10.3945/ajcn.2009.28553.

33. Cho E, Holmes M, Hankinson SE, Willett WC. Nutrients Involved in One-Carbon Metabolism and Risk of Breast Cancer among Premenopausal Women. *Cancer Epidemiology Biomarkers & Prevention*. 2007;16(12):2787-2790. doi:10.1158/1055-9965.EPI-07-0683.

34. Smith-Warner SA, Spiegelman D, Yaun S-S, et al. Alcohol and Breast Cancer in Women. *JAMA*. 1998;279(7):535. doi:10.1001/jama.279.7.535.

35. Prentice RL, BRESLOW NE. Retrospective studies and failure time models. *Biometrika*. 1978;65(1):153-158. doi:10.1093/biomet/65.1.153.

36. Anderson TW. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience; 2003.

37. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*. 1990;132(4):734-745. doi:10.1093/oxfordjournals.aje.a115715.

38. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*. 1989;8(9):1051–69–discussion1071–3.

39. Farvid MS, Spence ND, Holmes MD, Barnett JB. Fiber consumption and breast cancer incidence: A systematic review and meta-analysis of prospective studies. *Cancer*. 2020;22(3 suppl):1786. doi:10.1002/cncr.32816.

40. Zhang X. Carotenoid intakes and risk of breast cancer defined by estrogen receptor and progesterone receptor status: a pooled analysis of 18 prospective cohort studies. 2012. doi:10.4016/39352.01.

41. Travis RC, Key TJ. Oestrogen exposure and breast cancer risk. *Breast Cancer Res*. 2003;5(5):1-9. doi:10.1186/bcr628.

42. Goldin BR, Adlercreutz H, Gorbach SL, et al. Estrogen Excretion Patterns and Plasma Levels in Vegetarian and Omnivorous Women. *N Engl J Med*. 1982;307(25):1542-1547. doi:10.1056/NEJM198212163072502.

43. Goldin BR, Adlercreutz H, Dwyer JT, Swenson L, Warram JH, Gorbach SL. Effect of diet on excretion of estrogens in pre- and postmenopausal women. *Cancer Res*. 1981;41(9 Pt 2):3771-3773.

44. Adlercreutz H, Järvenpää P. Assay of estrogens in human feces. *J Steroid Biochem*. 1982;17(6):639-645.

45. Yue W, Wang JP, Li Y, et al. Effects of estrogen on breast cancer development: Role of estrogen receptor independent mechanisms. *Int J Cancer*. 2010;127(8):1748-1757. doi:10.1002/ijc.25207.

46. Siemes C, Visser LE, Coebergh J-WW, et al. C-reactive protein levels, variation in the C-reactive protein gene, and cancer risk: the Rotterdam Study. *J Clin Oncol*. 2006;24(33):5216-5222. doi:10.1200/JCO.2006.07.1381.

47. Ajani UA, Ford ES, Mokdad AH. Dietary fiber and C-reactive protein: findings from national health and nutrition examination survey data. *Journal of Nutrition*. 2004;134(5):1181-1185. doi:10.1093/jn/134.5.1181.

48. Bastid J, Bonnefoy N, Eliaou J-F, Bensussan A. Lymphocyte-derived interleukin-17A adds another brick in the wall of inflammation-induced breast carcinogenesis. *Oncoimmunology*. 2014;3(4):e28273. doi:10.4161/onci.28273.

49. Romieu I, Ferrari P, Chajès V, et al. Fiber intake modulates the association of alcohol intake with breast cancer. *Int J Cancer*. 2017;140(2):316-321. doi:10.1002/ijc.30415.

50. Aubertin-Leheudre M, Gorbach S, Woods M, Dwyer JT, Goldin B, Adlercreutz H. Fat/fiber intakes and sex hormones in healthy premenopausal women in USA. *J Steroid Biochem Mol Biol*. 2008;112(1-3):32-39. doi:10.1016/j.jsbmb.2008.08.002.

51. Berrino F, Bellati C, Secreto G, et al. Reducing Bioavailable Sex Hormones through a Comprehensive Change in Diet: the Diet and Androgens (DIANA) Randomized Trial. *Cancer Epidemiol Biomarkers Prev*. 2001;10(1):25-33. doi:10.1093/jnci/88.5.291.

52. Jensen OM, Wahrendorf J, Rosenqvist A, Geser A. The reliability of questionnaire-derived historical dietary information and temporal stability of food habits in individuals. *American Journal of Epidemiology*. 1984;120(2):281-290. doi:10.1093/oxfordjournals.aje.a113891.

53. Rohan TE, Potter JD. Retrospective assessment of dietary intake. *American Journal of Epidemiology*. 1984;120(6):876-887. doi:10.1093/oxfordjournals.aje.a113959.

Chapter 2

Dairy foods, calcium and breast cancer risk: a pooled analysis of 21 cohort studies

ABSTRACT

Background: Epidemiologic studies examining the relationships between dairy product and calcium intakes and breast cancer risk have been inconclusive, especially for tumor subtypes.

Methods: In 21 prospective cohorts, we pooled the individual-level data of over 1 million women who were followed for up to 8-20 years. Associations were evaluated for dairy product and calcium intakes and risk of incident invasive breast cancer overall (n = 37,861 cases) and by subtypes defined by hormone receptor status. Study-specific multivariable hazard ratios (HR) were estimated and then combined using random-effects models.

Results: Dairy products showed null or very weak inverse associations with risk of overall breast cancer (*p*-value, test for trend > 0.05 for all). Differences by estrogen receptor (ER) status were suggested for yogurt and cottage/ricotta cheese (*p*-value, test for common effects = 0.07 for each) with associations observed for ER-negative tumors only (pooled HR = 0.90, 95% confidence interval (CI) 0.83 to 0.98 comparing ≥60g/day with <1g/day of yogurt and 0.85, 95% CI 0.76 to 0.95 comparing ≥25g/day with <1g/day of cottage/ricotta cheese). Dietary calcium intake was only weakly associated with breast cancer risk (pooled HR = 0.98, 95% CI 0.97 to 0.99 for a 350mg/day increment).

Conclusion: We found no clear associations between consumption of specific dairy foods, calcium, and risk of overall breast cancer. Yogurt and cottage/ricotta cheese consumption were associated with modestly lower risk of ER-negative tumors. Future studies, focusing on fermented dairy products, ER-negative breast cancer, and different racial/ethnic populations may further elucidate the relation.

INTRODUCTION

Worldwide, breast cancer is the most commonly diagnosed malignancy and the leading cause of cancer death in women, accounting for 2.1 million cases each year and 15% of all cancer deaths[1]. Breast cancer is a heterogeneous disease with subtypes based on expression of hormone receptors having different etiologies, clinical characteristics, and survival rates[2-4]. One challenge in studying hormone receptor negative tumors is that they only account for < 20% of all breast cancers[5], so that many studies have inadequate statistical power to examine analyze them separately.

Dairy products have been hypothesized to influence breast carcinogenesis in conflicting ways. They are the main dietary sources of conjugated linoleic acid, calcium, and vitamin D (in fortified fluid milk and yogurt), all of which have been suggested to have anticarcinogenic properties by regulating cell proliferation, differentiation, and apoptosis [6-9]. Dairy products also contain branched chain amino acids and potentially increase circulating insulin-like growth factor-1 (IGF-1) concentrations[10] which may promote cell growth, elevate mitotic activity, and increase DNA replication errors [11,12]. Bovine sex hormones and hormone drugs used in dairy management practices (e.g. trenbolone acetate, zeranol, melengestrol acetate) might increase breast cancer risk as well[11,13]. A meta-analysis of 22 prospective cohort studies and 5 case-control studies reported that high total dairy consumption was associated with a lower risk of breast cancer (risk ratio = 0.90, 95% confidence interval [CI] 0.83-0.98, comparing >600g/day with <200g/day)[14]. The number of studies reporting on specific dairy products was limited, and results were not reported separately for breast cancer subtypes in that meta-analysis.

To evaluate the associations between intakes of specific dairy products and calcium and risk of

breast cancer overall and for subtypes defined by estrogen receptor (ER) and progesterone receptor (PR) status, we conducted a pooled analysis within the Pooling Project of Prospective Studies of Diet and Cancer (DCPP).

METHODS

Study Population. The DCPP is an international consortium of prospective cohort studies [15]. In this study, we analyzed 21[16-35] cohorts (Table 1) that met the following inclusion criteria: 1) at least one publication on any diet and cancer association; 2) assessed dairy product and calcium intake with a comprehensive long-term dietary assessment tool; 3) validated the dietary assessment tool or a closely related instrument; and 4) included at least 25 incident ER-negative breast cancer cases. Each included study was approved by their respective institutional review board.

Assessment of Dietary and Non-dietary Factors. Dietary intake was assessed at baseline by a validated study-specific food frequency questionnaire (FFQ), generally covering the past year. Total milk, reduced-fat milk, whole milk, hard cheese, cottage/ricotta cheese, yogurt, and ice cream were examined (see table 2 for the items in each group). All studies estimated dietary calcium intake (from foods); 12 studies also estimated total calcium intake that also included calcium from multivitamins and other supplements. Dietary and total calcium intakes were energy-adjusted using the residual method[36]. The Pearson correlation coefficients comparing intakes from the FFQ used in these studies or closely related FFQs with either multiple 24-hour recalls or dietary records generally ranged from 0.5-0.9 for intake of dairy products [26,37-43] and 0.6-0.8 for dietary calcium[37,39 40,41,44-47].

Each study collected age, height, and body weight at baseline. Most studies also assessed family history of breast cancer, educational attainment, physical activity, smoking habits, and several reproductive factors.

Case Ascertainment. Breast cancer was defined by International Classification of Diseases (ICD)-9 code 174.0 or ICD-10 code C50. Incident invasive breast cancer cases were identified by follow-up questionnaires and subsequent review of medical records, through linkage to cancer registries, or by both methods. Some cases also were identified using linkage to mortality registries. Follow-up generally exceeded 90% for the studies [15]. Hormone receptor status (obtained for 73.4% of all cases) was identified through cancer registries, pathology reports, medical records, or laboratory determinations. Cases with borderline hormone receptor status were considered as positive for that hormone receptor.

Statistical Analysis. We analyzed the primary participant-level data in each cohort. The Netherlands Cohort Study was analyzed as a case-cohort study[48], as required by its study design. The Nurses' Health Study was separated into two cohorts (1980-1986 Nurses' Health Study a; 1986-2006 Nurses' Health Study b) because of the more detailed dietary assessment after 1986 compared with 1980.

We excluded women who reported total energy intakes outside of three standard deviations from the mean $\log_e$-transformed energy intake in that study, who had been diagnosed with any cancer other than non-melanoma skin cancer prior to baseline, or who had missing values for dairy product or calcium intake.

The associations for dairy products, dietary calcium, and total calcium intake and risk of breast cancer overall and for subtypes defined by ER status and by ER/PR jointly, were evaluated for each study using Cox proportional hazards regression (SAS PROC PHREG). Dairy product and calcium intakes were modeled using categories defined by common absolute intake cut points. Calcium intake was modeled using study-specific quintiles as well. Most of the dairy products evaluated were comprised of a limited number of items and had relatively discrete intake distributions, thus we did not model them using quantiles. For each participant, we calculated person-years of follow-up from the age of the baseline questionnaire was returned to the age of diagnosis of incident breast cancer, death, loss to follow-up, or end of follow-up, whichever occurred first. We used age at baseline and year of baseline questionnaire return as stratification factors to account for age, calendar time, and time since study entry. In multivariable analyses, we adjusted for the confounding variables directly in the model for studies with more than 200 cases; or included propensity scores otherwise[49,50]. We created missing indicator variables for confounders with missing values. We evaluated the main exposures for divergences from the proportional hazards assumption by examining figures of Schoenfeld residuals[51], where we did not find evidence for significant violation.

We pooled the study-specific HRs using random-effects models[52]. Between-studies heterogeneity was evaluated using the Q[52] and I^2 statistics[53].

To test for a linear trend across categories of intake for each participant, we assigned the study-specific median value of their exposure category, modeled that variable as a continuous variable, and tested the coefficient using the Wald test. We compared nonparametric regression curves using

restricted cubic splines with the linear model using the likelihood ratio test [54] to test for nonlinearity in the associations for dairy products, dietary calcium, or total calcium. Intakes were modeled as continuous variables when evidence of nonlinearity was not found.

We investigated whether the associations of interest varied by menopausal status at diagnosis using a previously described algorithm [55](premenopausal, postmenopausal), age at diagnosis (<64, ≥64 years), body mass index (BMI, <25, ≥25 kg/m^2) and follow-up years (<5, ≥5 years) using a mixed-effects meta-regression model[56]. We used a contrast test to examine whether risk estimates varied by breast cancer subtype [57].

For each study that evaluated calcium intake in their validation study, we corrected for the bias in the estimated HRs due to measurement error in dietary calcium intake [16-19,21,24-26,28-34], using a regression calibration method[58,59].

For all tests of statistical hypotheses, two-sided Wald 95% CIs were calculated, and two-sided *p-values* < 0.05 were considered statistically significant. All analyses were conducted using SAS software versions 9.2-9.4 (SAS Institute, Cary, NC).

RESULTS

Across the 21 prospective studies with maximum follow-up ranging from 8 to 20 years, 37,861 incident invasive breast cancer (22,040 ER-positive and 5,367 ER-negative breast cancer cases) were diagnosed among 1,210,243 women (Table 2.1, Table S2.1).

Table 2.1 Characteristics of the cohort studies included in the pooled analysis of dietary fiber intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer

Study	Follow-up time, years	Baseline cohort size	Age range, years	No. of breast cancer cases
Breast Cancer Detection Demonstration Project Follow-up Study	1987-1999	42061	40-93	1305
The Black Women's Health Study	1995-2008	52576	20-70	670
Beta-Carotene and Retinol Efficacy Trial	1985-2005	6000	48-70	367
Campaign Against Cancer and Heart Disease	1989-2007	8279	18-93	288
Canadian National Breast Screening Study	1980-2000	45185	40-59	1240
Cancer Prevention Study II Nutrition Cohort	1992-2003	74137	40-87	2999
California Teachers Study	1995-2003	100067	22-104	2696
Iowa Women's Health Study	1986-2004	34584	52-71	1849
JPHC1, Japan Public Health Center-Based Study Cohort I	1990-2004	21609	40-59	289
Melbourne Collaborative Cohort Study	1990-2006	22456	31-75	799
Multiethnic Cohort	1993-2004	92435	45-78	3308
Nurses' Health Study II	1991-2003	93778	26-46	1331
Nurses' Health Study (part a)	1980-1986	88618	34-67	1122
Nurses' Health Study (part b)	1986-2006	68394	40-67	4467
NIH-AARP Diet and Health Study	1995-2003	200049	50-71	5972
Netherlands Cohort Study	1986-1999	62573	54-70	2013
New York University Women's Health Study	1985-2003	13257	31-70	919
Hormones and Diet in the Etiology of Breast Cancer Study	1987-2002	9044	34-70	283
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	1993-2007	28292	52-74	1090
Swedish Mammography Cohort	1987-2005	60950	38-76	2605
Women's Health Study	1993-2004	38385	38-89	1177
Women's Lifestyle and Health Study	1991-2006	47514	30-50	1072
Total		1210243		37861

Dairy product and calcium consumption varied substantially across studies. Median dietary calcium intake ranged from 555 to 853 mg/day. Median total calcium intakes ranged from 675 to 1173mg/day (Table 2.2). Dietary calcium intake was highly correlated with total milk intake (median Pearson correlation coefficient across studies = 0.74) and reduced-fat milk intake (median Pearson correlation = 0.73); correlations for total calcium intake with these two food items were weaker. Very weak correlations were observed between calcium and other dairy products (Table S2.2).

For all dairy products evaluated, null or weak associations were observed for risk of breast cancer overall (Table 2.3). The only statistically significant associations were observed for total milk (pooled HR comparing ≥500g/day (~16 oz/day) with <1 g/day = 0.94, 95% CI 0.90 to 0.99) and yogurt (pooled HR comparing ≥60g/day (~2 oz/day) with <1g/day = 0.96, 95% CI 0.92 to 0.99).

When we further estimated associations for subtypes of breast cancer defined by ER status (Table 2.3), differences between ER-positive and ER-negative tumors were suggested only for yogurt and cottage/ricotta cheese consumption, with statistically significant inverse associations being observed for ER-negative tumors only. The pooled HRs comparing ≥60g/day with <1g/day yogurt intake were 0.90 (95% CI 0.83 to 0.98) for ER-negative tumors and 0.98 (95% CI 0.94 to 1.03) for ER-positive tumors (*p*-value, test for common effects by ER status = 0.07). Similarly, higher cottage/ricotta cheese consumption was associated with a 15% lower risk of ER-negative (pooled HR comparing ≥25g/day to <1 g/day = 0.85, 95% CI 0.76 to 0.95) but not ER-positive breast cancer (pooled HR = 0.96, 95% CI 0.90 to 1.02; *p*-value, test for common effects by ER status = 0.07, Table 2.3). When intakes were modeled as continuous variables, we did not observe

Table 2.2 Daily median intakes of dairy products and calcium by cohort study

Study*	Median intake† (10th-90th percentile), g/day						
	Total milk‡	Hard cheese§	Cottage/ricotta cheese	Yogurt‖	Ice cream	Dietary calcium	Total calcium¶
BCDDP	189 (30-580)	8 (0-32)	5 (0-30)	-	7 (0-46)	794 (470-1338)	958 (518-1965)
BWHS	35 (0-298)	4 (0-27)	-	4 (0-96)	4 (0-29)	490 (238-981)	-
CARET	108 (0-675)	5 (0-26)	5 (0-23)	0 (0-50)	4 (0-29)	688 (414-1176)	-
CLUE II	60 (0-330)	7 (0-39)	-	-	9 (0-54)	708 (429-1181)	783 (453-1449)
CNBSS	148 (0-490)	17 (3-43)	4 (0-37)	2 (0-80)	4 (0-29)	642 (379-1007)	-
CPS II	206 (35-722)	5 (0-28)	-	15 (0-139)	0 (0-24)	812 (477-1393)	1003 (536-1918)
CTS	132 (0-709)	8 (0-28)	1 (0-20)	17 (0-121)	8 (0-56)	703 (426-1262)	-
IWHS	245 (0-613)	12 (2-23)	8 (0-45)	0 (0-32)	5 (0-28)	691 (436-1140)	943 (501-1701)
JPHC1	100 (0-300)	0 (0-4)	-	-	-	555 (330-880)	-
MCCS	-	13 (2-33)	0 (0-16)	7 (0-78)	4 (0-22)	652 (473-865)	-
MEC	85 (0-336)	2 (0-16)	0 (0-16)	0 (0-43)	4 (0-25)	609 (390-939)	750 (433-1568)
NHS II	216 (20-613)	12 (2-28)	8 (0-15)	18 (0-98)	5 (0-28)	739 (480-1155)	829 (511-1422)
NHSa	140 (0-613)	12 (2-28)	7 (0-45)	0 (0-98)	5 (0-28)	670 (393-1129)	675 (394-1144)
NHSb	196 (0-613)	12 (2-28)	8 (0-45)	0 (0-98)	5 (0-28)	670 (441-1071)	973 (515-1772)
NIH-AARP	128 (2-560)	1 (0-13)	1 (0-20)	5 (0-86)	1 (0-18)	665 (419-1131)	988 (505-1908)
NLCS	171 (16-383)	19 (2-42)	0 (0-21)	53 (0-139)	-	853 (557-1197)	-
NYUWHS	155 (0-465)	8 (0-45)	6 (0-40)	12 (0-106)	7 (0-46)	772 (465-1188)	849 (513-1289)

Table 2.2 (Continued)

ORDET	22 (0-128)	34 (8-88)	-	12 (0-48)	-	622 (407-954)	-
PLCO	174 (8-654)	5 (0-21)	3 (0-29)	7 (0-99)	4 (0-32)	752 (508-1178)	1173 (614-2006)
SMC	336 (0-763)	21 (6-60)	-	75 (0-218)	5 (0-12)	849 (560-1160)	-
WHS	196 (0-613)	4 (2-23)	8 (0-15)	18 (0-98)	5 (0-28)	681 (446-1081)	827 (483-1592)
WLHS	56 (0-393)	20 (0-59)	0 (0-6)	28 (0-197)	7 (0-28)	840 (515-1237)	-

[*]Abbreviations: CARET, Beta-Carotene and Retinol Efficacy Trial; BWHS, The Black Women's Health Study; BCDDP, Breast Cancer Detection Demonstration Project Follow-up Study; CTS, California Teachers Study; CLUE II: Campaign Against Cancer and Heart Disease; CNBSS, Canadian National Breast Screening Study; CPS II, Cancer Prevention Study II Nutrition Cohort; IWHS, Iowa Women's Health Study; JPHC1, Japan Public Health Center-Based Study Cohort I ; MCCS, Melbourne Collaborative Cohort Study; MEC, Multiethnic Cohort; NLCS, Netherlands Cohort Study; NYUWHS, New York University Women's Health Study; NIH-AARP, NIH-AARP Diet and Health Study; NHSa, Nurses' Health Study (part a); NHSb, Nurses' Health Study (part b); NHS II, Nurses' Health Study II; ORDET, Hormones and Diet in the Etiology of Breast Cancer Study; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SMC, Swedish Mammography Cohort; WHS, Women's Health Study; WLHS, Women's Lifestyle and Health Study.

[†]Milk: 8 oz serving is equivalent to 245 g; hard cheese: 1 oz serving is equivalent to 28 g; cottage cheese: 0.5 cup serving is equivalent to 105 g; yogurt: 1 cup serving is equivalent to 227 g; ice cream: 0.5 cup serving is equivalent to 66 g.

[‡]Total milk: whole milk, reduced-fat milk, buttermilk and evaporated milk.

[§]Total hard cheese: hard cheese, high-fat cheese, low-fat cheese, cheddar cheese, feta cheese, and unspecified cheese.

[ǁ]Yogurt: high-fat and low-fat yogurt but not frozen yogurt.
[¶]Total calcium intake is from dietary and supplemental sources.

Table 2.3 Pooled age and multivariable[*] adjusted hazard ratios (HR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER)[†] status according to intake of dairy products

Dairy groups[‡]		Categories of intake (g/day) Pooled HR (95% CI)						P_{diff}[§]	I^2[‖]	P_{het}[¶]	P_{trend}[#]
Total Milk		0	1-62.49	62.5-124.9	125-249.9	250-499.9	≥500				
Cases (ER+/-)		4774 (2835/740)	6485 (3737/940)	5750 (3302/844)	9150 (5366/1234)	5735 (3182/686)	5005 (3022/726)				
Total	Age-adjusted	1.00 (Ref)	0.94 (0.91-0.98)	1.00 (0.96-1.04)	0.98 (0.95-1.02)	0.94 (0.90-0.98)	0.94 (0.90-0.99)		<0.01	0.77	0.08
	Multivariable	1.00 (Ref)	0.94 (0.91-0.98)	1.00 (0.96-1.04)	0.99 (0.95-1.03)	0.94 (0.90-0.98)	0.94 (0.90-0.99)		<0.01	0.89	0.09
ER+	Multivariable	1.00 (Ref)	0.95 (0.90-1.00)	0.99 (0.93-1.04)	1.00 (0.94-1.06)	0.95 (0.89-1.01)	0.95 (0.90-1.01)		<0.01	0.63	0.31
ER-	Multivariable	1.00 (Ref)	0.96 (0.87-1.07)	1.09 (0.96-1.23)	0.99 (0.89-1.09)	0.89 (0.78-1.02)	1.02 (0.90-1.15)	0.34	<0.01	0.66	0.77
Whole Milk		0	1-62.49	62.5-124.9	≥125						
Cases (ER+/-)		27139 (15947/3664)	3722 (2115/575)	1272 (695/186)	3934 (2224/581)						
Total	Age-adjusted	1.00 (Ref)	0.94 (0.91-0.98)	0.93 (0.88-0.98)	0.92 (0.88-0.97)				0.15	0.28	<0.01
	Multivariable	1.00 (Ref)	0.97 (0.94-1.01)	0.97 (0.91-1.03)	0.96 (0.92-1.00)				<0.01	0.53	0.07
ER+	Multivariable	1.00 (Ref)	0.96 (0.90-1.01)	0.97 (0.90-1.05)	0.95 (0.88-1.03)				0.30	0.11	0.15
ER-	Multivariable	1.00 (Ref)	1.02 (0.93-1.12)	1.02 (0.88-1.19)	1.01 (0.90-1.13)			0.41	<0.01	0.64	0.88

Table 2.3 (Continued)

		Categories of intake (g/day) Pooled HR (95% CI)					P_{diff}[§]	I^2[‖]	P_{het}[¶]	P_{trend}[#]
Reduced Fat Milk		0	1-62.49	62.5-124.9	125-249.9	≥250				
Cases (ER+/-)		9428 (5179/1358)	5741 (3514/844)	5274 (3050/731)	8147 (4821/1127)	7552 (4399/961)				
Total	Age-adjusted	1.00 (Ref)	0.99 (0.95-1.03)	1.05 (1.01-1.10)	1.03 (1.00-1.06)	0.99 (0.96-1.03)		0.09	0.35	0.72
	Multivariable	1.00 (Ref)	0.97 (0.94-1.01)	1.03 (0.99-1.07)	1.01 (0.98-1.04)	0.97 (0.94-1.01)		0.08	0.36	0.35
ER+	Multivariable	1.00 (Ref)	0.98 (0.94-1.03)	1.01 (0.96-1.07)	1.03 (0.97-1.08)	0.98 (0.94-1.03)		<0.01	0.60	0.52
ER-	Multivariable	1.00 (Ref)	1.00 (0.91-1.10)	1.09 (0.99-1.21)	1.01 (0.91-1.11)	0.99 (0.89-1.11)	0.84	0.25	0.16	0.92
Hard Cheese		0	1-24.9	25-49.9	≥50					
Cases (ER+/-)		7369 (3792/946)	24602 (14833/3577)	3934 (2252/560)	1675 (974/241)					
Total	Age-adjusted	1.00 (Ref)	1.05 (1.01-1.09)	1.04 (0.96-1.12)	1.06 (0.97-1.16)			0.26	0.14	0.22
	Multivariable	1.00 (Ref)	1.02 (0.98-1.06)	0.99 (0.93-1.06)	1.01 (0.93-1.10)			0.13	0.29	0.76
ER+	Multivariable	1.00 (Ref)	1.05 (1.00-1.10)	1.00 (0.92-1.10)	1.03 (0.93-1.15)			0.08	0.36	0.82
ER-	Multivariable	1.00 (Ref)	0.96 (0.88-1.04)	0.94 (0.81-1.09)	1.00 (0.78-1.29)		0.84	0.23	0.18	0.56

Table 2.3 (Continued)

		Categories of intake (g/day) Pooled HR (95% CI)				P_{diff}[§]	I^2[‖]	P_{het}[¶]	P_{trend}[#]
Cottage Cheese		0	1-12.49	12.5-24.9	≥25				
Cases (ER+/-)		13217 (7271/1883)	10373 (6202/1377)	3277 (1945/448)	3596 (2073/474)				
Total	Age-adjusted	1.00 (Ref)	1.00 (0.97-1.03)	0.96 (0.92-1.01)	0.98 (0.94-1.02)		<0.01	0.79	0.21
	Multivariable	1.00 (Ref)	0.99 (0.96-1.02)	0.96 (0.92-1.00)	0.97 (0.93-1.01)		<0.01	0.53	0.19
ER+	Multivariable	1.00 (Ref)	1.00 (0.96-1.04)	0.96 (0.90-1.03)	0.96 (0.90-1.02)		0.09	0.35	0.10
ER-	Multivariable	1.00 (Ref)	0.91 (0.85-0.99)	0.93 (0.79-1.10)	0.85 (0.76-0.95)	0.07	<0.01	0.47	0.07
Yogurt		0	1-29.9	30-59.9	≥60				
Cases (ER+/-)		14616 (8534/2189)	10146 (5846/1344)	3675 (2049/513)	6997 (4163/961)				
Total	Age-adjusted	1.00 (Ref)	1.00 (0.97-1.03)	0.99 (0.95-1.03)	0.97 (0.94-1.00)		<0.01	0.45	0.07
	Multivariable	1.00 (Ref)	0.98 (0.96-1.01)	0.97 (0.93-1.01)	0.96 (0.92-0.99)		0.17	0.25	0.01
ER+	Multivariable	1.00 (Ref)	0.99 (0.96-1.03)	0.96 (0.91-1.02)	0.98 (0.94-1.03)		<0.01	0.94	0.20
ER-	Multivariable	1.00 (Ref)	0.92 (0.86-0.99)	0.93 (0.84-1.03)	0.90 (0.83-0.98)	0.07	<0.01	0.62	0.02

Table 2.3 (Continued)

		Categories of intake (g/day) Pooled HR (95% CI)				P_{diff}[§]	I^2[‖]	P_{het}[¶]	P_{trend}[#]
Ice Cream		0	1-16.9	17-33.9	≥34				
Cases (ER+/-)		11359 (6757/1494)	17381 (10294/2568)	3749 (2264/582)	2406 (1449/345)				
Total	Age-adjusted	1.00 (Ref)	1.01 (0.98-1.03)	1.02 (0.98-1.06)	0.99 (0.94-1.03)		<0.01	0.92	0.73
	Multivariable	1.00 (Ref)	1.01 (0.98-1.04)	1.03 (0.99-1.07)	1.00 (0.95-1.05)		<0.01	0.71	0.96
ER+	Multivariable	1.00 (Ref)	1.00 (0.96-1.04)	1.02 (0.97-1.08)	0.99 (0.92-1.06)	0.53	0.12	0.32	0.53
ER-	Multivariable	1.00 (Ref)	1.03 (0.96-1.11)	1.12 (0.99-1.26)	1.04 (0.90-1.21)		0.17	0.25	0.42

[*]Multivariable model includes race (White, African American, Hispanic, Asian, other), education (<high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m^2), height (<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75 m), alcohol consumption (0,>0-<5, 5-<15, 15-<30 and ≥30 g/day), energy intake (kcal/d, continuous), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 years except PLCO: <10, 10/11, 12/13, ≥14 years), hormone replacement therapy use (never, past, current), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25 years), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

[†]ER: estrogen receptor

[‡]The total milk group included whole milk, reduced-fat milk, buttermilk and evaporated milk; reduced fat milk included low-fat (e.g. 1%, 2%) milk and skim milk; total hard cheese contained hard cheese, high-fat cheese, low-fat cheese, cheddar cheese, feta cheese, and unspecified cheese; cottage cheese contained cottage and ricotta cheese; yogurt included high-fat and low-fat yogurt but not frozen yogurt.

[§]P_{diff}: p-value, test for common effects for different subtypes defined by estrogen receptor status for the highest category

[‖]I^2 for the highest category

[¶]P_{het}: p-value, test for between studies heterogeneity for the highest category

[#]P_{trend}: p-value, test for trend

significant associations for any dairy product with risk of breast cancer overall or for subtypes defined by ER (Table 2.5) or joint ER/PR status (data not shown).

Dietary calcium intake showed a significant inverse trend with risk of breast cancer overall (*p*-value, test for trend = 0.004), although the result for the highest intake category (≥1400mg/d) was not statistically significant (Table 2.4). Weak inverse associations were also observed when dietary calcium intake was modeled using study-specific quintiles (pooled HR comparing quintile 5 with 1 = 0.95, 95% CI 0.91 to 0.98, *p*-value, test for trend = 0.001, Table S2.3) or as a continuous variable (pooled HR for a 350mg/day increment = 0.98, 95% CI 0.97 to 0.99, Table 2.5). After correcting for measurement error, the pooled age- and energy-adjusted HR for a 350mg/d increment of dietary calcium changed from 0.98 (95% CI 0.96 to 1.00) to 0.96 (95% CI: 0.92 to 0.99) for overall breast cancer. Results for dietary calcium intake were similar in magnitude when we limited the analyses to only those studies included in the total calcium analyses or when limited to individuals with no supplemental calcium intake (results not shown). The associations between total calcium intake and risk of breast cancer were all weak and statistically non-significant when intake was modeled as a categorical variable (Table 2.4), as quintiles (Table S2.4), or continuously (Table 2.5). Results for dietary and total calcium intake did not differ by ER or ER/PR subtypes (*p-value*, test for common effects >0.4)

The associations between total milk, hard cheese, cottage/ricotta cheese, yogurt, dietary calcium, and total calcium intake and risk of overall, ER-positive, and ER-negative breast cancer generally did not vary significantly by BMI, menopausal status, age at diagnosis, or follow-up time (Table 2.5). The only exceptions included the association between hard cheese intake and risk of overall

Table 2.4 Pooled age and multivariable[*] adjusted hazard ratios (HR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER)[†] status according to calcium intake

Nutrients		Categories of intake Pooled HR (95% CI)						P_{diff}[‡]	I^2 [§]	P_{het}[‖]	P_{trend}[¶]
Dietary Calcium (mg/day)		<500	500-699.9	700-899.9	900-1099.9	1100-1399.9	≥ 1400				
Cases (ER+/-)		6976 (3999/1061)	11699 (6899/1708)	9394 (5449/1281)	5268 (3057/717)	3194 (1861/442)	1327 (774/158)				
Total	Age-adjusted								0.28	0.13	0.002
	Multivariable	1.00 (Ref)	0.98 (0.95-1.02)	0.97 (0.94-1.01)	0.94 (0.89-1.00)	0.93 (0.88-0.98)	0.97 (0.89-1.05)				
		1.00 (Ref)	0.99 (0.96-1.02)	0.98 (0.95-1.02)	0.95 (0.91-1.00)	0.93 (0.88-0.99)	0.99 (0.91-1.07)		0.22	0.19	0.004
ER+	Multivariable	1.00 (Ref)	1.00 (0.96-1.04)	0.98 (0.93-1.04)	0.96 (0.91-1.02)	0.93 (0.86-1.01)	1.00 (0.91-1.11)	0.71	0.21	0.21	0.08
ER-	Multivariable	1.00 (Ref)	0.99 (0.91-1.08)	0.97 (0.88-1.07)	0.96 (0.85-1.08)	0.97 (0.86-1.02)	1.04 (0.86-1.26)		0.11	0.32	0.18
Total Calcium (mg/day)		<500	500-699.9	700-899.9	900-1099.9	1100-1399.9	≥ 1400				
Cases (ER+/-)		2749 (1553/395)	5129 (3006/778)	4698 (2754/666)	3759 (2209/511)	3915 (2351/510)	5574 (3408/684)				
Total	Age-adjusted	1.00 (Ref)	0.98 (0.93-1.02)	0.97 (0.92-1.03)	0.96 (0.91-1.01)	0.98 (0.92-1.04)	1.00 (0.94-1.06)		0.26	0.20	0.95
	Multivariable										
		1.00 (Ref)	0.99 (0.95-1.04)	0.97 (0.93-1.02)	0.96 (0.92-1.02)	0.97 (0.91-1.04)	0.97 (0.91-1.04)		0.28	0.19	0.25

Table 2.4 (Continued)

Nutrients		Categories of intake Pooled HR (95% CI)						P_{diff}[‡]	I^2 [§]	P_{het}[‖]	P_{trend}[¶]
ER+	Multivariable	1.00 (Ref)	1.02 (0.96-1.08)	1.00 (0.93-1.06)	0.98 (0.92-1.05)	1.01 (0.94-1.10)	1.01 (0.93-1.09)		0.26	0.20	0.80
ER-	Multivariable	1.00 (Ref)	1.08 (0.95-1.22)	1.04 (0.89-1.22)	1.03 (0.86-1.23)	1.02 (0.88-1.18)	1.02 (0.88-1.17)	0.92	<0.01	0.82	0.57

[*]Multivariable model includes race (White, African American, Hispanic, Asian, other), education (<high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m^2), height (<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75 m), alcohol consumption (0,>0-<5, 5-<15, 15-<30 and ≥30 g/day), energy intake (kcal/d, continuous), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 years except PLCO: <10, 10/11, 12/13, ≥14 years), hormone replacement therapy use (never, past, current), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25 years), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

[†]ER: estrogen receptor

[‡]P_{diff}: p-value, test for common effects for different subtypes defined by estrogen receptor status for the highest category

[§]I^2 for the highest category

[‖]P_{het}: p-value, test for between studies heterogeneity for the highest category

[¶]P_{trend}: p-value, test for trend

Table 2.5 Pooled multivariable* hazard adjusted ratios (MVHR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER)† status stratified by participant characteristics

Increment	Total milk (250 g/day) MVHR (95% CI)	P_{het}‡	Hard cheese (14 g/day) MVHR (95% CI)	P_{het}‡	Cottage cheese (10 g/day) MVHR (95% CI)	P_{het}‡	Total yogurt (25 g/day) MVHR (95% CI)	P_{het}‡	Dietary calcium (350 mg/day) MVHR (95% CI)	P_{het}‡
Total Breast cancer (Cont.)	0.99 (0.98-1.00)	0.52	1.00 (0.99-1.01)	0.20	1.00 (0.99-1.00)	0.62	0.99 (0.99-1.00)	0.43	0.98 (0.97-0.99)	0.35
Premenopausal	0.96 (0.92-1.01)	0.28	1.00 (0.98-1.03)	0.53	1.00 (0.98-1.02)	0.49	1.00 (0.99-1.01)	0.99	0.96 (0.92-1.00)	0.89
Postmenopausal	1.00 (0.98-1.01)	0.67	1.00 (0.99-1.02)	0.24	1.00 (0.99-1.01)	0.45	0.99 (0.98-1.00)	0.12	0.99 (0.97-1.00)	0.69
p for interaction		0.12		0.81		0.82		0.24		0.15
Age ≥ 64 at diagnosis	0.99 (0.97-1.00)	0.99	1.00 (0.97-1.02)	0.01	1.00 (0.99-1.01)	0.34	0.99 (0.98-1.00)	0.06	0.98 (0.96-1.00)	0.55
Age < 64 at diagnosis	0.99 (0.97-1.01)	0.17	1.00 (0.98-1.01)	0.61	1.00 (0.99-1.01)	0.49	0.99 (0.99-1.00)	0.25	0.98 (0.96-1.01)	0.16
p for interaction		0.69		0.66		0.43		0.26		0.47
Follow up < 5 years	0.99 (0.97-1.00)	0.65	1.01 (0.99-1.03)	0.29	1.00 (0.99-1.01)	0.19	0.99 (0.99-1.00)	0.15	0.98 (0.96-1.00)	0.71
Follow up ≥ 5 years	1.00 (0.98-1.01)	0.43	0.99 (0.97-1.01)	0.05	1.00 (0.99-1.01)	0.30	0.99 (0.99-1.00)	0.80	0.98 (0.96-1.00)	0.43
p for interaction		0.45		0.03		0.97		0.53		0.91
BMI < 25 kg/m^2	0.99 (0.97-1.00)	0.60	1.01 (0.99-1.02)	0.38	1.00 (0.99-1.01)	0.60	0.99 (0.99-1.00)	0.37	0.98 (0.96-0.99)	0.69
BMI ≥ 25 kg/m^2	1.00 (0.98-1.02)	0.79	0.98 (0.97-1.00)	0.24	1.00 (0.99-1.00)	0.60	0.99 (0.99-1.00)	0.70	0.99 (0.97-1.01)	0.97
p for interaction		0.23		0.05		0.37		0.73		0.24

breast cancer being modified by follow-up time (p-value, test for interaction =0.03), and the association between cottage/ricotta cheese and risk of ER-negative breast cancer being modified by age at diagnosis and follow-up time (p-value, tests for interaction ≤ 0.03). However, the pooled HRs in each stratum were generally not statistically significant. The associations for dietary and total calcium did not vary by total vitamin D intake (p-value, test for interaction >0.19, results not shown).

DISCUSSION

In this pooled analysis, we found null or very weak inverse associations for consumption of total milk, yogurt, hard cheese, cottage/ricotta cheese, ice cream, dietary calcium, and total calcium with risk of overall and ER+ breast cancer. For ER-negative breast cancer, modest inverse associations were only observed for yogurt and cottage/ricotta cheese consumption when modeled as categorical variables; these associations were not statistically significant when their intakes were modeled as continuous variables. Results were generally consistent across studies and population subgroups defined by menopausal status at diagnosis, age at diagnosis, and BMI.

A recent meta-analysis of cohort and case-control studies showed a modest inverse relationship between dairy consumption and overall breast cancer risk and stronger inverse associations for yogurt and low-fat dairy products [14]. Of the 17 cohorts in that meta-analysis that examined diet during mid- to later-adulthood, seven were included in our pooled analysis but had 1-7 years longer follow-up. The meta-analysis included 10 studies that did not meet our inclusion criteria or were not yet participating in breast cancer analyses in our consortium, while our study included 14 cohorts which had not previously examined dairy products and breast cancer, minimizing the

influence of publication bias, a common limitation of meta-analyses of the published literature[60,61]. The meta-analysis did not examine breast cancer subtypes, while our study showed stronger associations for yogurt and cottage/ricotta cheese intake with risk of ER-negative than ER-positive breast cancer.

Our findings on calcium intake and overall breast cancer risk are consistent with a meta-analysis of 11 prospective cohort studies (of which 6 were included in our study) showing a significant but modest inverse relationship for dietary calcium[9]. Calcium has the potential to reduce breast cancer risk as it plays an important role in the regulation of cell proliferation, differentiation, and apoptosis [62]. Studies suggest that higher dietary calcium can markedly suppress Western-diet induced hyperproliferation of epithelial cells in mice[63,64], exert a pro-differentiation effect on mammary gland cells[65], and reduce the incidence of mammary tumors in rats[66]. Yet, in the Women's Health Initiative, calcium and vitamin D supplementation were not associated with breast cancer risk (HR=1.06, 95% CI 0.85 to 1.32)[67]. Mammographic density (higher breast density is associated with higher breast cancer risk[68]) also did not differ between the intervention and placebo groups [69]. The intervention dose, study duration, population studied (>60% were 60 years or older at baseline), and nonadherence may have contributed to the null findings for this trial. Our study and the meta-analysis[9] of calcium intake both found slightly stronger, although still weak, associations for dietary calcium than calcium from supplements, suggesting a synergistic effect of calcium and other nutrients in dairy foods and/or effects of other nutrients in dairy foods that are highly correlated with calcium. It is also possible that supplemental calcium reduces breast cancer risk only for women who are calcium deficient, resulting in supplementation having minimal benefit above and beyond adequate calcium intake from food.

There are a few explanations for the inverse association observed between yogurt intake and risk of ER-negative breast cancer. Yogurt consumption does not increase circulating IGF-1 as has been shown for other dairy products[70]. Probiotics - the beneficial living microorganisms enriched in yogurt – could also influence the mucosal immune system and its integration with the mammary glands[71,72]. Probiotics and fermented dairy products have been shown to boost intestinal microbiome richness, which might increase urinary estrogen[73]; induce apoptosis of breast cancer cell lines[74]; and counteract dietary and genetic predisposition to mammary cancer in mice[75]. Moreover, probiotics have been shown to be more enriched in controls than in breast cancer cases[76]. A lower risk of ER-negative breast cancer was also associated with cottage/ricotta cheese but not hard cheese. This may be attributable to the fluctuating viable bacterial counts during manufacturing and storage[77-79]; evidence has shown that the abundance of probiotic strains in low-fat hard cheese decreased over the ageing process[80].

The main strength of our study is that we analyzed the primary participant-level data from 21 prospective cohort studies, which made it possible to harmonize the definitions of the outcomes, exposures, and confounding variables, as well as the analytic strategy, allowing us to reduce potential sources of heterogeneity across studies due to different exposure and covariate definitions and use of different analytic approaches, as well as to estimate finer dose-response relationships than possible in meta-analyses of published studies. In addition, the large sample size gives us adequate statistical power to examine breast cancer subtypes defined by hormone receptor status, particularly less common subtypes. Lastly, the adjustments for known breast cancer risk factors minimized the likelihood of residual confounding strongly influencing our results. In fact, despite

the differences in the assessment methods used across studies for diet and confounding variables, there was no significant between-study heterogeneity in any of our main analyses.

Our study has limitations. Dietary intake was inevitably measured with error. However, moderate to high correlations between the measurements by FFQ and by dietary record or similar instruments have been reported in validation studies[26,37-47]. After correcting for measurement error, the association between calcium intake and risk of breast cancer changed only slightly. Because of the prospective study design, any measurement error should be nondifferential between the cases and non-cases, which would only bias the results toward the null. Since we only analyzed dietary data collected at baseline, the associations, if any, might be attenuated if diet changed substantially during follow-up. We were also not able to estimate consumption during earlier life periods, which could be biologically more relevant[81]. A recent study found that adolescent consumption of high-fat dairy products was positively associated with ER-negative-PR-negative breast cancer[82]. In a study of American women with generally lower milk consumption than reported in the studies in our analysis, milk consumption was associated with greater risk of overall and hormone receptor positive breast cancer[83] with the increase in risk being evident even at low consumption levels. Cheese and yogurt consumption were not associated with breast cancer risk in that study. We also could not further characterize other subtypes including luminal A, B, and basal-like subtypes. Lastly, our study population consisted predominantly of white women. The results might not be applicable to populations of other racial/ethnic compositions.

CONCLUSION

In summary, we found no clear associations between consumption of specific dairy foods, dietary

calcium, total calcium, and risk of overall breast cancer, while yogurt and cottage/ricotta cheese consumption were associated with a modestly lower risk of ER-negative tumors. Evaluation of these associations in more racially/ethnically diverse populations and in those with higher fermented dairy product consumption may help elucidate further any relation between dairy foods and breast cancer risk.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492.

2. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. *Cancer Epidemiology Biomarkers & Prevention*. 2004;13(10):1558-1568.

3. Yang XR, Chang-Claude J, Goode EL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst*. 2011;103(3):250-263. doi:10.1093/jnci/djq526.

4. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk Factors for Breast Cancer According to Estrogen and Progesterone Receptor Status. *JNCI Journal of the National Cancer Institute*. 2004;96(3):218-228. doi:10.1093/jnci/djh025.

5. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K. *SEER Cancer Statistics Review, 1975–2013*.

6. Zhang S, Zeng X, Ren M, Mao X, Qiao S. Novel metabolic and physiological functions of branched chain amino acids: a review. *J Anim Sci Biotechnol*. 2017;8(1):10-12. doi:10.1186/s40104-016-0139-z.

7. Białek A, Tokarz A. Conjugated linoleic acid as a potential protective factor in prevention of breast cancer. *Postepy Hig Med Dosw (Online)*. 2013;67:6-14. doi:10.5604/17322693.1028764.

8. Davoodi H, Esmaeili S, Mortazavian AM. Effects of Milk and Milk Products Consumption on Cancer: A Review. *Comprehensive Reviews in Food Science and Food Safety*. 2013;12(3):249-264. doi:10.1111/1541-4337.12011.

9. Hidayat K, Chen G-C, Zhang R, et al. Calcium intake and breast cancer risk: meta-analysis of prospective cohort studies. *Br J Nutr*. 2016;116(1):158-166. doi:10.1017/S000711 4516001768.

10. Harrison S, Lennon R, Holly J, et al. Does milk intake promote prostate cancer initiation or progression via effects on insulin-like growth factors (IGFs)? A systematic review and meta-analysis. *Cancer Causes Control*. 2017;28(6):497-528. doi:10.1007/s10552-017-0883-1.

11. Nachman KE, Smith TJS. Hormone Use in Food Animal Production: Assessing Potential Dietary Exposures and Breast Cancer Risk. *Curr Environ Health Rep*. 2015;2(1):1-14. doi:10.1007/s40572-014-0042-8.

12. Christopoulos PF, Msaouel P, Koutsilieris M. The role of the insulin-like growth factor-1 system in breast cancer. *MOLECULAR CANCER*. 2015;14(1):43. doi:10.1186/s12943-015-0291-7.

13. Farlow DW, Xu X, Veenstra TD. Quantitative measurement of endogenous estrogen metabolites, risk-factors for development of breast cancer, in commercial milk products by LC–MS/MS. *Journal of Chromatography B*. 2009;877(13):1327-1334. doi:10.1016/j.jchromb.2009.01.032.

14. Zang J, Shen M, Du S, Chen T, Zou S. The Association between Dairy Intake and Breast Cancer in Western and Asian Populations: A Systematic Review and Meta-Analysis. *Journal of Breast Cancer*. 2015;18(4):313-322. doi:10.4048/jbc.2015.18.4.313.

15. Smith-Warner SA, Spiegelman D, Ritz J, et al. Methods for Pooling Results of Epidemiologic Studies. *American Journal of Epidemiology*. 2006;163(11):1053-1064. doi:10.1093/aje/kwj127.

16. McCullough ML, Rodriguez C, Diver WR, et al. Dairy, Calcium, and Vitamin D Intake and Postmenopausal Breast Cancer Risk in the Cancer Prevention Study II Nutrition Cohort. *Cancer Epidemiol Biomarkers Prev*. 2005;14(12):2898-2904. doi:10.1158/1055-9965.EPI-05-0611.

17. Shin MH, Holmes MD, the SHJO. Intake of dairy products, calcium, and vitamin D and risk of breast cancer. *J Natl Cancer Inst*. 2002;94(17):1301-1311. doi:10.1093/jnci/94.17.1301".

18. Lin J, Manson JE, Lee I-M, Cook NR, Buring JE, Zhang SM. Intakes of Calcium and Vitamin D and Breast Cancer Risk in Women. *Arch Intern Med*. 2007;167(10):1050-1059. doi:10.1001/archinte.167.10.1050.

19. Omenn GS, Goodman G, Thornquist M, Grizzle J, Rosenstock L, Barnhart S. The β-carotene and retinol efficacy trial (CARET) for chemoprevention of lung cancer in high risk populations: Smokers and asbestos-exposed workers. *Lung Cancer*. 1994;11(5-6):423. doi:10.1016/0169-5002(94)92177-6.

20. Toniolo P, Riboli E, Shore RE, Pasternack BS. Consumption of meat, animal products, protein, and fat and risk of breast cancer: a prospective cohort study in New York. *Epidemiology*. 1994;5(4):391-397.

21. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Controlled Clinical Trials*. 2000;21(6):273S–309S. doi:10.1016/S0197-2456(00)00098-2.

22. Palmer JR, Rao RS, Adams-Campbell LL, Rosenberg L. Height and breast cancer risk: results from the Black Women's Health Study (United States). *Cancer Causes & Control*. 2001;12(4):343-348.

23. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ*. 2002;156:69-70.

24. Kumle M, Weiderpass E, Braaten T, Persson I, Adami H-O, Lund E. Use of oral contraceptives and breast cancer risk: The Norwegian-Swedish Women's Lifestyle and Health Cohort Study. *Cancer Epidemiol Biomarkers Prev*. 2002;11(11):1375-1381.

25. Terry P, Jain M, Miller AB, Howe GR, Rohan TE. Dietary carotenoids and risk of breast cancer. *Am J Clin Nutr*. 2002;76(4):883-888. doi:10.1093/ajcn/76.4.883.

26. Genkinger JM, Makambi KH, Palmer JR, Rosenberg L, Adams-Campbell LL. Consumption of dairy and meat in relation to breast cancer risk in the Black Women's Health Study. *Cancer Causes & Control*. 2013;24(4):675-684. doi:10.1007/s10552-013-0146-8.

27. Le Marchand L, Haiman CA, Wilkens LR, Kolonel LN, Henderson BE. MTHFR Polymorphisms, Diet, HRT, and Breast Cancer Risk: The Multiethnic Cohort Study. *Cancer Epidemiol Biomarkers Prev*. 2004;13(12):2071-2077.

28. Sieri S, Krogh V, Pala V, et al. Dietary patterns and risk of breast cancer in the ORDET cohort. *Cancer Epidemiol Biomarkers Prev*. 2004;13(4):567-572.

29. Velie EM, Schairer C, Flood A, He J-P, Khattree R, Schatzkin A. Empirically derived dietary patterns and risk of postmenopausal breast cancer in a large prospective cohort study. *Am J Clin Nutr*. 2005;82(6):1308-1319. doi:10.1093/ajcn/82.6.1308.

30. Sellers TA, Vierkant RA, Djeu J, et al. Unpasteurized milk consumption and subsequent risk of cancer. *Cancer Causes Control*. 2008;19(8):805-811. doi:10.1007/s10552-008-9143-8.

31. Umesawa M, Iso H, Ishihara J, et al. Dietary Calcium Intake and Risks of Stroke, Its Subtypes, and Coronary Heart Disease in Japanese. *Stroke*. 2008;39(9):2449-2456. doi:10.1161/STROKEAHA.107.512236.

32. Farvid MS, Cho E, Chen WY, Eliassen AH, Willett WC. Premenopausal dietary fat in relation to pre- and post-menopausal breast cancer. *Breast Cancer Res Treat*. 2014;145(1):255-265. doi:10.1007/s10549-014-2895-9.

33. Larsson SC, Andersson S-O, Johansson J-E, Wolk A. Cultured milk, yogurt, and dairy intake in relation to bladder cancer risk in a prospective study of Swedish women and men. *Am J Clin Nutr*. 2008;88(4):1083-1087. doi:10.1093/ajcn/88.4.1083.

34. Goldbohm RA, Chorus AMJ, Galindo Garre F, Schouten LJ, van den Brandt PA. Dairy consumption and 10-y total and cardiovascular mortality: a prospective cohort study in the Netherlands. *Am J Clin Nutr*. 2011;93(3):615-627. doi:10.3945/ajcn.110.000430.

35. Link LB, Canchola AJ, Bernstein L, et al. Dietary patterns and breast cancer risk in the California Teachers Study cohort. *Am J Clin Nutr*. 2013;98(6):1524-1532. doi:10.3945/ajcn.113.061184.

36. Willett W. *Nutritional Epidemiology*. Oxford University Press; 2013.

37. Sasaki S, Takahashi T, Iitoi Y, et al. Food and Nutrient Intakes Assessed with Dietary Records for the Validation Study of a Self-administered Food Frequency Questionnaire in JPHC Study Cohort I. *Journal of Epidemiology*. 2003;13(1sup):23-50. doi:10.2188/jea.13.1sup_23.

38. Romieu I, Stampfer MJ, Stryker WS, et al. Food Predictors of Plasma Meta-carotene and Alpha-tocopherol: Validation of a Food Frequency Questionnaire. *American Journal of Epidemiology*. 1990;131(5):864-876. doi:10.1093/oxfordjournals.aje.a115577.

39. Koralek DO, Bertone-Johnson ER, Leitzmann MF, et al. Relationship between calcium, lactose, vitamin D, and dairy products and ovarian cancer. *Nutr Cancer*. 2006;56(1):22-30. doi:10.1207/s15327914nc5601_4.

40. Goldbohm RA, van den Brandt PA. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *European journal of clinical nutrition*.

41. Flagg EW, Coates RJ, Calle EE, Potischman N, Thun MJ. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort Food Frequency Questionnaire. *Epidemiology*. 2000;11(4):462-468. doi:10.1097/00001648-200007000-00017.

42. Kaaks R. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol*. 1997;26(90001):26S–36. doi:10.1093/ije/26.suppl_1.S26.

43. Salvini S, Hunter DJ, Sampson L, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol*. 1989;18(4):858-867. doi:10.1093/ije/18.4.858.

44. Munger RG, Folsom AR, Kushi LH, Kaye SA, Sellers TA. Dietary Assessment of Older Iowa Women with a Food Frequency Questionnaire: Nutrient Intake, Reproducibility, and Comparison with 24-Hour Dietary Recall Interviews. *American Journal of Epidemiology*. 1992;136(2):192-200. doi:10.1093/oxfordjournals.aje.a116485.

45. Bassett JK, English DR, Fahey MT, et al. Validity and calibration of the FFQ used in the Melbourne Collaborative Cohort Study. *Public Health Nutrition*. 2016;19(13):2357-2368. doi:10.1017/S1368980016000690.

46. Bostick RM, Potter JD, Sellers TA, McKenzie DR, Kushi LH, Folsom AR. Relation of calcium, vitamin D, and dairy food intake to incidence of colon cancer among older women. The Iowa Women's Health Study. *American Journal of Epidemiology*. 1993;137(12):1302-1317. doi:10.1093/oxfordjournals.aje.a116640.

47. Kumanyika S. Relative Validity of Food Frequency Questionnaire Nutrient Estimates in the Black Women's Health Study. *Annals of Epidemiology*. 2003;13(2):111-118. doi:10.1016/S1047-2797(02)00253-3.

48. Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986;73:1-11. doi:10.1093/biomet/73.1.1","keywords“:[”Case-control.

49. Joffe MM, Rosenbaum PR. Invited Commentary: Propensity Scores. *American Journal of Epidemiology*. 1999;150(4):327-333. doi:10.1093/oxfordjournals.aje.a010011.

50. Cepeda MS. Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders. *American Journal of Epidemiology*. 2003;158(3):280-287. doi:10.1093/aje/kwg115.

51. GRAMBSCH PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526. doi:10.1093/biomet/81.3.515.

52. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2.

53. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560. doi:10.1136/bmj.327.7414.557.

54. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in Medicine*. 1989;8(5):551-561.

55. Smith-Warner SA, Spiegelman D, Yaun S-S, et al. Alcohol and Breast Cancer in Women. *JAMA*. 1998;279(7):535. doi:10.1001/jama.279.7.535.

56. Stram DO. Meta-analysis of published data using a linear mixed-effects model. *Biometrics*. 1996;52(2):536-544.

57. Wang M, Spiegelman D, Kuchiba A, et al. Statistical methods for studying disease subtype heterogeneity. *Statistics in Medicine*. 2016;35(5):782-800. doi:10.1002/sim.6793.

58. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*. 1990;132(4):734-745. doi:10.1093/oxfordjournals.aje.a115715.

59. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*. 1989;8(9):1051–69–discussion1071–3.

60. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ*. 2000;320(7249):1574-1577. doi:10.1136/bmj.320.7249.1574.

61. Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*. 2000;53(11):1119-1129. doi:10.1016/S0895-4356(00)00242-0.

62. Whitfield JF, Boynton AL, Macmanus JP, Sikorska M, Tsang BK. The regulation of cell proliferation by calcium and cyclic AMP. *Mol Cell Biochem*. 1979;27(3):155-179. doi:10.1007/BF00215364.

63. Xue L, Lipkin M, Newmark H, Wang J. Influence of dietary calcium and vitamin D on diet-induced epithelial cell hyperproliferation in mice. *JNCI Journal of the National Cancer Institute*. 1999;91(2):176-181. doi:10.1093/jnci/91.2.176.

64. McGrath CM, Soule HD. Calcium regulation of normal human mammary epithelial cell growth in culture. *In Vitro*. 1984;20(8):652-662.

65. Jacobson EA, James KA, Newmark HL, Carroll KK. Effects of dietary fat, calcium, and vitamin D on growth and mammary tumorigenesis induced by 7,12-dimethylbenz(a)anthracene in female Sprague-Dawley rats. *Cancer Res*. 1989;49(22):6300-6303.

66. Abou-Issa H, Moeschberger M, el-Masry W, Tejwani S, Curley RW, Webb TE. Relative efficacy of glucarate on the initiation and promotion phases of rat mammary carcinogenesis. *Anticancer Res*. 1995;15(3):805-810.

67. Prentice RL, Pettinger MB, Jackson RD, et al. Health risks and benefits from calcium and vitamin D supplementation: Women's Health Initiative clinical trial and cohort study. *Osteoporos Int*. 2013;24(2):567-580. doi:10.1007/s00198-012-2224-2.

68. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 1998;7(12):1133-1144.

69. Bertone-Johnson ER, McTiernan A, Thomson CA, et al. Vitamin D and calcium supplementation and one-year change in mammographic density in the women's health initiative calcium and vitamin D trial. *Cancer Epidemiol Biomarkers Prev*. 2012;21(3):462-473. doi:10.1158/1055-9965.EPI-11-1009.

70. Romo Ventura E, Konigorski S, Rohrmann S, et al. Association of dietary intake of milk and dairy products with blood concentrations of insulin-like growth factor 1 (IGF-1) in Bavarian adults. *Eur J Nutr*. 2019;1(2):147. doi:10.1007/s00394-019-01994-7.

71. Brandtzaeg P. The Mucosal Immune System and Its Integration with the Mammary Glands. *The Journal of Pediatrics*. 2010;156(2):S8-S15. doi:10.1016/j.jpeds.2009.11.014.

72. Ranjbar S, Seyednejad SA, Azimi H, Rezaeizadeh H, Rahimi R. Emerging Roles of Probiotics in Prevention and Treatment of Breast Cancer: A Comprehensive Review of Their Therapeutic Potential. *Nutr Cancer*. 2019;71(1):1-12. doi:10.1080/01635581.2018.1557221.

73. Flores R, Shi J, Fuhrman B, et al. Fecal microbial determinants of fecal and systemic estrogens and estrogen metabolites: a cross-sectional study. *J Transl Med*. 2012;10(1):253. doi:10.1186/1479-5876-10-253.

74. Hassan Z, Mustafa S, Rahim RA, Isa NM. Anti-breast cancer effects of live, heat-killed and cytoplasmic fractions of Enterococcus faecalis and Staphylococcus hominis isolated from human breast milk. *In Vitro Cell Dev Biol Anim*. 2016;52(3):337-348. doi:10.1007/s11626-015-9978-8.

75. Lakritz JR, Poutahidis T, Levkovich T, et al. Beneficial bacteria stimulate host immune cells to counteract dietary and genetic predisposition to mammary cancer in mice. *Int J Cancer*. 2014;135(3):529-540. doi:10.1002/ijc.28702.

76. Goedert JJ, Jones G, Hua X, et al. Investigation of the Association Between the Fecal Microbiota and Breast Cancer in Postmenopausal Women: a Population-Based Case-Control Pilot Study. *JNCI Journal of the National Cancer Institute*. 2015;107(8):995. doi:10.1093/jnci/djv147.

77. Baruzzi F, Morea M, Matarante A, Cocconcelli PS. Changes in the Lactobacillus community during Ricotta forte cheese natural fermentation. *J Appl Microbiol*. 2000;89(5):807-814. doi:10.1046/j.1365-2672.2000.01183.x.

78. Lovayová V, Dudriková E, Rimárová K, Siegfried L. Quantity of selected probiotic cultures in semi-hard cheese with low-cooking curd during the maturation process. *J Food Sci Technol*. 2015;52(8):4697-4702. doi:10.1007/s13197-014-1619-9.

79. Kok CR, Hutkins R. Yogurt and other fermented foods as sources of health-promoting bacteria. *Nutrition Reviews*. 2018;76(Supplement_1):4-15. doi:10.1093/nutrit/nuy056.

80. Ganesan B, Weimer BC, Pinzon J, et al. Probiotic bacteria survive in Cheddar cheese and modify populations of other lactic acid bacteria. *J Appl Microbiol*. 2014;116(6):1642-1656. doi:10.1111/jam.12482.

81. Hjartåker A, Laake P, Lund E. Childhood and adult milk consumption and risk of premenopausal breast cancer in a cohort of 48,844 women - the Norwegian women and cancer study. *Int J Cancer*. 2001;93(6):888-893.

82. Farvid MS, Eliassen AH, Cho E, Chen WY, Willett WC. Dairy Consumption in Adolescence and Early Adulthood and Risk of Breast Cancer. *Cancer Epidemiol Biomarkers Prev*. 2018;27(5):575-584. doi:10.1158/1055-9965.EPI-17-0345.

83. Fraser GE, Jaceldo-Siegl K, Orlich M, Mashchak A, Sirirat R, Knutsen S. Dairy, soy, and risk of breast cancer: those confounded milks. *Int J Epidemiol*. 2020;3:22. doi:10.1093/ije/dyaa007.

Chapter 3

Total red meat, unprocessed red meat, and processed meat and risk of breast cancer - a pooled analysis of substitution with alternative protein sources

ABSTRACT

Background: The relationship between red meat consumption and breast cancer has been evaluated in several epidemiological studies, yet there has been no clear scientific consensus on the association. To date, no study has comprehensively investigated the effect of substituting other major protein sources as alternatives to red meat.

Methods: This consortial study included 23 prospective cohort studies with over 1 million women among whom 46,176 breast cancer cases were diagnosed during follow-up. Study-specific multivariable hazard ratios (MVHR) and 95% confidence intervals (CI) were calculated by Cox proportional hazards regression with intakes of all groups of protein sources simultaneously in the model, then pooled using a random-effects model. In substitution analyses, results were presented for an increment of 200kcal/day for total red meat and unprocessed red meat, and of 50kcal/day for processed meat.

Results: Total red meat, processed meat, and unprocessed meat intakes were not significantly associated with risk of breast cancer when holding other protein sources constant. However, when substituting 200kcal/day of red meat with an energy-equivalent amount of mature beans, an inverse association was observed (pooled MVHR = 0.92, 95%CI: 0.87 - 0.98); the association was largely due to unprocessed red meat (pooled MVHR = 0.91, 95% CI: 0.85 – 0.97) not processed meat (pooled MVHR = 0.99, 95% CI: 0.97 – 1.00). When substituting 200kcal/day of red meat with an

energy-equivalent amount of dairy products, the pooled MVHR was 0.96 (95% CI: 0.94 - 0.99). There were no significant substitution effect replacing total or unprocessed red meat with poultry, seafood, eggs, or nuts. The results were similar for ER-positive and ER-negative breast cancer.

Conclusions: We observed slightly lower risks of breast cancer when substituting consumption of total red meat or unprocessed red meat with mature beans or dairy products. Replacing red/processed meat with healthy alternatives is recommended considering results from this study as well as the overall benefit for cancer prevention.

INTRODUCTION

Breast cancer accounts for more than 2 million cases each year. Despite its high 5-year survival rate, the high prevalence still makes it the leading cause of cancer death in women worldwide[1].

In the 2018 expert report on nutrition and cancer, the World Cancer Research Fund/American Institute for Cancer Research included limiting red and processed meat consumption as one of 10 modifiable lifestyle recommendations for cancer prevention[2]. The International Agency for Research on Cancer has also classified processed meat as class 1 carcinogen (causes cancer), and unprocessed red meat as class 2a carcinogen (probably causes cancer) [3]. Potential mechanisms as to why red meat may increase the risk of developing breast cancer include the presence of exogenous hormone residues in red meat which may be estrogenic [4]; carcinogenic heterocyclic aromatic amines generated from the cooking of meats at high temperature and for long duration [5-7]; and heme which has been shown to catalyze oxidative damage [8].

The relationship between red and processed meat and breast cancer risk has been evaluated in several epidemiological studies, yet there has been no clear scientific consensus [9-11]. One possible reason for the inconsistency points to the analytical approach - since red meat is a major protein source, any evaluation of risk between higher and lower red meat intake also inherently involves the effects from other protein or energy sources being substituted for red meat. It is critical that, when analyzing red meat or other major energy sources and any disease outcome, other protein sources be considered as confounders, or as alternative energy sources that could replace red meat intake, given the fact that individual's total energy intake remains relatively stable no matter their choice of dietary patterns [12,13]. To date, no study has comprehensively investigated the effect of

substituting other major protein sources as alternatives to red meat in relation to breast cancer risk. We conducted a pooled analysis within the Pooling Project of Prospective Studies of Diet and Cancer (DCPP) to comprehensively examine the association between red meat intake overall, as well as the substitutive effects of other major protein sources for red meat on risk of breast cancer.

METHODS

Study population. In this study, we analyzed the primary participant-level data from 23 studies in the DCPP, an international consortium of prospective cohort studies (Table 1). All participating studies met the following inclusion criteria: 1) at least one publication on any diet and cancer association; 2) assessed food groups of interest with a comprehensive long-term dietary assessment tool; 3) validated the dietary assessment tool or a closely related instrument; and 4) included at least 25 incident ER- breast cancer cases. Each included study was approved by their respective institutional review board.

Assessment of Dietary and Non-dietary Factors. Participant-level data were collected from each cohort, centralized, and harmonized before analyzing. Intakes of individual food items were assessed at baseline by validated study-specific food frequency questionnaires (FFQ) generally covering the past year in each study, then converted to grams consumed per day.

Food groups of interest included the following major protein sources: red meat (unprocessed and processed), poultry, seafood (fish and shellfish), eggs, dairy products, mature beans (dried beans excluding soybeans, peas, and lentils), soybeans, and nuts. Red meat refers to mammalian muscle meat, usually consists of animal fat, heme iron, and possibly ingested hormones. Processed meat

includes meat transformed through salting, curing, fermentation, smoking, or other processes to enhance flavor or improve preservation. Given that many cohorts did not distinguish between processed red and processed white meat on their FFQs, the processed meat group may contain some processed white meat. However, poultry consumption remained about one third to half of red meat consumption in countries such as the US at the time the FFQs were completed[14], and processed white meat comprised even smaller proportion of processed meat. Seafood was analyzed as a single group because only 12 studies asked fish and shellfish intake separately. We were able to assess dairy product intake as high-fat and low-fat groups in all studies except for the Japan Public Health Center-based Study Cohort I which only asked about "milk and dairy products (except cheeses) and cheese consumption. Overall, 10 studies assessed all food groups above, 10 studies assessed all but soy product intake, one study assessed all but nut intake, and two studies did not measure intakes of nuts or soy products. Although very few studies conducted food-based validation studies, for the studies that had assessed food items of our interest in their validation studies, the correlation coefficients between intakes obtained from FFQ or a closely related instrument and the reference method, generally ranged from 0.4 to 0.8 for women[15-17].

In order to account for differences in gram weights between foods with different liquid content, we converted intake of each food group from grams/day to kcal/day. Since the energy density varied substantially for high-fat versus low-fat dairy and between solid and fluid dairy products, we applied energy conversions for specific dairy foods rather than for the total dairy product food group. For each of the remaining food groups, an average energy intake conversion factor calculated from the energy content per gram of a few select foods in that group was applied to the

gram weight intake of the specific food groups. For example, 432 kcal was applied to every 100g of processed meat; 291 kcal was applied to every 100g of unprocessed red meat.

Each study collected data on age, height, body weight, and race at baseline. Most studies also assessed family history of breast cancer, educational attainment, physical activity, smoking status, age at menarche, oral contraceptive use, age at first, parity, history of benign breast disease, and hormone replacement therapy.

Case Ascertainment. During follow-up of each study, women diagnosed with incident invasive breast cancer (International Classification of Diseases (ICD)-9 code 174.0 or ICD-10 code C50) were identified through follow-up questionnaires with subsequent medical record review, linkage with cancer registries, or both methods. Hormone receptor status were obtained through cancer registries, pathology reports, medical records, or laboratory determinations. Cases with borderline hormone receptor status were considered as positive. Case ascertainment generally exceeded 90% across studies.

Statistical analysis. After applying the exclusion criteria used in each study, we further excluded women with extreme total energy intakes (beyond three standard deviations from the mean $\log_e$-transformed energy intake) or who had been diagnosed with any cancer other than non-melanoma skin cancer prior to baseline. If a study did not collect data on a given food group, participants in that study were not included in the model for that specific food group, but still contributed to the estimation of the other food groups.

We analyzed the primary participant-level data in each cohort. The Netherlands Cohort Study was analyzed as a case-cohort study, as required by its study design. The Nurses' Health Study was separated into two cohorts because of the more detailed dietary assessment after 1986 compared with 1980.

The pooled multivariable hazard ratios (MVHR) were obtained by a 2-stage approach. Cox proportional hazards regression (SAS PROC PHREG) was used to estimate the study-specific risk of breast cancer overall and the subtypes. Follow-up time was calculated for each individual from the age when the baseline questionnaire was returned to the age of breast cancer diagnosis, death, loss to follow-up, or the end of follow-up, whichever came first. In a previous study from this consortium, we did not observe significant nonlinearity in the association between total red meat, unprocessed red meat, or processed meat intake and risk of breast cancer [manuscript in preparation], therefore, we modeled intakes as continuous variables. We calculated study-specific MVHR and 95% confidence intervals (CI) by Cox proportional hazards regression. Intake of total red meat as a single food group or intakes of both unprocessed red meat and processed meat were included in separate models together with poultry, seafood, eggs, dairy products, mature beans, and nuts simultaneously, plus total energy and other breast cancer risk factors (see footnotes of Table 2 for details). We created missing indicator variables for covariates with missing values.

In sensitivity analyses, we 1) separated dairy products into high-fat and low-fat dairy products and 2) combined intakes of nuts, mature beans, and soy product intake into a plant protein group and included that group in the model. To assess the robustness of the models, we conducted two sets of analyses where we modeled the intake of each food group as (1) absolute energy substitution

(replacing with equivalent amount of calories), which is intuitive for result dissemination and public health messaging; or (2) energy density substitution (replacing with equivalent percent of total energy intake), which would be less influenced by between-person variation in total energy intake, but less interpretable. To obtain isocaloric substitution effects, in both settings, the difference in the β coefficients between an alternative protein food group and the red meat group were risk estimates for the substitution effect. We show the results for substituting 200 kcal/day of other protein sources for total or unprocessed red meat, and 50 kcal/day for processed meat based on their intake distributions.

We pooled the study-specific hazard ratios (HRs) reflecting the substitution effect using a random-effects model by inverse of variance weighting [18]. Between-studies heterogeneity was evaluated using the Q statistic [18] and I^2 statistic [19]. We used a contrast test to examine whether the risk estimates varied by breast cancer subtypes defined by estrogen receptor status [20].

To test for possible effect modification, we also conducted stratification analyses by the following factors: body mass index (BMI, < 25, 25 - <30, ≥30 kg/m^2), menopausal status at diagnosis (premenopausal, postmenopausal; estimated using a previously described algorithm based on age at diagnosis[21]), age at diagnosis (< 55, 55 - <65, ≥65 years), follow-up duration (<5, 5 - <10, ≥10 years), and region (North America, others). We estimated the exposure-breast cancer association within each category of the potential effect modifiers, and obtained the p-values for interaction by fitting the product term of the potential effect modifier and the exposure of interest in the model then pooling across the studies[22]. Differences by region were evaluated using a mixed-effects meta-regression model [23].

For all tests of statistical hypotheses, 95% confidence intervals (CIs) were two-sided Wald CIs, and two-sided $p < 0.05$ were considered statistically significant. All analyses were conducted using SAS software versions 9.4 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

Among the 1,273,551 participants in this study, 46,176 invasive breast cancer cases were diagnosed over the follow-up ranging 6 to 21 years. 28,870 cases were confirmed as ER+, 6354 as ER-, and the status was missing for the remaining 23.7% cases (Table 3.1).

The intake levels varied across the groups of protein sources and also across studies. Median red meat intake ranged from 62.6 kcal/day (roughly 1/5 serving of sirloin steak per day) in the California Teachers Study to 336.4 kcal/day (roughly 1 serving of sirloin steak per day) in the Canadian National Breast Screening Study (Table 3.2). Sample conversions of food items in other food groups are presented in Table S3.2. Median intake of poultry (range across studies: 14 - 129 kcal/day), seafood (8 - 40 kcal/day), eggs (6 - 37 kcal/day), total dairy (61 - 346 kcal/day), mature beans (0 - 46 kcal/day), and nuts (0 - 27 kcal/day) also reflect substantial between-study variation. Total red meat intake was highly correlated with unprocessed red meat intake and processed meat (median Pearson correlation coefficient across studies = 0.89 and 0.68, respectively). Poultry and seafood intake were moderately correlated (median Pearson correlation = 0.26). The correlations between other food groups were all < 0.20 (Table S3.1).

Table 3.1 Characteristics of the cohort studies included in the pooled analysis of red meat intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer

Study	Baseline cohort size	Mean follow-up time (yr)	Baseline age range (yr)	No. of cases		
				Total	ER+	ER-
NIH-AARP Diet and Health Study	200049	7.1	50-71	5972	2322	464
Breast Cancer Detection Demonstration Project Follow-up Study	42061	8.4	40-93	1305	793	166
Beta-Carotene and Retinol Efficacy Trial	6000	11.6	48-71	367	193	31
CLUE II: Campaign Against Cancer and Heart Disease	8279	14.9	18-93	288	198	50
Canadian National Breast Screening Study	45185	16.3	40-59	1240	367	125
Cancer Prevention Study II Nutrition Cohort	74137	9.3	40-87	2999	1835	323
California Teachers Study	100067	7.6	22-104	2696	1930	343
Iowa Women's Health Study	34584	15.9	52-71	1849	1329	238
Japan Public Health Center-based Study Cohort I	21609	13.7	40-59	289	111	69
Melbourne Collaborative Cohort Study	22456	13.2	31-75	799	493	171
Multiethnic Cohort Study	92435	10.1	45-78	3308	2169	543
Nurses' Health Study (a)	88618	6.3	34-67	1122	528	255
Nurses' Health Study (b)	68394	24	40-67	5667	4272	907
Nurses' Health Study II	93778	21.3	26-46	2921	2170	576
Netherlands Cohort Study	62573	6.8	54-70	2013	700	183
New York University Women's Health Study	13257	15.7	31-70	919	392	121
Hormones and Diet in the Etiology of Breast Cancer Risk	9044	12	34-70	283	206	67
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	28292	9.1	52-74	1090	858	137
Swedish Mammography Cohort	60950	15.2	38-76	2605	1605	384
VITamins And Lifestyle Cohort Study	30158	9.1	50-76	1203	728	124
Women's Health Initiative	85726	16.2	49-81	4992	3997	694
Women's Health Study	38385	9.7	38-89	1177	937	187
Women's Lifestyle and Health Study	47514	14.9	32-50	1072	737	196
Total	1443073			46176	32454	7485

Table 3.2 Distribution of total red meat, processed meat, and unprocessed meat in the pooled analysis of red meat intake and risk of breast cancer in the Pooling Project of Prospective Studies of Diet and Cancer

Study	Total red meat	Processed meat	Unprocessed red meat
	Median (10th - 90th), kcal/d		
NIH-AARP Diet and Health Study	118.3 (28.7, 317.9)	36.3 (6.7, 137.2)	68.3 (13.4, 205.3)
Breast Cancer Detection Demonstration Project Follow-up Study	100.1 (17, 259.4)	27.5(0, 122.5)	60.7(8.2, 157.2)
Beta-Carotene and Retinol Efficacy Trial	113.2 (32.5, 260.8)	29.3 (0, 102.2)	73.5 (20.3, 180.5)
CLUE II: Campaign Against Cancer and Heart Disease	111.7 (22.2, 274.1)	38.4 (0, 141.2)	61.1 (8.2, 155.1)
Canadian National Breast Screening Study	336.4 (160.8, 610)	59.1 (9.9, 151.7)	264.7 (118.8, 487.3)
Cancer Prevention Study II Nutrition Cohort	110.7 (25.6, 264)	24.5 (0, 102.2)	76.3 (17.3, 181.1)
California Teachers Study	62.6 (0, 178.6)	11.1 (0, 69.1)	41.5 (0, 121.4)
Iowa Women's Health Study	251.2 (92.4, 521.3)	20.8 (0, 73.6)	231.9 (72.6, 471.1)
Japan Public Health Center-based Study Cohort I	67.1 (23.7, 123.4)	18.5 (0, 43.2)	48.6 (23.7, 81.9)
Melbourne Collaborative Cohort Study	302.1 (101.7, 611.4)	57.5 (0, 159)	226.7 (67.2, 505.2)
Multiethnic Cohort Study	100.3 (13.4, 303.8)	36.6 (0, 134.6)	54.8 (0, 183.6)
Nurses' Health Study (a)	331.4 (127.2, 640.9)	29.9 (0, 95.7)	300.4 (106.4, 603.1)
Nurses' Health Study (b)	197.7 (77.1, 407.3)	29.4 (0, 81.9)	141 (67.6, 318.4)
Nurses' Health Study II	184.8 (54.7, 396.3)	24.9 (0, 70.2)	159.3 (52.4, 351)
Netherlands Cohort Study	212.9 (101.9, 351.7)	42 (5.2, 123.3)	160.7 (67.8, 270.7)
New York University Women's Health Study	103.5 (19.3, 267.5)	22 (0, 111.6)	71.3 (10.9, 175.4)
Hormones and Diet in the Etiology of Breast Cancer Risk	312.2 (127.5, 607.5)	85.5 (21.2, 224.2)	206.1 (64, 443.3)
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	122.5 (36.1, 307.2)	26.4 (6.1, 101.5)	86.2 (23.8, 223.5)
Swedish Mammography Cohort	167.4 (71.7, 296.9)	90.3 (23.2, 188.7)	69.1 (31, 141.7)
VITamins And Lifestyle Cohort Study	123.2 (23.5, 318.5)	18.8 (0, 96.7)	85.6 (12.7, 262.6)
Women's Health Initiative	98.2 (17.5, 279.6)	23.3 (0, 98.6)	61.8 (7.9, 206.8)
Women's Health Study	168.6 (52.4, 381.4)	15.6 (0, 54.6)	144.8 (39.6, 351)
Women's Lifestyle and Health Study	145.7 (61.5, 261.4)	15.2 (0, 53.1)	123.1 (50.6, 225.2)

Multivariable adjusted models showed that total red meat, processed meat, and unprocessed meat intakes were not significantly associated with risk of breast cancer overall or of subtypes defined by ER status when holding other alternative protein sources constant (Table 3.3). The pooled MVHR ranged from 1.00 to 1.02 for every 200-kcal/day increment of total or unprocessed red meat, and from 0.99 to 1.00 for every 50-kcal/day increment of processed red meat. Total dairy and mature bean intakes were associated with lower risk of breast cancer. The pooled MVHR of breast cancer overall associated with every 200-kcal/day increase in intake were 0.97 (95% CI 0.96-0.99) for total dairy and 0.94 (95% CI 0.89-0.99) for mature beans. The pooled MVHR ranged from 0.97 – 1.06 for other protein groups; all were not statistically significant. Associations for each food group were slightly stronger for ER- than ER+ tumors, although the differences in the associations between the two subtypes were not statistically significant.

Results for substitution analyses were generally null with a few exceptions (Figure 3.1). When substituting an increment of 200kcal/day of red meat with an energy-equivalent amount of mature beans, breast cancer risk was 8% lower (pooled MVHR = 0.92, 95% CI: 0.87 – 0.98); the association was largely due to unprocessed red meat (pooled MVHR = 0.91, 95% CI: 0.85 – 0.97). A weaker, nonsignificant association was observed for processed meat (pooled MVHR for a 50 kcal/d increment = 0.99, 95% CI: 0.97 – 1.00). The substitution effect of plant proteins (nuts, mature beans, and soybeans) for red meat consumption was in the same directly albeit smaller in magnitude, given the associations for nuts and soybeans were weaker than those observed for mature beans. When substituting 200kcal/day of red meat with an energy-equivalent amount of dairy products, the pooled MVHR was 0.96 (95% CI: 0.94 -–0.99). The effect was slightly stronger for high-fat dairy products (pooled MVHR = 0.95, 95% CI: 0.92 – 0.99) than that for low-fat

Table 3.3 Pooled multivariable[1] hazard ratio (MVHR) of breast cancer overall and subtypes defined by ER status by major protein sources

Breast cancer	**Red meat**		**Other protein sources**					
Overall	Total (200kcal/d)		Poultry (200kcal/d)	Seafood (200kcal/d)	Eggs (200kcal/d)	Total dairy (200kcal/d)	Mature beans (200kcal/d)	Nuts (200kcal/d)
Pooled MVHR (95% CI)	1.01 (0.99-1.03)		1.00 (0.98-1.03)	0.97 (0.90-1.05)	1.06 (0.98-1.15)	0.97 (0.96-0.99)	0.94 (0.89-0.99)	1.00 (0.96-1.04)
P_{het}[2]	0.527		0.733	0.110	0.865	0.132	0.840	0.921
	Unprocessed (200kcal/d)	Processed meat (50kcal/d)						
Pooled MVHR (95% CI)	1.02 (1.00-1.04)	1.00 (0.99-1.01)	1.00 (0.98-1.03)	0.97 (0.90-1.05)	1.06 (0.98-1.15)	0.97 (0.96-0.99)	0.94 (0.89-0.99)	1.00 (0.96-1.04)
P_{het}[2]	0.702	0.602	0.729	0.115	0.850	0.120	0.825	0.914
ER+	Total (200kcal/d)		Poultry (200kcal/d)	Seafood (200kcal/d)	Eggs (200kcal/d)	Total dairy (200kcal/d)	Mature beans (200kcal/d)	Nuts (200kcal/d)
Pooled MVHR (95% CI)	1.01 (0.99-1.03)		1.03 (0.98-1.07)	0.98 (0.89-1.09)	1.10 (0.98-1.24)	0.98 (0.95-1.00)	0.95 (0.88-1.02)	0.99 (0.94-1.04)
P_{het}[2]	0.627		0.313	0.088	0.264	0.081	0.931	0.961
	Unprocessed (200kcal/d)	Processed meat (50kcal/d)						
Pooled MVHR (95% CI)	1.02 (0.99-1.05)	1.00 (0.98-1.01)	1.03 (0.98-1.07)	0.98 (0.89-1.09)	1.10 (0.97-1.24)	0.98 (0.95-1.00)	0.95 (0.88-1.02)	0.99 (0.94-1.04)
P_{het}[2]	0.926	0.224	0.323	0.087	0.234	0.082	0.923	0.959

Table 3.3 (Continued)

ER-	Total (200kcal/d)		Poultry (200kcal/d)	Seafood (200kcal/d)	Eggs (200kcal/d)	Total dairy (200kcal/d)	Mature beans (200kcal/d)	Nuts (200kcal/d)
Pooled MVHR (95% CI)	1.00 (0.96-1.05)		0.95 (0.87-1.02)	0.98 (0.84-1.15)	0.91 (0.72-1.15)	0.95 (0.91-1.00)	0.92 (0.79-1.07)	1.03 (0.89-1.21)
P_{het}^2	0.416		0.563	0.502	0.558	0.300	0.730	0.016
	Unprocessed (200kcal/d)	Processed meat (50kcal/d)						
Pooled MVHR (95% CI)	1.00 (0.94-1.07)	0.99 (0.97-1.02)	0.95 (0.87-1.02)	0.98 (0.83-1.15)	0.92 (0.73-1.15)	0.95 (0.91-1.00)	0.92 (0.79-1.07)	1.03 (0.89-1.21)
P_{het}^2	0.103	0.774	0.551	0.495	0.598	0.304	0.717	0.015

1. Multivariable models include race (White, African American, Hispanic, Asian, other), education (< high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m2), height(<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75m), alcohol (0, 0-<5, 5-<15, 15-<30 and ≥30 g/d), energy intake (continuous, kcal/d), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 except PLCO <10,10/11,12/13, ≥14 years), hormone replacement therapy use (never user, past user, current user), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25y), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

2. P_{het}^2 : p for between-study heterogeneity for continuous analysis

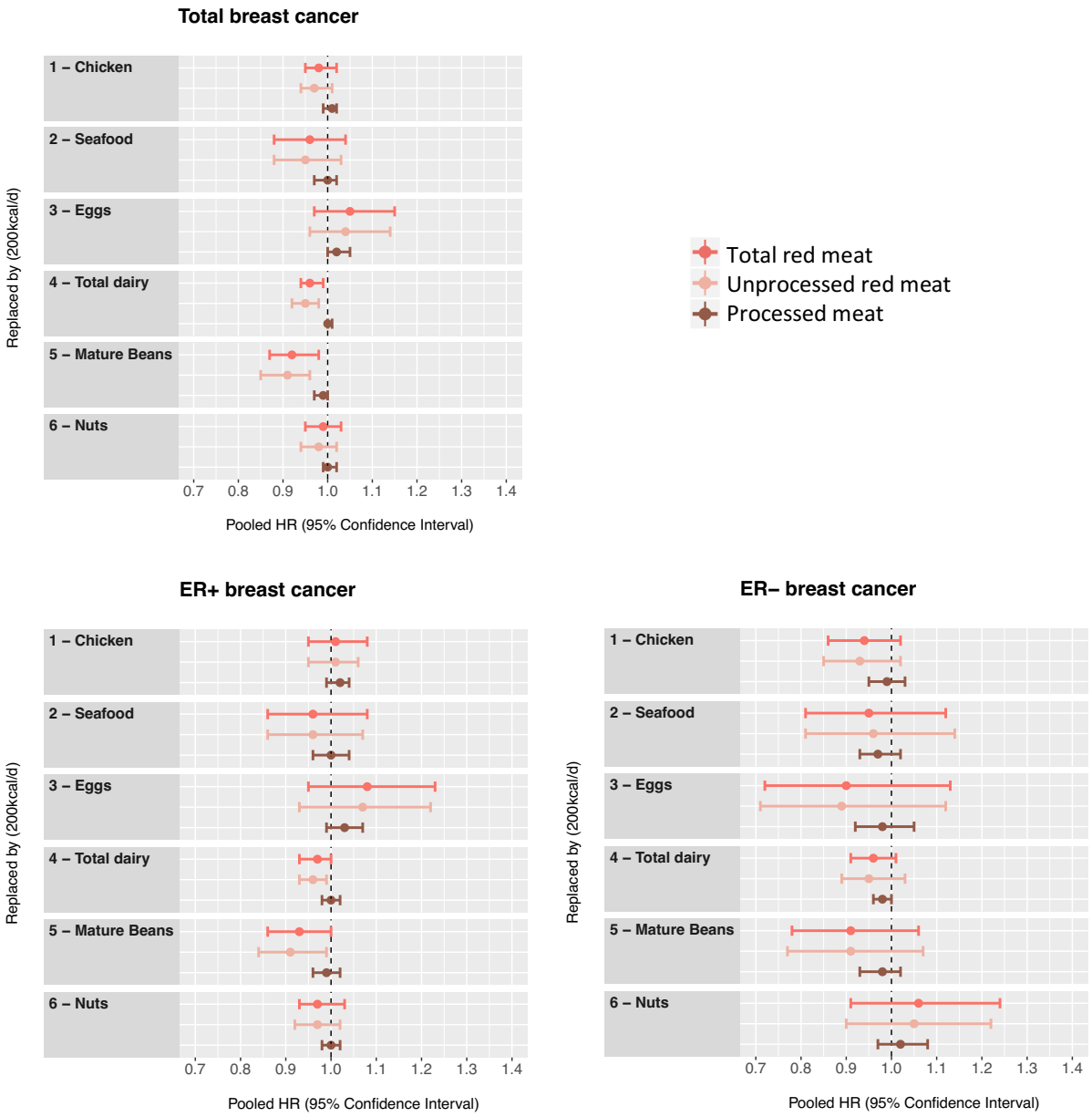

Figure 3.1 Substitution effect on risk of breast cancer overall and the subtypes defined by ER status when replacing red meat by alternative protein sources

dairy (pooled MVHR = 0.97, 95% CI: 0.94 – 0.99) products. There were no statistically significant associations for substituting total or unprocessed red meat with poultry, seafood, or eggs. When processed meat intake was replaced by eggs, the risk of breast cancer was slightly elevated (MVHR for 50 kcal/day substitution = 1.02, 95% CI: 1.00 - 1.21). No association was observed when replacing processed meat by poultry, seafood, or nuts consumption. The results were similar for ER-positive and ER-negative breast cancer (p for common effects >0.07), although the confidence intervals were much wider for ER-negative tumors due to lower statistical power.

The associations generally did not differ by body mass index, menopausal status at diagnosis, age at diagnosis, follow-up duration, or region (Table 3.4). The only test that reached statistical significance was observed for the effect of poultry replacing total red meat, stratified by age at diagnosis (p for interaction = 0.04), although the substitution effects were not significant within any of the strata.

DISCUSSION

In this large pooled analysis of over 46,000 breast cancer cases from 23 cohort studies, total red meat, unprocessed red meat, and processed meat were not associated with risk of breast cancer overall or of subtypes defined by ER status. However, isocaloric substitution of consumption of red meat with alternative protein sources such as mature beans or dairy products was associated with lower risk of breast cancer. Substituting consumption of poultry, seafood, eggs, or nuts for red meat consumption was not associated with breast cancer risk. The results were generally similar for breast cancer subtypes defined by ER status, and also across strata of body mass

Table 3.4 Substitution effect of alternative protein sources replacing total red meat[1] on risk of breast cancer by age at diagnosis, menopausal status, BMI, follow-up duration, and region

Alternative protein source	Poultry		Seafood		Eggs		Total dairy		Mature beans		Nuts	
	MVHR[2] (95% CI)	P_{het}[3]	MVHR[2] (95% CI)	P_{het}[3]	MVHR[2] (95% CI)	P_{het}[3]	MVHR[2] (95% CI)	P_{het}[3]	MVHR[2] (95% CI)	P_{het}[3]	MVHR[2] (95% CI)	P_{het}[3]
Menopausal status												
Pre-	1.04 (0.87-1.24)	0.08	1.10 (0.93-1.31)	0.69	0.91 (0.64-1.30)	0.20	0.95 (0.89-1.01)	0.61	0.95 (0.64-1.41)	0.08	0.97 (0.84-1.12)	0.58
Post-	1.01 (0.96-1.06)	0.73	1.01 (0.90-1.14)	0.05	1.05 (0.93-1.18)	0.86	0.98 (0.95-1.01)	0.44	0.95 (0.88-1.02)	0.90	1.00 (0.94-1.05)	0.86
p-interaction		0.58		0.59		0.48		0.44		0.82		0.82
Age at diagnosis (year-old)												
< 55	0.94 (0.87-1.01)	0.93	1.03 (0.91-1.18)	0.52	1.02 (0.79-1.31)	0.15	0.93 (0.89-0.98)	0.74	0.99 (0.76-1.29)	0.03	1.01 (0.91-1.11)	0.78
55- <65	1.05 (1.00-1.11)	0.58	0.98 (0.88-1.09)	0.62	1.09 (0.94-1.26)	0.86	0.99 (0.95-1.02)	0.65	0.93 (0.84-1.03)	0.91	1.04 (0.97-1.11)	0.63
≥ 65	0.97 (0.92-1.02)	0.83	0.97 (0.87-1.07)	0.41	1.06 (0.94-1.20)	0.91	0.96 (0.92-1.01)	0.04	0.92 (0.85-0.99)	0.68	0.96 (0.90-1.02)	0.62
p-interaction		0.04		0.72		0.84		0.33		0.70		0.29
BMI (kg/m^2)												
<25	0.96 (0.92-1.01)	0.44	0.99 (0.88-1.10)	0.10	1.03 (0.92-1.16)	0.94	0.94 (0.91-0.97)	0.22	0.89 (0.82-0.96)	0.51	0.97 (0.92-1.02)	1.00
25- <30	1.02 (0.95-1.08)	0.51	1.00 (0.89-1.12)	0.64	1.10 (0.93-1.29)	0.48	0.98 (0.94-1.03)	0.33	0.97 (0.87-1.09)	0.57	1.05 (0.97-1.13)	0.59
≥30	1.03 (0.95-1.11)	0.88	0.83 (0.68-1.02)	0.14	1.15 (0.95-1.39)	0.16	1.01 (0.95-1.07)	0.28	0.98 (0.83-1.17)	0.33	0.97 (0.82-1.15)	0.04
p-interaction		0.22		0.18		0.62		0.07		0.34		0.26

Table 3.4 (Continued)

Alternative protein source	Poultry		Seafood		Eggs		Total dairy		Mature beans		Nuts	
Follow up (years)												
<5	0.97 (0.92-1.03)	0.71	0.99 (0.88-1.12)	0.18	1.10 (0.96-1.26)	0.38	0.96 (0.92-0.99)	0.31	0.95 (0.86-1.06)	0.30	1.02 (0.95-1.08)	0.98
5 - <10	1.01 (0.94-1.09)	0.23	0.97 (0.85-1.11)	0.17	1.03 (0.89-1.19)	0.76	0.97 (0.94-1.01)	0.70	0.90 (0.82-1.00)	0.80	0.98 (0.91-1.05)	0.76
≥ 10	0.97 (0.88-1.06)	0.21	0.95 (0.85-1.07)	0.46	1.05 (0.88-1.24)	0.95	0.97 (0.91-1.03)	0.06	0.88 (0.76-1.01)	1.00	0.96 (0.88-1.04)	0.65
p-interaction		0.28		0.73		0.75		0.79		0.48		0.51
Region												
North America	0.99 (0.95-1.02)	0.45	0.99 (0.91-1.07)	0.13	1.07 (0.98-1.17)	0.83	0.96 (0.94-0.99)	0.19	0.93 (0.87-0.98)	0.91	0.98 (0.94-1.03)	0.83
Others	1.00 (0.67-1.48)	0.13	0.81 (0.61-1.06)	0.87	0.82 (0.59-1.15)	0.99	0.95 (0.86-1.05)	0.26	0.91 (0.61-1.36)	0.12	1.11 (0.86-1.43)	0.83
p-interaction		0.92		0.17		0.13		0.76		0.65		0.36

1. Results presented for substitution of very 200kcal/day
2. Pooled multivariable hazard ratio
3. p for between-study heterogeneity for substitution analysis
4. p for between-study heterogeneity not available because only JPHC1 was included

index, menopausal status at diagnosis, age at diagnosis, follow-up duration, or region. Results for unprocessed meat were consistent with that of total red meat consumption, while the associations for processed meat were in similar direction but appeared weaker due to its lower intake level.

Accumulating evidence from published epidemiologic studies and meta-analyses has shown that the association between red meat intake and breast cancer is weak or close to null, as we also observed. For example, a meta-analysis of 13 cohort studies (6 were included in our analyses), 3 nested case–control studies (1 was included in our analyses) and two clinical trials suggested a 6% higher breast cancer risk (pooled RR, 1.06; 95% CI, 0.99–1.14; I^2 = 56.3%) comparing the highest versus lowest total red meat consumption; processed meat consumption was associated with a 9% higher breast cancer risk (pooled RR, 1.09; 95% CI, 1.03–1.16; I^2 = 44.4%) comparing the highest versus lowest consumption.

There are two major limitations of most prior published studies – the effects of different protein sources were studied in isolation (not mutually adjusted); no comparison or substitution effect between two food groups were evaluated. People eat a food item (or a group of food) not in isolation but in combination with other foods. Since daily energy intake in humans is relatively stable [24], we evaluated the impact of replacing red meat intake, an important source of protein in Western diets, with alternative protein sources. These types of analyses are of great public health significance because replacing food groups with those of better quality is intuitive and practical to adopt as dietary recommendations.

Although this study does not suggest that red meat consumption *per se* increases the risk of breast

cancer, it nevertheless highlights the impact of recommendations to lower red meat consumption on breast cancer risk while taking into consideration the alternative protein sources from other animal sources as well as plant sources. Beans are a major source of plant protein, as well as a source of dietary fiber, which has been shown to reduce risk of breast cancer in this consortium [manuscript in preparation]. Soy products are hypothesized to decrease breast cancer risk due to their phytoestrogen content. However, only a few studies in this pooled analysis had data on soy product intake, and intake was low and extremely skewed. For these reasons we did not include soy beans in the mature bean group, nor as a separate group. In sensitivity analysis, however, the effect of plant protein in replacement of red meat was in the same directly although smaller in magnitude, since they were averaged out by the null association between nuts and breast cancer.

Dairy is an important source of protein. A previous analysis in the DCPP suggested weak associations between milk and yogurt, and risk of breast cancer (pooled HR comparing ≥500g/day (~16 oz/day) with <1 g/day of total milk = 0.94, 95% CI 0.90 to 0.99, pooled HR comparing ≥60g/day (~2 oz/day) with <1g/day of yogurt = 0.96, 95% CI 0.92 to 0.99) [manuscript under review]. In this study, we observed 4% lower risk of breast cancer overall when substituting 200 kcal/day of total red meat intake with dairy products.

This study has several strengths. First, we harmonized the participant-level exposure, covariate, and outcome data in each of the studies to improve the comparability of the data being analyzed. In addition, we standardized the analytical methods across studies. As a result, we reduced potential sources of between-study heterogeneity. In addition, because we analyzed the participant-level data from each cohort, we were able to examine the main effect of each protein food group

as well as the impact of substituting consumption of other protein sources for red meat. The large sample size allowed investigation of tumor subtypes and population subgroups. Lastly, since most of the studies included in this analysis had not published on this topic, publication bias was not of a concern.

The results should be interpreted with the following limitations taken into consideration. We only analyzed exposure data collected at baseline from each study, therefore we were not able to study the effect of diet during earlier life periods. However, a previous study showed that replacing one serving/day of adolescent red meat intake with poultry, fish, legume, and nut intake combined was associated with a 15% lower risk of breast cancer[25]. Any changes in dietary habits during follow-up were also not taken into account – an important consideration because there has been a secular trend of decreasing meat consumption in some high-income countries including countries in this analysis [26]. The consumption of food groups of interest had wide ranges within and across studies, meaning that some substitution models might have extrapolated into data ranges of less certainty in some studies. Another limitation is that we applied a common conversion factor to groups of foods rather than using each food's energy intake. Lastly, the study population was comprised of predominantly White women, which limits the generalizability of the results.

CONCLUSION

Total red meat, processed meat, and unprocessed meat intakes were not significantly associated with risk of breast cancer when holding other protein sources constant. However, we observed slightly lower risks of breast cancer when substituting consumption of total red meat or unprocessed red meat with mature beans or dairy products. These associations were similar for

ER+ and ER- breast cancer. Replacing red/processed meat with healthy alternatives is recommended considering results from this study on breast cancer as well as the overall benefit for cancer prevention.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492.

2. World Cancer Research Fund/American Institute for Cancer Research. *Continuous Update Project Expert Report 2018. Recommendations and Public Health and Policy Implications.*

3. Bouvard V, Loomis D, Guyton KZ, et al. Carcinogenicity of consumption of red and processed meat. *Lancet Oncol*. 2015;16(16):1599-1600. doi:10.1016/S1470-2045(15)00444-1.

4. Jeong S-H, Kang D, Lim M-W, Kang CS, Sung HJ. Risk assessment of growth hormones and antimicrobial residues in meat. *Toxicol Res*. 2010;26(4):301-313. doi:10.5487/TR.2010.26.4.301.

5. Mignone LI, Giovannucci E, Newcomb PA, et al. Meat consumption, heterocyclic amines, NAT2, and the risk of breast cancer. *Nutr Cancer*. 2009;61(1):36-46. doi:10.1080/01635580802348658.

6. Wu K, Sinha R, Holmes MD, Giovannucci E, Willett W, Cho E. Meat mutagens and breast cancer in postmenopausal women--a cohort analysis. *Cancer Epidemiol Biomarkers Prev*. 2010;19(5):1301-1310. doi:10.1158/1055-9965.EPI-10-0002.

7. Zheng W, Lee S-A. Well-done meat intake, heterocyclic amine exposure, and cancer risk. *Nutr Cancer*. 2009;61(4):437-446. doi:10.1080/01635580802710741.

8. Hong C-C, Ambrosone CB, Ahn J, et al. Genetic variability in iron-related oxidative stress pathways (Nrf2, NQ01, NOS3, and HO-1), iron intake, and risk of postmenopausal breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2007;16(9):1784-1794. doi:10.1158/1055-9965.EPI-07-0247.

9. Farvid MS, Stern MC, Norat T, et al. Consumption of red and processed meat and breast cancer incidence: A systematic review and meta-analysis of prospective studies. *Int J Cancer*. 2018;143(11):2787-2799. doi:10.1002/ijc.31848.

10. Anderson JJ, Darwis NDM, Mackay DF, et al. Red and processed meat consumption and breast cancer: UK Biobank cohort study and meta-analysis. *Eur J Cancer*. 2018;90:73-82. doi:10.1016/j.ejca.2017.11.022.

11. De Cicco P, Catani MV, Gasperi V, Sibilano M, Quaglietta M, Savini I. Nutrition and Breast Cancer: A Literature Review on Prevention, Treatment and Recurrence. *Nutrients*. 2019;11(7):1514. doi:10.3390/nu11071514.

12. Kipnis V, Freedman LS, Brown CC, Hartman A, Schatzkin A, Wacholder S. Interpretation of energy adjustment models for nutritional epidemiology. *American Journal of Epidemiology*. 1993;137(12):1376-1380. doi:10.1093/oxfordjournals.aje.a116647.

13. Willett WC. Issues in analysis and presentation of dietary data. In: *Nutritional Epidemiology*. Oxford University Press; 2013:321-346.

14. Daniel CR, Cross AJ, Koebnick C, Sinha R. Trends in meat consumption in the USA. *Public Health Nutrition*. 2011;14(4):575-583. doi:10.1017/S1368980010002077.

15. Tsubono Y, Kobayashi M, Sasaki S, Tsugane S. Validity and reproducibility of a self-administered food frequency questionnaire used in the baseline survey of the JPHC Study Cohort I. *Journal of Epidemiology*. 2003;13(1 Suppl):S125-S133. doi:10.2188/jea.13.1sup_125.

16. Salvini S, Hunter DJ, Sampson L, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol*. 1989;18(4):858-867. doi:10.1093/ije/18.4.858.

17. Kaaks R. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol*. 1997;26(90001):26S–36. doi:10.1093/ije/26.suppl_1.S26.

18. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2.

19. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560. doi:10.1136/bmj.327.7414.557.

20. Wang M, Spiegelman D, Kuchiba A, et al. Statistical methods for studying disease subtype heterogeneity. *Statistics in Medicine*. 2016;35(5):782-800. doi:10.1002/sim.6793.

21. Smith-Warner SA, Spiegelman D, Yaun S-S, et al. Alcohol and Breast Cancer in Women. *JAMA*. 1998;279(7):535. doi:10.1001/jama.279.7.535.

22. Smith-Warner SA, Spiegelman D, Ritz J, et al. Methods for Pooling Results of Epidemiologic Studies. *American Journal of Epidemiology*. 2006;163(11):1053-1064. doi:10.1093/aje/kwj127.

23. Stram DO. Meta-analysis of published data using a linear mixed-effects model. *Biometrics*. 1996;52(2):536-544.

24. Willett W. *Nutritional Epidemiology*. Oxford University Press; 2013.

25. Farvid MS, Cho E, Chen WY, Eliassen AH, Willett WC. Adolescent meat intake and breast cancer risk. *Int J Cancer*. 2014;136(8):1909-1920. doi:10.1002/ijc.29218.

26. Mathijs E. Exploring future patterns of meat consumption. *Meat Sci*. 2015;109:112-116. doi:10.1016/j.meatsci.2015.05.007.

Chapter 4

Examining the influence of timing of measurement on assessing preventability of breast cancer by lifestyle risk factors

ABSTRACT

Background: A wide range of population attributable risk fraction (PAR%) of cancer incidence by modifiable risk factors has been reported, yet there is no consensus on what contributes to the variation.

Method: Using repeated measurements from the Nurses' Health Study, we examined the PAR% of postmenopausal breast cancer by modifiable risk factors. Cox proportional hazard model was used to estimate the relative risks. Prevalence and the variance-covariance matrix were obtained for all combinations of the risk factors. We calculated the partial PAR% by alcohol consumption, fruit and vegetable intake, physical activity and weight change since age 18, while holding unmodifiable risk factors unchanged. We also estimated the PAR% had everyone been at the optimal level for all four risk factors (low risk method). For each method, we modeled the exposures at baseline only, as simple updates, and as cumulative averages.

Results: We identified 6510 invasive breast cancer among 85,035 women during 28 years of follow-up. Multivariable-adjusted hazard ratios were higher in models based on repeated measures. The estimated PAR% was 16.2% (95% CI 9.6% - 22.6%) if only baseline information was used, 25.3% (95% CI 13.8% - 36.2%) if used simple-updated exposures, and 23.9% (95% CI 10.8% - 36.1%) if cumulative averaged exposures were in the model, holding the other covariates unchanged. PAR% estimated by the low-risk method were higher but in similar pattern. The

proportion of preventable breast cancer cases by switching everyone to overall low-risk status was 22.5% (95% CI 4.5% - 39.1%) if used baseline exposures, 35.5% (95% CI 20.2% - 49.1%) if used simple updates, and 38.3% (95% CI 16.9% - 56.2%) if used cumulative averages, while holding the other covariates unchanged.

Conclusion: Given the same statistical method for PAR% calculation, models based on repeated measures yielded greater estimated PAR%. The PAR% of breast cancer incidence by modifiable risk factors in current literatures that relied on studies with baseline data has likely underestimated the preventable fraction.

INTRODUCTION

Cancer is a leading cause of disease burden and mortality across the countries [1,2]. Evidence is clear that cancer occurrence is not simply a random event, and a substantial proportion could be prevented by primary intervention[3]. Population attributable risk in percentage (PAR%), a key epidemiologic indicator, represents the percentage of disease cases in a target population that would not have occurred in the absence of a risk factor, or if a risk factor was set to the optimal level. Intuitively, the preventable fraction depends on both the relative risk and the prevalence of any given risk factor; a risk factor of large effect size but low prevalence would result in similar PAR% of a common risk factor of a smaller effect in a population. Therefore, quantifying the preventable fraction of cancer helps us understand the public health impact of exposures, and is essential for personal decisions, priority setting in primary prevention, and healthcare policy.

There are various methods to calculate PAR%. The conventional method applies a standard formula incorporating exposure prevalence, normally from a national or regional representative population, and relative risk data, usually from published literatures. Another method involves comparing the incidence rate of a pre-defined low-risk group to a high-risk group, thus the hypothetical incidence had everyone been in the low-risk group can be derived. Within each method, the data sources of relative risk and prevalence can also differ. To estimate the prevalence, some studies used cross-sectional survey [4,5], while others used empirical data observed in cohort studies [6]. To estimate the overall relative risk, some studies based on case-control design [7], some used cohort data but had baseline exposures only[5], while others used repeated and time-varying measurements[6]. In this study, we used postmenopausal breast cancer as an example to evaluate the

degree to which PAR% can vary by different choices of exposure measurement in the same population.

METHODS

Study population. The Nurses' Health Study was established in 1976. A total of 121,701 female nurses aged 30-55 years returned the initial questionnaire [8] and have been followed up biennially to collect their medical, lifestyle, and other health-related information. To obtain dietary information, a semiquantitative food frequency questionnaire (FFQ) was first sent to the participants in 1980[9], and later extended to a more comprehensive version in 1986[10]. Since our main goal of this study was to compare the PAR% based on lifestyle measurements with different timings, we considered 1986 (or 1990 for those who did not return their 1986 food frequency questionnaire) as the baseline to make any cut points of high/low risk groups comparable over time. The participants were excluded if they had a history of cancer, except for nonmelanoma skin cancer, at baseline. We further excluded women with extreme total energy intake (below 600 or above 3,500 kcal/day). The overall response rate has been greater than 90% through 2010.

Assessment of exposures. In 2018, World Cancer Research Fund and American Institute for Cancer Research (WCRF/AICR) updated the Cancer Prevention Recommendations based on the latest Expert Report, a part of the Continuous Update Project[11]. Combining the overall recommendations for cancer prevention and the summarized evidence for diet, nutrition, physical activity and breast cancer[12], we included the following lifestyle risk factors in the PAR% calculation: alcohol consumption, fruit and vegetable intake, physical activity, and weight change since age 18.

Every 4 years, the participants returned FFQ covering their usual diet in the past year. Alcohol consumption and fruit and vegetables intake were estimated based on the quantity and frequency of all relevant food items each person consumed [13,14]. Every 2-4 years, the participants reported their average time per week spent engaging in various types of physical activity, which were converted into metabolic equivalent task hours per week (METs-hr/week) [15]. Weight at age 18 was reported in 1980, and current weight was reported biennially afterwards. Breastfeeding was not included in PAR% calculation because total lifetime breastfeeding duration was only assessed once in 1986 (no repeated measures).

Based on the WCRF/AICR recommendations, we defined high-risk and low-risk level with the following cut points: alcohol consumption (drinkers vs non-drinker), fruit and vegetable intake (<5 vs ≥5 servings/day), physical activity (<18 vs ≥18 METs-hr/week), weight change since age 18 (≥5 vs <5kg).

Assessment of covariates. Age at each questionnaire return was calculated by the return date and date of birth. Height and age at menarche were collected in 1976. Menopausal status and confirmed benign breast disease were updated biennially. Family history of breast cancer obtained in 1982 and updated every 4 years beginning in 1988. Oral contraceptive use was assessed in 1980, 1982, and 1984. Age at first birth was asked in 1976 and updated biennially until 1982. Parity was asked biennially until 1996. We cross-classified age at first birth and parity into 9 categories. Postmenopausal hormone therapy use was asked biennially until 2004. Body mass index (BMI, kg/m^2) at age 18 was calculated from weight at age 18 and height.

Statistical analysis. We modelled the exposures of interest in three ways: (1) using baseline only, where we classified the participants according to their measurements in 1986; (2) using simple updated exposures, where the measurement from any questionnaire return only relates to the person-time between then and the next follow-up cycle; and (3) using cumulative averaged exposures, where for each 2-year cycle, we allocated the person-time to the cumulative averaged exposures up to the assessment just before diagnosis, loss to follow- up, or the last assessment before the end of follow-up. In sensitivity analyses, we also evaluated the effect of incorporating latency period into the models. Specifically, we implemented 4-8 years, 8-12 years, 12-16 years, 16-20 years, and 20-24 years lags, under the assumption that person-time during follow-up being allocated to exposure categories from 4-8 years, 8-12 years, 12-16 years, 16-20 years, or 20-24 years before, respectively.

The exposures of interest were modeled as categorical variables first, and then as binary variables. We did not impute missing data for the four main exposures; missing values were grouped into a separate missing category. Age and multivariable-adjusted relative risks (and 95% confidence intervals [CI]) were estimated using Cox proportional hazards models. The covariates were included in the models as time-varying covariates whenever possible. Multivariable models include the exposures of interest simultaneously, adjusting for height, age at menarche, BMI at age 18, duration of oral contraceptive use, age at first birth and parity, benign breast cancer, family history of breast cancer, and postmenopausal hormone therapy use. Missing data for some covariates, such as benign breast disease and family history of breast cancer, were filled in by

carrying-forward. Other missing values were handled as missing indicators, where missing observations were grouped into a separate category.

The PAR% and the 95% CIs were estimated using the %PAR SAS macro developed by Spiegelman et al[16]. Briefly, the macro uses the relative risks comparing the high-risk versus the low-risk group for each exposure of interest, their variance-covariance matrix, and the observed prevalence for each variable. The PAR% reflects the proportional reduction expected in the number of incidence invasive breast cancer if all of the risk factors of interest were set to the optimal (low-risk) level in the targeted population. In this study, we calculated the composite partial PAR%, which estimate the PAR% associated with the four modifiable risk factors combined, while holding the other covariates unchanged.

We presented the PAR% calculated by two methods. The first method considered the four risk factors separately and then combined to obtain a composite PAR%. The second method classifies people into two overall high-risk and low-risk groups, where being low-risk is defined as being at the optimal level for all four risk factors, while all other were considered having high-risk. Participants with missingness for no more than one factor, while being 'low-risk' for all other factors are considered overall 'low-risk', while participants with missingness for two or more of the risk factors were excluded from the analyses. Sensitivity analysis was conducted where participants with missingness for two risk factors were considered 'high-risk'.

For all hypothesis tests, a p-value <0.05 was considered statistically significant, and all tests of statistical significance were 2-sided. All analyses were conducted using SAS, version 9.2 (SAS Institute, Inc., Cary, NC).

RESULTS

After applying the exclusion criteria, we identified 6510 invasive breast cancer during 28 years of follow-up. The majority of women were parous, and their parity mostly remained the same over the course of follow-up. Half of the women had never used oral contraceptive. 28.6% had confirmed benign breast disease at baseline, while by the end of follow-up there were 38.4%. Family history of breast cancer was known for 7.7% women at baseline, then during follow-up, about another 10% women reported breast cancer diagnosis for their mother or sisters (Table 4.1).

To compare the exposure at baseline and their respective cumulative average, we present the percentages for cumulative averaged value by summing up all person-time associated with each category and divided by total person-time in the cohort. Fewer people remained as non-drinker when cumulative average was calculated. Changes in total physical activity were seen over time. As expected, more women were classified into higher categories of weight gain as time went by. Fruit and vegetable intake, on the other hand, remained about the same through the study follow-up (Table 4.1).

Results of the Cox regression models confirmed the relationships between the exposures of interest and risk of breast cancer (Table 4.2). Higher alcohol consumption, lower fruit and vegetable intake, lower physical activity, and greater weight gain since age 18 were all associated with

Table 4.1 Risk factors of breast cancer for postmenopausal women in the Nurses' Health Study (n=85,035)

Risk factors	Frequency (%)		Covariates	Frequency (%) or mean(sd)	
Modifiable risk factors in WCRF recommendations			Non-modifiable risk factors and covariates		
	Baseline (1986)	Cumulative average		Baseline (1986)	End-of-follow-up
Alcohol consumption			Age, years	52.4 (7.2)	
non drinker	36.5%	20.3%	Height, m	1.64 (0.06)	
>0.0-4.9 g/d	32.3%	45.9%	BMI at age 18	21.3 (2.8)	
5.0-9.9 g/d	10.6%	14.0%	Age at menarche, years	12.4 (1.8)	
10.0-14.9 g/d	9.2%	8.0%	Parity and age at first birth		
15.0-29.9 g/d	6.5%	8.8%	Nulliparous	5.6%	5.5%
30+ g/d	4.9%	3.0%	1-2 children AFB <25'	13.8%	13.3%
Fruit and vegetable intake			1-2 children AFB 25-29	15.0%	14.5%
0-2 servings/d	8.0%	5.5%	1-2 children AFB 30+'	6.1%	6.0%
2-4 servings/d	32.9%	33.9%	3-4 children AFB <25'	25.4%	24.8%
4-6 servings/d	31.4%	35.9%	3-4 children AFB 25-29	16.2%	15.8%
6-8 servings/d	16.8%	16.7%	3-4 children AFB 30+'	2.6%	2.5%
8+ servings/d	11.0%	8.0%	5+ children	9.1%	9.0%
Physical activity			Missing	6.3%	8.6%
<3 total METs-hr/week	28.0%	12.3%	Oral contraceptive use		
3-9 total METs-hr/week	26.9%	26.1%	No OC use	53.0%	52.4%
9-18 total METs-hr/week	19.3%	27.9%	>0-2 yrs of OC use	17.8%	17.2%
18-27 total METs-hr/week	10.7%	15.5%	>2-5yrs of OC use	12.9%	12.4%
27+ total METs-hr/week	15.0%	18.2%	>5-10 yrs of OC use	11.3%	10.9%
Weight change since age 18			>10 yrs of OC use	5.0%	4.9%
<5kg	33.5%	27.2%	Benign breast disease history	29.7%	38.5%
5 - <10kg	21.3%	17.3%	Family history of breast cancer	8.1%	17.4%
10 - <15kg	16.4%	16.3%	Menopausal hormone therapy use		
15 - <20kg	11.0%	13.0%	pre/missing menopause	33.6%	4.5%
>=20kg	17.9%	26.3%	Never user	35.6%	38.6%
Total breastfeeding duration			current user	16.9%	14.2%
None	38.6%	38.4%	past user	13.9%	42.8%
yes, < 6mo	35.3%	35.6%	Smoking status		
yes, 6mo - 2yrs	20.1%	20.2%	Never smoker	44.5%	42.6%
yes, >2yrs	5.9%	5.9%	Past smoker	33.8%	46.7%
			Current smoker	21.7%	7.8%

Table 4.2 Hazard ratio and 95% CI for breast cancer overall by WCRF cancer prevention recommendations using baseline, simple update, or cumulative average

	Baseline 1		Baseline 2		Simple update		Cumulative average	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Alcohol consumption								
non drinker	Ref		Ref		Ref		Ref	
>0.0-4.9 g/d	1.03	(0.97-1.10)	1.02	(0.95-1.09)	1.06	(1.00-1.13)	1.05	(0.99-1.12)
5.0-9.9 g/d	1.06	(0.97-1.16)	1.05	(0.95-1.15)	1.12	(1.03-1.22)	1.11	(1.03-1.21)
10.0-14.9 g/d	1.16	(1.06-1.27)	1.16	(1.06-1.28)	1.12	(1.02-1.23)	1.20	(1.10-1.32)
15.0-29.9 g/d	1.16	(1.05-1.29)	1.14	(1.02-1.27)	1.17	(1.06-1.29)	1.28	(1.17-1.40)
30+ g/d	1.40	(1.26-1.57)	1.39	(1.23-1.56)	1.38	(1.23-1.54)	1.50	(1.32-1.70)
Fruit and vegetable intake								
0-2 servings/d	Ref		Ref		Ref		Ref	
2-4 servings/d	1.01	(0.91-1.12)	1.03	(0.92-1.14)	0.95	(0.87-1.03)	0.91	(0.82-1.01)
4-6 servings/d	1.00	(0.90-1.11)	1.02	(0.91-1.14)	0.92	(0.85-1.01)	0.91	(0.82-1.01)
6-8 servings/d	1.00	(0.89-1.12)	1.00	(0.89-1.13)	0.91	(0.82-1.00)	0.89	(0.80-1.00)
8+ servings/d	0.98	(0.87-1.12)	0.99	(0.87-1.13)	0.90	(0.81-1.01)	0.86	(0.75-0.98)
Physical activity								
<3 total METs-hr/week	Ref		Ref		Ref		Ref	
3-9 total METs-hr/week	1.02	(0.95-1.09)	1.01	(0.95-1.08)	1.05	(0.98-1.12)	1.02	(0.94-1.10)
9-18 total METs-hr/week	1.01	(0.94-1.09)	1.00	(0.93-1.08)	1.00	(0.93-1.07)	0.98	(0.91-1.06)
18-27 total METs-hr/week	1.03	(0.95-1.13)	1.03	(0.94-1.13)	1.01	(0.93-1.10)	0.99	(0.90-1.08)
27+ total METs-hr/week	1.00	(0.92-1.09)	1.00	(0.92-1.08)	0.93	(0.87-1.00)	0.93	(0.85-1.01)
Weight change since age 18								
<5kg	Ref		Ref		Ref		Ref	
5 - <10kg	1.12	(1.05-1.21)	1.12	(1.05-1.21)	1.16	(1.08-1.24)	1.16	(1.08-1.25)
10 - <15kg	1.18	(1.09-1.27)	1.19	(1.10-1.28)	1.12	(1.04-1.21)	1.13	(1.05-1.22)
15 - <20kg	1.30	(1.20-1.42)	1.31	(1.21-1.43)	1.32	(1.22-1.43)	1.33	(1.23-1.43)
>=20kg	1.36	(1.26-1.47)	1.39	(1.29-1.49)	1.44	(1.35-1.54)	1.45	(1.36-1.55)

Baseline1: model includes baseline covariates.

Baseline2: model includes updated covariates (benign breast disease, family history of breast cancer, postmenopausal hormone use).

increased risk of breast cancer. Results of the two baseline models suggested that whether the other time-varying covariates were updated did not affect the magnitude of the association substantially. Thus, we only presented the results of one baseline model for the remaining analyses. Compared to the models of repeated measures, the strengths of the associations tended to be weaker when we used baseline information only. For dietary factors (alcohol consumption and fruit and vegetable intake), the multivariable-adjusted hazard ratio comparing the highest versus the lowest category was slightly stronger in magnitude in models including cumulative averages than in models including simple updates. There was no such apparent difference for physical activity or weight change since age 18. We did not observe clear pattern in the hazard ratios in latency analysis (Table S4.1).

Similarly, in models that included the exposures as binary variables, the hazard ratios were slightly higher in those based on repeated measures (Table 4.3). The percentages of total person-time allocated to high-risk categories were generally higher for simple updated and cumulative averaged exposures, except for physical activity. Taken together, had everyone switched their alcohol consumption, fruit and vegetable intake, physical activity, and weight change since age 18 to the low-risk levels, the estimated PAR% is 16.2% (95% CI 9.6% - 22.6%) if only baseline information was used, 25.3% (95% CI 13.8% - 36.2%) if used simple-updated exposures, and 23.9% (95% CI 10.8% - 36.1%) if cumulative averaged exposures were in the model, holding the other covariates unchanged.

The PAR% estimated by the low-risk method were higher but in similar pattern. The results suggested that had everyone who was not in the overall low-risk group switched to being low-

Table 4.3 Prevalence, multivariable-adjusted[1] hazard ratio, and partial PAR associated with high-risk groups of the risk factors

		Baseline (1986)	Simple update	Cumulative average
	No. of cases	6510	6510	6510
		Breast cancer overall		
Alcohol (> 0g/day)				
	High-risk%	69.0%	57.7%	79.7%
	HR (95% CI)	1.07 (1.01-1.12)	1.10 (1.04-1.16)	1.09 (1.02-1.16)
Fruit & vegetable (< 5 servings/day)				
	High-risk%	58.2%	62.3%	60.0%
	HR (95% CI)	1.04 (0.99-1.10)	1.07 (1.01-1.13)	1.04 (0.99-1.10)
Physical activity (< 18 total METs-hr/week)				
	High-risk%	73.9%	66.5%	66.3%
	HR (95% CI)	1.01 (0.95-1.07)	1.05 (1.00-1.11)	1.03 (0.98-1.09)
Weight gain since age 18 (≥ 5kg)				
	High-risk%	66.6%	72.8%	72.8%
	HR (95% CI)	1.20 (1.13-1.27)	1.23 (1.16-1.31)	1.24 (1.17-1.31)
Total	Composite PAR%	16.2%	25.3%	23.9%
		(9.6% - 22.6%)	(13.8% - 36.2%)	(10.8% - 36.1%)

1. Multivariable model include: height (<1.6m, 1.60-1.64m, 1.65-1.69m, 1.70-1.74m, ≥1.75m), age at menarche (<12 yo, 12, 23, 24, >14 yo), oral contraceptive use (No OC use, >0-2 yrs, >2-5yrs, >5-10 yrs, >10 yrs), joint classification of age at first birth and parity (Nulliparous, 1-2 children AFB <25, 1-2 children AFB 25-29, 1-2 children AFB 30+, 3-4 children AFB <25, 3-4 children AFB 25-29, 3-4 children AFB 30+, 5+ children AFB <25, 5+ children AFB 25+), menopausal status and post-menopausal hormone therapy use (premenopausal/missing menopause, no history of postmenopausal hormone use, current postmenopausal hormone use, past postmenopausal hormone use), confirmed benign breast disease (yes, no), family history of breast cancer (yes, no), body mass index at age 18 (<18.5, 18.5-<20, 20-<22.5, 22.5-<25, 25-<30, 30+), total energy intake.

risk for all four factors, the percentage of breast cancer cases that would have been averted are 22.5% (95% CI 4.5% - 39.1%) if only baseline information was used, 35.5% (95% CI 20.2% - 49.1%) if simple-updated exposures was included in the model, and 38.3% (95% CI 16.9% - 56.2%) if based on cumulative average of the exposures, while holding the other covariates unchanged (Table 4.4).

DISCUSSION

In this study with 85,035 participants and over 6500 invasive breast cancer cases, we estimated that the preventable fraction by alcohol consumption, fruit and vegetable intake, physical activity, and weight change since age 18 varied according to the data sources and the methods. When only baseline data is available, PAR% around 16% was observed; when repeated measures were analyzed, the number went up about 7%. Results from the models including the four risk factors simultaneously appeared lower than that from the models including an indicator summarizing overall risk, likely because the latter grouped more extreme people as the reference level. The differences between the baseline model, the simple-update model, and the cumulative-average model were more apparent when the low-risk method was used, yet the pattern was comparable. Among the four risk factors, weight change since 18 contributed most substantially to the PAR%, followed by alcohol consumption. This is consistent with a previously published study, where the highest PAR% (21%) among all modifiable risk factors was reported for weight gain more than 5kg since age 18[17].

Currently published PAR% of breast cancer by a combination of modifiable lifestyle risk factors ranged from 26.0% to 40.7% [4,6,7,17-21]. The variation might be due to the varying effect sizes of

Table 4.4 Prevalence, multivariable-adjusted hazard ratio, and PAR associated with overall high-risk[1] profile

	Baseline		Simple update		Cumulative average	
No. cases	6510		6462		6499	
Low-risk%	1.5%		1.8%		0.9%	
High-risk%	98.5%		98.2%		99.1%	
HR (95% CI)	1.29	(1.03, 1.63)	1.56	(1.24, 1.96)	1.63	(1.18, 2.25)
PAR	22.5%	(4.5%, 39.1%)	35.5%	(20.2%, 49.1%)	38.3%	(16.9%, 56.2%)

the relative risks. Most of the modifiable risk factors, such as diet and physical activity, change over the life course. Repeated measures of lifestyle factors are time-integrated, thus more representative of one's long-time exposures and more relevant to long-time changes and preventability. The variation could also be a result of applying prevalence data from population other than which the relative risks were derived from. Taken together, it is unclear whether the variation is due to only the differences between populations and the risk profiles, or the inconsistency in PAR% calculation methods.

In this study, we estimated the preventability of breast cancer by four modifiable risk factors with two methods. Each method was applied to three types of exposure measurements: baseline, simple update, and cumulative average. Despite numerical discrepancies between the two methods, the overall patterns comparing the three types of exposure measurements were consistent – repeated measures resulted in greater estimated PAR%.

As previously mentioned, a greater PAR% could be explained by higher relative risk, or higher prevalence of the high-risk group. Since modifiable risk factors may vary over time, models with

repeated measures data generally capture the level of exposures more precisely, thus yield greater relative risks comparing the exposed and the non-exposed groups. If only baseline data was used, we inevitably assumed that the exposure levels remained the same throughout the follow-up, therefore any change in the risk factors later in time would lead to misclassification and bias the risk estimates toward the null. Empirically, in this study, the prevalence of high and low risk did not differ much across the three types of exposure measurements when we transformed the person-time of each observation into multiple cycles of questionnaire return. However, this consistency might not always hold in other populations, under which circumstances we would also recommend using repeated measures to accurately estimate the dynamic distribution of the risk factors.

To our knowledge, this study is the first to compare PAR% estimates derived from the same study population, with the only difference being how we model the exposures. We were able to assess modifiable lifestyle risk factors separate from other established risk factors by calculating the partial PAR%, while still adjusting for confounding in the statistical models. This is a major strength because many of the PAR% from the literature did not keep the unmodifiable confounders unchanged (which led to biased estimates) or computed the full PAR% including lifestyle risk factors and confounders such as age at menarche and family history of breast cancer (which is not realistic to intervene on, thus becomes unattainable). There are also limitations to be recognized. We only analyzed postmenopausal breast cancer data as an example. The risk factors of postmenopausal breast cancer could be very different from other tumors. For instance, weight change since age 18 contributed the most PAR% in our analysis; whereas red/processed meat consumption, another risk factor identified by WCRF, was not considered for breast cancer but could have considerable impact on other outcomes, such as colorectal cancer. Moreover, the study

population consists predominately White female nurses. Evidence has shown that the prevalence of being at high risk in the Nurses' Health Study is lower than the nationally representative samples[3]. Therefore, the estimate PAR% of postmenopausal breast cancer by the same set of risk factors is expected to be higher in the general population. Nevertheless, the purpose of this study is not to quantify the preventability on a national or global level, but rather to demonstrate the degree of variation in the computation, emphasize the importance of high-quality data source in PAR% calculation, and call for cautious interpretation of PAR% in the current literature.

CONCLUSION

Repeated measures are necessary to accurately estimate PAR% of cancer incidence. The PAR% of postmenopausal breast cancer in current literatures using relative risks from studies with baseline data likely underestimated the preventable fraction. Intervening on the modifiable risk factors could potentially prevent more breast cancer incidence than previously expected.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492.

2. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1736-1788. doi:10.1016/S0140-6736(18)32203-7.

3. Song M, Giovannucci E. Preventable Incidence and Mortality of Carcinoma Associated With Lifestyle Factors Among White Adults in the United States. *JAMA Oncol*. 2016;2(9):1154-1161. doi:10.1001/jamaoncol.2016.0843.

4. Wilson LF, Page AN, Dunn NAM, Pandeya N, Protani MM, Taylor RJ. Population attributable risk of modifiable risk factors associated with invasive breast cancer in women aged 45-69 years in Queensland, Australia. *Maturitas*. 2013;76(4):370-376. doi:10.1016/j.maturitas.2013.09.002.

5. Islami F, Goding Sauer A, Miller KD, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin*. 2018;68(1):31-54. doi:10.3322/caac.21440.

6. Tamimi RM, Spiegelman D, Smith-Warner SA, et al. Population Attributable Risk of Modifiable and Nonmodifiable Breast Cancer Risk Factors in Postmenopausal Breast Cancer. *American Journal of Epidemiology*. 2016;184(12):884-893. doi:10.1093/aje/kww145.

7. Barnes BBE, Steindorf K, Hein R, Flesch-Janys D, Chang-Claude J. Population attributable risk of invasive postmenopausal breast cancer and breast cancer subtypes for modifiable and non-modifiable risk factors. *Cancer Epidemiol*. 2011;35(4):345-352. doi:10.1016/j.canep.2010.11.003.

8. Colditz GA, Martin P, Stampfer MJ, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *American Journal of Epidemiology*. 1986;123(5):894-900. doi:10.1093/oxfordjournals.aje.a114319.

9. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire. *American Journal of Epidemiology*. 1985;122(1):51-65. doi:10.1093/oxfordjournals.aje.a114086.

10. Willett WC, Reynolds RD, Cottrell-Hoehner S, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J Am Diet Assoc*. 1987;87(1):43-47.

11. World Cancer Research Fund/American Institute for Cancer Research. *Continuous Update Project Expert Report 2018. Recommendations and Public Health and Policy Implications.*

12. World Cancer Research Fund/American Institute for Cancer Research. *World Cancer Research Fund American Institute for Cancer Research. Continuous Update Project Expert Report 2018.*

13. Hirko KA, Chen WY, Willett WC, et al. Alcohol consumption and risk of breast cancer by molecular subtype: Prospective analysis of the nurses' health study after 26 years of follow-up. *Int J Cancer*. 2016;138(5):1094-1101. doi:10.1002/ijc.29861.

14. Farvid MS, Chen WY, Rosner BA, Tamimi RM, Willett WC, Eliassen AH. Fruit and vegetable consumption and breast cancer incidence: Repeated measures over 30 years of follow-up. *Int J Cancer*. 2019;144(7):1496-1510. doi:10.1002/ijc.31653.

15. Eliassen AH, Hankinson SE, Rosner B, Holmes MD, Willett WC. Physical Activity and Risk of Breast Cancer Among Postmenopausal Women. *Arch Intern Med*. 2010;170(19). doi:10.1001/archinternmed.2010.363.

16. Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control*. 2007;18(5):571-579. doi:10.1007/s10552-006-0090-y.

17. Sprague BL, Trentham-Dietz A, Egan KM, Titus-Ernstoff L, Hampton JM, Newcomb PA. Proportion of invasive breast cancer attributable to risk factors modifiable after menopause. *American Journal of Epidemiology*. 2008;168(4):404-411. doi:10.1093/aje/kwn143.

18. Park B, Park S, Shin H-R, et al. Population attributable risks of modifiable reproductive factors for breast and ovarian cancers in Korea. *BMC CANCER*. 2016;16(1):5. doi:10.1186/s12885-015-2040-0.

19. van Gemert WA, Lanting CI, Goldbohm RA, et al. The proportion of postmenopausal breast cancer cases in the Netherlands attributable to lifestyle-related risk factors. *Breast Cancer Res Treat*. 2015;152(1):155-162. doi:10.1007/s10549-015-3447-7.

20. Clarke CA, Purdie DM, Glaser SL. Population attributable risk of breast cancer in white women associated with immediately modifiable risk factors. *BMC CANCER*. 2006;6(1):170. doi:10.1186/1471-2407-6-170.

21. Hayes J, Richardson A, Frampton C. Population attributable risks for modifiable lifestyle factors and breast cancer in New Zealand women. *Intern Med J*. 2013;43(11):1198-1204. doi:10.1111/imj.12256.

Supplemental Materials

Table S1.1 Case status defined by PR or jointly by ER and PR and intake of dietary fiber from fruits, vegetables, and grains

Study	No. of cases				
	PR+	PR-	ER+/PR+	ER+/PR-	ER-/PR-
The NIH-AARP Diet and Health Study	1917	786	1853	381	404
Breast Cancer Detection Demonstration Project Follow-up Study	667	270	635	135	132
Beta-Carotene and Retinol Efficacy Trial	163	48	160	22	26
Campaign Against Cancer and Heart Disease	168	78	159	37	41
Cancer Prevention Study II Nutrition Cohort	1483	561	1440	282	272
California Teachers Study	1544	625	1509	316	308
Canadian National Breast Screening Study	309	140	272	48	76
The European Prospective Investigation into Cancer and Nutrition	3478	2092	3286	966	1108
Iowa Women's Health Study	1117	388	1082	191	196
The Japan Public Health Center-Based Study Cohort I	87	82	74	30	52
Multiethnic Cohort Study	1767	773	1698	309	464
Melbourne Collaborative Cohort Study	420	240	393	96	144
Nurses' Health Study (part a)	389	304	345	104	186
Nurses' Health Study (part b)	3494	1588	3375	810	778
Nurses' Health Study II	1936	784	1862	287	496
The Netherlands Cohort Study	361	199	348	100	96
New York University Women's Health Study	296	204	272	111	92
The Hormones and Diet in the Etiology of Breast Cancer Study	180	92	159	46	46
The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	758	227	750	101	126
Swedish Mammography Cohort	1308	673	1199	394	276
The Vitamin D and Omega-3 trial	637	213	630	96	117
Women's Health Initiative	3395	1253	3338	618	634
Women's Health Study	819	288	796	124	164
Women's Lifestyle and Health Study	613	309	571	155	153
Total	27306	12217	26206	5759	6387

Table S1.2a Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for total dietary fiber intake

		Pooled HR (95% CI)					p for between-study heterogeneity for top category	I^2	p for common-effect by ER status for top category
Breast cancer	Total fiber intake (g/day)	<10	10-15	15-20	20-25	25+			
Overall	*No. of cases*	*4390*	*16036*	*19079*	*11025*	*6314*			
	Age-adjusted	1.00 (0.96-1.04)	1 ref	0.98 (0.96-1.00)	0.93 (0.90-0.96)	0.89 (0.86-0.93)	0.31	12.1%	
	Multivariable	0.99 (0.95-1.03)	1 ref	0.99 (0.96-1.01)	0.94 (0.91-0.97)	0.92 (0.88-0.96)	0.36	8.1%	
ER+	*No. of cases*	*2573*	*9776*	*11807*	*6608*	*3620*			
	Multivariable	0.99 (0.93-1.04)	1 ref	1.01 (0.97-1.04)	0.96 (0.91-1.01)	0.94 (0.89-0.98)	0.90	0%	
ER-	*No. of cases*	*651*	*2256*	*2643*	*1506*	*772*			0.15
	Multivariable	1.04 (0.93-1.16)	1 ref	0.97 (0.92-1.03)	0.94 (0.87-1.02)	0.86 (0.78-0.96)	0.91	0%	

Multivariable models include race (White, African American, Hispanic, Asian, other), education (< high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m2), height(<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75m), alcohol (0, 0-<5, 5-<15, 15-<30 and ≥30 g/d), energy intake (continuous, kcal/d), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 except PLCO <10,10/11,12/13, ≥14 years), hormone replacement therapy use (never user, past user, current user), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25y), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

Table S1.2b Pooled multivariable HRs (95% CIs) of breast cancer overall and subtypes defined by ER status for sources of dietary fiber

Breast cancer		Pooled HR (95% CI) <3	3-6	6-9	9-12	12+	p for between-study heterogeneity for top category	I^2	p for common-effect by ER status for top category
		Fiber from fruits							
Overall	Age-adjusted	1.01 (0.98-1.04)	1 ref	0.95 (0.92-0.98)	0.92 (0.88-0.97)	0.86 (0.78-0.96)	0.04	41.3%	
	Multivariable	1.00 (0.97-1.04)	1 ref	0.96 (0.93-0.99)	0.95 (0.91-1.00)	0.90 (0.81-1.00)	0.06	37.7%	
ER+	Multivariable	0.99 (0.95-1.03)	1 ref	0.95 (0.91-0.98)	0.92 (0.87-0.99)	0.91 (0.80-1.02)	0.18	24.5%	
ER-	Multivariable	1.02 (0.95-1.09)	1 ref	0.96 (0.89-1.04)	0.93 (0.81-1.07)	0.95 (0.79-1.14)	0.84	0%	0.68
		Fiber from vegetables							
Overall	Age-adjusted	1.03 (1.00-1.06)	1 ref	0.99 (0.97-1.02)	0.98 (0.92-1.04)	0.90 (0.83-0.97)	0.02	46.6%	
	Multivariable	1.03 (1.00-1.06)	1 ref	0.99 (0.97-1.02)	1.00 (0.95-1.04)	0.93 (0.88-0.99)	0.26	17.9%	
ER+	Multivariable	1.02 (0.97-1.07)	1 ref	1.00 (0.97-1.03)	1.01 (0.96-1.07)	0.96 (0.88-1.05)	0.12	30.2%	
ER-	Multivariable	1.13 (1.03-1.24)	1 ref	0.97 (0.90-1.04)	0.96 (0.86-1.05)	0.91 (0.79-1.04)	0.90	0%	0.46

Table S1.2b (Continued)

				Fiber from grains					
Overall	Age-adjusted	1.00 (0.98-1.03)	1 ref	0.98 (0.96-1.01)	1.00 (0.95-1.06)	0.96 (0.91-1.00)	0.10	34.2%	
	Multivariable	0.99 (0.96-1.02)	1 ref	0.99 (0.96-1.01)	1.02 (0.96-1.08)	0.97 (0.92-1.02)	0.61	0%	
ER+	Multivariable	1.00 (0.96-1.03)	1 ref	1.00 (0.96-1.03)	1.02 (0.96-1.09)	0.99 (0.91-1.08)	0.21	22.1%	
ER-	Multivariable	0.93 (0.86-1.00)	1 ref	1.00 (0.93-1.07)	0.99 (0.89-1.10)	0.91 (0.79-1.04)	0.44	1.2%	0.29

Multivariable models include race (White, African American, Hispanic, Asian, other), education (< high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m2), height(<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75m), alcohol (0, 0-<5, 5-<15, 15-<30 and ≥30 g/d), energy intake (continuous, kcal/d), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 except PLCO <10,10/11,12/13, ≥14 years), hormone replacement therapy use (never user, past user, current user), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25y), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

Table S1.3 Pooled multivariable HRs (95% CIs) of total breast cancer for total dietary fiber intake adjusting for fruit and vegetable intake, vitamin C, and carotenoids

		Pooled MVHR (95% CI)					*p*-trend	*p*-het[2]	I^2
Breast cancer		Quintile1	Quintile2	Quintile3	Quintile4	Quintile5			
Overall	Model 1[1]	1	0.98	0.97	0.96	0.93	< 0.001	0.04	37.8%
		ref	(0.95-1.01)	(0.94-1.00)	(0.93-1.00)	(0.90-0.97)			
	Model 1 + Fruit and vegetable intake	1	0.99	0.98	0.96	0.93	0.001	0.06	36.7%
		ref	(0.95-1.02)	(0.95-1.02)	(0.92-1.01)	(0.88-0.98)			
	Model 1 + vegetable intake	1	0.98	0.98	0.96	0.92	0.001	0.02	44.4%
		red	(0.95-1.02)	(0.95-1.01)	(0.92-1.00)	(0.87-0.97)			
	Model 1 + Vitamin C[3]	1	0.99	0.98	0.96	0.93	< 0.001	0.19	22.5%
		ref	(0.95-1.02)	(0.95-1.02)	(0.93-1.00)	(0.89-0.97)			
	Model 1 + α-carotene	1	1.00	0.99	0.98	0.95	0.009	0.06	37.0%
		ref	(0.96-1.03)	(0.96-1.03)	(0.93-1.02)	(0.90-0.99)			
	Model 1 + β-carotene	1	0.99	0.98	0.97	0.94	0.004	0.06	37.5%
		ref	(0.95-1.03)	(0.95-1.02)	(0.92-1.02)	(0.89-0.99)			
	Model 1 + β-cryptoxanthin	1	1.00	1.00	0.98	0.96	0.01	0.18	24.1%
		ref	(0.96-1.04)	(0.97-1.03)	(0.94-1.03)	(0.92-1.00)			
	Model 1 + Lycopene	1	1.00	0.99	0.98	0.95	0.001	0.24	18.1%
		ref	(0.96-1.03)	(0.96-1.03)	(0.94-1.02)	(0.91-0.99)			
	Model 1 + Lutein	1	0.98	0.98	0.96	0.93	0.001	0.18	24.8%
		ref	(0.95-1.02)	(0.95-1.02)	(0.92-1.00)	(0.89-0.98)			
ER+	Model 11	1	0.98	0.98	0.98	0.95	0.08	0.01	46.3%
		ref	(0.94-1.02)	(0.94-1.03)	(0.93-1.03)	(0.90-1.00)			
	Model 1 + Fruit and vegetable intake	1	0.99	1.00	0.98	0.95	0.13	0.004	52.5%
		ref	(0.95-1.03)	(0.95-1.05)	(0.92-1.03)	(0.87-1.02)			
	Model 1 + vegetable intake	1	0.99	0.99	0.97	0.93	0.02	0.06	37.0%
		ref	(0.95-1.03)	(0.95-1.04)	(0.92-1.02)	(0.87-0.99)			
	Model 1 + Vitamin C	1	0.99	1.00	0.98	0.95	0.10	0.02	45.5%
		ref	(0.95-1.03)	(0.96-1.05)	(0.93-1.04)	(0.88-1.01)			

Table S1.3 (Continued)

	Model 1 + α-carotene	1	1.00	1.01	0.99	0.96	0.18	0.11	31.5%
		ref	(0.95-1.05)	(0.96-1.06)	(0.93-1.05)	(0.91-1.02)			
	Model 1 + β-carotene	1	0.99	1.0	0.98	0.96	0.13	0.15	26.4%
		ref	(0.95-1.04)	(0.95-1.05)	(0.92-1.05)	(0.90-1.02)			
	Model 1 + β-cryptoxanthin	1	1.00	1.02	1.00	0.98	0.53	0.03	43.5%
		ref	(0.96-1.05)	(0.97-1.07)	(0.94-1.06)	(0.92-1.05)			
	Model 1 + Lycopene	1	1.00	1.01	0.99	0.97	0.33	0.07	36.0%
		ref	(0.96-1.04)	(0.97-1.06)	(0.94-1.05)	(0.92-1.03)			
	Model 1 + Lutein	1	0.98	0.99	0.96	0.95	0.06	0.23	20.3%
		ref	(0.94-1.03)	(0.95-1.04)	(0.92-1.02)	(0.90-1.01)			
ER-	Model 1[1]	1	0.93	0.96	0.94	0.87	< 0.001	0.43	10.5%
		ref	(0.86-1.01)	(0.88-1.04)	(0.86-1.03)	(0.80-0.95)			
	Model 1 + Fruit and vegetable intake	1	0.95	0.97	0.96	0.92	0.13	0.81	0%
		ref	(0.87-1.03)	(0.89-1.07)	(0.88-1.06)	(0.83-1.02)			
	Model 1 + vegetable intake	1	0.96	1.01	1.00	0.97	0.61	0.70	0%
		ref	(0.88-1.05)	(0.92-1.11)	(0.91-1.10)	(0.88-1.07)			
	Model 1 + Vitamin C	1	0.95	0.98	0.94	0.89	0.02	0.62	0%
		ref	(0.86-1.04)	(0.88-1.08)	(0.86-1.03)	(0.81-0.98)			
	Model 1 + α-carotene	1	0.96	0.99	0.96	0.93	0.07	0.26	16.9%
		ref	(0.88-1.05)	(0.90-1.10)	(0.87-1.05)	(0.83-1.04)			
	Model 1 + β-carotene	1	0.97	1.01	0.98	0.95	0.37	0.09	33.8%
		ref	(0.88-1.07)	(0.91-1.12)	(0.88-1.09)	(0.83-1.09)			
	Model 1 + β-cryptoxanthin	1	0.95	0.98	0.95	0.92	0.07	0.61	0%
		ref	(0.87-1.04)	(0.90-1.07)	(0.87-1.04)	(0.83-1.01)			
	Model 1 + Lycopene	1	0.95	0.99	0.95	0.92	0.06	0.30	13.5%
		ref	(0.87-1.04)	(0.90-1.09)	(0.87-1.04)	(0.83-1.02)			
	Model 1 + Lutein	1	0.94	0.97	0.98	0.93	0.36	0.14	28.0%
		ref	(0.86-1.04)	(0.88-1.06)	(0.88-1.09)	(0.82-1.06)			

Table S1.3 (Continued)

1. Multivariable models include race (White, African American, Hispanic, Asian, other), education (< high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m2), height(<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75m), alcohol (0, 0-<5, 5-<15, 15-<30 and ≥30 g/d), energy intake (continuous, kcal/d), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 except PLCO <10,10/11,12/13, ≥14 years), hormone replacement therapy use (never user, past user, current user), oral contraceptive use (ever, never), add categories: parity (0, 1-2, ≥3), age at first birth (≤25, >25y), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

2. *p* for between-study heterogeneity for quintile 5.

3. Vitamin C from foods

Table S1.4 Median Pearson correlations[1] for dietary fiber, carotenoids, and vitamin C across all studies included in the breast cancer analyses in the Pooling Project of Prospective Studies of Diet and Cancer

	Dietary fiber	Fruit fiber	Vegetable fiber	Grains fiber	α-carotene	β-carotene	β-crypto-xanthin	Lycopene	Lutein	Vitamin C
Dietary fiber	1									
Fruit fiber	0.65	1								
Vegetable fiber	0.75	0.32	1							
Grains fiber	0.66	0.18	0.20	1						
α-carotene	0.49	0.27	0.62	0.17	1					
β-carotene	0.65	0.42	0.81	0.19	0.83	1				
β-cryptoxanthin	0.45	0.58	0.27	0.16	0.20	0.34	1			
Lycopene	0.38	0.22	0.44	0.15	0.22	0.33	0.19	1		
Lutein	0.52	0.31	0.70	0.14	0.39	0.70	0.26	0.25	1	
Vitamin C	0.62	0.62	0.52	0.19	0.36	0.52	0.79	0.38	0.45	1

1. Median correlation value was calculated over all studies that measured that dairy food or nutrient. Studies that did not measure that particular food or nutrient were excluded from that specific analysis.

Table S2.1 Number of cases of breast cancer subtypes defined by estrogen receptor (ER) and progesterone receptor (PR) status for cohort study included in the pooled analyses of dairy products and breast cancer risk

Study*	Country	ER+	ER-	Missing ER status, %	ER+/PR+	ER+/PR-	ER-/PR+	ER-/PR-
CARET	USA	193	31	38.96	160	22	3	26
BWHS	USA	416	254	0	315	92	11	239
BCDDP	USA	793	166	26.51	635	135	31	132
CTS	USA	1930	343	15.69	1509	316	32	308
CLUE II	USA	198	50	13.89	159	37	9	41
CNBSS	Canada	367	125	60.32	272	48	21	76
CPSII	USA	1835	323	28.04	1440	282	37	272
IWHS	USA	1329	238	15.25	1082	191	34	196
JPHC1	Japan	111	69	37.72	74	30	13	52
MCCS	Australia	493	171	16.9	393	96	26	144
MEC	USA	2169	543	18.02	1698	309	69	464
NLCS	The Netherlands	700	183	56.14	348	100	13	96
NYUWHS	USA	392	121	44.18	272	111	24	92
AARP	USA	2322	464	53.35	1853	381	56	404
NHSa	USA	528	255	30.21	345	104	33	186
NHSb	USA	3075	757	14.22	2358	595	97	638
NHS II	USA	846	303	13.67	710	119	49	239
ORDET	Italy	206	67	3.53	159	46	19	46
PLCO	USA	858	137	8.72	750	101	8	126
SMC	Sweden	1605	384	23.65	1199	394	106	276
WHS	USA	937	187	4.5	796	124	23	164
WLHS	Sweden	737	196	12.97	571	155	42	153
Total		22040	5367		17098	3788	756	4370

*Abbreviations: CARET, Beta-Carotene and Retinol Efficacy Trial; BWHS, The Black Women's Health Study; BCDDP, Breast Cancer Detection Demonstration Project Follow-up Study; CTS, California Teachers Study; CLUE II, CLUE II: Campaign Against Cancer and Heart Disease; CNBSS, Canadian National Breast Screening Study; CPS II, Cancer Prevention Study II Nutrition Cohort; IWHS, Iowa Women's Health Study; JPHC1, Japan Public Health Center-Based Study Cohort I ; MCCS, Melbourne Collaborative Cohort Study; MEC, Multiethnic Cohort; NLCS, Netherlands Cohort Study; NYUWHS, New York University Women's Health Study; NIH-AARP, NIH-AARP Diet and Health Study; NHSa, Nurses' Health Study (part a); NHSb, Nurses' Health Study (part b); NHS II, Nurses' Health Study II; ORDET, Hormones and Diet in the Etiology of Breast Cancer Study; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SMC, Swedish Mammography Cohort; WHS, Women's Health Study; WLHS, Women's Lifestyle and Health Study.

Table S2.2 Median Pearson correlations* for dairy products and nutrients across all studies included in the breast cancer analyses in the Pooling Project of Prospective Studies of Diet and Cancer

	Total milk	Whole milk	Reduced-fat milk	Hard cheese	Cottage/ ricotta cheese	Yogurt	Ice cream	Dietary calcium	Total calcium
Total milk	1								
Whole milk	0.29	1							
Reduced-fat milk	0.90	-0.09	1						
Hard cheese	0.03	0.03	0.01	1					
Cottage cheese	0.07	0.00	0.07	0.10	1				
Yogurt	0.07	-0.02	0.07	0.05	0.14	1			
Ice cream	0.03	0.05	0.01	0.10	0.02	0.01	1		
Dietary calcium	0.74	0.12	0.73	0.19	0.12	0.28	0.00	1	
Total calcium	0.44	0.02	0.44	0.07	0.09	0.22	-0.03	0.59	1

*Median correlation was calculated over all studies that measured that dairy food or nutrient. Studies that did not measure that particular food or nutrient were excluded from that specific analysis.

Table S2.3 Pooled age and multivariable[*] adjusted hazard ratios (HR) and 95% confidence intervals (CI) for breast cancer overall and for subtypes defined by estrogen receptor (ER)[†] status according to calcium intake

Nutrients	Quintiles of intake MVHR (95% CI)					P_{diff}[‡]	I^2 [§]	P_{het}[‖]	P_{trend}[¶]
Dietary Calcium	Q1	Q2	Q3	Q4	Q5				
Total	1.00 (Ref)	0.98 (0.95-1.02)	0.99 (0.95-1.02)	0.97 (0.93-1.00)	0.95 (0.91-0.98)		0.06	0.39	0.001
ER+	1.00 (Ref)	0.97 (0.92-1.01)	0.98 (0.93-1.03)	0.95 (0.90-1.01)	0.93 (0.87-0.99)	0.63	0.44	0.02	0.04
ER-	1.00 (Ref)	1.03 (0.94-1.12)	0.99 (0.91-1.08)	0.97 (0.88-1.06)	0.96 (0.87-1.04)		<0.001	0.53	0.14
Total Calcium									
Total	1.00 (Ref)	1.00 (0.96-1.04)	0.98 (0.94-1.02)	0.99 (0.94-1.03)	0.96 (0.91-1.02)		0.37	0.11	0.24
ER+	1.00 (Ref)	1.02 (0.96-1.07)	0.98 (0.93-1.04)	1.01 (0.96-1.07)	0.99 (0.91-1.07)	0.98	0.51	0.03	0.83
ER-	1.00 (Ref)	1.05 (0.91-1.21)	1.01 (0.85-1.20)	1.01 (0.88-1.16)	0.99 (0.89-1.10)		<0.001	0.70	0.50

[*]Multivariable model includes race (White, African American, Hispanic, Asian, other), education (<high school, high school, >high school), BMI (<23, 23-<25, 25-<30, ≥30 kg/m^2), height (<1.60, 1.60-<1.65, 1.65-<1.70, 1.70-<1.75, ≥1.75 m), alcohol consumption (0,>0-<5, 5-<15, 15-<30 and ≥30 g/day), energy intake (kcal/d, continuous), smoking status (never, past, current), physical activity (low, medium, high), age at menarche (<11, 11/12, 13/14, ≥15 years except PLCO: <10, 10/11, 12/13, ≥14 years), hormone replacement therapy use (never, past, current), oral contraceptive use (ever, never), parity (0, 1-2, ≥3), age at first birth (≤25, >25 years), history of benign breast disease (yes, no), family history of breast cancer (yes, no). Age in years and year of questionnaire return were included as stratification variables.

[†]ER: estrogen receptor

[‡]P_{diff}: p-value, test for common effects for different subtypes defined by estrogen receptor status for the highest quintile

[§]I^2 for the highest quintile

[‖]P_{het}: p-value, test for between studies heterogeneity for the highest quintile

[¶]P_{trend}: p-value, test for trend

Table S3.1 Median Pearson correlations[1] for food groups across all studies included in the breast cancer analyses in the Pooling Project of Prospective Studies of Diet and Cancer

	Total red meat	Processed meat	Unprocessed red meat	Poultry	Seafoods	Eggs	Nuts	Mature beans	Total dairy
Total red meat	1								
Processed meat	0.68	1							
Unprocessed red meat	0.89	0.28	1						
Poultry	0.06	0.04	0.06	1					
Seafoods	0.00	0.01	0.00	0.26	1				
Eggs	0.19	0.18	0.16	0.06	0.07	1			
Nuts	0.04	0.04	0.03	0.02	0.04	0.04	1		
Mature beans	0.03	0.03	0.03	0.04	0.07	0.03	0.07	1	
Total dairy	0.06	0.05	0.04	0.05	0.08	0.08	0.09	0.06	1

1. Median correlation value was calculated over all studies that measured that food group. Studies that did not measure that particular protein source were excluded from that specific analysis.

Table S3.2 Sample conversions from grams to calories of common items in the food groups

	Food item	Gram	Calories	Sample conversion
Red meat				
	Beef brisket, cooked	100	251	1 piece = 320g or 800kcal
	Pork chop, bone-in, cooked	100	209	1 chop = 160g or 330kcal
	Lamb, ground, cooked	100	283	1 unit = 310g or 890kcal
Processed meat				
	Pork, bacon, cooked	100	548	1 slice = 8g or 44kcal
	Sausage, beef, cooked	100	332	1 serving = 43g or 143kcal
Poultry				
	Chicken, meat and skin, roasted	100	239	From 1lb raw = 178g or 425kcal
	Chicken, meat only, roasted	100	190	From 1lb raw = 146g or 277kcal
	Chicken, meat only, fried	100	219	From 1lb raw = 155g or 339kcal
Seafood				
	Salmon, cooked	100	182	1 fillet = 154g or 280kcal
	Cod, cooked	100	84	6oz = 170g or 142kcal
	Shrimp, cooked	100	99	6oz = 170g or 170kcal
Eggs				
	Egg, whole, poached	100	143	1 large = 50g or 72kcal
High-fat dairy				
	Whole milk	100	61	1 cup = 245g or 150kcal
	Yogurt, plain, whole milk	100	61	4oz container = 113g or 70kcal
	Cheese, monterey	100	373	1 cup = 113g or 421kcal
	Cheese, mozzarella	100	299	1 cup = 113g or 335kcal
	Cheese, ricotta	100	150	1/2 cup = 124g or 186kcal
Low-fat dairy				
	Nonfat milk	100	35	1 cup = 245g or 86kcal
	Yogurt, plain, skim milk	100	56	4oz container = 113g or 63kcal
Mature beans				
	Baked beans, canned	100	105	1 cup = 253g or 266kcal
Soy				
	Tofu, silken, firm	100	62	1 slice = 84g or 52kcal
	Soybean, mature boiled	100	172	1 cup = 172g or 296kcal

Table S4.1 Multivariable-adjusted[1] hazard ratio and 95% CI for breast cancer overall by WCRF cancer prevention recommendations using baseline, simple update, or cumulative average

	0-4 years		4-8 years		8-12 years		12-16 years	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Alcohol consumption								
non drinker	Ref		Ref		Ref		Ref	
>0.0-4.9 g/d	1.06	(1.00-1.13)	0.98	(0.92-1.05)	1.06	(0.99-1.14)	1.01	(0.93-1.09)
5.0-9.9 g/d	1.12	(1.03-1.22)	1.06	(0.96-1.16)	1.10	(1.00-1.22)	1.05	(0.94-1.18)
10.0-14.9 g/d	1.12	(1.02-1.23)	1.16	(1.06-1.28)	1.17	(1.06-1.30)	1.13	(1.00-1.27)
15.0-29.9 g/d	1.17	(1.06-1.29)	1.25	(1.13-1.39)	1.19	(1.06-1.34)	1.31	(1.15-1.49)
30+ g/d	1.38	(1.23-1.54)	1.26	(1.12-1.43)	1.33	(1.16-1.53)	1.31	(1.12-1.54)
Fruit and vegetable intake								
0-2 servings/d	Ref		Ref		Ref		Ref	
2-4 servings/d	0.95	(0.87-1.03)	0.99	(0.91-1.09)	0.88	(0.79-0.97)	0.98	(0.87-1.11)
4-6 servings/d	0.92	(0.85-1.01)	0.96	(0.87-1.06)	0.90	(0.81-1.00)	0.96	(0.85-1.09)
6-8 servings/d	0.91	(0.82-1.00)	0.91	(0.82-1.02)	0.88	(0.78-0.99)	0.92	(0.80-1.06)
8+ servings/d	0.90	(0.81-1.01)	0.91	(0.81-1.03)	0.85	(0.75-0.97)	0.92	(0.79-1.07)
Physical activity								
<3 total METs-hr/week	Ref		Ref		Ref		Ref	
3-9 total METs-hr/week	1.05	(0.98-1.12)	1.09	(1.02-1.18)	1.05	(0.97-1.14)	1.01	(0.92-1.10)
9-18 total METs-hr/week	1.00	(0.93-1.07)	1.07	(0.99-1.15)	1.01	(0.94-1.10)	1.04	(0.95-1.13)
18-27 total METs-hr/week	1.01	(0.93-1.10)	1.10	(1.01-1.20)	1.00	(0.91-1.10)	1.07	(0.96-1.19)
27+ total METs-hr/week	0.93	(0.87-1.00)	1.06	(0.98-1.14)	1.05	(0.96-1.14)	1.03	(0.94-1.14)
Weight change since age 18								
<5kg	Ref		Ref		Ref		Ref	
5 - <10kg	1.16	(1.08-1.24)	1.15	(1.07-1.24)	1.16	(1.07-1.26)	1.2	(1.10-1.31)
10 - <15kg	1.12	(1.04-1.21)	1.18	(1.10-1.28)	1.26	(1.16-1.37)	1.29	(1.17-1.42)
15 - <20kg	1.32	(1.22-1.43)	1.37	(1.26-1.48)	1.41	(1.29-1.54)	1.46	(1.33-1.62)
≥20kg	1.44	(1.35-1.54)	1.49	(1.39-1.60)	1.48	(1.37-1.60)	1.56	(1.43-1.70)

1. Multivariable model include: height (<1.6m, 1.60-1.64m, 1.65-1.69m, 1.70-1.74m, ≥1.75m), age at menarche (<12 yo, 12, 23, 24, >14 yo), oral contraceptive use (No OC use, >0-2 yrs, >2-5yrs, >5-10 yrs, >10 yrs), joint classification of age at first birth and parity (Nulliparous, 1-2 children AFB <25, 1-2 children AFB 25-29, 1-2 children AFB 30+, 3-4 children AFB <25, 3-4 children AFB 25-29, 3-4 children AFB 30+, 5+ children AFB <25, 5+ children AFB 25+), menopausal status and postmenopausal hormone therapy use (premenopausal/missing menopause, no history of postmenopausal hormone use, current postmenopausal hormone use, past postmenopausal hormone use), confirmed benign breast disease (yes, no), family history of breast cancer (yes, no), body mass index at age 18 (<18.5, 18.5-<20, 20-<22.5, 22.5-<25, 25-<30, 30+), total energy intake.