Title
The role of diet on breast cancer survival: The Pathways Study

Author
Ergas, Isaac

Publication Date
2021

Peer reviewed|Thesis/dissertation

The role of diet on breast cancer survival: The Pathways Study

By

Isaac Joshua Ergas

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Patrick Bradshaw, Chair
Professor Barbara Laraia
Professor Kristine Madsen

Summer 2021

Abstract

The role of diet on breast cancer survival: The Pathways Study

By

Isaac Joshua Ergas

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Patrick Bradshaw, Chair

Breast cancer survivors are at higher risk of death, cancer recurrence, and other comorbidities after diagnosis, as compared to cancer-free women. Dietary guidelines designed to improve prognosis, a lifestyle treatment strategy for newly diagnosed breast cancer patients, are primarily based on research findings related to breast cancer incidence, not survival. Due to post diagnostic factors such as altered physiology, cancer treatments, increased awareness of symptoms, and modified dietary and lifestyle behaviors, it is not likely that the mechanisms contributing to breast cancer etiology are the same as those that lead to recurrence and death. This dissertation provides critical and timely information to support or warrant modification of current dietary recommendations for breast cancer survivors, which will ultimately benefit the estimated 3.8 million women currently living with breast cancer in the United States.

The three analyses outlined in this dissertation leveraged data from the Pathways Study, a long-standing prospective cohort study of a diverse population of breast cancer survivors at Kaiser Permanente Northern California (KPNC), an integrated health care delivery system. Recruitment of study participants began in 2006 and continued through 2013, enrolling 4,505 breast cancer survivors within an average of 2 months after diagnosis. Surveys were administered to participants at enrollment (baseline), 6, 24 and 72 months. In addition to the surveys, these analyses utilized KPNC's rich clinical and administrative databases, including both demographic and clinical characteristics such as tumor staging, tumor size, hormone receptor status, and treatment. These databases also provided ascertainment of both breast cancer recurrence and mortality.

In chapter 1, the relationship between four *a priori* dietary quality indices consistent with healthy eating recommendations around the time of breast cancer diagnosis and breast cancer recurrence, cause-specific mortality, and all-cause mortality were evaluated. The dietary quality indices included an index based on the American Cancer Society nutrition guidelines (ACS), the alternate Mediterranean Diet Index (aMED), an index based on the Dietary Approaches to Stop Hypertension diet (DASH), and the 2015 Healthy Eating Index (HEI). Minimally and fully adjusted outcome models including each index were evaluated, and sub-analyses were also conducted to examine the independent associations between the individual food components from each index and

all-cause mortality. Assessments of effect measure modification were also considered between the indices and select covariates. Overall, this analysis showed that participants who reported consuming diets that were more concordant with healthy eating patterns, were found to be at lower risk of non-breast cancer-specific and all-cause mortality. No clear patterns emerged when examining the associations between the dietary quality indices and breast cancer recurrence or breast cancer-specific mortality.

In chapter 2, the potential effects of interventions on diet quality and lifestyle factors on survival after a breast cancer diagnosis was examined using a causal inference approach. In this study, the parametric g-formula was applied to observational data from the Pathways Study to estimate the risk of all-cause mortality under several hypothetical interventions related to dietary quality, physical activity, and smoking. Each intervention was assumed to begin at the time of the breast cancer diagnosis and maintained over a 13-year follow-up period. Hypothetical interventions at modest levels of intensity were first considered and then interventions with progressively increased intensity to their maximum levels were pursued. Joint interventions on combinations of diet and lifestyle factors were evaluated, as well as the expected risks under no intervention (natural course). This analysis showed hypothetical interventions that increased diet quality, increased physical activity, and stopped participants from smoking, each reduced the risk of death among breast cancer survivors. It also showed that increasing the intensity of the intervention on diet and physical activity was directly related to the strength of the associations. Joint interventions on combinations of diet and lifestyle factors were also evaluated and conveyed the greatest reductions in risk.

In chapter 3, a hierarchical modelling approach was used to examine the relationship between survival and intake of multiple food items assessed at baseline, including whole grains, refined grains, dairy, vegetables, legumes, nuts, processed meat, red meat, poultry and added sugars. A second level model was specified to explain drivers of the food level effects *via* nutrients considered to be related to survival (*e.g.* carbohydrates, protein, fiber, calcium, iron, isoflavones, vitamin C, vitamin D, and others). This approach allowed estimation of the mutually adjusted associations between multiple food items on breast cancer survival, as well as the role of specific nutrients in these foods. This study showed a decreased risk in all-cause mortality with increased consumption of whole grains, soy products and nuts and seeds at baseline. Among the fixed effects, iron, isoflavone and fiber consumption were associated with a decreased risk of all-cause mortality, though the estimates for iron and fiber were most imprecise.

These analyses demonstrated that diet, whether evaluated in the context of an overall dietary pattern, or though hypothetical interventions in combination with other lifestyle factors, or examined as individual food items in conjunction with their associated nutrients, plays a critical role in survival after a breast cancer diagnosis. A clearer understanding of diet and its impact on breast cancer prognosis is an essential contribution to the development of guidelines specifically designed for woman surviving breast cancer.

TABLE OF CONTENTS

I. DEDICATION

For Iris and David Ergas

II. ACKNOWLEDGEMENTS

I am extremely grateful to everyone in my life who has supported and encouraged me to believe in myself and aim high.

First and foremost, a big thank you to my academic advisor and committee chair, Patrick Bradshaw, for spending so much time and unyielding attention to helping me make this dissertation great. I have learned so much from you over the past four years and feel so lucky to have found you when I did.

To Barbara Laraia and Kristine Madsen, my committee members, thank you for your attention, guidance, and support throughout this dissertation process. Your time and feedback on both my dissertation and academic decision making has been invaluable. Barbara, I so much enjoyed working with you as your GSI and am always so honored when you call on me to return to your class to give a talk.

To Jack Colford, you have been a mentor to me for the last 20 years! Thank you for always supporting me and offering up your teachings, wisdom and advice which often extended beyond academics. I am truly fortunate to have known you all these years.

To Art Reingold, from day one you worked to help me make my time at UC Berkeley financially feasible. My family and I thank you so much for this, I could not have completed this program without your help.

To Sandi McCoy, I learned so much from being your lead GSI. You are a great teacher and a major asset to the program. Thank you for being so dedicated and supportive. There were some tough moments there for me and your thoughtfulness and empathy helped to get me through it.

To Mahasin Muhahid, thank you for giving me the opportunity to teach the SAS course, it was a great experience and gave me good insights to the life of a professor.

To Lisa Barcellos, thank you for teaching me genetic epidemiology and inviting me to speak to your study group! You opened an entire world to me.

To Janene Martinez, Lauren Krupa and Sumaiya Elahi, you are each amazing and I do not know how you do it. Thank you for responding so quickly to all my emails and for helping me to navigate this program!

To Larry Kushi, Marilyn Kwan, and Janise Roh, I do not even know where to begin thanking each of you for being incredible role models, teachers, colleagues, and friends. We have worked for so long and have been through so much together, I cannot imagine my career without you. Thank you for constantly putting your faith in me and for setting me up for success.

To Elizabeth Feliciano, Cecile Laurent, Valerie Lee, Emily Valice, Lia D'Addario, Julie Munneke, Rosa Hippler, Lisa Moy, Jen Cadenhead, Alyssa Troeschel, Margarita Torres, and the rest of the folks at or affiliated with the Division of Research at Kaiser Permanente, thank you all so much for being brilliant colleagues. Elizabeth, thank you especially for always taking the time to offer your insights, feedback, and wisdom into my research. You have been an amazing resource and I feel so lucky to have known you.

To the participants of the Pathways Study, thank you for offering your time and energy to take part in this study. Without you, there would be no study, no data, and no dissertation.

To my amazing and brilliant epi PhD cohort crew, you guys rock! I cannot wait to see what each of you accomplish in your careers. Thank you for being so extremely awesome and supportive. I am so fortunate to have known you.

To my uncle Hy and his descendants, thank you for helping to financially support this endeavor. You provide a great service to our family.

To Mara, Sanders, Whitney, Ken, Seta, Barrett, Michelle, and everyone in their families, I love you all so much and thank you for always being in my corner.

To my lovely wife Ojig and my two amazing boys, Nshan and Jano, you made all this possible. Thank you for being on my team, I love life with you.

III. INTRODUCTION

Other than skin cancer, breast cancer is the most commonly diagnosed cancer among women living in the United States (US) and the second leading cause of cancer death.[1] The age adjusted rates of new breast cancer cases and deaths are 129.1 and 20.1 per 100,000 women per year, respectively.[1] It is also estimated that there will be 281,550 newly diagnosed cases of invasive breast cancer by the end of 2021 and that approximately 43,600 women will die from this disease.[2] Mortality rates from breast cancer have dropped by 41% between 1989 and 2018, which is believed to be the result of earlier screening, increased awareness, and improved treatments.[2] This decrease in mortality translates into approximately 403,200 fewer deaths than would have otherwise been expected without the improvements.[2] According to the American Cancer Society (ACS), survival rates for women diagnosed with breast cancer relative to women not diagnosed with breast cancer are 91%, 84%, and 80% at 5, 10 and 15 years after diagnosis, respectively.[3] As a result, breast cancer survivors are now a growing population with an estimated 3.8 million women currently living with breast cancer in the US.[3]

The term “breast cancer survivor,” in the context of this dissertation, refers specifically to biologically identified women, even though prior research has not historically made this distinction. Furthermore, while the risk of breast cancer in biologically identified men has increased slightly over the past 30 years, it continues to be rare (approximately 1.2 new cases per 100,000 during 2012-2016),[3] and is not the focus of this work. Typically, and as is the case here, the term “breast cancer survivor” also includes cases at all stages of survival, including diagnosis, treatment, recovery and through the end of life.[4] However, this term alone does not do justice to the heterogeneity of survivorship, whereby the impact of both modifiable and fixed exposures can vary across the continuum of survival.[4] In the diagnostic phase, breast cancer can present itself as both non-invasive and invasive disease. When non-invasive, the cells that line the milk ducts of the breast have become cancerous, but they have not spread into surrounding tissue. This type of cancer usually does not spread outside the breast and is known as ductal carcinoma *in situ* (DCIS).[3] Alternatively, invasive breast cancer develops beyond the milk ducts into normal, healthy tissue and can be categorized into four general groups of increasing severity. The first is *localized*, in which cancer is found only in the part of the body where it started. *Regional* is more advanced, where the cancer has spread to the regional lymph nodes. The most progressed is *distant,* in which cancer has metastasized to other parts of the body, and there is *unknown* staging when it is not clear where the cancer started or the extent of its spread.[4]

There is substantial variability in the prognosis of breast cancer. About 63% of breast cancer cases are localized and the five-year survival rate for these women is approximately 99%. The five-year survival rate is only slightly less for regional (86%), but then drops considerably for both distant (29%) and unknown (58%).[1] These rates may highlight the importance of screening and early detection. There are several other demographic and clinical factors which are known to be associated with survival,

including age, race,[5] lymph node involvement, tumor size, tumor grade, hormone receptor status and conventional therapies.[6-9]

While cardiovascular disease (CVD) is a major public health issue for both men and women living in the US,[10] it is of growing concern for women living with breast cancer.[11, 12] A recent study of 754,270 women diagnosed with breast cancer between 2000 and 2015 from the Surveillance, Epidemiology, and End Results (SEER) program showed that breast cancer-specific deaths accounted for 65.4%, 58.6%, 38.2% and 23.6% of all deaths during the first year, 1 to 5 years, 5 to 10 years and more than 10 years after their breast cancer diagnosis, respectively. While the proportion of breast cancer-specific deaths decreased, deaths due to heart disease increased with proportions of 10%, 11.2%, 15.7% and 19.3% of all deaths over the same four study periods.[13] Breast cancer surivors have also been shown to be at greater risk of dying from CVD as compared to women without breast cancer approximately 7 years after diagnosis.[11] This elevated risk of CVD among breast cancer survivors may be explained by cardiotoxic cancer treatments and exacerbated by associated CVD risk factors.[14, 15]

However, after accounting for these known risk factors, substantial variation in survivorship outcomes remain.[4] For this reason, there is now a great deal of attention focused on diet, physical activity, and obesity, and their role in survival,[4] though the body of research surrounding these risk factors within the context of survival is limited. Because these are modifiable risk factors, their importance and potential public health impact cannot be understated. Significant interventions may be possible if we can come to understand how these risk factors impact breast cancer prognosis.[4]

Approaches to the study of diet and breast cancer

Dietary research on breast cancer etiology has historically focused on the assessment of specific nutrients and the foods from which they are consumed.[16, 17] This may stem from the growing evidence of the impact of diet and nutrition on the biological mechanisms that underly the cellular changes required for cancer development.[4] Nutritional factors have been shown to adversely and beneficially influence mechanisms involved in DNA repair, influence pathways by which carcinogens are metabolized and prompt epigenetic changes in cells.[18-20] Fruits and vegetables, for example, containing folate, fiber and vitamins A, C and E have been shown to decrease inflammation and support apoptosis as well as genomic stability.[4] Conversely, higher intakes of red and processed meats provide elevated exposure to nitrites and endogenous N-nitroso compounds which may lead to oxidative stress and genomic instability.[4] While these singular food and nutrient level relationships are important pieces of the dietary puzzle and their investigation is warranted, diet as a whole is a complex system of interacting components and therefore may also require a more holistic approach.

Recently, there has been a growing interest in dietary patterns and their impact on breast cancer survival.[21-25] The reason being that people do not eat foods in isolation, but rather in combination with other foods according to various patterns. There is also a growing consensus that different patterns of diet, when combined with physical activity,

can lead to metabolic dispositions which are conducive to the cellular alterations marked by cancer.[4] Dietary patterns, which reflect the intakes of all foods, drinks and nutrients in one's overall diet, [23, 26] can be quantified with either data driven or *a priori* indices designed to measure adherence to a given pattern.[27] Empirically driven indices will often utilize a principal components or a factor-analytic approach to help derive dietary patterns within the cohort.[28] An example of this approach might be a pattern corresponding to a "western" diet, typically characterized by high intakes of meat, refined grains and high-fat foods, and comparing this to a "prudent" diet, which is typically characterized by high intakes of fruits and vegetables as well as whole grains.[25] A more prominent method of measuring dietary patterns is through the use of *a priori* numerical indices, which measure adherence to a predefined dietary pattern derived from prior research. Some examples of the most utilized dietary indices are the Healthy Eating Index (HEI), Dietary Approaches to Stop Hypertension index (DASH) and the alternate Mediterranean Diet Index (aMED).[29] While HEI measures adherence to the dietary guidelines published by Health and Human Services and the United States Department of Agriculture, DASH assesses adherence to a diet that was originally developed to treat and prevent hypertension.[29] Conversely, aMED measures adherence to a diet which emphasizes foods based on the dietary traditions of Crete, Greece and Southern Italy, where chronic disease has been historically low.[29] There are also two newly developed indices, the first based on the American Cancer Society's nutritional guidelines for cancer prevention (ACS),[30] and the second is the Healthful Plant Based Diet index (hPDI).[31] While the hPDI emphasizes consumption of healthy plant based foods, ACS reflects the most current scientific evidence related to diet and cancer prevention.[30]

While dietary indices are a strategic means of capturing the entire dietary experience, they too have their limitations. Overall dietary index scores alone cannot distinguish the main dietary drivers for observed associations, since the unidimensional nature of the index means that each of the food components of the score are treated interchangeably. It is also possible that any observed dietary association might be confounded by food components that were not included in the index. The aMED, for example, while emphasizing limited consumption of red and processed meats as well as adequate consumption of fruits and vegetables, provides an overall index score on adherence to these behaviors. However, it does not separate out the contribution these individual components have on the observed association.[32] The aMED also does not account for added sugars or more specifically sugar sweetened beverages, for example, which could confound observed associations.[32] It is therefore important to supplement findings with a more rigorous examination of the jointly adjusted individual food components that make up the index. Educating the public contextually using dietary patterns may seem overly complicated and abstract to some people in need of an intervention. Because advising on specific food and nutrient consumption has historically been more straightforward and actionable, there may now be a need for more education to help the public better understand dietary patterns.

Epidemiology of diet and breast cancer survival

Breast cancer survivors are highly motivated to make lifestyle changes after diagnosis, and diet and nutrition are important factors for these women.[33-35] In one study, women who changed their diets after diagnosis reported that they did so in order to help treat their disease and promote longer survival.[34] They also reported that most advice related to diet came from friends, family and the media and expressed a desire for more evidence-based information from health professionals.[34] Another study showed that the majority of women who changed their diets after a breast cancer diagnosis reduced their consumption of meat and sugar and increased their consumption of fruits and vegetables.[36] Institutions such as the ACS and the World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR) publish recommended dietary guidelines for breast cancer survivors based on the most up to date research, however the majority of this research is based on the etiology of disease, not studies of survival.[33, 37] Generally speaking, guidelines for long term survival recommend diets that are low in fat and high in fruits and vegetables, and to avoid processed meats and limit red meat consumption for optimal health.[33, 37] However, because the majority of studies examining diet and breast cancer survival are informed from etiological studies of breast cancer, they are unable to assess the impact of dietary behaviors after diagnosis.

While there is an abundance of published epidemiological studies examining diet and the etiology of breast cancer,[38-41] the research on diet and breast cancer prognosis is limited.[4] Some prior studies have shown that that diets rich in high fat foods are associated with worse survival outcomes, specifically diets rich in high fat dairy products.[26, 42-44] Conversely, diets high in vegetable, fiber and fruit intake have been found to be associated with improved prognosis and lower mortality.[45-47] Other studies have failed to find a relationship between alcohol consumption after diagnosis and survival outcomes.[48-50] Importantly, research on the relationship between dietary patterns and mortality among breast cancer survivors is lacking. Two studies of empirically defined dietary patterns found associations between western vs prudent (or healthy vs unhealthy) diets and non-breast cancer mortality. Two other observational studies found associations between the HEI and DASH and non-breast cancer mortality, as well as with all-cause mortality. It is important to note that these studies were heterogeneous in terms of the timing of the baseline dietary assessment (pre vs post-diagnosis) and none of them included follow-up measures. Only one study has assessed multiple *a priori* dietary quality indices at baseline among a cohort of breast cancer survivors enrolled near the time of their diagnosis.[51] The Shanghai Breast Cancer Survival Study (SBCSS), reported that comparisons of extreme quartiles from both the Chinese Food Pagoda and DASH were associated with a lower risk of total mortality and breast cancer-specific events.[51] No published studies in the US have reported on dietary patterns among breast cancer survivors enrolled at the time of their diagnosis and make use of repeated measures.

The use of randomized controlled trials (RCT) to examine diet and breast cancer survival is rare. This may be due to the difficulty of effectively implementing such a study that would require strict adherence to dietary interventions for long follow-up

periods, and which, could ultimately fail the required standards of equipoise. Until now, there have been only two RCTs, each examining low-fat diets on breast cancer recurrence.[52, 53] These are the Women's Healthy Eating and Living (WHEL) study[53] and the Women's Intervention Nutrition (WIN) study,[52] both multisite RCTs examining the effectiveness of low-fat diets among women with early stage breast cancer. While the WHEL study was unable to demonstrate a low-fat diet having any effect on survival [HR=0.91, 95% CI=0.72-1.15], the WIN study did [HR=0.82, 95% CI=0.70-0.96].

There are several research gaps that need to be addressed around the role of diet on breast cancer survival. Among them are the need to better understand the impact of overall dietary patterns on survival, because national dietary guidelines are typically presented as dietary patterns, and not individual foods.[37] Also, due to the paucity of dietary data on breast cancer survival from RCTs and the need for stronger causal evidence, more sophisticated epidemiological techniques need to be employed in order to make use of existing observational data. These techniques include causal approaches that incorporate g-computational methods to initiate hypothetical interventions when actual interventions are not possible, as well as Bayesian approaches that can incorporate prior knowledge of nutrient and constituent effects into the analysis of mutually adjusted foods on survival.

Study Overview

To help address these gaps, data from the Pathways Study, a prospective cohort of women with invasive breast cancer, was used to evaluate the role of diet on survival. Recruitment began in 2006 and continued through 2013, enrolling women an average of 2 months after their diagnosis from Kaiser Permanente Northern California (KPNC), an integrated health care delivery system. Study surveys were administered to participants at baseline, 6, 24 and 72 months. Data from these surveys were augmented with data from KPNC's rich clinical and administrative databases, enabling access to both demographic and clinical characteristics such as tumor staging, tumor size, hormone receptor status, and treatment. These databases also provided rapid ascertainment of both breast cancer recurrence and mortality.

First, the associations between four *a priori* dietary quality indices (ACS, aMED, DASH and HEI) estimated from dietary data collected soon after breast cancer diagnosis and breast cancer recurrence, cause-specific and all-cause mortality, were evaluated. Then a parametric g-formula was used on repeated dietary indices (hPDI, ACS and HEI) and behavioral measures (physical activity and smoking) to estimate the 13-year risk of all-cause mortality under several hypothetical interventions on dietary quality, physical activity and smoking at the time of breast cancer diagnosis. Finally, the relationship between the consumption of multiple food groups assessed at baseline, as well as the role of specific nutrients within these foods, was evaluated on all-cause mortality. By making use of nutrients, foods, and overall dietary patterns, these three analyses together provided a comprehensive assessment of the role of diet on breast cancer survival.

Data from 4,505 female breast cancer survivors enrolled in the Pathways Study was used to address the following aims and hypotheses:

Aim 1. Estimate and compare the associations between four major dietary indices at baseline (ACS, aMED, DASH and HEI) and breast cancer outcomes (recurrence, breast cancer-specific mortality and all-cause mortality) and to assess the magnitude of the food components that drive those associations.

Hypothesis: Increased concordance with each of the four major dietary indices is associated with a *decreased* risk of each of the breast cancer outcomes.

Aim 2. Estimate the impact of hypothetical dietary and lifestyle interventions at varying levels of intensity at the time of breast cancer diagnosis on the risk of all-cause mortality.

Hypothesis 1: Increased adherence to each of the diet, lifestyle and combined hypothetical interventions are associated with a *decreased* risk of all-cause mortality.

Hypothesis 2: The hypothetical interventions characterized by the combined diet and lifestyle factors will have a *greater* impact than each of hypothetical interventions will individually.

Aim 3. Simultaneously estimate the relationship between the intake of multiple food groups assessed at baseline, in the context of their constituent nutrients on breast cancer survival.

Hypothesis 1: Fruits, vegetables, legumes, nuts and seeds are each associated with a *decreased* risk of all-cause mortality.

Hypothesis 2: Red and processed meats, sugar-sweetened beverages and sweets are associated with an *increased* risk of all-cause mortality.

Hypothesis 3: Consumption of foods with high fiber, high soy and low total fat content are associated with a *decreased* risk of all-cause mortality.

Kaiser Permanente Northern California (KPNC)

Founded in 1945, Kaiser Permanente is one of the nation's largest not-for-profit health plans, and its northern region serves over 4.5 million members in 21 hospitals and 264 medical offices. It employs over 80,000 employees, with over 9,500 of them physicians. Headquartered in Oakland, KPNC oversees 19 county regions in Northern California, including the San Francisco Bay and Sacramento metropolitan areas. KPNC has a number of electronic databases with data available for the full membership dating back to 1996 for clinical, administrative, and research purposes, which include inpatient and outpatient encounters, diagnoses, procedures, laboratory and pathology orders and

results, and pharmacy prescriptions. These rich data enable efficient data collection on breast cancer treatment and outcomes.

Pathways Study

The Pathways Study (NCI R01 CA105274, PI: Kushi; U01 CA195565, MPIs: Kushi, Ambrosone), is a prospective cohort study of a diverse population of breast cancer survivors in KPNC.[54] The primary aims of the study are to examine the effect on recurrence, survival, and other outcomes of 1) lifestyle factors such as diet, physical activity, and use of complementary and alternative therapies; 2) molecular factors such as biomarkers of inflammation, germline genomic variables, and tumor characteristics; and 3) geospatial and neighborhood factors. Eligibility criteria for the study were: age ≥21 years; current KPNC member; recently diagnosed with invasive breast cancer; no prior history of malignant cancer; spoke English, Spanish, Cantonese, or Mandarin; and lived within a 65-mile radius of a field interviewer. Field interviewers lived throughout the KPNC service area. Recruitment into the cohort was from January 2006 to May 2013. Baseline data collection was by in-person interview (on average two months post-diagnosis), with active follow-up via mailed or phone questionnaires at 6, 24, and 72 months and outcomes and comorbidities at 12, 24, 48 and 72 months post-enrollment.[54] Passive follow-up is through KP electronic health records as well as linkage to external databases for some characteristics such as mortality through the National Death Index. Interviewer administered questionnaires collected information on demographics, family health history, pregnancy health history, menstrual history, history of breast care screening procedures, smoking history, hormone use, medication history, and vitamin and mineral use. Self-administered questionnaires collected information on diet, physical activity, CAM use, lymphedema, and psychosocial and quality of life measures.[54]

Dietary assessment

Dietary measures were assessed using data from the Pathways food frequency questionnaires. This survey was administered to each participant at baseline, 6, 24 and 72 months and is a 139-item modified version of the Block 2005 food frequency questionnaire (FFQ).[55] Food items for this questionnaire were selected by identifying the top population contributors of each nutrient among whites, African Americans, and Hispanics in the National Health and Nutrition Examination Survey (1999-2002). The food items and additional questions were selected to be representative of a wide range of dietary factors. The questionnaire was modified through inclusion of items to better capture foods that are popular in Hispanic and Asian populations. The FFQ assessed dietary consumption based on two primary questions. The first being “how often, on average, did you eat the food during the past 6 months?”, which could have been answered from 8 possible choices of “never”, “a few times a month”, “once per month”, “2-3 times per month”, “once per week”, “2 times per week”, “3-4 times per week”, “5-6 times per week” and “every day”.[55] The second question asked was “how much do you usually eat of the food?”. In many cases (bacon and breakfast sausage, for example) participants could answer this question by providing the number of “pieces” consumed.

In other cases where the size of the item was not clear (meat loaf or steak, for example) participants were shown pictures of food portions which are equated to ¼, ½ , 1 and 2 cups of food to choose from. Completed questionnaires were sent to NutritionQuest for scanning using a nutrient database developed primarily from the USDA Food and Nutrient Database for Dietary Studies.[55]

Breast cancer outcomes

The primary outcomes in this study included breast cancer recurrence, breast cancer-specific deaths, non-breast cancer deaths and all deaths including deaths due to breast cancer. During active follow-up, participants of the Pathways Study reported new breast or other cancers and conditions. Reports of death and causes of death came from relatives and additional contacts. In addition, electronic medical records were searched on a monthly basis to identify re-initiation of chemotherapy or other evidence of a potential recurrence using codes from the *International Classification of Diseases, Night Revision (ICD-9*), or *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10).*[56-59] These findings were then confirmed by medical record review. Additional death information came from medical records and linkage with the KPNC mortality file, which included data from KPNC sources, the State of California, the Social Security Administration and the National Death Index. Causes of death came from the National Death Index, death certificates, hospital discharge summaries, autopsy or coroner's report, and physician notes. For all analyses, delayed entry was used to identify the beginning of follow-up, whereby Pathways participants began follow-up at the time of completing their baseline questionnaire. Enrolled participants were then followed for outcomes up until their death or the end of the analytic follow-up period, whichever came first.

Covariate selection

Confounders and potential effect modifiers were determined *a priori* for each analysis from an extensive literature review as well as the application of a Directed Acyclic Graph (DAG).[60] A more detailed assessment of covariates was also performed at each stage of analysis and then specifically tailored depending on the research question being considered. The set of covariates under consideration included the following demographic and behavioral factors: age at diagnosis, race-ethnicity, education, marital status, household income, menopausal status, total energy, body mass index (BMI), physical activity and smoking, each of which were ascertained from the participant questionnaires and supplemented by electronic medical records where possible. Clinical factors were also considered, including cancer stage, estrogen receptor status (ER), progesterone receptor status (PR), human epidermal growth factor receptor 2 status (HER2), tumor size, surgery, chemotherapy, radiation therapy and hormone use, which were each obtained from the KPNC Cancer Registry and other electronic clinical and administrative databases.

Statistical Methods

The statistical approaches used to address each of the three aims proposed in this dissertation varied widely and depended on the research question being asked. Here we provide an overview, with additional detail found in the corresponding papers.

For Aim 1, three separate delayed-entry Cox proportional hazard models were used to calculate hazard ratios (HRs) and 95% confidence intervals (CIs) to assess the association between four dietary quality indices measured at baseline and recurrence, breast cancer-specific, non-breast cancer-specific and all-cause mortality. The first model adjusted for age at diagnosis and total energy intake, the second for the variables in the first model as well as race/ethnicity, education level, menopausal status, physical activity, smoking, cancer stage at diagnosis, ER, PR and HER2, and the third for all variables in Model 2 as well as BMI, type of surgery, chemotherapy, radiation, and hormonal therapies. Sub analyses were conducted to examine the independent associations between the individual food components from each index and all-cause mortality, as well as the interaction between each dietary quality index score and each of chemotherapy, radiation, ER, PR and HER2.

For Aim 2, parametric g-computation was used to estimate the 13-year risks of death under hypothetical interventions on three dietary quality indices (the healthy Plant-Based Dietary Index (hPDI), ACS and HEI), physical activity and smoking. For dietary quality and physical activity, four hypothetical longitudinal interventions of increasing intensity were imposed, beginning at baseline and maintained for the entire duration of the 13-year follow-up. Estimates of the 13-year risk for the natural course (no intervention), where also calculated, whereby no changes were made to any of the values of the predicted covariates. Risk ratios and risk differences were calculated by comparing the final population risk associated with each intervention to the estimated risk associated with the natural course. To estimate 95% confidence intervals for the risk ratios and risk differences, non-parametric bootstrapping was performed on 1000 repeated samples.

For Aim 3, hierarchical Cox proportional hazard models were fit with a 2-stage specification to estimate the HRs and CIs associated with the joint intake of several individual food groups on all-cause mortality. The second level model was specified to explain these effects in terms of nutrients considered to be related to survival (*e.g.* carbohydrates, protein, fiber, calcium, iron, lycopene, vitamin C, vitamin D, and others). This approach allowed estimation of the mutually adjusted associations between multiple food items on breast cancer survival, as well as the role of specific nutrients in these foods.

Human subjects and potential risks

This observational study did not involve an intervention, nor did it involve any patient contact. The main risk to the subjects was a breach of confidentiality. However, we followed standard practices to protect confidentiality. All study activities were approved

by both the Institutional Review Boards of KPNC and University of California, Berkeley. Written informed consent was obtained from all study participants.

1 CHAPTER 1. Diet quality and breast cancer survival

1.1 Abstract

Background

Prior research suggests a relationship between overall diet quality and breast cancer survival, though few studies have reported on this topic. We evaluated whether four dietary quality indices consistent with healthy eating recommendations around the time of breast cancer diagnosis were associated with risk of recurrence, cause-specific and all-cause mortality.

Methods

A total of 3,660 women diagnosed with invasive breast cancer were included. Diet was assessed an average of 2.3 (range = 0.7-18.7) months after diagnosis, from which four dietary quality indices were derived: the American Cancer Society guidelines (ACS), the alternate Mediterranean Diet Index (aMED), the Dietary Approaches to Stop Hypertension (DASH) and the 2015 Healthy Eating Index (HEI). Over 40,888 person-years of follow-up, 461 breast cancer recurrences and 655 deaths were ascertained. Cox models were used to estimate hazards ratios (HRs) and 95% confidence intervals (CIs).

Results

Adjusted comparisons between extreme quintiles showed all four dietary quality indices to be inversely associated with all-cause mortality, suggesting a 21%-27% lower risk (ACS HR = 0.73 [95%CI = 0.56-0.95], aMED HR = 0.79 [95%CI = 0.61-1.03], DASH HR = 0.76 [95%CI = 0.58-1.00], HEI HR = 0.77 [95%CI = 0.60-1.01]). Similar patterns were noted for non-breast cancer mortality (ACS HR = 0.69 [95%CI = 0.48-0.98], aMED HR = 0.73 [95%CI = 0.50-1.05], DASH HR = 0.55 [95%CI = 0.38-0.79], HEI HR = 0.67 [95%CI = 0.48-0.94]). None of the dietary quality indices were associated with recurrence or breast cancer-specific mortality.

Conclusion

Food intake patterns concordant with dietary quality indices consistent with recommendations for healthy eating may be beneficial for women with breast cancer.

1.2 Introduction

It is estimated that there are more than 3.5 million breast cancer survivors living in the United States (US).[3] After a breast cancer diagnosis, women are highly motivated to make lifestyle changes and have expressed a desire for more evidence-based information from health-professionals.[34, 35] A growing body of research has evaluated diet and breast cancer survival, focusing mostly on individual foods and nutrients.[43, 61-63]

While providing insight into biological mechanisms, national dietary guidelines are typically presented as dietary patterns, not individual foods.[37]

A dietary pattern index is the composite measure of the quantities and portions of all foods, drinks and nutrients in one's diet as well as the frequency with which they are consumed.[27] Studies have used both data driven and *a priori* dietary quality indices to examine concordance with healthy dietary patterns and breast cancer survival.[23, 24, 64-66] However, methodological issues concerning timing of dietary assessment relative to diagnosis, inconsistent exposure assessment, and absence of secondary outcomes have made comparability challenging.

To our knowledge, only one other study has assessed multiple *a priori* dietary quality indices among a cohort of breast cancer survivors enrolled near the time of their diagnosis.[51] The Shanghai Breast Cancer Survival Study (SBCSS), reported that comparisons of extreme quartiles from both the Chinese Food Pagoda and Dietary Approaches to Stop Hypertension dietary indices were associated with a lower risk of total mortality and breast cancer-specific events.[51]

Our study examined four *a priori* dietary quality indices estimated from dietary data collected soon after breast cancer diagnosis, with breast cancer recurrence, cause-specific and all-cause mortality, among a cohort of breast cancer survivors in the US. The dietary quality indices of interest were an index based on the American Cancer Society nutrition guidelines (ACS),[30] the alternate Mediterranean Diet Index (aMED),[67] an index based on the Dietary Approaches to Stop Hypertension diet (DASH),[68] and the 2015 Healthy Eating Index (HEI).[69]

1.3 Participants and Methods

Study Cohort

The Pathways Study is a prospective cohort of 4,505 female breast cancer survivors diagnosed with breast cancer between the years of 2005 and 2013 from Kaiser Permanente Northern California (KPNC), further details on this study are provided elsewhere.[54] Briefly, diet was assessed an average of 2.3 (range = 0.7-18.7) months after diagnosis. Eligibility criteria included: being female, 21 years or older, KPNC membership, speaking English, Spanish, Cantonese or Mandarin, living within a 65-mile radius of a field interviewer, diagnosis of incident invasive breast cancer, and no prior history of other invasive cancers. The enrollment rate was 40.3% of those eligible (4,505 of 11,174), and participants received an in-person baseline interview administered by field staff.

This study was approved by both the Institutional Review Boards of KPNC and University of California, Berkeley. Written informed consent was obtained from all study participants.

Dietary Assessment

Dietary intake was assessed at baseline with a modified version of the Block 2005 Food Frequency Questionnaire (FFQ).[55] Its 139 food items and additional questions were selected to be representative of a wide range of dietary factors, as well as to capture foods that are popular in Hispanic and Asian populations.

Among the 4,505 women in the cohort, 782 (17.4%) were excluded from this analysis because they did not complete the dietary assessment at baseline. An additional 63 (1.4%) participants were excluded due to estimated daily total energy intake (kcal/d) being less than 400 or greater than 4000. While no statistically significant differences were observed in regard to survival outcomes, excluded participants were less likely to be older, white, educated, post-menopausal, non-smokers and ER-positive as compared to those included in the analysis (**Table S1**). These exclusions brought the final sample size to 3,660.

Four *a priori* dietary quality indices were created to assess concordance with dietary patterns at baseline: ACS, aMED, DASH and HEI (**Table S2**). While ACS was selected due to its direct relevance to cancer specific outcomes, the others were chosen because of their prior demonstrated associations with cancer prevention,[38, 70, 71] and survival.[23, 65, 72] For all four dietary quality indices, a higher score is indicative of a food and nutrient intake that is more concordant with a healthful dietary pattern.

The ACS score ranges from 0 to 9 and has three main components: total fruits and vegetables (which rewards variety for those consuming at least 5 different fruits or vegetables per month), whole grains as a percent of total grains and total red and processed meats. Each component is worth 0 (lowest) to 3 (highest) points and is based on cohort-specific quartiles for that component (red and processed meat score is reversed).[30]

The aMED score is comprised of 9 dietary components, 7 to be encouraged, and 2 to be moderated. The encouraged components include vegetables, legumes, fruits, nuts, whole grains, seafood, and the ratio of monosaturated to saturated fats, while the moderated components are alcohol and red and processed meats. Intakes above the population median for encouraged components received 1 point and all other intakes received 0 points. The red and processed meat component is reversed scored and those consuming alcohol between 5 and 15 grams per day, received 1 point and all others 0 points. Scores for this index range between 0 and 9 points.[32]

The DASH score was calculated by creating 8 dietary components worth 5 points each from population quintiles ranging from 1 (lowest) to 5 (highest). Total scores range from 0 to 40 by combining scores from the favorable components: fruits, vegetables, nuts, grains and low fat dairy; and reverse scored adverse components: sodium, red and processed meats and sugar sweetened beverages.[68]

The Health Eating Index (HEI), designed to align with the 2010-2015 Dietary Guidelines for Americans, is scored from a total of 13 dietary subcomponents.[69] Six of these (total fruits, whole fruits, total vegetables, greens and beans, total protein foods and seafood and plant protein foods) are worth 5 points each and seven others (whole grains, dairy, fatty acids, refined grains, sodium, added sugars and saturated fats) are worth 10 points each, for a total possible score of 100 points. The HEI is the only index in this analysis for which every component is scored on a density basis (per 1000 kcal or percentage of energy), except for fatty acids which is the ratio of unsaturated to saturated fats.[69]

Covariates

Demographic and behavioral factors including age, race/ethnicity, education, menopausal status, smoking status, physical activity, and body mass index (BMI) were collected using the baseline questionnaire at time of entry. Where possible, missing data were supplemented with data obtained from the KPNC electronic health records (EHR) and medical chart review (MCR), except in the case of BMI, where the EHR data took precedence over self-reported values. Diagnostic and clinical data, which included tumor stage, estrogen receptor status (ER), progesterone receptor status (PR), human epidermal growth factor receptor 2 status (HER2), type of surgery, and receipt of chemotherapy, radiation and hormonal therapies, were ascertained from a combination of the KPNC Cancer Registry and other clinical databases.

Outcomes Ascertainment

The primary outcomes for this study include breast cancer recurrence, breast cancer-specific mortality, non-breast cancer-specific mortality and all-cause mortality. Recurrences were ascertained either during follow-up interviews with participants or from monthly algorithmic searches of KPNC electronic databases. All recurrences were confirmed by MCR. Deaths and causes of death were identified during follow up interviews with relatives of participants, and then confirmed by MCR, or from linkages with data from the State of California, the Social Security Administration, and the National Death Index. Over the course of 40,888 person-years, there were 461 (12.6%) recurrences, 324 (8.9%) deaths due to breast cancer, 331 (9.0%) deaths due to causes other than breast cancer and 655 (17.9%) deaths due to any cause.

Statistical Analysis

Spearman correlation coefficients were used to compare the total scores of the dietary quality indices and cohort-specific quintiles of each dietary quality index were calculated. Cox models were used to calculate hazard ratios (HRs) and 95% confidence intervals (CIs) to assess the association between each dietary quality index and recurrence, breast cancer-specific, non-breast cancer-specific and all-cause mortality. In all models, the lowest scoring group was used as the reference group for each index. Models with dietary quality indices expressed as linear splines were considered,[73] but a simple linear term was favored in formal comparisons based on the Bayes Information

Criterion. Trend results corresponding to a 1-unit increase in each index using the non-categorized variable are presented.

Because women entered the cohort after their initial breast cancer diagnosis, they were not considered at risk for a possible outcome before their baseline dietary assessment. Therefore, delayed-entry models were used, and person-time was calculated from the date of completion of the baseline questionnaire to the date of first confirmed breast cancer recurrence or date of death, depending on the analysis. Those participants without an event were censored at the end of the study period, December 31st, 2018.

Three separate models for each index were evaluated: the first (Model 1) was adjusted for age at diagnosis and total energy intake. Because FFQs are known to result in greater variance in estimates of food and nutrient intake than may be biologically plausible, we adjust for total energy in all models in order to diminish the impact of this extraneous variance.[74] The second model (Model 2) was adjusted for the variables in Model 1 and race/ethnicity, education level, menopausal status, physical activity, smoking, cancer stage at diagnosis, ER, PR and HER2. These are variables that, to our knowledge, unambiguously satisfy the criteria for confounding the relationship between diet and survival (**Figure S1**). The third model (Model 3) included all variables in Model 2 and BMI, type of surgery, chemotherapy, radiation, and hormonal therapies. These additional variables were identified as factors that may lie on the causal pathway between diet at breast cancer diagnosis and survival (**Figure S1**).

Sub analyses were conducted to examine the independent associations between the individual food components from each index and all-cause mortality. Separate models were estimated for each food component while adjusting for all variables in Model 2, plus the other components in the index being evaluated. Adjusted models containing interaction terms for each dietary quality index score and each of chemotherapy, radiation, ER, PR and HER2 were also assessed.

All statistical analyses were conducted using SAS 9.4 software (SAS Institute, Cary, NC). Figures were generated using R software.[75]

1.4 Results

Baseline Characteristics

The mean age of participants at diagnosis was 59.7 (range = 24-94) years and women in the highest category of the ACS score as compared to the lowest were more likely to be older, White, more educated and physically active, post-menopausal, non-smokers, have lower BMI and lower reported energy intake at baseline. They were also more likely to be PR-negative and to have received radiation therapy, and less likely to have received chemotherapy (**Table 1**). Differences in participant characteristics across categories of the other dietary quality indices were qualitatively similar to those seen for ACS, except in the case of aMED and DASH, where respondents in the highest category reported higher energy intake as compared to those in the lowest.

Dietary Quality Indices and Study Outcomes

All four dietary quality indices were inversely associated with all-cause mortality when adjusting for age at diagnosis and total energy intake (**Table 2**). Tests for linear trend were statistically significant for ACS and HEI. While the results in Model 2 were somewhat attenuated when compared to Model 1, they were relatively consistent. The direction and magnitude of each dietary quality index in Model 2 were associated with a lower risk of all-cause mortality when comparing high and low score categories (ACS HR = 0.73 [95%CI = 0.56-0.95], aMED HR = 0.79 [95%CI = 0.61-1.03], DASH HR = 0.76 [95%CI = 0.58-1.00], HEI HR = 0.77 [95%CI = 0.60-1.01]). Additionally, accounting for BMI and treatment variables in Model 3 did not notably change the results.

All four dietary quality indices were inversely associated with non-breast cancer-specific mortality when comparing high to low categories and adjusting for the variables in Model 2 (ACS HR = 0.69 [95%CI = 0.48-0.98], aMED HR = 0.73 [95%CI = 0.50-1.05], DASH HR = 0.55 [95%CI = 0.38-0.79], HEI HR = 0.67 [95%CI = 0.48-0.94]) (**Table 3**). The tests for linear trend were statistically significant for all dietary quality indices except for aMED. No associations were observed between the dietary quality indices and breast cancer specific outcomes.

Sub-Analyses

Table 4 presents the associations for each index-specific dietary component and all-cause mortality after adjusting for variables in Model 2 and the other dietary components in each index. Greater intake of whole grains in the case of ACS (HR = 0.91, 95% CI = 0.85–0.99) and nuts in the case of aMED (HR = 0.82, 95% CI = 0.68–0.98) were each associated with a lower risk of all-cause mortality. In the case of HEI and all-cause mortality, a decreased intake of refined grains (HR = 0.95, 95% CI = 0.90–0.99) and sodium (HR = 0.96, 95% CI = 0.92–0.99) were each associated with lower risk and higher intake of total fruit (HR = 1.12, 95% CI = 1.02–1.23) was associated with higher risk. Examination of the interaction between the dietary quality indices and ER on all-cause mortality suggested a stronger association among patients with ER-positive breast cancer when comparing the highest to lowest quartile for each of the dietary quality index scores. For example, for women with ER-positive breast cancer, the hazard ratio comparing highest to lowest ACS categories was 0.68 (95% CI = 0.51-0.91), whereas for women with ER-negative breast cancer, it was 1.05 (95% CI = 0.59-1.89). The p-values for the interaction terms were not statistically significant (**Table 5**). No statistically significant interactions were observed when evaluating the dietary quality indices by each of PR, HER2, chemotherapy and radiation therapy.

1.5 Discussion

In this prospective cohort study of 3,660 breast cancer survivors, participants who reported consuming diets that were more concordant with healthy eating patterns, as measured by four major dietary quality indices, were at lower risk of non-breast cancer-

specific and all-cause mortality. No clear patterns emerged when examining the associations between the dietary quality indices and breast cancer recurrence or breast cancer-specific mortality.

The ACS, aMED, DASH and HEI were each associated with a lower risk for all-cause mortality when comparing the highest to lowest categories and adjusting for all variables in Model 2. These results are consistent with the findings from the SBCSS,[51] in that they found a 34% lower risk of all-cause mortality when comparing extreme quartiles of DASH among breast cancer survivors at 5 years post diagnosis. However, they did not report statistically significant associations in their assessment of HEI.[51] One reason for this may be that HEI was developed to correspond to the recommendations of the US dietary guidelines and these guidelines may not be as directly applicable to dietary patterns in Shanghai.

The ACS, aMED, DASH and HEI were each associated with a lower risk for non-breast cancer-specific mortality, while none were associated with breast cancer-specific outcomes. Our findings are consistent with prior literature,[23, 25, 65] with the exception of SBCSS, which reported an association between DASH and breast cancer-specific events.[51] One explanation for our findings could be related to diet quality playing a more important role in cardiovascular disease (CVD) than on breast cancer prognosis and women surviving breast cancer being at greater risk for CVD as compared to women without breast cancer.[11] One prior study found breast cancer patients to be at higher risk of dying from CVD than from their breast cancer at 10 years after diagnosis,[12] which could be explained by cardiotoxic cancer treatments and exacerbated by associated CVD risk factors.[14, 15]

The correlation coefficients comparing each of the dietary quality indices suggest some overlap in their assessment of diet, however each appears to have some independent characteristics (**Table S3**). To investigate the main drivers of each index, we explored the associations of the individual dietary component scores on all-cause mortality, while adjusting for the other dietary components within each index. We found independent associations with all-cause mortality for whole grains in the case of ACS and nuts in the case of aMED. Interestingly, these findings are consistent with prior research on the impact of these individual food items and breast cancer incidence.[76, 77]

The adjusted interaction between each of the four dietary quality indices and ER on all-cause mortality, suggested a stronger association among patients with ER-positive breast cancer when comparing the highest to lowest quartile of the dietary quality index score. However, the p-values for the interaction terms were not statistically significant. These findings are consistent with one prior study,[65] and may be due to ER-positive survivors generally having a better prognosis as compared to those who are ER-negative, and therefore more likely to die of other causes.[12] For these ER-positive patients, diet quality may have had a stronger impact on causes of death other than breast cancer.

There are several strengths to this study, including drawing from a large population of women newly diagnosed with breast cancer, prospective data collection with a long follow-up period, and comprehensive measures of dietary exposures, outcomes, and covariates. The main limitations of this study are: the use of a single dietary measure at baseline and not addressing dietary changes that could occur after that point, our population being predominantly white and therefore underpowered to examine differences by race and ethnicity, and apart from HEI, the associations of each of the dietary quality indices being restricted to the highest quintile when comparing to the lowest, suggesting that participants in the highest group could have some unidentified behaviors that are not accounted for in the analysis. Finally, it is always possible that participants who chose to enroll in this study were systematically different than those who did not. However, when comparing the enrolled with the unenrolled, the differences in age, race, ethnicity, BMI, and cancer stage were minimal.

In summary, our study found that the ACS, aMED, DASH and HEI were each associated with a decreased risk of non-breast cancer-specific and all-cause mortality. However, the dietary quality indices were not associated with breast cancer-specific outcomes. These findings highlight the importance of an overall healthy dietary pattern for breast cancer survivors.

1.6 Tables

Table 1. Baseline characteristics of the Pathways Study participants across quintiles of ACS (n=3,660)

	Quintiles of ACS-score					
Characteristic	Q1 (n=1,282)	Q2 (n=655)	Q3 (n=586)	Q4 (n=503)	Q5 (n=634)	
Continuous, Mean (SD)						P*
Age at diagnosis (years)	57.6 (11.9)	58.9 (12.2)	60.6 (11.7)	61.9 (11.8)	62.0 (11.0)	<.0001
Physical activity (MET h/wk)‡	45.3 (31.1)	53.9 (34.5)	55.7 (36.5)	56.4 (34.7)	67.3 (41.0)	<.0001
BMI (kg/m^2)	29.9 (7.3)	28.8 (7.1)	28.2 (6.4)	27.4 (5.7)	26.3 (5.2)	<.0001
Energy intake (kcal/d)	1531.5 (612.7)	1499.7 (616.0)	1453.5 (560.8)	1341.9 (510.6)	1410.1 (439.5)	<.0001
Categorical, No. (%)						P†
Race/ethnicity						0.0011
White	817 (63.7)	431 (65.8)	397 (67.7)	371 (73.8)	475 (74.9)	
Black	91 (7.1)	43 (6.6)	43 (7.3)	28 (5.6)	35 (5.5)	
Asian/Pacific Islander	182 (14.2)	93 (14.2)	77 (13.1)	60 (11.9)	63 (9.9)	
Hispanic	163 (12.7)	72 (11.0)	56 (9.6)	35 (7.0)	52 (8.2)	
American Indian/Alaska Native	29 (2.3)	16 (2.4)	13 (2.2)	9 (1.8)	9 (1.4)	
Education						<.0001
High school or less	240 (18.7)	115 (17.6)	77 (13.1)	59 (11.7)	56 (8.8)	
Some college	467 (36.4)	228 (34.8)	197 (33.6)	169 (33.6)	184 (29.0)	
College graduate	375 (29.3)	178 (27.2)	155 (26.5)	145 (28.8)	171 (27.0)	
Post graduate	199 (15.5)	133 (20.3)	157 (26.8)	130 (25.8)	223 (35.2)	
Unknown	1 (0.1)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	
Menopausal status						<.0001
Premenopausal	453 (35.3)	214 (32.7)	144 (24.6)	114 (22.7)	135 (21.3)	
Postmenopausal	829 (64.7)	441 (67.3)	442 (75.4)	389 (77.3)	499 (78.7)	
Smoking status						<.0001
Never	699 (54.5)	385 (58.8)	361 (61.6)	301 (59.8)	346 (54.6)	
Former	490 (38.2)	247 (37.7)	206 (35.2)	192 (38.2)	273 (43.1)	
Current	92 (7.2)	22 (3.4)	17 (2.9)	9 (1.8)	14 (2.2)	
Unknown	1 (0.1)	1 (0.2)	2 (0.3)	1 (0.2)	1 (0.2)	
Cancer stage						0.2212
I	679 (53.0)	337 (51.5)	329 (56.1)	300 (59.6)	363 (57.3)	
II	447 (34.9)	243 (37.1)	197 (33.6)	155 (30.8)	208 (32.8)	
III	133 (10.4)	66 (10.1)	52 (8.9)	44 (8.7)	51 (8.0)	
IV	23 (1.8)	9 (1.4)	8 (1.4)	4 (0.8)	12 (1.9)	
ER status						0.8934
Positive	1079 (84.2)	548 (83.7)	491 (83.8)	428 (85.1)	526 (83.0)	
Negative	201 (15.7)	107 (16.3)	95 (16.2)	75 (14.9)	108 (17.0)	
Unknown	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
PR status						0.0170
Positive	837 (65.3)	433 (66.1)	370 (63.1)	335 (66.6)	372 (58.7)	
Negative	441 (34.4)	221 (33.7)	216 (36.9)	168 (33.4)	262 (41.3)	
Unknown	4 (0.3)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	
HER2 status						0.3601
Positive	166 (12.9)	85 (13.0)	88 (15.0)	53 (10.5)	80 (12.6)	
Negative	1063 (82.9)	547 (83.5)	484 (82.6)	429 (85.3)	522 (82.3)	
Unknown	53 (4.1)	23 (3.5)	14 (2.4)	21 (4.2)	32 (5.0)	
Surgery type						0.1883
Lumpectomy	755 (58.9)	364 (55.6)	361 (61.6)	321 (63.8)	379 (59.8)	
Mastectomy	484 (37.8)	265 (40.5)	207 (35.3)	166 (33.0)	239 (37.7)	
None	43 (3.4)	26 (4.0)	18 (3.1)	15 (3.0)	15 (2.4)	
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.2)	
Chemotherapy						<.0001
No	616 (48.0)	341 (52.1)	322 (54.9)	291 (57.9)	368 (58.0)	
Yes	662 (51.6)	311 (47.5)	263 (44.9)	211 (41.9)	264 (41.6)	
Unknown	4 (0.3)	3 (0.5)	1 (0.2)	1 (0.2)	2 (0.3)	
Radiation therapy						0.0087
No	752 (58.7)	372 (56.8)	326 (55.6)	252 (50.1)	334 (52.7)	
Yes	530 (41.3)	283 (43.2)	260 (44.4)	251 (49.9)	300 (47.3)	
Hormonal therapy						0.2191
No	308 (24.0)	147 (22.4)	154 (26.3)	119 (23.7)	174 (27.4)	
Yes	963 (75.1)	504 (76.9)	427 (72.9)	382 (75.9)	457 (72.1)	
Unknown	11 (0.9)	4 (0.6)	5 (0.9)	2 (0.4)	3 (0.5)	

Abbreviations: Q (quintile), SD (standard deviation), MET (metabolic equivalent of task) ACS (American Cancer Society nutrition guidelines score), BMI (body mass index), ER (estrogen receptor), PR (progesterone receptor), HER2 (human epidermal growth factor receptor 2).

*The p-value is from analysis of variance.

†The p-value is from the Pearson X^2 test.

‡There were 4 participants with unknown physical activity.

Table 2. Hazards ratios and 95% confidence intervals of quintiles of dietary quality indices and all-cause mortality

Dietary Index	Range	Model 1* (n=3,660)				Model 2† (n=3,506)				Model 3‡ (n=3,471)			
		Events	PT, y	HR (95% CI)	HR (P_{trend})	Events	PT, y	HR (95% CI)	HR (P_{trend})	Events	PT, y	HR (95% CI)	HR (P_{trend})
ACS					0.93 (<0.001)				0.96 (0.04)				0.96 (0.07)
Q1	0-3	237	11792.3	ref		232	11271.5	ref		226	11148.5	ref	
Q2	4	128	5954.7	1.00 (0.81, 1.25)		124	5746.4	1.00 (0.80, 1.25)		122	5683.9	1.02 (0.81, 1.27)	
Q3	5	99	5363.6	0.80 (0.64, 1.02)		98	5211.6	0.90 (0.71, 1.15)		95	5172.1	0.90 (0.71, 1.16)	
Q4	6	96	4639.8	0.84 (0.66, 1.07)		93	4425.8	1.01 (0.78, 1.29)		92	4409.8	1.03 (0.80, 1.33)	
Q5	7-9	95	5817.2	0.67 (0.53, 0.85)		87	5516.2	0.73 (0.56, 0.95)		86	5466.2	0.77 (0.59, 1.01)	
aMED					0.91 (<.001)				0.96 (0.10)				0.97 (0.27)
Q1	0-2	177	7349.3	ref		171	7011	ref		170	6908.0	ref	
Q2	3	116	5628.9	0.84 (0.67, 1.07)		114	5425.2	1.07 (0.84, 1.36)		111	5380.1	1.08 (0.85, 1.39)	
Q3	4	119	5983.4	0.83 (0.65, 1.05)		114	5743.0	1.02 (0.79, 1.30)		112	5694.5	1.05 (0.82, 1.35)	
Q4	5	117	5584.4	0.87 (0.68, 1.11)		113	5401.7	1.13 (0.87, 1.46)		107	5346.2	1.17 (0.90, 1.53)	
Q5	6-9	126	9021.7	0.56 (0.43, 0.71)		122	8590.6	0.79 (0.61, 1.03)		121	8551.5	0.87 (0.66, 1.14)	
DASH					0.96 (<.001)				0.98 (0.049)				0.98 (0.10)
Q1	10-20	131	5751.9	ref		127	5536.3	ref		124	5490.4	ref	
Q2	21-23	187	9440.6	0.82 (0.66, 1.03)		184	9062.5	0.93 (0.74, 1.18)		180	8965.2	0.94 (0.75, 1.19)	
Q3	24-25	119	5772.7	0.78 (0.61, 1.00)		117	5541.5	0.96 (0.74, 1.24)		114	5450.4	1.00 (0.77, 1.31)	
Q4	26-27	95	4455.1	0.78 (0.60, 1.02)		88	4240.9	0.99 (0.74, 1.31)		87	4225.9	1.02 (0.77, 1.36)	
Q5	28-37	123	8147.4	0.53 (0.42, 0.68)		118	7790.2	0.76 (0.58, 1.00)		116	7748.6	0.80 (0.61, 1.05)	
HEI					0.98 (<.001)				0.99 (0.04)				0.99 (0.12)
Q1	42.1-63.2	131	6104.4	ref		127	5841.8	ref		124	5741.9	ref	
Q2	63.3-69.2	131	6019.3	1.01 (0.79, 1.28)		130	5859.8	0.85 (0.66, 1.10)		127	5815.7	0.85 (0.66, 1.10)	
Q3	69.3-73.9	131	6494.9	0.88 (0.69, 1.12)		128	6241.0	0.94 (0.73, 1.20)		124	6196.6	0.94 (0.73, 1.22)	
Q4	74.0-79.9	131	7411.4	0.74 (0.58, 0.94)		125	7068.9	0.77 (0.60, 1.00)		124	7010.1	0.82 (0.63, 1.06)	
Q5	80.0-95.4	131	7537.7	0.64 (0.51, 0.82)		124	7160.0	0.77 (0.60, 1.01)		122	7116.0	0.81 (0.62, 1.06)	

*Adjusted for age at diagnosis and total energy. Abbreviations: Q (quintile), ACS (American Cancer Society nutrition guidelines score), aMED (alternate Mediterranean Diet score), DASH (Dietary Approaches to Stop Hypertension score), HEI (Healthy Eating Index score), PT (person-time), HR (hazard ratio), CI (confidence interval), ref (referent).
†Adjusted for age at diagnosis, total energy, race/ethnicity, education, menopausal status, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2.
‡Adjusted for all variables in model 2, plus body mass index, surgery type, chemotherapy, radiation and hormonal therapies.

Table 3. Hazards ratios and 95% confidence intervals of quintiles of dietary quality indices and recurrence, breast cancer-specific death and non-breast cancer-specific death (No.=3,506) *

Dietary Index	Range	Recurrence				Breast Cancer-Specific Mortality				Non-Breast Cancer-Specific Mortality			
		Events	PT (yrs)	HR (95% CI)	HR (P_{trend})	Events	PT (yrs)	HR (95% CI)	HR (P_{trend})	Events	PT (yrs)	HR (95% CI)	HR (P_{trend})
ACS					1.01 (0.55)				0.97 (0.29)				0.94 (0.03)
Q1	0-3	153	10731.6	ref		114	11271.5	ref		118	11271.5	ref	
Q2	4	80	5522.2	1.01 (0.77, 1.33)		61	5746.4	1.14 (0.83, 1.56)		63	5746.4	0.98 (0.72, 1.34)	
Q3	5	73	5015.5	1.05 (0.79, 1.39)		54	5211.6	1.09 (0.78, 1.52)		44	5211.6	0.72 (0.51, 1.02)	
Q4	6	60	4237.6	1.10 (0.81, 1.50)		41	4425.8	1.13 (0.78, 1.63)		52	4425.8	0.89 (0.63, 1.25)	
Q5	7-9	83	5179.4	1.19 (0.89, 1.57)		42	5516.2	0.75 (0.52, 1.09)		45	5516.2	0.69 (0.48, 0.98)	
aMED					1.02 (0.46)				0.96 (0.25)				0.94 (0.08)
Q1	0-2	97	6688.1	ref		75	7011.0	ref		96	7011.0	ref	
Q2	3	71	5188.8	1.06 (0.78, 1.45)		56	5425.2	1.25 (0.88, 1.79)		58	5425.2	0.83 (0.59, 1.15)	
Q3	4	82	5533.9	1.17 (0.86, 1.59)		60	5743.0	1.26 (0.88, 1.80)		54	5743.0	0.80 (0.57, 1.14)	
Q4	5	89	5081.4	1.35 (0.99, 1.84)		62	5401.7	1.31 (0.91, 1.90)		51	5401.7	0.91 (0.63, 1.31)	
Q5	6-9	110	8194.0	1.08 (0.79, 1.47)		59	8590.6	0.79 (0.54, 1.16)		63	8590.6	0.73 (0.50, 1.05)	
DASH					1.00 (0.95)				0.99 (0.68)				0.96 (0.002)
Q1	10-20	75	5276.1	ref		55	5536.3	ref		72	5536.3	ref	
Q2	21-23	140	8609.2	1.13 (0.85, 1.51)		102	9062.5	1.16 (0.83, 1.62)		82	9062.5	0.73 (0.53, 1.00)	
Q3	24-25	74	5324.4	0.98 (0.71, 1.37)		56	5541.5	1.03 (0.70, 1.52)		61	5541.5	0.80 (0.56, 1.14)	
Q4	26-27	64	4043.8	1.13 (0.80, 1.60)		43	4240.9	1.15 (0.76, 1.74)		45	4240.9	0.76 (0.51, 1.12)	
Q5	28-37	96	7432.8	1.02 (0.73, 1.41)		56	7790.2	0.93 (0.63, 1.39)		62	7790.2	0.55 (0.38, 0.79)	
HEI					1.01 (0.30)				0.99 (0.44)				0.98 (0.006)
Q1	42.1-63.2	60	5630.1	ref		49	5841.8	ref		78	5841.8	ref	
Q2	63.3-69.2	101	5577.1	1.39 (1.01, 1.93)		69	5859.8	1.02 (0.70, 1.48)		61	5859.8	0.77 (0.55, 1.08)	
Q3	69.3-73.9	96	5948.4	1.39 (1.00, 1.93)		72	6241.0	1.30 (0.90, 1.88)		56	6241.0	0.67 (0.47, 0.95)	
Q4	74.0-79.9	105	6720.9	1.33 (0.96, 1.85)		72	7068.9	1.03 (0.71, 1.50)		53	7068.9	0.55 (0.39, 0.79)	
Q5	80.0-95.4	87	6809.8	1.24 (0.88, 1.75)		50	7160.0	0.84 (0.56, 1.27)		74	7160.0	0.67 (0.48, 0.94)	

*All models adjusted for age at diagnosis, total energy, race/ethnicity, education, menopausal status, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2. Abbreviations: Q (quintile), ACS (American Cancer Society nutrition guidelines score), aMED (alternate Mediterranean Diet score), DASH (Dietary Approaches to Stop Hypertension score), HEI (Healthy Eating Index score), PT (person-time), HR (hazard ratio), CI (confidence interval), ref (referent

Table 4. Index-specific hazard ratios and 95% confidence intervals on all-cause mortality, for each component of the dietary quality index (n=3,506) *†

Dietary Component	ACS HR (95% CI)	aMED HR (95% CI)	DASH HR (95% CI)	HEI HR (95% CI)
Fruits				
Total fruit	—	1.02 (0.86, 1.21)	1.03 (0.96, 1.10)	1.12 (1.02, 1.23)
Whole fruits		—	—	0.94 (0.85, 1.04)
Vegetables				
Total vegetables	—	0.97 (0.81, 1.16)	1.01 (0.94, 1.08)	1.06 (0.92, 1.23)
Greens and beans	—	—	—	0.92 (0.83, 1.02)
Total fruits and vegetables	0.98 (0.91, 1.06)	—	—	—
Grains				
Whole grains	0.91 (0.85, 0.99)	0.92 (0.77, 1.09)	0.96 (0.90, 1.02)	0.98 (0.96, 1.01)
Refined grains	—	—	—	0.95 (0.90, 0.99)
Diary				
Total dairy	—	—	—	1.00 (0.96, 1.03)
Low-fat dairy	—	—	0.95 (0.89, 1.00)	—
Protein foods				
Total protein foods	—	—	—	1.02 (0.91, 1.14)
Seafood and plant proteins	—	—	—	1.01 (0.92, 1.11)
Red and processed meats	0.99 (0.91, 1.08)	0.96 (0.80, 1.15)	1.02 (0.95, 1.10)	—
Fish	—	1.08 (0.91, 1.28)	—	—
Legumes	—	0.95 (0.80, 1.14)	—	—
Nuts	—	0.82 (0.68, 0.98)	—	—
Nuts and legumes	—	—	0.93 (0.87, 1.00)	—
Fat				
Unsaturated fats	—	1.01 (0.85, 1.19)	—	1.02 (0.98, 1.06)
Saturated fats	—		—	0.98 (0.94, 1.02)
Sodium	—	—	0.93 (0.82, 1.04)	0.96 (0.92, 0.99)
Sugar				
Added sugar	—	—	—	0.96 (0.92, 1.01)
Sweetened beverages	—	—	0.98 (0.92, 1.04)	—
Alcohol	—	0.99 (0.79, 1.25)	—	—

*All models adjusted for age at diagnosis, total energy, race/ethnicity, education, menopausal status, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2 and all other components within the index. Abbreviations: ACS (American Cancer Society nutrition guidelines score), DASH (Dietary Approaches to Stop Hypertension score), aMED (alternate Mediterranean Diet score), HEI (Healthy Eating Index score), HR (hazard ratio), CI (confidence interval).

†All hazard ratios calculated as a 1-unit change within the dietary quality index component-specific score.

Table 5. Hazards ratios and 95% confidence intervals of quintiles of dietary quality indices and all-cause mortality stratified by ER status

Dietary Index	Range	ER-Positive (n=2,942)				ER-Negative (n=564)				$P_{interaction}$
		Events	PT, y	HR (95% CI)	HR (P_{trend})	Events	PT, y	HR (95% CI)	HR (P_{trend})	
ACS					0.94 (0.01)				1.02 (0.63)	0.24
Q1	0-3	194	9554.9	ref		38	1716.6	ref		
Q2	4	90	4892.3	0.87 (0.68, 1.13)		34	854.1	1.74 (1.07, 2.83)		
Q3	5	77	4402.2	0.83 (0.64, 1.09)		21	809.5	1.29 (0.75, 2.22)		
Q4	6	74	3804.1	0.90 (0.68, 1.19)		19	621.7	1.55 (0.85, 2.83)		
Q5	7-9	67	4626.8	0.68 (0.51, 0.91)		20	889.4	1.05 (0.59, 1.89)		
aMED					0.95 (0.08)				0.98 (0.72)	0.17
Q1	0-2	141	5727.4	ref		30	1283.6	ref		
Q2	3	85	4703.4	0.91 (0.69, 1.20)		29	721.7	1.55 (0.91, 2.63)		
Q3	4	96	4863.2	0.99 (0.75, 1.30)		18	879.8	1.04 (0.56, 1.92)		
Q4	5	84	4618.7	0.95 (0.71, 1.26)		29	783	1.92 (1.07, 3.44)		
Q5	6-9	96	7367.6	0.75 (0.56, 1.01)		26	1223	0.92 (0.49, 1.71)		
DASH					0.98 (0.02)				1.01 (0.55)	0.08
Q1	10-20	107	4615.9	ref		20	920.4	ref		
Q2	21-23	140	7575.2	0.89 (0.69, 1.15)		44	1487.3	1.33 (0.77, 2.32)		
Q3	24-25	96	4789	0.89 (0.67, 1.19)		21	752.5	1.36 (0.71, 2.60)		
Q4	26-27	65	3577.5	0.91 (0.66, 1.26)		23	663.4	1.60 (0.84, 3.04)		
Q5	28-37	94	6722.7	0.70 (0.52, 0.95)		24	1067.5	1.25 (0.64, 2.43)		
HEI					0.99 (0.03)				1.00 (0.99)	0.32
Q1	42.1-63.2	104	5044.3	ref		23	797.6	ref		
Q2	63.3-69.2	101	4786.9	0.91 (0.69, 1.21)		29	1072.9	0.71 (0.40, 1.24)		
Q3	69.3-73.9	102	5297.3	1.00 (0.76, 1.32)		26	943.7	0.85 (0.48, 1.52)		
Q4	74.0-79.9	90	5966.2	0.72 (0.54, 0.97)		35	1102.7	1.06 (0.61, 1.83)		
Q5	80.0-95.4	105	6185.6	0.80 (0.60, 1.06)		19	974.3	0.73 (0.38, 1.40)		

*All models adjusted for age at diagnosis, total energy, race/ethnicity, education, menopausal status, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2. Abbreviations: Q (quintile), ACS (American Cancer Society nutrition guidelines score), aMED (alternate Mediterranean Diet score), DASH (Dietary Approaches to Stop Hypertension score), HEI (Healthy Eating Index score), PT (person-time), HR (hazard ratio), CI (confidence interval), ref (referent).

1.7 Supplementary Tables and Figures

Table S1. Baseline characteristics and survival outcomes comparing excluded participants with included participants from the Pathways Study (n=4,505)

Characteristic	Excluded (n=3,660)	Included (n=845)	P
Continuous, Mean (SD)[a]			
Age at diagnosis (years)	56.7 (12.6)	59.7 (11.9)	<.001
Physical activity (MET h/wk)[b]	54.4 (44.4)	53.8 (35.8)	0.68
Categorical, No. (%)[c]			
Race/ethnicity			<.001
White	460 (54.4)	2491 (68.1)	
Black	111 (13.1)	240 (6.6)	
Asian/Pacific Islander	124 (14.7)	475 (13.0)	
Hispanic	134 (15.9)	378 (10.3)	
American Indian/Alaska Native	16 (1.9)	76 (2.1)	
Education			<.001
High school or less	160 (18.9)	547 (14.9)	
Some college	323 (38.2)	1245 (34.0)	
College graduate	217 (25.7)	1024 (28.0)	
Postgraduate	139 (16.4)	842 (23.0)	
Unknown	6 (0.7)	2 (0.1)	
Menopausal status			0.001
Premenopausal	292 (34.6)	1060 (29.0)	
Postmenopausal	553 (65.4)	2600 (71.0)	
Smoking status			<.001
Never	446 (52.8)	2092 (57.2)	
Former	315 (37.3)	1408 (38.5)	
Current	75 (8.9)	154 (4.2)	
Unknown	9 (1.1)	6 (0.2)	
Cancer stage			0.09
I	426 (50.4)	2008 (54.9)	
II	308 (36.4)	1250 (34.2)	
III	94 (11.1)	346 (9.5)	
IV	17 (2.0)	56 (1.5)	
ER status			0.01
Positive	680 (80.5)	3072 (83.9)	
Negative	165 (19.5)	586 (16.0)	
Unknown	0 (0.0)	2 (0.1)	
PR status			0.08
Positive	515 (60.9)	2347 (64.1)	
Negative	330 (39.1)	1308 (35.7)	
Unknown	0 (0.0)	5 (0.1)	
HER2 status			0.43
Positive	116 (13.7)	472 (12.9)	
Negative	685 (81.1)	3045 (83.2)	
Unknown	44 (5.2)	143 (3.9)	
Recurrence			0.74
No	735 (87.0)	3199 (87.4)	
Yes	110 (13.0)	461 (12.6)	
Breast Cancer-Specific Mortality			0.57
No	765 (90.5)	3336 (91.1)	
Yes	80 (9.5)	324 (8.9)	
Non-Breast Cancer-Specific Mortality			0.78
No	766 (90.7)	3329 (91.0)	
Yes	79 (9.3)	331 (9.0)	
All-Cause Mortality	. (.)	. (.)	0.53
No	686 (81.2)	3005 (82.1)	
Yes	159 (18.8)	655 (17.9)	

[a]Analysis of variance. Abbreviations: FFQ (food frequency questionnaire), SD (standard deviation), MET (metabolic equivalent of task), ER (estrogen receptor), PR (progesterone receptor), HER2 (human epidermal growth factor receptor 2).
[b]There were 42 participants who did not complete the FFQ and 4 participants who did complete the FFQ with unknown physical activity.
[c]Pearson X^2 test.

Table S2. Scoring methods and optimal quantities for each component of ACS, DASH, aMED and HEI

	ACS (total = 9 points)	aMED (total = 9 points)	DASH (total = 40 points)	HEI (total = 100 points)
Fruits				
Total Fruit		≥ Median	= Highest quartile	≥ 0.8 cups/1,000 kcal
Whole Fruits				≥ 0.4 cups/1,000 kcal
Vegetables				
Total Vegetables*		≥ Median	= Highest quartile	≥ 1.1 cups/1,000 kcal
Greens and Beans				≥ 0.2 cups/1,000 kcal
Total Fruits and Vegetables	= Highest tertile†			
Grains				
Whole Grains	= Highest tertile‡	≥ Median	= Highest quartile	> 1.5 ounces/1,000 kcal
Refined Grains				≤ 1.8 ounces/1,000 kcal
Diary				
Total Dairy				≥ 1.3 cups/1,000 kcal
Low-fat Dairy			= Highest quartile	
Protein Foods				
Total Protein Foods				≥ 2.5 ounces/1,000 kcal
Seafood and Plant Proteins				≥ 0.8 ounces/1,000 kcal
Red and Processed Meats	= Lowest tertile	< Median	= Lowest quartile	
Fish		≥ Median		
Legumes		≥ Median		
Nuts		≥ Median		
Nuts and Legumes			= Highest quartile	
Fat				
Unsaturated Fats		MUFAs:SFAs ≥ median		PUFAs + MUFAs:SFAs ≥ 2
Saturated Fats				≤ 8% of energy
Sodium			= Lowest quartile	≤ 1.1 grams/1,000 kcal
Sugar				
Added Sugar				≤ 6.5% of energy
Sweetened beverages			= Lowest quartile	
Alcohol		5-15 grams/d		

Abbreviations: ACS (American Cancer Society nutrition guidelines score), aMED (Alternate Mediterranean Diet score), DASH (Dietary Approaches to Stop Hypertension score), HEI (2015 Healthy Eating Index score), MUFA (monounsaturated fat), SFA (saturated fat), PUFA (polyunsaturated fat).

*aMED excludes potatoes; DASH excludes potatoes and legumes.

†Excludes fruit juices and potatoes. Includes partial variety score for consumption of ≥5 different fruits or vegetables per month.

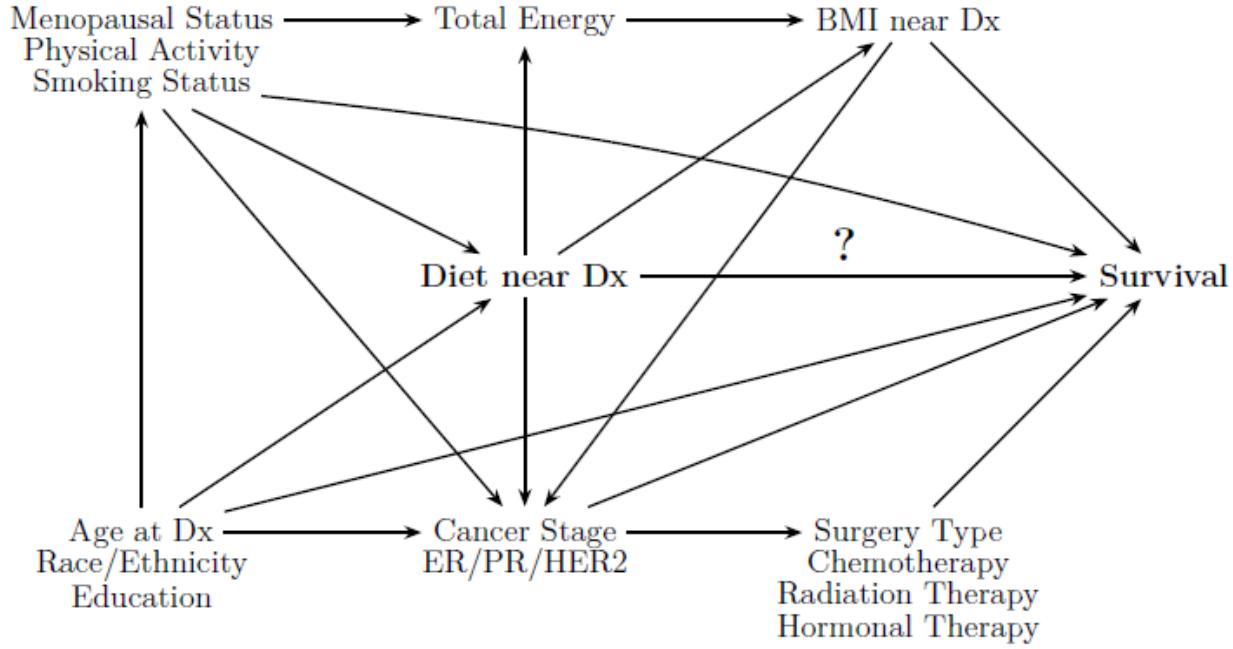
‡Whole grains calculated as a percent of total grains.

Table S3. Spearman correlation coefficients among total summary scores for ACS, aMED, DASH and HEI (n=3,660) [a]

	ACS	aMED	DASH	HEI
ACS	1.00	0.50	0.69	0.65
aMED		1.00	0.66	0.59
DASH			1.00	0.74
HEI				1.00

[a]All p < 0.001. Abbreviations: ACS (American Cancer Society nutrition guidelines score), aMED (alternate Mediterranean Diet score), DASH (Dietary Approaches to Stop Hypertension score), HEI (Healthy Eating Index score).

Figure S1. Causal diagram of the relationship between diet near diagnosis and breast cancer survival [a]



[a]Related variables have been grouped together to simplify presentation.

2 CHAPTER 2. Hypothetical interventions on diet quality and lifestyle factors to improve breast cancer survival

2.1 Abstract

Background

Cancer survivors are motivated to make lifestyle changes, yet evidence to inform recommendations is limited. We estimated the impact of hypothetical diet and lifestyle interventions on breast cancer survival.

Methods

The Pathways Study is a prospective cohort study of women diagnosed with breast cancer in the Kaiser Permanente Northern California health system, with women enrolled soon after diagnosis between 2006-2013; a total of 3,660 women were included in these analyses. Using data from questionnaires completed shortly after diagnosis, and at 6, 24, and 72-month follow-ups, we estimated five behavioral risk factors: moderate to vigorous physical activity (PA, hours per week); smoking (SM, pack-years); and three *a priori* diet quality indices, healthful plant-based diet index (hPDI), American Cancer Society nutrition guidelines for cancer prevention (ACS), and the 2015 Healthy Eating Index (HEI). Follow-up was through December 2018, with 655 (17.9%) deaths. Using parametric g-computation, we estimated the 13-year risk ratios (RR) and 95% confidence intervals (CI) for all-cause mortality under hypothetical interventions on diet quality, physical activity and smoking compared to the natural course (no intervention).

Results

The RRs suggested a range of a 5 to 34% decreased risk of all-cause mortality when all participants consumed and maintained a dietary pattern with an hPDI score of at least 60 or 90 (the maximal score attainable), respectively, as compared to their natural course [hPDI≥60 RR=0.95, 95% CI 0.85-1.02; hPDI=90 RR=0.66, 95% CI 0.40-1.08]}. The RRs suggested a range of a 4 to 11% decreased risk of all-cause mortality when all participants consumed and maintained a dietary pattern with an ACS score of at least 6 or 9 (the maximal score attainable), respectively, as compared to their natural course [ACS≥6 RR=0.96, 95% CI 0.86-1.03; ACS=9 RR=0.89, 95% CI 0.74-1.05]. The RRs suggested a range of a 0 to 6% decreased risk of all-cause mortality when all participants consumed and maintained a dietary pattern with an HEI score of at least 70 or 100 (the maximal score attainable), respectively, as compared to their natural course [HEI≥70 RR=1.00, 95% CI 0.96-1.02; HEI=100 RR=0.94, 95% CI 0.80-1.13]. All three dietary estimates (RRs for hPDI, ACS and HEI) each lacked precision and had confidence intervals that contained the null value.

For moderate to vigorous physical activity, the RRs suggested a range of an 8 to 39% decreased risk of all-cause mortality when all participants engaged and maintained 5 or

20 hours per week of moderate to vigorous activity, respectively, as compared to their natural course [PA≥5 RR=0.92, 95% CI 0.88-0.9; PA≥20 RR=0.61, 95% CI 0.45-0.86].

For smoking, the RR suggested an 8% reduction in risk of all-cause mortality if all smokers stopped smoking and all non-smokers never started as compared to their smoking status in the natural course (SM RR=0.92, 95% CI 0.88-0.96).

In the hypothetical case where all three behaviors had been achieved by cohort members, the RR suggested a range of a 19 to 64% reduction in risk of all-cause mortality when all participants consumed and maintained a dietary pattern with an hPDI score of at least 60 or 90 in conjunction with all participants engaging and maintaining 5 or 20 hours per week of moderate to physical activity and all smokers quit smoking and non-smokers never start, as compared to their natural course [hPDI≥6 + PA≥5 + No SM RR=0.81, 95% CI 0.73-0.90; hPDI≥9 + PA≥20 + No SM RR=0.36, 95% CI 0.21-0.69]. Similar but slightly attenuated impacts on all-cause mortality were estimated when considering ACS and HEI jointly with PA and SM.

Conclusion

Interventions that improve diet quality, increase physical activity, or reduce smoking may improve survival among breast cancer survivors. We estimate that almost two-thirds of deaths could be delayed by joint adoption of these behaviors.

2.2 Introduction

Excluding skin-cancers, breast cancer is the most commonly-diagnosed cancer among woman globally.[78] Due to improved screening and more effective treatments, women with breast cancer are living longer than ever before, and there are currently over 3.8 million breast cancer survivors living in the United States (US).[3] These women are highly motivated to make lifestyle changes, and have expressed a desire for more evidence-based information from health professionals.[34, 35, 79] However, there is limited evidence to help inform recommendations, and therefore women with breast cancer are typically offered guidance based on prevention, not survival.[4]

Associations of diet quality, physical activity, and smoking with improved breast cancer survival have each been independently shown,[25, 80-82] yet their combined effects have been relatively unexplored.[4] Dietary patterns (a composite measure of the quantities and portions of all foods, drinks and nutrients in one's diet as well as the frequency with which they are consumed)[27] concordant with healthy eating have been previously demonstrated to be inversely associated with mortality among breast cancer survivors.[25, 51, 80, 83] Physical activity both before[81, 84] and after[85, 86] a breast cancer diagnosis has also been shown to be related to longer survival. Relatively few studies have examined smoking at or around the time of breast diagnosis and its impact on survival.[82, 87] Overall, the existing evidence comes from a limited number of observational studies that have varied in their assessment of lifestyle factors in the post-diagnosis period.

Although evidence to support lifestyle recommendations in cancer patients and their impact on survival could be evaluated in a randomized controlled trial (RCT) setting, there are few such studies due to the perceived expense, possible non-compliance, and long-term follow-up required. As an alternative, the parametric g-formula can be used with observational data to estimate and compare risks under hypothetical interventions on either individual or combined risk factors.[88] Furthermore, this approach facilitates appropriate adjustment for time-varying confounders affected by prior exposures that might otherwise lead to bias when using standard statistical methods.[89]

This study is among the first to evaluate rigorously the effect of lifestyle factors on breast cancer survival using a causal inference approach. In the Pathways Study, a prospective cohort study of women with breast cancer, we applied the parametric g-formula to estimate the risk of all-cause mortality under several hypothetical interventions on dietary quality, physical activity and smoking cessation at the time of breast cancer diagnosis.

2.3 Participants and Methods

Study Cohort

A total of 4,505 female breast cancer survivors diagnosed with invasive breast cancer between the years of 2005 and 2013 at Kaiser Permanente Northern California (KPNC), were enrolled in the Pathways Study, a longitudinal prospective cohort. Further details on this study are provided elsewhere.[54] Briefly, enrollment occurred an average of 2.3 (range:0.7-18.7) months after diagnosis and eligibility criteria included: female, 21 years or older, KPNC member, English, Spanish, Cantonese or Mandarin speaker, living within a 65-mile radius of a field interviewer, diagnosis of incident invasive breast cancer, and no prior history of other invasive cancers. Participants underwent an in-person baseline interview administered by field staff, as well as follow-up interviews by mail, by phone, on the web, or in person.

This study was approved by the Institutional Review Boards of KPNC and University of California, Berkeley. Written informed consent was obtained from all study participants.

Dietary Assessment

Dietary intake was assessed at baseline and at approximately 6, 24 and 72 months following baseline, with a modified version of the Block 2005 Food Frequency Questionnaire (FFQ).[55] The FFQ asked questions regarding consumption of foods and beverages in the 6 months preceding each assessment time point.. The FFQ included a list of 139 food items and additional questions about aspects of food choices, with food items selected to be representative of a wide range of dietary factors, as well as to capture foods that are popular in Hispanic and Asian populations.

Diet Quality Indices

Three *a priori* diet quality indices were created to assess concordance with healthful dietary patterns at baseline and each subsequent follow-up: a healthful plant-based diet index (hPDI),[31] an indices based on the American Cancer Society nutrition guidelines for cancer prevention (ACS),[30] and the 2015 Healthy Eating Index (HEI).[69] (Supplemental Table 2). For all three diet quality indices, a higher score is indicative of a food and nutrient intake that is more concordant with a healthful dietary pattern.

hPDI score

The hPDI score ranges from 18 to 90 points and is created from the sum of 18 dietary components, each of which combine foods that are similar in nutrient and culinary characteristics.[31] Greater consumption of 7 of these (whole grains, fruits, vegetables, nuts, legumes, vegetable oils and teas and coffees) result in higher hPDI scores, while for the other 11 (fruit juices, refined grains, potatoes, sugar sweetened beverages, sweets and desserts, animal fats, dairy, eggs, fish or seafood, meats and miscellaneous animal-based foods), greater consumption results in lower hPDI scores. Each component is worth up to 5 points, with follow-up specific quintiles of intake resulting in the points allocated for a given dietary component.[31]

ACS score

The ACS score ranges from 0 to 9 and has three main components: total fruits and vegetables (which rewards variety for those consuming at least 5 different fruits or vegetables per month), whole grains as a percent of total grains, and total red and processed meats. Each component is worth 0 (lowest) to 3 (highest) points and is based on cohort-specific quartiles for that component (red and processed meat score is reversed).[30]

2015 HEI Score

The Health Eating Index (HEI), designed to align with the 2010-2015 Dietary Guidelines for Americans, is scored from a total of 13 dietary subcomponents.[69] Six of these (total fruits, whole fruits, total vegetables, greens and beans, total protein foods and seafood and plant protein foods) account for up to 5 points each, and seven others (whole grains, dairy, fatty acids, refined grains, sodium, added sugars and saturated fats) are worth up to 10 points each, for a total possible score of 100 points. The HEI is the only index in this analysis for which every component is scored on a density basis (per 1000 kcal or percentage of energy), except for fatty acids which is the ratio of unsaturated to saturated fats.[69]

Physical Activity

Physical activity was assessed at baseline and at 6-, 24- and 72-months following baseline, using the Arizona Activity Questionnaire.[90] The baseline and 6-month

questionnaire addressed the prior 6-month periods, the 24-month questionnaire addressed the prior 12-month period, and the 72-month questionnaire addressed the prior 24-month period. The questionnaire asked about the frequency, duration and intensity of activities in several different domains, including work related (e.g., standing, walking, stooping), recreational (e.g., sports, exercise, hobbies), transportation related (e.g., walking, biking, climbing stairs), not related to paid or volunteer work (e.g., household chores, care giving, home maintenance), and sedentary behaviors (e.g., reading, socializing). Using the Compendium of Physical Activities,[91] each activity was assigned a standard metabolic equivalent task value (MET), according to the level of intensity. To examine moderate to vigorous physical activity, a summary variable in hours/week was created by multiplying each activity by frequency and duration and summing over all activities at each follow-up with MET values of 3 or more.

Cigarette Smoking

For cigarette smoking, the baseline questionnaire addressed the lifetime of the participant up until that point. The 6-month questionnaire addressed the prior 6-month period, the 24-month questionnaire addressed the prior 12-month period, and the 72-month questionnaire addressed the prior 24-month period. Participants were asked if they currently or ever smoked cigarettes and those who answered yes to currently smoking, were then asked on average how many cigarettes per day they smoked. Those who answered no to currently smoking and yes to ever smoking, were asked for the approximate date they quit smoking, as well as the average number of cigarettes per day smoked when they did smoke. To examine smoking in pack-years, the average number of cigarettes per day was divided by 20 (typical number of cigarettes in a pack) and then multiplied by the number of years the participant smoked.

Other Covariates

Demographic factors including age, race/ethnicity and education were collected using a baseline questionnaire at the time of entry. Where possible, missing data were supplemented with data obtained from the KPNC electronic health records (EHR). Both body mass index (BMI) and comorbid conditions were obtained from the EHR. A search for comorbid conditions was performed as far back as the EHR allowed and was then used to calculate an Elixhauser Comorbidity Score.[92] Diagnostic and clinical data, which included tumor stage, estrogen receptor status (ER), progesterone receptor status (PR), human epidermal growth factor receptor 2 status (HER2), type of surgery, and receipt of chemotherapy, radiation therapy and hormonal therapy were obtained from the KPNC Cancer Registry and other clinical databases.

All-Cause Mortality

Deaths from all causes up through December 31st, 2018 were identified primarily through linkage with KPNC mortality databases, which includes data compiled from the health system including the KPNC Cancer Registry, as well as linkages with data from the State of California, the Social Security Administration, and the National Death Index.

Deaths may also be reported to us by relatives of participants, and then confirmed by medical chart review or the VDW mortality file.

Final Analytic Sample

Among the 4,505 women in the cohort, 782 (17.4%) were excluded from this analysis because they did not complete the baseline dietary assessment. An additional 63 (1.4%) participants were excluded due to estimated daily total energy intake at baseline being less than 400 kcal/day or greater than 4,000 kcal/day. Participants excluded for these reasons were overall somewhat younger and less likely to be non-Hispanic white, educated, post-menopausal, non-smokers or diagnosed with ER-positive breast cancer compared to those included in the analysis (Supplemental Table 1). The final analytic cohort included 3,660 Pathways Study participants, of whom 655 (17.9%) had died during the follow-up period.

Hypothetical Interventions

For each of the 3 diet quality indices (hPDI, ACS and HEI) and for physical activity, we specified 4 hypothetical longitudinal interventions of increasing concordance with these measures, beginning at baseline. For each of these hypothetical interventions, participants who were observed to be at or above the intervention level maintained their observed value, while those who were observed to be under the intervention level had their values set to the level of the intervention. For example, for a hypothetical intervention on hPDI at the 80-point level, all participants were assigned their observed score if they scored 80 or above on their baseline hPDI index, whereas scores were set to 80 for participants who scored under 80 points on their baseline hPDI index. For the smoking intervention, we set smoking levels after baseline to the total accumulated pack-years of cigarettes smoked as measured at baseline, thereby simulating quitting smoking at baseline for current smokers and never starting for non-smokers. We additionally considered combinations of joint interventions on diet quality, physical activity, and smoking, as well as the expected risks under the standard course (no intervention). All hypothetical interventions were maintained over the entire up to 13-years follow-up period (Table 1).

Statistical Analysis

We estimated the 13-year risks of death under each intervention using parametric g-computation, a generalization of direct standardization. Briefly: g-computation fits a series of parametric regression models for the outcome and its predictors, then uses these models to simulate expected outcomes under specified values of predictor variables or exposures of interest.[89] These specified values can be interpreted as similar to an intervention that achieved that level of the predictor variable. This approach allows estimation of the effects of a range of exposure comparisons on the outcome of interest, in this case all-cause mortality, with the ability to properly address time-varying confounding or dynamic interventions.[89] Under standard modeling assumptions (no unmeasured confounding, no measurement error and no model misspecification), as

well as meeting the positivity assumption that any individual has a positive probability of receiving all values of the treatment variable, this approach can consistently estimate the effect of longitudinal interventions.[89]

We calculated effect estimates using a SAS macro developed by the Harvard Program on Causal Inference,[93] whereby regression models were fit to estimate the joint distribution of each covariate at each time interval, and then a pooled logistic regression model for the outcome was fit across all time periods, each conditional on the covariate history. Using Monte Carlo simulations, 3,660 covariate histories (equal to the sample size) consistent with the intervention were generated for each iteration. Using the estimated coefficients from the regression models, the values of each of the covariates were then calculated, and for the covariates that underwent the intervention, their values were then changed according to the intervention specification. When calculating the risk under the natural course (no intervention), no changes were made to any of the values of the predicted covariates. Risk ratios and risk differences were calculated by comparing the final population risk associated with each intervention to the estimated risk associated with the natural course. To estimate 95% confidence intervals for the risk ratios and risk differences, non-parametric bootstrapping was performed on 1,000 repeated samples.

The fully adjusted models included time-varying data on diet quality, physical activity, and smoking, depending on which intervention was being examined. When evaluating the individual effect of each primary exposure, the other two exposures were included in the models as covariates (for physical activity and smoking, HEI was used as the covariate). Other time-varying data included total energy intake, BMI and comorbidity scores for all models. Additionally, the following baseline characteristics were also included: age at diagnosis, race/ethnicity, education level, cancer stage, ER, PR, HER-2, surgery type, chemotherapy, radiation, and hormonal therapy (variables were categorized as specified in Table 2). These variables were regarded as potential confounders of the relationship between diet and lifestyle factors and all-cause mortality (Supplementary Figure 1).

2.4 Results

Baseline Characteristics

A total of 3,660 participants were followed for 40,888 person-years. Their mean age at the time of their breast cancer diagnosis was 59.7 (range:24-94) years and diverse in race/ethnicity, including whites (68.1%), Asians/Pacific-Islanders (13.0%), Hispanics (10.3%), Blacks (6.6%) and American Indians/Alaska natives (2.1%). They were also mostly educated, with 85% of participants having at least some college education. Most participants had early-stage breast cancer, with 89.0% diagnosed at stage I or II. They were also predominantly ER-positive (83.9%), PR-positive (64.1%) and HER-2 negative (83.2%). Most participants also had some form of surgery (96.8%), and nearly half (46.7%) had chemotherapy and/or radiation therapy (44.4%) and 74.7% had hormonal therapy. The mean hPDI, ACS and HEI scores were 54.0 (standard deviation (SD)=6.7),

4.4 (SD=2.1) and 72.0 (SD=9.5), respectively. The average amount of moderate to vigorous physical activity was 5.9 (SD=5.5) hours/week and for smoking it was 7.4 (SD=15.6) total pack-years (Table 2).

Hypothetical Interventions and All-Cause Mortality

The fully adjusted estimated 13-year risk of all-cause mortality under the natural course (no intervention) was 21.6% (19.6-24.1). Table 3 presents the adjusted estimated risks of all-cause mortality over the follow-up period under each hypothetical intervention as well as the estimated risk ratios and risk differences compared to the natural course. For each hypothetical intervention on the diet quality indices and physical activity, the estimated risk ratios of each intervention as compared to the natural course decreased with increased intensity of the intervention. For example, the estimated risk ratios comparing the hypothetical interventions on hPDI alone in the form of hPDI≥60, hPDI≥70, hPDI≥80 and hPDI=90, as compared to the natural course were 0.95 (0.85-1.02), 0.90 (0.74-1.03), 0.86 (0.63-1.04) and 0.66 (0.40-1.08), respectively. Likewise, the estimated risk ratios comparing the hypothetical interventions on PA alone in the form of PA≥5, PA≥10, PA≥15 and PA≥20, as compared to the natural course were 0.92 (0.88-0.98), 0.82 (0.72-0.93), 0.72 (0.57-0.89) and 0.61 (0.45-0.86), respectively. The risk ratio comparing the hypothetical intervention of no smoking with the natural course also showed an inverse association (RR: 0.92 (0.88,0.96).

When examining the effects of moderate hypothetical interventions on diet quality and physical activity with no smoking, there was a lower risk of mortality as compared to the natural course [hPDI≥60 + PA≥5 + SM RR=0.81, 0.73-0.90; ACS≥6 + PA≥5 + SM RR=0.80, 0.72-0.88; HEI≥70 + PA≥5 + SM HR=0.85, 0.78-0.90]. When we combined the most intense hypothetical interventions on diet quality and physical activity together with no smoking, risk ratios suggested substantially lower mortality risk compared to the natural course [hPDI=90 + PA≥20 + SM RR=0.40, 0.23-0.73, ACS=9 + PA≥20 + SM RR=0.49, 0.35-0.71; HEI=100 + PA≥20 + SM HR=0.52, 0.37-0.77].

2.5 Discussion

This study is the first to use a rigorous causal inference approach to demonstrate that higher diet quality, increased physical activity, and no smoking each could hypothetically reduce risk of mortality after breast cancer diagnosis. Increasing the intensity of diet and physical activity interventions strengthened these associations, suggesting a dose-response relationship. Similarly, a joint intervention on all three lifestyle factors conveyed the greatest protection.

To our knowledge, there have been no RCTs intervening jointly on diet quality, physical activity and smoking on all-cause mortality among breast cancer survivors enrolled at the time of their diagnosis. Thus, the existing evidence base consists primarily of observational studies that measure lifestyle factors at a single timepoint without control for time-varying confounding or accounting for change in lifestyle behaviors following a breast cancer diagnosis. To fill this gap, we simulated a multi-arm RCT of lifestyle

interventions with long-term follow-up through application of the g-computation. The application of these causal methods to observational studies enables estimation of the effects of hypothetical interventions of varying intensity.

To our knowledge, there have also been no published studies that have examined hPDI and breast cancer survival. However, plant-based diets have been observed to decrease the risk of incident breast, colon and prostate cancers,[31, 94-98] as well as cardiovascular disease.[31, 96] Furthermore, there is a growing consensus that the demand for plant-based foods, including dairy and meat alternatives, has been steadily increasing in the US.[99, 100] Because certain foods such as refined grains and sugar-sweetened beverages are plant-based and associated with adverse chronic disease outcomes,[101-105] we chose the hPDI for its emphasis on healthful plant-based food items. ACS was selected due to its direct relevance to cancer specific outcomes as well as its association with decreased mortality,[106] and HEI because of its prior demonstrated associations with cancer prevention,[38, 70, 71] and survival.[23, 65, 72]

Our findings are consistent with prior research on diet quality and breast cancer survival. Two previous studies which used data driven dietary quality indices,[25, 83] and four studies which used *a priori* dietary quality indices to each examine concordance with healthful dietary patterns and survival after a breast cancer diagnosis, concluded that a healthful dietary pattern could be a potentially beneficial point of intervention to help improve breast cancer prognosis.[51, 65, 66, 80] Another study from the Women's Health Initiative concluded that a ≥15% decrease in HEI from before to after diagnosis was associated with an increased risk of death from breast cancer.[21]

Our results suggest that greater physical activity could decrease the overall risk of death after breast cancer, independent of either diet quality or smoking. This finding is supported by a recent study showing patients meeting the highest level of Physical Activity Guidelines for Americans both before and 1 year after diagnosis had a 49% reduced risk in mortality when compared to those who met them the least.[107] An additional time-dependent sub-analysis, which also factored in physical activity measured during treatment and 2 years after diagnosis, revealed that the most active patients experienced a 69% reduced risk when compared to the least active.[107] Similar associations have been found among women with early stage as well as metastatic breast cancers.[108]

A hypothetical intervention where smokers stopped smoking after diagnosis and non-smokers never started, resulted in an 8% reduction in risk of all-cause mortality when assuming smokers stopped and non-smokers did not start, as compared to their natural course. Consistent with these findings, a meta-analysis from 2014 that examined cigarette smoking at the time of diagnosis and all-cause mortality pooled data across 9 studies and reported a 33% increased risk when comparing smokers at the time of diagnosis to never smokers. Two studies prospectively examined post-diagnosis changes in smoking status and its impact on survival, both of which supported an increased risk of all-cause mortality with continued smoking.[82, 109]

This study has several strengths, including a large sample size which draws from a population of women with newly-diagnosed invasive breast cancer, prospective longitudinal data collection with a long follow-up period, and repeated measures on a dietary and lifestyle exposures and covariates. Our causal inference-based approach allowed for testing several hypothetical individual and joint interventions that would otherwise be difficult to implement in practice. This approach also appropriately adjusted for time-varying confounding that could lead to biased estimates in more traditional analytic techniques.

It should be noted that despite this strong causal inference framework, this approach does not directly address limitations arising from measurement error that is known to influence assessment of lifestyle factors, especially food intake.[110] This would result in imprecision of estimating the effects of the hypothetical interventions, perhaps underestimating their effects. As a non-interventional study, there may be residual confounding from unmeasured sources, although we accounted for several principal confounders that likely influence the effects of lifestyle factors on risk of death. It is also always possible that women with breast cancer who enrolled in this study might be somehow systematically different from those who did not. However, other data (not shown) shows that the Pathways Study cohort reflects the eligible population, with slight shifts toward slightly younger ages and earlier stages at diagnosis. In regard to the most extreme interventions on diet and physical activity we recognize that adhering to such limits may be unrealistic in practice, however reporting a range of estimates that describes the entire spectrum of possible interventions may be helpful in policy and decision making. Finally, the reference condition – the natural course – was based on the experience of the Pathways Study cohort. The impacts of these interventions may differ if the characteristics of a cohort of breast cancer survivors differs substantially from those of the Pathways Study.

We found that hypothetical interventions that increase diet quality, increase physical activity, or stop smoking, reduced the risk of death among breast cancer survivors. Furthermore, the joint intervention on all three risk factors conveyed the highest survival benefit. These findings highlight the lifesaving potential of diet and lifestyle interventions on breast cancer survivors at the time of their diagnosis and add new rigor to the evidence base for lifestyle guidelines for women with breast cancer.

2.6 Tables

Table 1. Descriptions of hypothetical interventions at baseline for Pathways Study participants*

Intervention	Description
(1) Natural course	Observed data: no intervention on treatment at any time
(2) hPDI ≥ 60	Any hPDI < 60 is set to 60 and any hPDI ≥ 60 is not intervened on
(3) hPDI ≥ 70	Any hPDI < 70 is set to 70 and any hPDI ≥ 70 is not intervened on
(4) hPDI ≥ 80	Any hPDI < 80 is set to 80 and any hPDI ≥ 80 is not intervened on
(5) hPDI = 90	All hPDI is set to 90 (highest possible score)
(6) ACS ≥ 6	Any ACS < 6 is set to 6 and any ACS ≥ 6 is not intervened on
(7) ACS ≥ 7	Any ACS < 7 is set to 7 and any ACS ≥ 7 is not intervened on
(8) ACS ≥ 8	Any ACS < 8 is set to 8 and any ACS ≥ 8 is not intervened on
(9) ACS = 9	All ACS is set to 9 (highest possible score)
(10) HEI ≥ 70	Any HEI < 70 is set to 70 and any HEI ≥ 70 is not intervened on
(11) HEI ≥80	Any HEI < 80 is set to 80 and any HEI ≥ 80 is not intervened on
(12) HEI ≥ 90	Any HEI < 90 is set to 90 and any HEI ≥ 90 is not intervened on
(13) HEI = 100	All HEI is set to 100 (highest possible score)
(14) PA ≥ 5	Any PA < 5 is set to 5 and any PA ≥ 5 is not intervened on
(15) PA ≥ 10	Any PA < 10 is set to 10 and any PA ≥ 10 is not intervened on
(16) PA ≥ 15	Any PA < 15 is set to 15 and any PA ≥ 15 is not intervened on
(17) PA ≥ 20	Any PA < 20 is set to 20 and any PA ≥ 20 is not intervened on
(18) No Smoking	Smokers quit and non-smokers never start
(20) (2) + (14) + (18)	Joint intervention of hPDI ≥ 60 and PA ≥ 5 and No Smoking
(21) (5) + (17) + (18)	Joint intervention of hPDI = 90 and PA ≥ 20 and No Smoking
(22) (6) + (14) + (18)	Joint intervention of ACS ≥ 6 and PA ≥ 5 and No Smoking
(23) (9) + (17) + (18)	Joint intervention of ACS = 9 and PA ≥ 20 and No Smoking
(24) (10) + (14) + (18)	Joint intervention of HEI ≥ 70 and PA ≥ 5 and No Smoking
(25) (13) + (17) + (18)	Joint intervention of HEI = 100 and PA ≥ 20 and No Smoking

Abbreviations: hPDI (healthy plant-based dietary index score), ACS (American Cancer Society nutrition guidelines score), HEI (Healthy Eating Index score), PA (physical activity in MET hrs/wk)

*All hypothetical interventions are assumed to be maintained over the entire 13-year follow-up period.

Table 2. Baseline characteristics of the Pathways Study participants (n=3,660)

Demographic characteristics	N (%)	Mean (SD)
Age at diagnosis (years)		58.9 (12.2)
Race/ethnicity		
White	817 (63.7)	
Black	91 (7.1)	
Asian/Pacific Islander	182 (14.2)	
Hispanic	163 (12.7)	
American Indian/Alaska Native	29 (2.3)	
Education		
High school or less	240 (18.7)	
Some college	467 (36.4)	
College graduate	375 (29.3)	
Postgraduate	199 (15.5)	
Unknown	1 (0.1)	
Physical activity (hours/week) ‡		5.9 (5.5)
Smoking (total pack-years)		7.4 (15.6)
BMI (kg/m^2)		28.8 (7.1)
Energy intake (kcal/day)		1499.7 (616.0)
hPDI Score		54.0 (6.7)
ACS Score		4.4 (2.1)
HEI Score		72.0 (9.5)
Clinical characteristics		
AJCC cancer stage		
I	679 (53.0)	
II	447 (34.9)	
III	133 (10.4)	
IV	23 (1.8)	
ER status		
Positive	1079 (84.2)	
Negative	201 (15.7)	
Unknown	2 (0.2)	
PR status		
Positive	837 (65.3)	
Negative	441 (34.4)	
Unknown	4 (0.3)	
HER2 status		
Positive	166 (12.9)	
Negative	1063 (82.9)	
Unknown	53 (4.1)	
Surgery type		
Lumpectomy	755 (58.9)	
Mastectomy	484 (37.8)	
None	43 (3.4)	
Unknown	0 (0.0)	
Chemotherapy		
No	616 (48.0)	
Yes	662 (51.6)	
Unknown	4 (0.3)	
Radiation therapy		
No	752 (58.7)	
Yes	530 (41.3)	
Hormonal therapy		
No	308 (24.0)	
Yes	963 (75.1)	
Unknown	11 (0.9)	

Abbreviations: SD (standard deviation), BMI (body mass index), hPDI (healthy plant-based dietary index), ACS (American Cancer Society nutrition guidelines), HEI (Healthy Eating Index), ER (estrogen receptor), PR (progesterone receptor), HER2 (human epidermal growth factor receptor 2).

‡There were 4 participants with unknown physical activity.

Table 3. Hazards ratios and 95% confidence intervals of quintiles of dietary quality indices and all-cause mortality*

Intervention	Risk (95% CI)	RR (95% CI)	RD (95% CI)
(1) Natural course	21.6 (19.6, 24.1)	1.00	0.00
(2) hPDI ≥ 60	20.6 (18.0, 23.4)	0.95 (0.85, 1.02)	-0.98 (-3.09, 0.39)
(3) hPDI ≥ 70	19.5 (16.3, 23.4)	0.90 (0.74, 1.03)	-2.16 (-5.44, 0.63)
(4) hPDI ≥ 80	18.6 (13.9, 23.6)	0.86 (0.63, 1.04)	-3.07 (-7.68, 0.82)
(5) hPDI = 90	14.4 (8.5, 23.7)	0.66 (0.40, 1.08)	-7.25 (-13.74, 1.60)
(6) ACS ≥ 6	20.8 (18.0, 22.9)	0.96 (0.86, 1.03)	-0.90 (-3.09, 0.63)
(7) ACS ≥ 7	19.9 (17.8, 22.6)	0.92 (0.83, 1.04)	-1.73 (-4.11, 0.85)
(8) ACS ≥ 8	19.6 (16.9, 22.6)	0.90 (0.78, 1.06)	-2.07 (-4.97, 1.24)
(9) ACS = 9	19.3 (16.2, 22.6)	0.89 (0.74, 1.05)	-2.36 (-5.95, 0.92)
(10) HEI ≥ 70	21.6 (19.1, 24.3)	1.00 (0.96, 1.02)	-0.06 (-0.89, 0.54)
(11) HEI ≥80	21.3 (18.2, 23.8)	0.98 (0.90, 1.05)	-0.38 (-2.30, 1.04)
(12) HEI ≥ 90	20.8 (17.1, 24.4)	0.96 (0.84, 1.08)	-0.88 (-3.49, 1.89)
(13) HEI = 100	20.2 (15.9, 25.0)	0.94 (0.80, 1.13)	-1.40 (-4.47, 2.86)
(14) PA ≥ 5	20.0 (17.6, 23.3)	0.92 (0.88, 0.98)	-1.68 (-2.56, -0.47)
(15) PA ≥ 10	17.8 (14.8, 21.9)	0.82 (0.72, 0.93)	-3.88 (-6.02, -1.62)
(16) PA ≥ 15	15.6 (11.8, 20.4)	0.72 (0.57, 0.89)	-6.16 (-8.95, -2.45)
(17) PA ≥ 20	13.3 (9.2, 19.0)	0.61 (0.45, 0.86)	-8.40 (-11.59, -3.27)
(18) No Smoking	19.9 (17.9, 22.3)	0.92 (0.88, 0.96)	-1.80 (-2.66, -0.98)
(20) 2 + 14 + 18	17.4 (15.0, 20.3)	0.81 (0.73, 0.90)	-4.20 (-5.93, -2.30)
(21) 5 + 17 + 18	7.7 (4.4, 14.5)	0.36 (0.21, 0.69)	-13.90 (-17.66, -6.84)
(22) 6 + 14 + 18	17.4 (15.1, 20.4)	0.80 (0.72, 0.88)	-4.23 (-6.21, -2.73)
(23) 9 + 17 + 18	10.7 (7.8, 15.7)	0.49 (0.35, 0.71)	-10.95 (-13.84, -6.54)
(24) 10 + 14 + 18	18.3 (15.9, 21.0)	0.85 (0.78, 0.90)	-3.34 (-4.83, -2.21)
(25) 13 + 17 + 18	11.2 (7.9, 17.3)	0.52 (0.37, 0.77)	-10.48 (-13.84, -5.13)

Abbreviations: hPDI (healthy plant-based dietary index score), ACS (American Cancer Society nutrition guidelines score), HEI (Healthy Eating Index score), PA (physical activity in MET hrs/wk), RR (risk ratio), RD (risk difference), CI (confidence interval).

*Adjusted for age at diagnosis, total energy, race/ethnicity, education, menopausal status, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2.

2.7 Supplementary Tables and Figures

Table S1. Baseline characteristics and survival outcomes comparing excluded participants with included participants from the Pathways Study (n=4,505)

Characteristic	Excluded (n=3,660)	Included (n=845)	
Continuous, Mean (SD)			P*
Age at diagnosis (years)	56.7 (12.6)	59.7 (11.9)	<.0001
Physical activity (MET h/wk)‡	54.4 (44.4)	53.8 (35.8)	0.6871
Categorical, No. (%)			P†
Race/ethnicity			<.0001
White	460 (54.4)	2491 (68.1)	
Black	111 (13.1)	240 (6.6)	
Asian/Pacific Islander	124 (14.7)	475 (13.0)	
Hispanic	134 (15.9)	378 (10.3)	
American Indian/Alaska Native	16 (1.9)	76 (2.1)	
Education			<.0001
High school or less	160 (18.9)	547 (14.9)	
Some college	323 (38.2)	1245 (34.0)	
College graduate	217 (25.7)	1024 (28.0)	
Post graduate	139 (16.4)	842 (23.0)	
Unknown	6 (0.7)	2 (0.1)	
Menopausal status			0.0014
Premenopausal	292 (34.6)	1060 (29.0)	
Postmenopausal	553 (65.4)	2600 (71.0)	
Smoking status			<.0001
Never	446 (52.8)	2092 (57.2)	
Former	315 (37.3)	1408 (38.5)	
Current	75 (8.9)	154 (4.2)	
Unknown	9 (1.1)	6 (0.2)	
Cancer stage			0.0894
I	426 (50.4)	2008 (54.9)	
II	308 (36.4)	1250 (34.2)	
III	94 (11.1)	346 (9.5)	
IV	17 (2.0)	56 (1.5)	
ER status			0.0137
Positive	680 (80.5)	3072 (83.9)	
Negative	165 (19.5)	586 (16.0)	
Unknown	0 (0.0)	2 (0.1)	
PR status			0.0753
Positive	515 (60.9)	2347 (64.1)	
Negative	330 (39.1)	1308 (35.7)	
Unknown	0 (0.0)	5 (0.1)	
HER2 status			0.4293
Positive	116 (13.7)	472 (12.9)	
Negative	685 (81.1)	3045 (83.2)	
Unknown	44 (5.2)	143 (3.9)	
Recurrence			0.7396
No	735 (87.0)	3199 (87.4)	
Yes	110 (13.0)	461 (12.6)	
Breast Cancer-Specific Mortality			0.5728
No	765 (90.5)	3336 (91.1)	
Yes	80 (9.5)	324 (8.9)	
Non-Breast Cancer-Specific Mortality			0.7809
No	766 (90.7)	3329 (91.0)	
Yes	79 (9.3)	331 (9.0)	
All-Cause Mortality	. (.)	. (.)	0.5308
No	686 (81.2)	3005 (82.1)	
Yes	159 (18.8)	655 (17.9)	

Abbreviations: FFQ (food frequency questionnaire), SD (standard deviation), MET (metabolic equivalent of task), ER (estrogen receptor), PR (progesterone receptor), HER2 (human epidermal growth factor receptor 2).

*The p-value is from analysis of variance.

†The p-value is from the Pearson X^2 test.

‡There were 42 participants who did not complete the FFQ and 4 participants who did complete the FFQ with unknown physical activity.

Table S2. Scoring methods and optimal quantities for each component of hPDI, ACS and HEI

	ACS	ACS	HEI
Theoretical range	0 to 9	0 to 9	0 to 100
Fruits			
Total Fruit			≥ 0.8 cups/1,000 kcal
Whole Fruits	= Highest quintile		≥ 0.4 cups/1,000 kcal
Fruit Juices	= Lowest quintile		
Vegetables			
Total Vegetables*	= Highest quintile		≥ 1.1 cups/1,000 kcal
Non-Starchy Vegetables	= Highest quintile		
Starchy Vegetables	= Lowest quintile		
Greens and Beans			≥ 0.2 cups/1,000 kcal
Total Fruits and Vegetables		= Highest tertile†	
Grains			
Whole Grains	= Highest quintile	= Highest tertile‡	> 1.5 ounces/1,000 kcal
Refined Grains	= Lowest quintile		≤ 1.8 ounces/1,000 kcal
Diary			
Total Dairy	= Lowest quintile		≥ 1.3 cups/1,000 kcal
Low-fat Dairy			
Protein Foods			
Total Protein Foods			≥ 2.5 ounces/1,000 kcal
Seafood and Plant Proteins			≥ 0.8 ounces/1,000 kcal
Red and Processed Meats	= Lowest quintile*	= Lowest tertile	
Fish	= Lowest quintile		
Eggs	= Lowest quintile		
Legumes	= Highest quintile		
Nuts	= Highest quintile		
Nuts and Legumes			
Fat			
Unsaturated Fats			PUFAs + MUFAs:SFAs ≥ 2.5
Saturated Fats			≤ 8% of energy
Animal Fats	= Lowest quintile		
Vegetable Oils	= Highest quintile		
Sodium			≤ 1.1 grams/1,000 kcal
Sugar			
Added Sugar			≤ 6.5% of energy
Sweetened beverages	= Lowest quintile		
Sweets	= Lowest quintile		
Teas and Coffees	= Highest quintile		

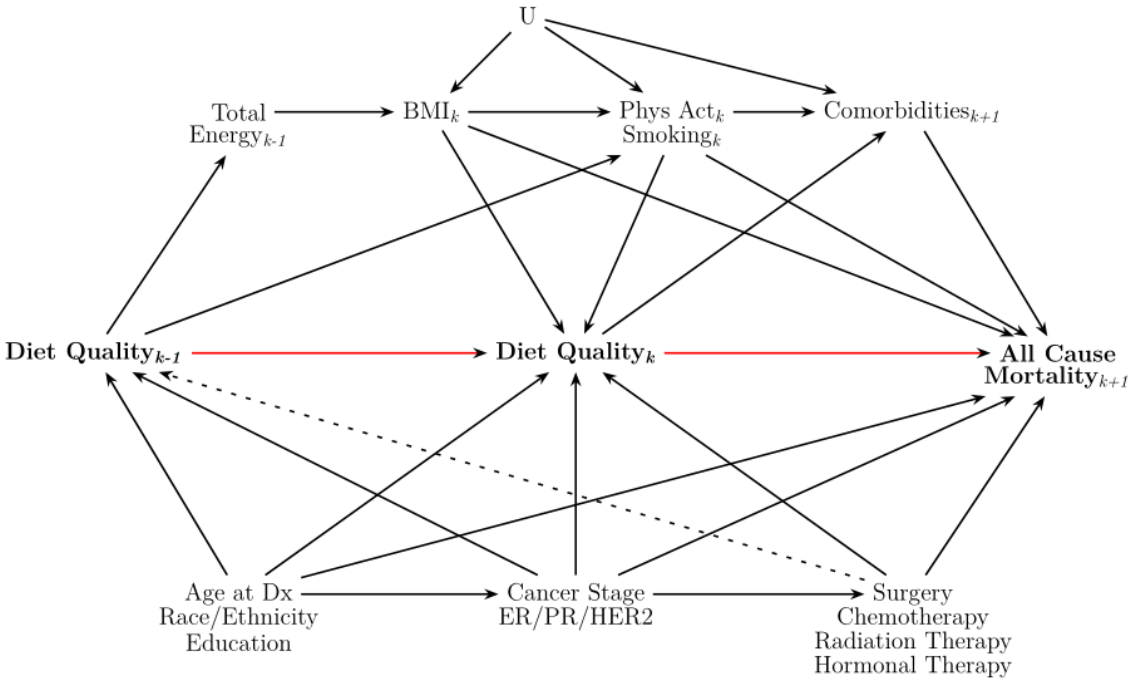
Abbreviations: hPDI (healthy plant-based diet score), ACS (American Cancer Society nutrition guidelines score), HEI (2015 Healthy Eating Index score), MUFA (monounsaturated fat), SFA (saturated fat), PUFA (polyunsaturated fat).

*Includes poultry and organ meat.

†Excludes fruit juices and potatoes. Includes partial variety score for consumption of ≥5 different fruits or vegetables per month.

‡Whole grains calculated as a percent of total grains.

Figure S1. Causal diagram of the relationship between baseline and subsequent changes in diet and breast cancer survival*



*Related variables have been grouped together to simplify presentation.

3 CHAPTER 3. Hierarchical regression analysis of baseline food consumption and breast cancer

3.1 Abstract

Background

Previous work on the relationship between intakes of individual foods and breast cancer survival have been limited in number and have not accounted for co-exposure confounding or nutrient composition of the foods. We applied a 2-stage hierarchical model to estimate mutually adjusted associations between individual food items and all-cause mortality after a breast cancer diagnosis accounting for the role of select nutrients driving those associations.

Methods

Using a 139-item food frequency questionnaire, diet was assessed among 3,660 women an average of 2.3 (range: 0.7-18.7) months after an invasive breast cancer diagnosis. There were 28 individual food groups constructed from the items on the questionnaire and 17 food constituents having known associations with breast cancer were chosen to be included in the analysis. Over 40,888 person-years of follow-up, 655 deaths were ascertained. Minimally and fully adjusted Cox proportional hazards models were used to estimate hazards ratios (HR) and 95% confidence intervals (CI) for the food groups included simultaneously. We specified a two-stage hierarchical model where the food group effects varied according to their constituents and compared the results to conventional multivariable models.

Results

Overall, the estimates associated with the hierarchical models had better precision as compared to the estimates from the conventional models. We also observed a decreased risk of all-cause mortality with greater baseline consumption of whole grains [conventional per 100g HR=0.91, 95% CI=0.80-1.03; hierarchical per 100g HR=0.85, 95% CI=0.73-0.98]. Soy products [conventional HR=1.08, 95% CI=0.93-1.25; hierarchical HR=0.62, 95% CI=0.39-1.00] and nuts and seeds [conventional per 100g HR=0.70, 95% CI=0.36-1.39; hierarchical per 100g HR=0.70, 95% CI=0.39-1.25] were also associated with an decreased risk in all-cause mortality, though the estimates were less precise. Additionally, we observed an increased risk of all-cause mortality with an increased consumption of eggs, though confidence intervals were wide [conventional per 100g HR=1.50, 95% CI=0.90-2.51; hierarchical per 100g HR=1.42, 95% CI=0.86-2.33]. Among the second-stage constituent effects, iron, isoflavone and fiber consumption were associated with the greatest decreased risk of all-cause mortality [iron per 10mg HR=0.64, 95% CI=0.25,1.62; isoflavones per 10mg HR=0.76, 95% CI=0.57-1.00; fiber per 10g HR=0.85, 95% CI=0.57-1.25], though the estimates for iron and fiber were imprecise.

Conclusion

The results from this two-stage model suggests that whole grains, soy products, nuts and seeds, as well as iron and isoflavones may be inversely associated with all-cause mortality, while eggs may be adversely associated. Our hierarchical approach highlights the importance of taking nutrient level effects into consideration when simultaneously modeling multiple foods on all-cause mortality among breast cancer survivors.

3.2 Introduction

Breast cancer is the second leading cause of cancer death among women in the United States (US). However, due to increased awareness, widespread screening and improved treatments, women with breast cancer are living longer. It is estimated that there are now more than 3.8 million breast cancer survivors living in the US.[3] After a breast cancer diagnosis, women are highly motivated to make lifestyle changes and have expressed a desire for more evidence-based information from health-professionals.[34, 35]

While there have been a great number of epidemiological studies examining diet and risk of breast cancer, not nearly as many have considered diet in relation to breast cancer prognosis. According to the 2018 World Cancer Research Fund/American Institute for Cancer Research report, the available evidence on the effect of diet, nutrition and physical activity on cancer survivors is limited, and that the amount of quality research in this area is insufficient to make firm conclusions.[4] However, their report also highlights limited evidence supporting foods high in fiber and soy content and low in saturated fat content decreasing the risk of death after a breast cancer diagnosis.[4] Findings on several other foods including fruits, vegetables, fish and meat have been either lacking or inconsistent. The report expresses a need for more randomized controlled trials, as well as a greater understanding of the underlying mechanisms which link diet and survival outcomes.[4]

While dietary patterns (the quantities and proportions of all foods, drinks and nutrients in one's diet as well as the frequency with which they are consumed[27]) may provide important insights into the totality of one's diet and its impact on breast cancer survival, they have some limitations for establishing specific dietary guidelines and yielding insight into factors which drive effects of individual foods. Dietary patterns and the indices that measure them are often vague, and their translation into actionable guidelines can be challenging. They also treat each food component equivalently and are unable to distinguish the most important dietary factors, those foods which may be of most interest to the public.[67-69]

For valid inference of the association between individual food items on mortality, one must account for all significant food intakes to address "co-exposure confounding," (i.e. confounding by other foods) which is typically accomplished by mutually adjusting for other foods within the same model. Notably, conventional analyses of individual food items have rarely considered co-exposure confounding, or issues of collinearity and

multiple comparisons.[111] Hierarchical regression, which allows for the incorporation of multiple levels of information into a single analysis, can address these issues by incorporating a 2nd stage model which explains the effects of the primary exposures by factors (such as nutrients) believed to drive their relationships with the outcome. These types of hierarchical regression techniques have been shown to provide estimates of effect that are more precise and plausible than those from more traditional methods of analysis in both dietary epidemiological studies,[112, 113] as well as in other scientific fields of study.[114-116]

With data from a large, prospective cohort of women with breast cancer, we applied a two-stage hierarchical model to estimate the association between intake of multiple food items (first stage exposures) measured at or around the time of diagnosis and breast cancer mortality. The second stage model expressed the first level (foods) associations in terms of their nutrient composition. This framework allows us to estimate the mutually adjusted associations between the individual food items on breast cancer survival and the role of specific nutrients within these associations. To our knowledge, this is the first study to apply a hierarchical modelling approach to examining food components on all-cause mortality among breast cancer survivors.

3.3 Participants and Methods

Study Cohort

The Pathways Study is a prospective cohort of 4,505 female breast cancer survivors diagnosed with breast cancer between the years of 2005 and 2013 from Kaiser Permanente Northern California (KPNC), details of this study are previously published.[54] Briefly, eligibility criteria included: being female, 21 years or older, KPNC membership, speaking English, Spanish, Cantonese or Mandarin, living within a 65-mile radius of a field interviewer, diagnosis of incident invasive breast cancer, and no prior history of other invasive cancers. The enrollment rate was 40.3% of those eligible (4,505 of 11,174), and participants received an in-person baseline interview administered by field staff.

This study was approved by both the Institutional Review Boards of KPNC and University of California, Berkeley. Written informed consent was obtained from all study participants.

Dietary Assessment

Diet was assessed an average of 2.3 (range:0.7-18.7) months after diagnosis. Dietary intake 6 months preceding the interview was assessed at the baseline visit with a modified version of the Block 2005 Food Frequency Questionnaire (FFQ).[55] Its 139 food items and additional questions were selected to be representative of a wide range of dietary factors, as well as to capture foods that are popular in Hispanic and Asian populations. Completed questionnaires were sent to NutritionQuest (Berkeley, CA) for scanning, as well as food and nutrient identification, using a nutrient database

developed primarily from the United States Department of Agriculture (USDA) Food and Nutrient Database for Dietary Studies (FNDDS).[55]

Among the 4,505 women in the cohort, 782 (17.4%) were excluded from this analysis because they did not complete the dietary assessment at baseline. An additional 63 (1.4%) participants were excluded due to estimated daily total energy intake (kcal/d) being less than 400 or greater than 4000. While no significant differences were observed in regard to survival outcomes, excluded participants were less likely to be older, white, educated, post-menopausal, non-smokers and ER-positive as compared to those included in the analysis (Supplemental Table 1). These exclusions brought the final sample size to 3,660.

Food groups

Distinct food groups were created as the primary exposures in this analysis, by allocating individual food items into groups similar to those listed in the USDA MyPlate dietary guidelines.[117] In some cases, where certain foods did not directly correspond to MyPlate food groups, new groups were created based on prior literature. For example, MyPlate does not recommend the consumption of sugar sweetened beverages as part of their dietary guidelines, and therefore, a new group was created to represent those food items. Some food items from the FFQ were not included in the final analysis: those that were rarely consumed (<=20% of the population), as well as any food item that was not clearly classifiable into a specific food group (e.g. if a participant reported eating “a breakfast sandwich with egg or meat”, it was not possible to know if this food item should belong in the egg or meat). Alcohol and water consumption were also excluded. In all, the following 28 food groups were calculated as total consumption in 100 gram servings per day and included in the final analysis: dairy (butter; cheese; milk; yogurt), fruit (berries; melons; other fruit; fruit juice), grains (refined grains; whole grains), oils and dressings, protein (eggs; fish with high omega-3 fatty acids; fish with low omega-3 fatty acids; processed fish; shellfish; red meat; poultry; nuts and seeds; soy products), sugar sweetened beverages, sweets, vegetables (allium vegetables; beans, peas and lentils; dark green vegetables; red and orange vegetables; starchy vegetables; other vegetables).

Nutrient intakes

The reported frequencies and portion sizes from the FFQ were used to estimate the following 17 daily nutrient intakes attributed to the food groups described above: beta-carotene (µg/day), calcium (mg/day), carbohydrate (g/day), cholesterol (g/day), choline (mg/day), total fat (g/day), fiber (g/day), folate (µg/day), iron (mg/day), isoflavones (mg/day), protein (g/day), retinol (µg/day), selenium (µg/day), vitamin B-12 (µg/day), vitamin C (mg/day), vitamin D (µg/day) and vitamin E (α-tocopherol, mg/day). These variables represent many, though not all, of the known factors that drive the relationship between intake of the foods and all-cause mortality.

Covariates

Demographic and behavioral factors including age, race/ethnicity, education, menopausal status, smoking status, physical activity, and body mass index (BMI) were collected using the baseline questionnaire at time of entry. Where possible, missing data were supplemented with data obtained from the KPNC electronic health records (EHR) and medical chart review (MCR), except in the case of BMI, where the EHR data took precedence over self-reported values. Diagnostic and clinical data, which included cancer stage, estrogen receptor status (ER), progesterone receptor status (PR), human epidermal growth factor receptor 2 status (HER2), type of surgery, and receipt of chemotherapy, radiation and hormonal therapies, were ascertained from a combination of the KPNC Cancer Registry and other clinical databases.

All-Cause Mortality

Because women entered the cohort after their initial breast cancer diagnosis, they were not considered at risk for a possible outcome before their baseline dietary assessment. Therefore, delayed-entry models were used, and person-time was calculated from the date of completion of the baseline questionnaire to the date of death.[118] Those participants without an event were censored at the end of the study period on December 31, 2018. Deaths from all causes were identified during follow up interviews with relatives of participants, and then confirmed by medical chart review. Linkages with data from the State of California, the Social Security Administration, and the National Death Index were also performed. Included in these analyses, were 655 (17.9%) deaths due to any cause.

Statistical Analysis

Conventional models. Conventional minimally and fully adjusted multivariable Cox proportional hazards models were used to calculate hazard ratios (HR) and 95% confidence intervals (CI) to assess the mutually adjusted associations between the food groups (all food groups included in the models) and all-cause mortality. This was modeled as:

$$\log[h(t|X,W)] = \log[h_0(t)] + X\beta + W\gamma \quad (1)$$

where $h(t)$ is the hazard of death at time t, $h_0(t)$ is the referent hazard when all covariates are set to 0, X is the vector of 32 food items expressed as number of 100g servings from each food group, and W is the vector of additional covariates (*e.g.* confounders) included in the model. This model yields 32 coefficients on food items (β, the log-HRs) representing the association between each food and mortality. Both the minimally and fully adjusted conventional models include all 32 food items and age at diagnosis, while the fully adjusted conventional model also include race/ethnicity, education, menopausal status, smoking status, physical activity, BMI, ER status, PR status, HER2 status, and type of surgery, chemotherapy, radiation and hormonal therapies.

Hierarchical model. To improve the accuracy of our estimates, we also specified a second-level model to incorporate prior knowledge of factors (nutrients and other constituents of the foods) that may drive the associations between the food groups and mortality. This provides estimates of the food-level effects given their nutrient composition, as well as estimates of the associations between the nutrients in these foods and all-cause mortality. This second level of the hierarchical regression for the log-hazard coefficients β of the food items was:

$$\beta = Z\pi + \delta \quad (2)$$

where Z represents the second-stage design matrix, which contained the respective nutrients for each food (each row of Z corresponds to a given food, each column a nutrient). The vector π are the coefficients that relate each nutrient to each food group's association with all-cause mortality, and δ is the vector of residual errors for food items on all-cause mortality (*i.e.* other factors not expressed in Z). This second stage model implies that foods with similar nutrient values have similar effects on mortality, but also allows for residual effects specific to each food. The δ are assumed to be distributed normally with mean 0 and variance τ^2, which we discuss below. Each cell of the Z matrix represents the average amount of nutrient consumed per 100-gram serving of food. For example, the cell in Table 3 associated with both butter and cholesterol indicates that in the study population there was on average 214.2 mg of cholesterol consumed per 100-gram serving of butter.

Substituting equation (2) into equation (1) yields the equivalent mixed model representation:[114]

$$\log[h(t|X,Z,W)] = \log[h_0(t)] + XZ\pi + X\delta + W\gamma, \quad (3)$$

which we used for this analysis. Note that the term XZ in this model represents the total intakes of each nutrient in Z across all foods in X, with fixed effects of nutrients (π) and random coefficients for foods (δ).[111] For the model fitting, given the available data, we used estimates of the nutrient intakes that varied according to each subject's consumption of specific items within each food group (essentially allowing a slightly different Z matrix for each participant). Although this approach deviates slightly from the above presentation, it should yield a more accurate characterization of the nutrient-specific effects of these food groups. We calculated posterior point estimates and 95% confidence intervals of each food's effect on breast cancer survival, given the effects of their respective nutrient composition, by substituting estimates of π and δ from equation (3) into equation (2), with Z representing the average nutrient profile for a 100g serving of each food. Model fitting was performed using penalized maximum likelihood (ridge regression) with only the δ terms subject to the quadratic penalty.[119] The residual standard deviation (τ) associated with the food items was specified *a priori* using a semi-Bayes approach.[112] As in previous applications of this method,[111, 113, 114] the food-level variance (τ^2) was set to 0.1225 (τ=0.35), implying that the prior 95% confidence

intervals for a 100-gram serving increase of any food was within a four-fold range of its mean ($e^{(3.92*\tau)} \approx 4.0$).[112]

All data cleaning and manipulations, as well as general frequencies and distributions were performed using SAS 9.4 software (SAS Institute, Cary, NC). The analyses associated with both the conventional and hierarchical modeling was performed using R software[75] in conjunction with the *survival* package.[120]

3.4 Results

The baseline characteristics for the women in this study are presented in Table 1. Their mean age at the time of their breast cancer diagnosis was 59.7 (range:24-94) years and though predominantly white (68.1%), Asian/Pacific-Islanders (13.0%), Hispanics (10.3%), Blacks (6.6%) and American Indian/Alaska natives (2.1%) were also represented. Most participants had a less advanced cancer stage, with 89.0% identified as either stage I or II and were predominantly ER-positive (83.9%), PR-positive (64.1%) and HER-2 negative (83.2%). Food consumption is shown in Table 2 in grams per day, whereby women consumed on average more milk (140.3 g/d, range=0-966.2) than other diary items, more fruits (118.3 g/d, range=0-649.6) other than berries (12.8 g/d, range=0-152.0) or melons (15.6 g/d, range=0-374.7), more refined grains (81.8 g/d, range=0-589.3) as compared to whole grains (65.6 g/d, range=0-631.4), more red meat (38.7 g/d, range=0-477.9) as compared to other sources of protein and more dark-green vegetables (97.6 g/d, range=0-571.2) as compared to other types of vegetables. Specific food items allotted to each of the food groups are shown in Supplemental Table 1.

In Table 3 we present the average amount of each nutrient consumed per 100 grams of each food item per day, which we use for the second stage predictor matrix for the hierarchical model estimates. The most sparsely populated food constituent was isoflavones, found only in soy products, beans, peas and lentils, other vegetables, yogurt, and sweets. Among the most widely populated constituents were beta-carotene, calcium, carbohydrate, choline, folate, and protein. The foods containing the highest amounts of cholesterol were eggs (330 mg/100g/d) and butter (214.2 mg/100g/d), and the highest amounts of total fat were butter (81.1 g/100g/d), oils and dressings (57.8 g/100g/d) and nuts and seeds (50.9 g/100g/d). The foods containing the highest amounts protein were fish high in omega-3 fatty acids (23.1 µg/100g/d), nuts and seeds (20.6 µg/100g/d), and red meat (19.7 µg/100g/d), and those foods with highest amounts of fiber were nuts and seeds (7.8 g/100g/d), beans, peas and lentils (3.9 g/100g/d), refined grains (3.9 g/100g/d), and whole grains (3.5 g/100g/d).

As shown in Table 4, the estimates that were generated from the hierarchical regression models as compared to the conventional models were generally more precise. The confidence limit ratios (CLR) associated with the fully adjusted hierarchical models for the majority of food items were considerably lower, and in some cases by more than 2-fold (e.g. butter conventional CLR = 43.25 vs. butter hierarchical CLR = 8.14).

Whole grain consumption was inversely associated with each of a 9 and 15% decreased risk of all-cause mortality per 100-gram increase of consumption, as shown by the fully adjusted conventional and hierarchical regression models, respectively [conventional HR=0.91, 95% CI=0.80-1.03; hierarchical HR=0.85, 95% CI=0.73-0.98]. The fully adjusted association for soy product consumption on all-cause mortality was strengthened and flipped by the hierarchical models as compared to the conventional models [conventional HR=1.08, 95% CI=0.93-1.25; hierarchical HR=0.62, 95% CI=0.39-1.00]. The minimally adjusted estimates were also similar. Intake of nuts and seeds was inversely associated with all-cause mortality in both conventional and hierarchical models, and though the estimate for the hierarchical models was more precise, the CI contained the null value [nuts and seeds conventional HR=0.70, 95% CI=0.36-1.39; nuts and seeds hierarchical HR=0.70, 95% CI=0.39-1.25]. Other food items such as fish with high omega-3 fatty acids, and allium vegetables, berries and sweets had estimates suggesting possible associations, though the confidence intervals were wide and, in all cases, contained the null values. While butter consumption had the strongest association with all-cause mortality among both the conventional and hierarchical models, the confidence intervals were extremely wide and in the case of the hierarchical model, included the null [conventional HR=0.14, 95% CI=0.02-0.94; hierarchical HR=0.66, 95% CI=0.23-1.88].

Egg consumption was the food most strongly associated with greater risk of all-cause mortality. The minimally adjusted conventional and hierarchical models suggested a 50 and 42% increased risk of all-cause mortality per 100-gram increase in consumption, respectively [conventional HR=1.50, 95% CI=0.90-2.51; hierarchical HR=1.42, 95% CI=0.86-2.33]. The fully adjusted models attenuated the estimates suggesting a 30 and 19% increased risk of all-cause mortality per 100-gram increase in consumption [conventional HR=1.30, 95% CI=0.76,2.24; hierarchical HR=1.19, 95% CI=0.70-2.03]. Though the hierarchical models in both cases slightly increased the precision, the CIs were wide and contained the null values.

The nutrient-specific fixed effects from the minimally and fully adjusted hierarchical models are shown in Table 5. Among the fully adjusted models, the most notable inverse associations were observed on iron, isoflavones and fiber, each suggesting a 36, 24 and 15% decreased risk of all-cause mortality per 10-milligram increase of consumption per day, respectively [iron per 10mg HR=0.64, 95% CI=0.25,1.62; isoflavones per 10mg HR=0.76, 95% CI=0.57-1.00; fiber per 10g HR=0.85, 95% CI=0.57-1.25]. However, the estimates for iron and fiber were more imprecise. The strongest adverse association observed among the fully adjusted models was on vitamin B12 [per 10μg HR=2.35, 95% CI=0.62-8.85], though the confidence interval was also wide and contained the null value.

3.5 Discussion

This, to our knowledge, is the first study to apply a hierarchical modelling approach to examine the relationship between specific foods items and breast cancer survival, which allowed us to account for the effect of the nutrients contained in the food exposures. In

this prospective cohort study of 3,660 women diagnosed with invasive breast cancer, we observed a decrease in risk of all-cause mortality with increased consumption of whole grains, soy products and nuts and seeds at baseline. We also noted other associations, such as an inverse association between butter and all-cause mortality, and an increased risk of all-cause mortality with a greater consumption of eggs, though for these the confidence intervals were also wide. Among the nutrients, iron and isoflavone consumption were associated with a decreased risk of all-cause mortality and vitamin B12 was associated with an increased risk.

Post-diagnostic consumption of whole grains, nuts, tuna and other fish have each been shown to be associated with a reduced risk of all-cause mortality after a breast cancer diagnosis in other works.[61, 80, 121] We did not observe notable associations in regard to fish in our population, however we did find noteworthy relationships between whole grains and nuts and seeds and survival after breast cancer. As shown in Table 3, both whole grains and nuts and seeds contain high levels of fiber, which suggested a 15% decreased risk on all-cause mortality in our fully adjusted hierarchical model (though the estimate was imprecise). However, this finding is supported by a recent meta-analysis of 7 prospective cohort studies which together showed a 37% reduced risk of all-cause mortality among breast cancer survivors.[122] There are several mechanistic pathways by which dietary fiber might mitigate breast cancer progression, including a reduction of insulin-like growth factor bioactivity, decreasing circulating estrogens, and reducing inflammation.[123] The whole grains and nuts and seeds groups also contained high levels of folate and selenium, both nutrients which have been shown to reduce the risk of death after breast cancer,[124-126] and nuts and seeds alone also contained the highest levels of vitamin E α-tocopherol, which has been previously shown to induce apoptosis in cancer cells.[127] While selenium did suggest a weak association in our analysis [per 10μg HR=0.91, 95% CI=0.77-1.07], neither folate or vitamin E α-tocopherol showed any signs of an association.

Prior research evaluating food items and their relationships with breast cancer survival outcomes is limited, however perhaps the most prolific study results from any food group evaluated among breast cancer survivors are soybean products.[128-130] Soybeans contain isoflavones, which have antiestrogenic and anticancer properties and their consumption both pre and post diagnosis has been consistently shown to be associated with reduced risk of breast cancer survival outcomes.[4, 128, 131, 132] As shown in Table 3, the soy products from this study is the only food group that contained a significant amount of isoflavones (24.3 mg/100g), other than beans, peas and lentils, which had the second highest amount (3.4 mg/100g). This together with the finding that isoflavones suggested a 24% decreased risk in all-cause mortality from our fully adjusted hierarchical model, may explain this inverse relationship that we observed with soy products on all-cause mortality.

Though egg consumption was associated with an increased risk of death from any cause after breast cancer diagnosis, there is limited research to support this finding. However, egg consumption in our cohort contained on average 330 mg of cholesterol per 100 grams of egg consumed, the highest of any food group (Table 3). The dietary

cholesterol limits of 300 mg/day was recently removed from the 2015 Dietary Guidelines for Americans[133] due to recent findings that showed eggs to be unassociated with cardiovascular disease (CVD) among healthy individuals.[134, 135] However, it is still not clear if dietary cholesterol could lead to CVD among unhealthy individuals and there is evidence to support continued concern around dietary cholesterol consumption and breast cancer risk.[136-138]

The strengths of this study included drawing from a large population of women newly diagnosed with breast cancer, prospective data collection with a long follow-up period, and comprehensive measures of dietary exposures, outcomes, and covariates. We also used hierarchical regression, a Bayesian approach that allowed us to incorporate prior knowledge into our parameter estimates.[111] This technique improves upon conventional modelling by incorporating multiple correlated food items into a single model while also considering the nutrient effects. By doing so, we were also able to address issues of multiple comparisons and collinearity that often plague traditional approaches.[111] Limitations of this study include the potential for measurement error from the use of FFQs as well the possibility of residual confounding from unmeasured sources. It is also always possible that participants who chose to enroll in this study were systematically different than those who did not. However, when comparing the enrolled with the unenrolled, the differences in age, race, ethnicity, BMI, and cancer stage were minimal. While our study only made use of a single dietary measure at baseline, future studies may want to consider performing a similar analysis incorporating multiple dietary measures to address dietary changes that could occur after diagnosis.

In summary, this analysis is the first to examine the associations of multiple foods on breast cancer survival while simultaneously taking the effect of nutrients into consideration. Our findings suggest that consuming increased amounts of whole grains, nuts and seeds and soy products (foods with higher levels of fiber, isoflavones, folate and selenium) at or around the time of a breast cancer diagnosis may lead to longer survival.

3.6 Tables

Table 1. Baseline demographic and clinical characteristics (n=3,660)

Continuous factors	Mean (SD)
Age at diagnosis (years)	59.7 (11.9)
Physical activity (MET h/wk)‡	53.8 (35.8)
BMI (kg/m^2)	28.5 (6.7)
Energy intake (kcal/dy)	1466.2 (568.4)
Categorical factors	**No. of Women (%)**
Race/ethnicity	
White	2491 (68.1)
Black	240 (6.6)
Asian/Pacific Islander	475 (13.0)
Hispanic	378 (10.3)
American Indian/Alaska Native	76 (2.1)
Education	
High school or less	547 (14.9)
Some college	1245 (34.0)
College graduate	1024 (28.0)
Post graduate	842 (23.0)
Unknown	2 (0.1)
Menopausal status	
Premenopausal	1060 (29.0)
Postmenopausal	2600 (71.0)
Smoking status	
Never	2092 (57.2)
Former	1408 (38.5)
Current	154 (4.2)
Unknown	6 (0.2)
Cancer stage	
I	2008 (54.9)
II	1250 (34.2)
III	346 (9.5)
IV	56 (1.5)
ER status	
Positive	3072 (83.9)
Negative	586 (16.0)
Unknown	2 (0.1)
PR status	
Positive	2347 (64.1)
Negative	1308 (35.7)
Unknown	5 (0.1)
HER2 status	
Positive	472 (12.9)
Negative	3045 (83.2)
Unknown	143 (3.9)
Surgery type	
Lumpectomy	2180 (59.6)
Mastectomy	1361 (37.2)
None	117 (3.2)
Unknown	2 (0.1)
Chemotherapy	
No	1938 (53.0)
Yes	1711 (46.7)
Unknown	11 (0.3)
Radiation therapy	
No	2036 (55.6)
Yes	1624 (44.4)
Hormonal therapy	
No	902 (24.6)
Yes	2733 (74.7)
Unknown	25 (0.7)

Abbreviations: SD (standard deviation), MET (metabolic equivalent task), (BMI (body mass index), ER (estrogen receptor), PR (progesterone receptor), HER2 (human epidermal growth factor receptor 2).
‡There were 4 participants with unknown physical activity.

Table 2. Distribution of food intake at baseline (N=3,660)

Food group (g/day)	Mean	SD	Min	5%	%25	%50	75%	95%	Max
Dairy									
Butter	3.0	4.7	0.0	0.0	0.1	0.7	4.1	12.8	42.6
Cheese	19.2	26.4	0.0	0.0	3.4	11.3	24.0	62.5	460.4
Milk	140.3	162.1	0.0	3.0	31.2	88.2	179.6	528.4	966.2
Yogurt	32.1	43.4	0.0	0.0	1.9	16.1	49.0	113.0	227.0
Fruit									
Berries	12.8	19.0	0.0	0.0	1.2	5.5	15.8	52.8	152.0
Melons	15.6	26.6	0.0	0.0	2.3	5.9	17.1	63.6	374.7
Other	118.3	93.3	0.0	14.8	45.0	93.8	170.0	296.4	649.6
Fruit juice	65.9	103.6	0.0	0.0	4.1	19.1	90.2	253.1	980.0
Grains									
Refined	81.8	63.8	0.0	14.1	37.8	65.3	106.5	209.2	589.3
Whole	65.6	71.3	0.0	1.6	16.8	42.0	90.7	208.4	631.4
Oils and dressings	20.8	13.7	0.0	2.8	11.0	18.2	28.2	46.3	104.6
Protein									
Eggs	12.6	15.1	0.0	0.0	2.9	7.7	14.3	43.3	150.0
Fish (high omega-3)	7.7	12.6	0.0	0.0	0.0	3.8	8.8	32.5	167.0
Fish (low omega-3)	5.9	10.0	0.0	0.0	0.0	1.9	8.1	32.5	98.7
Fish (processed)	8.6	13.0	0.0	0.0	1.7	4.0	9.9	31.2	188.9
Fish (shellfish)	3.6	6.1	0.0	0.0	0.5	2.0	4.6	12.7	104.0
Meat (red)	38.7	38.1	0.0	1.9	13.8	28.7	51.6	108.6	477.9
Meat (poultry)	31.4	38.5	0.0	1.2	8.3	19.6	40.8	98.1	427.5
Nuts and seeds	11.4	13.8	0.0	0.3	2.0	6.2	16.0	38.2	127.6
Soy products	13.8	52.5	0.0	0.0	0.0	1.2	5.3	66.8	891.7
Sweetened beverages	239.3	323.7	0.0	0.0	17.9	106.3	340.0	948.9	2870.3
Sweets	22.2	23.3	0.0	1.3	7.0	15.6	29.7	65.1	234.0
Vegetables									
Allium	10.6	13.3	0.0	0.0	2.0	7.6	14.9	38.2	208.7
Beans, peas and lentils	35.2	42.2	0.0	2.9	11.2	22.7	42.9	108.9	636.1
Dark-green	97.6	79.5	0.0	9.4	39.1	75.5	132.4	266.6	571.2
Red and orange	41.9	45.8	0.0	3.1	12.6	27.5	53.2	130.5	488.7
Starchy	30.9	27.8	0.0	3.5	12.1	23.5	40.9	82.5	255.2
Other	84.8	92.4	0.0	8.1	26.5	54.6	108.9	254.4	1143.6

Table 3. Second stage design matrix (Z) for hierarchical models

		Nutrient (per day)																
Food group (100g/day)	N	Beta-carotene (µg)	Calcium (mg)	Carbohydrate (g)	Cholesterol (mg)	Choline (mg)	Fat (Total, g)	Fiber (g)	Folate (µg)	Iron (mg)	Isoflavones (mg)	Protein (g)	Retinol (µg)	Selenium (µg)	Vitamin B12 (µg)	Vitamin C (mg)	Vitamin D (µg)	Vitamin E: (α-tocopherol, mg)
Dairy																		
Butter	2783	157.0	23.2	0.1	214.2	18.9	81.1	0.0	1.0	0.0	0.0	0.8	670.0	0.9	0.2	0.0	1.4	2.3
Cheese	3471	49.6	537.6	4.7	60.2	20.0	19.1	0.1	13.7	0.4	0.0	19.1	156.9	13.3	1.1	0.1	0.8	0.3
Milk	3575	11.3	114.4	8.6	13.6	17.5	3.7	0.2	5.4	0.1	0.0	3.4	64.5	3.0	0.6	0.4	0.9	0.1
Yogurt	3023	3.0	154.0	15.6	5.0	15.1	1.2	0.0	10.0	0.1	0.2	4.5	29.0	3.7	0.5	1.4	0.5	0.0
Fruit																		
Berries	3450	19.0	14.0	9.7	0.0	6.1	0.3	2.4	19.0	0.4	0.0	0.7	0.0	0.3	0.0	43.2	0.0	0.4
Melons	3306	1377.3	8.2	7.9	0.0	6.3	0.2	0.7	14.3	0.2	0.0	0.8	0.0	0.4	0.0	26.0	0.0	0.1
Other	3647	85.9	12.2	15.9	0.0	6.7	0.3	2.1	12.4	0.2	0.0	0.7	0.0	0.4	0.0	16.3	0.0	0.2
Fruit juice	3008	15.2	68.9	12.3	0.4	5.1	0.2	0.3	13.1	0.2	0.0	0.6	0.9	0.2	0.0	29.0	0.2	0.1
Grains																		
Refined	3652	56.2	115.4	43.0	9.6	17.8	6.1	3.9	137.2	5.4	0.0	7.7	96.3	15.8	1.2	4.2	1.2	1.1
Whole	3503	7.2	77.6	26.5	0.8	14.5	2.5	3.5	34.8	1.9	0.0	5.9	38.0	16.4	0.1	0.2	0.1	0.3
Oils and dressings	3650	62.9	12.5	6.3	7.8	6.9	57.8	0.1	1.2	0.4	0.0	0.5	72.8	1.2	0.1	0.4	0.1	7.3
Protein																		
Eggs	3404	51.0	87.0	2.2	330.0	195.6	13.8	0.1	32.0	1.5	0.0	12.3	155.0	26.9	0.9	0.9	1.6	1.2
Fish (high omega-3)	2703	9.0	45.0	1.6	61.0	98.9	8.5	0.1	9.0	0.8	0.0	23.1	50.0	41.9	3.6	1.6	10.8	1.1
Fish (low omega-3)	2601	94.0	27.0	3.5	51.0	59.8	4.5	0.3	15.0	0.7	0.0	17.2	25.0	30.7	1.8	4.4	2.5	0.9
Fish (processed)	3165	52.9	33.7	9.4	35.9	38.4	8.8	0.5	16.8	1.1	0.0	15.7	21.1	39.2	1.6	1.0	2.2	0.9
Fish (shellfish)	2982	135.4	62.1	7.7	92.0	62.3	5.9	0.6	25.8	1.8	0.0	12.3	36.4	29.7	3.1	4.6	0.1	1.3
Meat (red)	3508	62.0	57.1	7.3	64.9	68.3	11.4	0.7	22.0	1.9	0.0	19.7	81.9	21.9	1.9	1.6	0.3	0.4
Meat (poultry)	3499	121.1	24.6	6.6	62.2	49.6	8.3	0.5	15.1	1.0	0.0	19.5	16.5	20.4	0.2	1.9	0.1	0.5
Nuts and seeds	3548	7.9	101.2	22.8	0.0	55.9	50.9	7.8	81.3	3.5	0.0	20.6	2.7	28.2	0.0	0.5	0.0	12.4
Soy products	2216	75.1	89.4	12.6	0.5	34.1	5.1	2.3	47.6	1.5	24.3	8.6	5.0	8.3	0.3	1.4	0.1	0.2
Sweetened beverages	3302	7.6	4.5	5.2	0.0	0.4	0.0	0.0	0.6	0.1	0.0	0.1	1.4	0.0	0.0	4.4	0.0	0.0
Sweets	3615	185.4	45.1	61.2	12.4	17.3	12.0	1.6	25.5	1.5	0.1	3.5	40.7	4.5	0.1	1.9	0.1	0.8
Vegetables																		
Allium	3519	46.6	37.2	11.5	0.1	7.8	0.5	1.7	18.8	0.4	0.0	1.6	2.9	1.7	0.0	8.8	0.0	0.1
Beans, peas & lentils	3633	214.5	44.5	12.7	3.6	22.3	3.5	3.9	40.4	1.5	3.4	5.2	4.7	3.6	0.1	3.7	0.0	0.5
Dark-green	3646	1537.4	45.0	3.9	0.7	14.3	0.8	1.9	63.7	0.8	0.0	1.6	3.0	0.7	0.0	21.7	0.0	0.8
Red and orange	3646	3967.8	21.0	9.1	0.4	9.6	0.9	2.1	15.0	0.5	0.0	1.1	3.4	0.2	0.0	11.8	0.0	0.7
Starchy	3639	29.4	23.3	25.2	7.2	19.9	9.4	2.3	24.7	0.8	0.0	3.3	16.0	2.3	0.1	9.5	0.1	1.4
Other	3644	457.6	24.7	7.7	4.4	15.0	5.0	2.3	34.2	0.6	0.7	2.5	5.4	1.8	0.1	13.1	0.0	0.8

Abbreviations: g (grams), mg (milligrams), µg (microgram)

Table 4. Minimally and fully adjusted hazard ratios and 95% confidence intervals from both conventional and hierarchical Cox proportional hazards models for food group intake in association with all-cause mortality

	Minimally Adjusted				Fully Adjusted			
	Conventional		Hierarchical		Conventional		Hierarchical	
Food group (100g/day)	HR (95% CI)	CLR	HR (95% CI)	CLR	HR (95% CI)	CLR	HR (95% CI)	CLR
Dairy								
Butter	0.29 (0.05,1.81)	38.90	1.30 (0.46,3.64)	7.89	0.14 (0.02,0.94)	43.25	0.66 (0.23,1.88)	8.14
Cheese	0.98 (0.74,1.31)	1.78	1.12 (0.76,1.64)	2.15	0.96 (0.70,1.32)	1.89	1.00 (0.67,1.50)	2.24
Milk	0.96 (0.91,1.01)	1.10	1.00 (0.94,1.07)	1.14	0.98 (0.93,1.03)	1.11	0.98 (0.92,1.04)	1.14
Yogurt	1.01 (0.85,1.21)	1.44	1.01 (0.85,1.21)	1.42	1.03 (0.85,1.24)	1.46	1.01 (0.85,1.22)	1.44
Fruit								
Berries	0.68 (0.40,1.14)	2.83	0.78 (0.52,1.18)	2.29	0.77 (0.45,1.32)	2.93	0.83 (0.54,1.27)	2.33
Melons	1.23 (0.93,1.62)	1.75	1.21 (0.92,1.58)	1.71	1.11 (0.83,1.48)	1.79	1.09 (0.83,1.43)	1.73
Other	1.02 (0.92,1.12)	1.21	1.02 (0.93,1.12)	1.21	1.01 (0.92,1.11)	1.21	1.01 (0.92,1.12)	1.21
Fruit juice	1.06 (0.99,1.14)	1.16	1.06 (0.98,1.14)	1.16	1.02 (0.94,1.10)	1.17	1.02 (0.94,1.11)	1.18
Grains								
Refined	0.99 (0.85,1.15)	1.35	1.02 (0.86,1.19)	1.38	1.07 (0.91,1.26)	1.39	1.08 (0.91,1.28)	1.41
Whole	0.91 (0.80,1.03)	1.28	0.83 (0.72,0.97)	1.35	0.91 (0.80,1.03)	1.28	0.85 (0.73,0.98)	1.36
Oils and dressings	1.29 (0.68,2.47)	3.65	1.12 (0.66,1.91)	2.88	1.02 (0.52,2.02)	3.91	0.94 (0.54,1.63)	3.00
Protein								
Eggs	1.50 (0.90,2.51)	2.78	1.42 (0.86,2.33)	2.71	1.30 (0.76,2.24)	2.95	1.19 (0.70,2.03)	2.90
Fish (high omega-3)	0.60 (0.26,1.40)	5.48	0.76 (0.43,1.37)	3.21	0.63 (0.27,1.46)	5.37	0.91 (0.51,1.62)	3.20
Fish (low omega-3)	0.80 (0.31,2.04)	6.57	0.84 (0.48,1.46)	3.02	1.02 (0.39,2.64)	6.74	0.93 (0.53,1.62)	3.07
Fish (processed)	1.43 (0.76,2.72)	3.59	1.07 (0.65,1.77)	2.74	1.44 (0.72,2.88)	4.02	1.04 (0.61,1.76)	2.87
Fish (shellfish)	1.00 (0.98,1.01)	1.03	1.16 (0.77,1.75)	2.26	1.00 (0.99,1.02)	1.03	1.00 (0.66,1.50)	2.26
Meat (red)	1.05 (0.83,1.33)	1.60	1.06 (0.85,1.33)	1.57	0.92 (0.73,1.17)	1.61	0.95 (0.76,1.20)	1.58
Meat (poultry)	1.09 (0.88,1.35)	1.53	1.07 (0.86,1.33)	1.55	1.12 (0.90,1.38)	1.54	1.07 (0.85,1.34)	1.57
Nuts and seeds	0.51 (0.27,0.99)	3.71	0.51 (0.29,0.89)	3.06	0.70 (0.36,1.39)	3.88	0.70 (0.39,1.25)	3.19
Soy products	1.06 (0.92,1.22)	1.33	0.61 (0.38,0.96)	2.53	1.08 (0.93,1.25)	1.34	0.62 (0.39,1.00)	2.57
Sweetened beverages	1.03 (1.00,1.05)	1.05	1.03 (1.00,1.05)	1.05	1.02 (0.99,1.04)	1.05	1.01 (0.98,1.04)	1.06
Sweets	0.81 (0.56,1.19)	2.13	0.83 (0.60,1.16)	1.94	0.84 (0.57,1.25)	2.19	0.87 (0.62,1.22)	1.97
Vegetables								
Allium	0.88 (0.47,1.64)	3.50	0.88 (0.56,1.39)	2.49	0.92 (0.49,1.75)	3.60	0.91 (0.57,1.45)	2.54
Beans, peas and lentils	0.99 (0.80,1.23)	1.53	0.90 (0.73,1.12)	1.54	1.03 (0.84,1.27)	1.51	0.94 (0.76,1.17)	1.53
Dark-green	0.87 (0.76,1.00)	1.31	0.89 (0.79,1.01)	1.29	0.92 (0.80,1.06)	1.32	0.94 (0.82,1.08)	1.31
Red and orange	1.02 (0.85,1.22)	1.43	1.05 (0.87,1.28)	1.48	1.08 (0.90,1.30)	1.44	1.10 (0.90,1.34)	1.50
Starchy	1.20 (0.90,1.60)	1.76	1.12 (0.86,1.47)	1.71	1.04 (0.77,1.40)	1.82	1.08 (0.82,1.43)	1.74
Other	0.99 (0.90,1.10)	1.23	0.99 (0.89,1.10)	1.24	1.03 (0.93,1.14)	1.23	1.02 (0.92,1.13)	1.23

Abbreviations: g (grams), HR (hazard ratio), CI (confidence interval) CLR (conficence limit ratio)

*All models adjusted for age at diagnosis, race/ethnicity, education, body mass index, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2, surgery type, chemotherapy, radiation and hormonal therapy

Table 5. Minimally and fully adjusted hazard ratios and 95% confidence intervals from hierarchical Cox proportional hazards models for nutrient intakes in association with all-cause mortality, among Pathways participants

	Minimally Adjusted	Fully Adjusted
Nutrient (10 units/day)	HR (95% CI)*	HR (95% CI)†
Beta-carotene (µg)	1.00(1.00,1.00)	1.00(1.00,1.00)
Calcium (mg)	1.00(0.99,1.01)	1.00(0.99,1.01)
Carbohydrate (g)	1.01(0.97,1.05)	1.00(0.95,1.04)
Cholesterol (mg)	1.04(0.99,1.09)	1.02(0.97,1.07)
Choline (mg)	0.96(0.89,1.03)	0.98(0.91,1.06)
Fat (Total) (g)	1.03(0.91,1.17)	0.97(0.85,1.10)
Fiber (g)	0.78(0.53,1.14)	0.85(0.57,1.25)
Folate (µg)	1.01(0.97,1.05)	1.02(0.99,1.06)
Iron (mg)	0.93(0.38,2.30)	0.64(0.25,1.62)
Isoflavone (mg)	0.75(0.57,0.98)	0.76(0.57,1.00)
Protein (g)	1.04(0.75,1.45)	1.02(0.73,1.44)
Retinol (µg)	0.99(0.98,1.00)	0.99(0.98,1.01)
Selenium (µg)	0.90(0.77,1.06)	0.91(0.77,1.07)
Vitamin B12 (µg)	3.17(0.85,11.73)	2.35(0.62,8.85)
Vitamin C (mg)	0.99(0.96,1.03)	0.97(0.94,1.01)
Vitamin D (µg)	0.97(0.64,1.46)	1.19(0.79,1.80)
Vitamin E: α-tocopherol (mg)	0.74(0.37,1.50)	1.08(0.53,2.20)

Abbreviations: g (grams), mg (miligrams), µg (micrograms), HR (hazard ratio), CI (confidence interval).

*All models adjusted for age at diagnosis.

†All models adjusted for age at diagnosis, race/ethnicity, education, body mass index, physical activity, smoking, cancer stage, estrogen receptor status, progesterone receptor status and human epidermal growth factor receptor 2, surgery type, chemotherapy, radiation and hormonal therapy.

3.7 Supplementary Tables

Supplemental Table S1. Categorization of individual food items into groups

Food group (g/day)	Food Items (g/day)
Dairy	
Butter	Butter at table, butter for cooking, ghee
Cheese	Cottage cheese, ricotta cheese, low-fat cheese, full-fat cheese
Milk	Ice cream (low and full fat), milk (whole, reduced fat, low fat, non-fat), creams
Yogurt	All yogurts
Fruit	
Berries	All berries
Melons	All melons
Other	Bannanas, apples, pears, peaches, nectarines, fruit salad, canned fruit, apple sauce
Fruit juice	All fruit juices
Grains	
Refined	Pancakes, waffles, biscuits, scones, croissants, non-whole grain breads, bagels, sandwich buns, tortillas, pastas, noodles, non-whole grain cereals, rice milk, chips (not potato), crackers, pizza
Whole	Brown rice, whole grain breads, oatmeal, grits
Oils and dressings	Salad dressings (low and high fat), margarine, mayonnaise, olive oil, vegetable oil
Protein	
Eggs	All eggs
Fish (high omega-3)	All high omega-3 fish
Fish (low omega-3)	All other fish
Fish (processed)	Tuna fish, fish sticks, fish sandwich, fried fish
Fish (shellfish)	Oysters, other shellfish
Meat (red)	Hamburger, meat loaf, steak, taco/burrito/enchilada meat, ribs, pork chops, liver, feet, neck, tail, tongue, lamb, goat, game, menudo
Meat (poultry)	Chicken, turkey, all other poultry
Nuts and seeds	Almonds, pecans, cashew, sesame, sunflower, peanuts
Soy products	Tofu, tempeh, meat substitutes, veggie meats, soy milk, miso soup, soy nuts, roasted soy beans
Sugar sweetened beverages	Hi-C, Tang, drinks with some juice, ice tea, koolaid, soda, diet soda
Sweets	Donuts, cakes, cookies, pies, chocolate, candy, jam, jelly, sugar, honey
Vegetables	
Allium	Onions, leeks, green onions, garlic
Beans, peas and lentils	Green beans, peas, refried beans, hummus, lentils, split peas
Dark-green	Green salad, spinach, broccoli, bok choy, cabbages
Red and orange	Carrots, tomatoes, tomato juice, sweet potatos, yams, winter squash, butternut squash
Starchy	Corn, french fries, fried potatoes, white potatoes, potato chips, potato salad
Other	Cauliflower, sprouts, cole slaw, avocados, summer squash, zucchini, vegetable soup/stew

Abbreviations: g (grams)

IV. CONCLUSION

Aside from skin cancer, breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among women globally.[78] In recent years there has been much progress in early detection and improved treatments, which has led to a dramatic increase in the number of survivors living in the US and around the world.[139] Research has shown that women are motivated to make lifestyle changes at or around the time of their breast cancer diagnosis and desire more information grounded in science on diet and lifestyle choices that may improve their prognosis.[33, 34]

Current dietary and lifestyle guidelines designed to improve prognosis are primarily based on research findings related to breast cancer incidence, not survival.[4] Due to post diagnostic factors such as altered physiology, cancer treatments, increased awareness of symptoms, and modified dietary and lifestyle behaviors, it is not likely that the mechanisms contributing to breast cancer etiology are the same as those that lead to recurrence and death. While their currently exists some evidence supporting the consumption of certain foods with high amounts of fiber and soy, as well as low amounts of total and saturated fats, leading to improved survival for women with breast cancer, it must be evaluated in light of its limitations.[4] Most notably, are the lack of data from randomized controlled trials, the absence of causal or other advanced statistical approaches to analysis, as well the scarcity of relevant factors such as repeated dietary measures and clinical information such as cancer subtypes and treatments.

This dissertation leveraged data from the Pathways Study, a prospective cohort of 4,505 breast cancer survivors enrolled at Kaiser Permanente Northern California at or around the time of their diagnosis. Enrollment began in 2006 and participants were followed for 13 years, during which surveys were administered at enrollment, 6, 24 and 72 months. In addition to the surveys, these analyses utilized KPNC's rich clinical and administrative databases, including both demographic and clinical characteristics such as tumor staging, tumor size, hormone receptor status, and treatment. These databases also provided rapid ascertainment of both breast cancer recurrence and mortality.

The first chapter examined the relationship between four *a priori* dietary quality indices consistent with healthy eating recommendations around the time of breast cancer diagnosis and breast cancer recurrence, cause-specific mortality, and all-cause mortality. The dietary quality indices included an index based on the American Cancer Society nutrition guidelines (ACS),[30] the alternate Mediterranean Diet Index (aMED),[32] an index based on the Dietary Approaches to Stop Hypertension diet (DASH),[140] and the 2015 Healthy Eating Index (HEI).[69] Participants who reported consuming diets that were more concordant with aforementioned healthy eating patterns, were found to be at lower risk of non-breast cancer-specific and all-cause mortality. No clear patterns emerged when examining the associations between the dietary quality indices and breast cancer recurrence or breast cancer-specific mortality. The food components of each index were also exmained and we found independent inverse associations with all-cause mortality for increased consumption of whole grains in the case of ACS and nuts in the case of aMED, as well as decreased consumption of refined grains and

sodium in the case of HEI. A higher risk was also observed for higher consumption of total fruit in the case of HEI. The adjusted interaction between each of these four dietary quality indices and ER status on all-cause mortality was also evaluated, and suggested a stronger association among patients with ER-positive breast cancer when comparing the highest to lowest quartile of the dietary quality index score. These findings were consistent with one prior study.[65]

The second chapter examined the potential effects of hypothetically intervening on diet quality and lifestyle factors and survival after a breast cancer diagnosis using a causal inference approach. We applied the parametric g-formula to observational data from the Pathways Study to estimate the risk of all-cause mortality under several hypothetical interventions. We first examined the independent effects of dietary quality, as measured by the following indices: the Healthy Plant Based Diet Index (hPDI),[31] ACS and HEI. We then examined the independent effects of hypothetically intervening on physical activity, and smoking. Each intervention was assumed to begin at the time of breast cancer diagnosis and maintained over a 13-year follow-up period. The expected risks under the hypothetical interventions were then compared to the expected risks under no intervention (natural course). Interventions at modest levels of intensity were first considered and then interventions which progressively increased their intensity up to maximum levels were evaluated. We found that the hypothetical interventions that increased diet quality, increased physical activity, and stopped participants from smoking, each reduced the risk of death among breast cancer survivors. We also found that increasing the intensity of the intervention on diet and physical activity was directly related to the strength of the associations. Joint interventions on combinations of diet and lifestyle factors were also evaluated and conveyed the greatest reductions in risk.

In the third chapter, we applied a hierarchical modelling approach to examine the relationship between survival and baseline intake of multiple food items, including diary, fruit, grains, oils and dressings, proteins, added sugars and vegetables. A second level model was specified to explain drivers of the food level effects *via* constituents considered to be related to survival (*e.g.* carbohydrates, protein, fiber, calcium, iron, isoflavones, vitamin C, vitamin D, and others). This approach allowed estimation of the mutually adjusted associations between multiple food items on breast cancer survival, as well as the role of specific nutrients contained in these foods. In this study we observed a decreased risk in all-cause mortality with increased consumption of whole grains, soy products and nuts and seeds at baseline. We also observed an increased risk of all-cause mortality with an increased consumption of eggs. Among the fixed effects, we observed that iron, isoflavone and fiber consumption to be associated with a decreased risk of all-cause mortality, though the estimates for iron and fiber were most imprecise.

Overall, this dissertation has tried to shed light on several of the research gaps that typify the role of diet on breast cancer survival. While the first paper demonstrated that overall diet at the time of a breast cancer diagnosis decreased the risk of all-cause mortality, the third paper highlighted the associations of the individual food items and the constituents that drove those associations. Together, these two papers not only

support the importance of diet on breast cancer prognosis, but also emphasize the need for evaluating diet on both the macro and micro levels. By doing so, we may be able to better inform survivors on specific dietary choices and do so within the context of their overall diets. The second paper attempted to explain what is hypothetically possible if we were to intervene on diet and other lifestyle factors at the time of breast cancer diagnosis. In a field where data from RCTs are mostly unavailable, this analysis provided some insights into the possible benefits of dietary interventions, especially when considered jointly with other lifestyle factors such as physical activity and smoking. Together, these three papers demonstrate at the very least, that considering dietary factors at all levels (overall patterns, foods and nutrients), and in conjunction with other lifestyle factors are an essential part of ensuring the best possible prognosis for women surviving breast cancer.

Future Directions

This dissertation, in conjunction with prior research, provides compelling evidence that nutritional factors at all levels of diet, including overall dietary patterns, specific food groups, as well as the nutrients contained within those foods, predict important outcomes for women living with breast cancer. However, there is still much to be done. Perhaps the most important research gap is the scarcity of randomized controlled trials (RCT) on diet and lifestyle factors and breast cancer survival. RCT's will be critical for producing conclusive results regarding diet and lifestyle choices that can be confidently translated into public guidelines. Designing such studies will require access to financial resources and technical expertise to ensure that the most relevant study questions are being addressed and answered. Because each phase of survivorship (e.g. diagnosis, treatment, recovery, recurrence, etc.) is unique, carefully designed and adequately powered prospective cohort studies will also be necessary to address dietary decision making in the context of the heterogeneous nature of the breast cancer survival experience. Given the complexity of current care and the precision treatments related to cancer subtypes, dietary choices will need to be nuanced to optimize a patient's prognosis at each stage of recovery. In this study we described the associations between diet and breast cancer survival at both the macro (dietary patterns) and micro (foods and nutrient) dietary levels, which provided a more in depth understanding of the impact that diet has on breast cancer outcomes as compared to confining the research to only one dietary level. However, we must also strive to understand how diet interacts with the underlying metabolic and genetic pathways that ultimately lead to the biological changes required for cancer progression and development. Finally, to ensure the accuracy of our reporting, we must strive to utilize the most advanced and up to date statistical methods at our disposal. To do this successfully, we must be willing to detach from dogmatic practices which have proven faulty, continue to educate our researchers on best practices, and borrow and learn from other fields of science.

Summary

Diet as measured by an overall dietary pattern, or hypothetically intervened upon, or examined as individual food items in conjunction with their associated nutrients, plays a

critical role in survival after a breast cancer diagnosis. This dissertation provides important and timely information to support or warrant modification of current dietary recommendations which will ultimately benefit the estimated 3.8 million women currently living with breast cancer in the United States.

V. REFERENCES

1. Institute NC. Cancer Stat Facts: Female Breast Cancer. National Cancer Institute; 2019 [cited 2019]; Available from: https://seer.cancer.gov/statfacts/html/breast.html.
2. Society AC. Cancer Facts & Figures 2021. 2021 [cited 2021 5/31/2021]; Available from: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html.
3. Society AC. Breast Cancer Facts & Figures 2019-2020. 2020 [cited 2021 04/24/2021]; Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf.
4. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective: World Cancer Research Fund/American Institute for Cancer Research2018.
5. Breast cancer causes. American Cancer Society; 2018 [cited 2018]; Available from: https://www.cancer.org/content/dam/CRC/PDF/Public/8577.00.pdf.
6. Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet (London, England). 1998; 351:1451-67.
7. Hortobagyi GN. Treatment of breast cancer. N Engl J Med. 1998; 339:974-84.
8. Miller BA, Hankey BF, Thomas TL. Impact of sociodemographic factors, hormone receptor status, and tumor grade on ethnic differences in tumor stage and size for breast cancer in US women. Am J Epidemiol. 2002; 155:534-45.
9. Saphner T, Tormey DC, Gray R. Annual hazard rates of recurrence for breast cancer after primary therapy. J Clin Oncol. 1996; 14:2738-46.
10. Prevention CfDCa. Heart Disease Facts. 2018 [cited 2021]; Available from: https://www.cdc.gov/heartdisease/facts.htm.
11. Bradshaw PT, Stevens J, Khankari N, Teitelbaum SL, Neugut AI, Gammon MD. Cardiovascular Disease Mortality Among Breast Cancer Survivors. Epidemiology. 2016; 27:6-13.
12. Patnaik JL, Byers T, DiGuiseppi C, Dabelea D, Denberg TD. Cardiovascular disease competes with breast cancer as the leading cause of death for older females diagnosed with breast cancer: a retrospective cohort study. Breast Cancer Res. 2011; 13:R64.
13. Afifi AM, Saad AM, Al-Husseini MJ, Elmehrath AO, Northfelt DW, Sonbol MB. Causes of death after breast cancer diagnosis: A US population-based analysis. Cancer. 2020; 126:1559-67.
14. Accordino MK, Neugut AI, Hershman DL. Cardiac effects of anticancer therapy in the elderly. J Clin Oncol. 2014; 32:2654-61.
15. Yeh ET, Bickford CL. Cardiovascular complications of cancer therapy: incidence, pathogenesis, diagnosis, and management. J Am Coll Cardiol. 2009; 53:2231-47.
16. Willett WC. Micronutrients and cancer risk. The American journal of clinical nutrition. 1994; 59:1162S-5S.
17. Willett WC. Diet and cancer: one view at the start of the millennium. Cancer Epidemiol Biomarkers Prev. 2001; 10:3-8.

18. Elsamanoudy AZ, Mohamed Neamat-Allah MA, Hisham Mohammad FA, Hassanien M, Nada HA. The role of nutrition related genes and nutrigenetics in understanding the pathogenesis of cancer. J Microsc Ultrastruct. 2016; 4:115-22.
19. Bull C, Fenech M. Genome-health nutrigenomics and nutrigenetics: nutritional requirements or 'nutriomes' for chromosomal stability and telomere maintenance at the individual level. Proc Nutr Soc. 2008; 67:146-56.
20. Kaput J. Diet-disease gene interactions. Nutrition. 2004; 20:26-31.
21. Sun Y, Bao W, Liu B, Caan BJ, Lane DS, Millen AE, et al. Changes in Overall Diet Quality in Relation to Survival in Postmenopausal Women with Breast Cancer: Results from the Women's Health Initiative. J Acad Nutr Diet. 2018; 118:1855-63 e6.
22. Vrieling A, Buck K, Seibold P, Heinz J, Obi N, Flesch-Janys D, et al. Dietary patterns and survival in German postmenopausal breast cancer survivors. British journal of cancer. 2013; 108:188-92.
23. Izano MA, Fung TT, Chiuve SS, Hu FB, Holmes MD. Are diet quality scores after breast cancer diagnosis associated with improved breast cancer survival? Nutrition and cancer. 2013; 65:820-6.
24. Kim EH, Willett WC, Fung T, Rosner B, Holmes MD. Diet quality indices and postmenopausal breast cancer survival. Nutrition and cancer. 2011; 63:381-8.
25. Kwan ML, Weltzien E, Kushi LH, Castillo A, Slattery ML, Caan BJ. Dietary patterns and breast cancer recurrence and survival among women with early-stage breast cancer. J Clin Oncol. 2009; 27:919-26.
26. Xing MY, Xu SZ, Shen P. Effect of low-fat diet on breast cancer survival: a meta-analysis. Asian Pac J Cancer Prev. 2014; 15:1141-4.
27. Cespedes EM, Hu FB. Dietary patterns: from nutritional epidemiologic analysis to national guidelines. The American journal of clinical nutrition. 2015; 101:899-900.
28. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol. 2002; 13:3-9.
29. The Nutrition Source. Harvard T.H. Chan School of Public Health; [cited 2019]; Available from: https://www.hsph.harvard.edu/nutritionsource/.
30. Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV, et al. American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin. 2012; 62:30-67.
31. Satija A, Bhupathiraju SN, Spiegelman D, Chiuve SE, Manson JE, Willett W, et al. Healthful and Unhealthful Plant-Based Diets and the Risk of Coronary Heart Disease in U.S. Adults. J Am Coll Cardiol. 2017; 70:411-22.
32. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, et al. Diet and overall survival in elderly people. BMJ. 1995; 311:1457-60.
33. Velentzis LS, Keshtgar MR, Woodside JV, Leathem AJ, Titcomb A, Perkins KA, et al. Significant changes in dietary intake and supplement use after breast cancer diagnosis in a UK multicentre study. Breast cancer research and treatment. 2011; 128:473-82.
34. Salminen E, Heikkila S, Poussa T, Lagstrom H, Saario R, Salminen S. Female patients tend to alter their diet following the diagnosis of rheumatoid arthritis and breast cancer. Prev Med. 2002; 34:529-35.

35. Salminen EK, Lagstrom HK, Heikkila S, Salminen S. Does breast cancer change patients' dietary habits? Eur J Clin Nutr. 2000; 54:844-8.
36. Thomson CA, Flatt SW, Rock CL, Ritenbaugh C, Newman V, Pierce JP. Increased fruit, vegetable and fiber intake and lower fat intake reported among women previously treated for invasive breast cancer. J Am Diet Assoc. 2002; 102:801-8.
37. Runowicz CD, Leach CR, Henry NL, Henry KS, Mackey HT, Cowens-Alvarado RL, et al. American Cancer Society/American Society of Clinical Oncology Breast Cancer Survivorship Care Guideline. J Clin Oncol. 2016; 34:611-35.
38. Kord-Varkaneh H, Salehi-Sahlabadi A, Zarezade M, Rahmani J, Tan SC, Hekmatdoost A, et al. Association between Healthy Eating Index-2015 and Breast Cancer Risk: A Case-Control Study. Asian Pac J Cancer Prev. 2020; 21:1363-7.
39. Xiao Y, Xia J, Li L, Ke Y, Cheng J, Xie Y, et al. Associations between dietary patterns and the risk of breast cancer: a systematic review and meta-analysis of observational studies. Breast Cancer Res. 2019; 21:16.
40. Fung TT, Hu FB, Hankinson SE, Willett WC, Holmes MD. Low-carbohydrate diets, dietary approaches to stop hypertension-style diets, and the risk of postmenopausal breast cancer. Am J Epidemiol. 2011; 174:652-60.
41. Prentice RL, Caan B, Chlebowski RT, Patterson R, Kuller LH, Ockene JK, et al. Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA. 2006; 295:629-42.
42. Chlebowski RT, Aragaki AK, Anderson GL, Simon MS, Manson JE, Neuhouser ML, et al. Association of Low-Fat Dietary Pattern With Breast Cancer Overall Survival: A Secondary Analysis of the Women's Health Initiative Randomized Clinical Trial. JAMA Oncol. 2018; 4:e181212.
43. Printz C. Link between high-fat dairy consumption and poor breast cancer survival. Cancer. 2013; 119:2517.
44. Makarem N, Chandran U, Bandera EV, Parekh N. Dietary fat in breast cancer survival. Annu Rev Nutr. 2013; 33:319-48.
45. Pierce JP, Natarajan L, Caan BJ, Parker BA, Greenberg ER, Flatt SW, et al. Influence of a diet very high in vegetables, fruit, and fiber and low in fat on prognosis following treatment for breast cancer: the Women's Healthy Eating and Living (WHEL) randomized trial. JAMA. 2007; 298:289-98.
46. Thomson CA, Rock CL, Thompson PA, Caan BJ, Cussler E, Flatt SW, et al. Vegetable intake is associated with reduced breast cancer recurrence in tamoxifen users: a secondary analysis from the Women's Healthy Eating and Living Study. Breast cancer research and treatment. 2011; 125:519-27.
47. McEligot AJ, Largent J, Ziogas A, Peel D, Anton-Culver H. Dietary fat, fiber, vegetable, and micronutrients are associated with overall survival in postmenopausal women diagnosed with breast cancer. Nutrition and cancer. 2006; 55:132-40.
48. Kowalski A, Striley CW, Varma DS, Egan KM, Yaghjyan L. Interactions between Alcohol Consumption and Adjuvant Hormone Therapy in Relation to Breast Cancer-Free Survival. Journal of breast cancer. 2018; 21:158-64.
49. Lowry SJ, Kapphahn K, Chlebowski R, Li CI. Alcohol Use and Breast Cancer Survival among Participants in the Women's Health Initiative. Cancer Epidemiol Biomarkers Prev. 2016; 25:1268-73.

50. Ali AM, Schmidt MK, Bolla MK, Wang Q, Gago-Dominguez M, Castelao JE, et al. Alcohol consumption and survival after a breast cancer diagnosis: a literature-based meta-analysis and collaborative analysis of data for 29,239 cases. Cancer Epidemiol Biomarkers Prev. 2014; 23:934-45.
51. Wang F, Cai H, Gu K, Shi L, Yu D, Zhang M, et al. Adherence to Dietary Recommendations among Long-Term Breast Cancer Survivors and Cancer Outcome Associations. Cancer Epidemiol Biomarkers Prev. 2020; 29:386-95.
52. Chlebowski RT, Blackburn GL, Thomson CA, Nixon DW, Shapiro A, Hoy MK, et al. Dietary fat reduction and breast cancer outcome: interim efficacy results from the Women's Intervention Nutrition Study. Journal of the National Cancer Institute. 2006; 98:1767-76.
53. Pierce JP, Faerber S, Wright FA, Rock CL, Newman V, Flatt SW, et al. A randomized trial of the effect of a plant-based dietary pattern on additional breast cancer events and survival: the Women's Healthy Eating and Living (WHEL) Study. Control Clin Trials. 2002; 23:728-56.
54. Kwan ML, Ambrosone CB, Lee MM, Barlow J, Krathwohl SE, Ergas IJ, et al. The Pathways Study: a prospective study of breast cancer survivorship within Kaiser Permanente Northern California. Cancer causes & control : CCC. 2008; 19:1065-76.
55. NutritionQuest. Questionnaires and Screeners. Available from: https://nutritionquest.com/assessment/list-of-questionnaires-and-screeners/.
56. Kroenke CH, Chubak J, Johnson L, Castillo A, Weltzien E, Caan BJ. Enhancing Breast Cancer Recurrence Algorithms Through Selective Use of Medical Record Data. Journal of the National Cancer Institute. 2016; 108.
57. Haque R, Shi J, Schottinger JE, Ahmed SA, Chung J, Avila C, et al. A hybrid approach to identify subsequent breast cancer using pathology and automated health information data. Med Care. 2015; 53:380-5.
58. Hassett MJ, Ritzwoller DP, Taback N, Carroll N, Cronin AM, Ting GV, et al. Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. Med Care. 2014; 52:e65-73.
59. Chubak J, Yu O, Pocobelli G, Lamerato L, Webster J, Prout MN, et al. Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer. Journal of the National Cancer Institute. 2012; 104:931-40.
60. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999; 10:37-48.
61. Khankari NK, Bradshaw PT, Steck SE, He K, Olshan AF, Shen J, et al. Dietary intake of fish, polyunsaturated fatty acids, and survival after breast cancer: A population-based follow-up study on Long Island, New York. Cancer. 2015; 121:2244-52.
62. Pierce JP, Stefanick ML, Flatt SW, Natarajan L, Sternfeld B, Madlensky L, et al. Greater survival after breast cancer in physically active women with high vegetable-fruit intake regardless of obesity. J Clin Oncol. 2007; 25:2345-51.
63. Parada H, Jr., Steck SE, Bradshaw PT, Engel LS, Conway K, Teitelbaum SL, et al. Grilled, Barbecued, and Smoked Meat Intake and Survival Following Breast Cancer. Journal of the National Cancer Institute. 2017; 109.

64. McCullough ML, Gapstur SM, Shah R, Campbell PT, Wang Y, Doyle C, et al. Pre- and postdiagnostic diet in relation to mortality among breast cancer survivors in the CPS-II Nutrition Cohort. Cancer causes & control : CCC. 2016; 27:1303-14.
65. George SM, Ballard-Barbash R, Shikany JM, Caan BJ, Freudenheim JL, Kroenke CH, et al. Better postdiagnosis diet quality is associated with reduced risk of death among postmenopausal women with invasive breast cancer in the women's health initiative. Cancer Epidemiol Biomarkers Prev. 2014; 23:575-83.
66. George SM, Irwin ML, Smith AW, Neuhouser ML, Reedy J, McTiernan A, et al. Postdiagnosis diet quality, the combination of diet quality and recreational physical activity, and prognosis after early-stage breast cancer. Cancer causes & control : CCC. 2011; 22:589-98.
67. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med. 2003; 348:2599-608.
68. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. Arch Intern Med. 2008; 168:713-20.
69. Krebs-Smith SM, Pannucci TE, Subar AF, Kirkpatrick SI, Lerman JL, Tooze JA, et al. Update of the Healthy Eating Index: HEI-2015. J Acad Nutr Diet. 2018; 118:1591-602.
70. Buckland G, Travier N, Cottet V, Gonzalez CA, Lujan-Barroso L, Agudo A, et al. Adherence to the mediterranean diet and risk of breast cancer in the European prospective investigation into cancer and nutrition cohort study. International journal of cancer. 2013; 132:2918-27.
71. Heidari Z, Mohammadi E, Aghamohammadi V, Jalali S, Rezazadeh A, Sedaghat F, et al. Dietary Approaches to Stop Hypertension (DASH) diets and breast cancer among women: a case control study. BMC Cancer. 2020; 20:708.
72. Di Maso M, Maso LD, Augustin LSA, Puppo A, Falcini F, Stocco C, et al. Adherence to the Mediterranean Diet and Mortality after Breast Cancer. Nutrients. 2020; 12.
73. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology. 1995; 6:356-65.
74. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. The American journal of clinical nutrition. 1997; 65:1220S-8S; discussion 9S-31S.
75. R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: https://www.R-project.org/.
76. Sharif Y, Sadeghi O, Benisi-Kohansal S, Azadbakht L, Esmaillzadeh A. Legume and Nuts Consumption in Relation to Odds of Breast Cancer: A Case-Control Study. Nutrition and cancer. 2020:1-10.
77. Sangaramoorthy M, Koo J, John EM. Intake of bean fiber, beans, and grains and reduced risk of hormone receptor-negative breast cancer: the San Francisco Bay Area Breast Cancer Study. Cancer Med. 2018; 7:2131-44.
78. Fund WCR. Breast cancer statistics. 2021 [cited 2021 07202021]; Available from: https://www.wcrf.org/dietandcancer/breast-cancer-statistics/.

79. McBride CM, Ostroff JS. Teachable moments for promoting smoking cessation: the context of cancer care and survivorship. Cancer Control. 2003; 10:325-33.
80. Ergas IJ, Cespedes Feliciano EM, Bradshaw PT, Roh JM, Kwan ML, Cadenhead J, et al. Diet Quality and Breast Cancer Recurrence and Survival: The Pathways Study. JNCI Cancer Spectr. 2021; 5:pkab019.
81. Abrahamson PE, Gammon MD, Lund MJ, Britton JA, Marshall SW, Flagg EW, et al. Recreational physical activity and survival among young women with breast cancer. Cancer. 2006; 107:1777-85.
82. Parada H, Jr., Bradshaw PT, Steck SE, Engel LS, Conway K, Teitelbaum SL, et al. Postdiagnosis Changes in Cigarette Smoking and Survival Following Breast Cancer. JNCI Cancer Spectr. 2017; 1.
83. Kroenke CH, Fung TT, Hu FB, Holmes MD. Dietary patterns and survival after breast cancer diagnosis. J Clin Oncol. 2005; 23:9295-303.
84. de Glas NA, Fontein DB, Bastiaannet E, Pijpe A, De Craen AJ, Liefers GJ, et al. Physical activity and survival of postmenopausal, hormone receptor-positive breast cancer patients: results of the Tamoxifen Exemestane Adjuvant Multicenter Lifestyle study. Cancer. 2014; 120:2847-54.
85. Bradshaw PT, Ibrahim JG, Khankari N, Cleveland RJ, Abrahamson PE, Stevens J, et al. Post-diagnosis physical activity and survival after breast cancer diagnosis: the Long Island Breast Cancer Study. Breast cancer research and treatment. 2014; 145:735-42.
86. Irwin ML, McTiernan A, Manson JE, Thomson CA, Sternfeld B, Stefanick ML, et al. Physical activity and survival in postmenopausal women with breast cancer: results from the women's health initiative. Cancer Prev Res (Phila). 2011; 4:522-9.
87. Duan W, Li S, Meng X, Sun Y, Jia C. Smoking and survival of breast cancer patients: A meta-analysis of cohort studies. Breast. 2017; 33:117-24.
88. Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The parametric g-formula for time-to-event data: intuition and a worked example. Epidemiology. 2014; 25:889-97.
89. Taubman SL, Robins JM, Mittleman MA, Hernan MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. Int J Epidemiol. 2009; 38:1599-611.
90. Staten LK, Taren DL, Howell WH, Tobar M, Poehlman ET, Hill A, et al. Validation of the Arizona Activity Frequency Questionnaire using doubly labeled water. Med Sci Sports Exerc. 2001; 33:1959-67.
91. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Jr., Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc. 2011; 43:1575-81.
92. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998; 36:8-27.
93. Inference HPoC. GFORMULA 3.0 - The parametric g-formula in SAS. [cited 2021 4/24/2021]; Available from: https://www.hsph.harvard.edu/causal/software/.
94. Saxe GA, Major JM, Nguyen JY, Freeman KM, Downs TM, Salem CE. Potential attenuation of disease progression in recurrent prostate cancer with plant-based diet and stress reduction. Integr Cancer Ther. 2006; 5:206-13.

95. Nguyen JY, Major JM, Knott CJ, Freeman KM, Downs TM, Saxe GA. Adoption of a plant-based diet by patients with recurrent prostate cancer. Integr Cancer Ther. 2006; 5:214-23.
96. Huang T, Yang B, Zheng J, Li G, Wahlqvist ML, Li D. Cardiovascular disease mortality and cancer incidence in vegetarians: a meta-analysis and systematic review. Ann Nutr Metab. 2012; 60:233-40.
97. Singh PN, Fraser GE. Dietary risk factors for colon cancer in a low-risk population. Am J Epidemiol. 1998; 148:761-74.
98. Tantamango-Bartley Y, Jaceldo-Siegl K, Fan J, Fraser G. Vegetarian diets and the incidence of cancer in a low-risk population. Cancer Epidemiol Biomarkers Prev. 2013; 22:286-94.
99. Plant-based food retail sales grow 27% to $7 billion in 2020. Plant Based Foods Association; 2021 [updated 4/6/2021; cited 2021 4/12/2021]; Available from: www.plantbasedfoods.org/2020-retail-sales-data-announcement/.
100. The Vegan Society Statistics. The Vegan Society; 2021 [cited 2021 4/12/2021]; Available from: www.vegansociety.com/news/media/statistics.
101. Hu EA, Pan A, Malik V, Sun Q. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. BMJ. 2012; 344:e1454.
102. Gaesser GA. Whole Grains, Refined Grains, and Cancer Risk: A Systematic Review of Meta-Analyses of Observational Studies. Nutrients. 2020; 12.
103. Huang C, Huang J, Tian Y, Yang X, Gu D. Sugar sweetened beverages consumption and risk of coronary heart disease: a meta-analysis of prospective studies. Atherosclerosis. 2014; 234:11-6.
104. Romanos-Nanclares A, Toledo E, Gardeazabal I, Jimenez-Moleon JJ, Martinez-Gonzalez MA, Gea A. Sugar-sweetened beverage consumption and incidence of breast cancer: the Seguimiento Universidad de Navarra (SUN) Project. Eur J Nutr. 2019; 58:2875-86.
105. Koyratty N, McCann SE, Millen AE, Nie J, Trevisan M, Freudenheim JL. Sugar-Sweetened Soda Consumption and Total and Breast Cancer Mortality: The Western New York Exposures and Breast Cancer (WEB) Study. Cancer Epidemiol Biomarkers Prev. 2021.
106. McCullough ML, Patel AV, Kushi LH, Patel R, Willett WC, Doyle C, et al. Following cancer prevention guidelines reduces risk of cancer, cardiovascular disease, and all-cause mortality. Cancer Epidemiol Biomarkers Prev. 2011; 20:1089-97.
107. Cannioto RA, Hutson A, Dighe S, McCann W, McCann SE, Zirpoli GR, et al. Physical Activity Before, During, and After Chemotherapy for High-Risk Breast Cancer: Relationships With Survival. Journal of the National Cancer Institute. 2021; 113:54-63.
108. Delrieu L, Jacquet E, Segura-Ferlay C, Blanc E, Febvey-Combes O, Friedenreich C, et al. Analysis of the StoRM cohort reveals physical activity to be associated with survival in metastatic breast cancer. Sci Rep. 2020; 10:10757.
109. Passarelli MN, Newcomb PA, Hampton JM, Trentham-Dietz A, Titus LJ, Egan KM, et al. Cigarette Smoking Before and After Breast Cancer Diagnosis: Mortality From Breast Cancer and Smoking-Related Diseases. J Clin Oncol. 2016; 34:1315-22.
110. Willett W. Nutritional epidemiology: issues and challenges. Int J Epidemiol. 1987; 16:312-7.

111. Greenland S. When should epidemiologic regressions use random coefficients? Biometrics. 2000; 56:915-21.
112. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology. 1994; 5:612-21.
113. Bradshaw PT, Khankari NK, Teitelbaum SL, Xu X, Fink BN, Steck SE, et al. Nutrient pathways and breast cancer risk: the Long Island Breast Cancer Study Project. Nutrition and cancer. 2013; 65:345-54.
114. Witte JS, Greenland S, Kim LL, Arab L. Multilevel modeling in epidemiology with GLIMMIX. Epidemiology. 2000; 11:684-8.
115. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect. 1994; 102 Suppl 8:33-9.
116. Hamra GB, Buckley JP. Environmental exposure mixtures: questions and methods to address them. Curr Epidemiol Rep. 2018; 5:160-5.
117. United States Department of Agriculture: MyPlate. 2021 [cited 2021 07/11/2021]; Available from: https://www.myplate.gov/eat-healthy/food-group-gallery.
118. Lash TL, Cole SR. Immortal person-time in studies of cancer outcomes. J Clin Oncol. 2009; 27:e55-6.
119. Cole SR, Chu H, Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. Am J Epidemiol. 2014; 179:252-60.
120. Terry M. Therneau PMG. Modeling Survival Data: Extending the Cox Model. Springer, New York: R Software; 2000.
121. Farvid MS, Spence ND, Rosner BA, Willett WC, Eliassen AH, Holmes MD. Post-diagnostic coffee and tea consumption and breast cancer survival. British journal of cancer. 2021; 124:1873-81.
122. Jayedi A, Emadi A, Khan TA, Abdolshahi A, Shab-Bidar S. Dietary Fiber and Survival in Women with Breast Cancer: A Dose-Response Meta-Analysis of Prospective Cohort Studies. Nutrition and cancer. 2020:1-11.
123. Deschasaux M, Zelek L, Pouchieu C, His M, Hercberg S, Galan P, et al. Prospective association between dietary fiber intake and breast cancer risk. PLoS One. 2013; 8:e79718.
124. Szwiec M, Marciniak W, Derkacz R, Huzarski T, Gronwald J, Cybulski C, et al. Serum Selenium Level Predicts 10-Year Survival after Breast Cancer. Nutrients. 2021; 13.
125. Lubinski J, Marciniak W, Muszynska M, Huzarski T, Gronwald J, Cybulski C, et al. Serum selenium levels predict survival after breast cancer. Breast cancer research and treatment. 2018; 167:591-8.
126. McEligot AJ, Ziogas A, Pfeiffer CM, Fazili Z, Anton-Culver H. The association between circulating total folate and folate vitamers with overall survival after postmenopausal breast cancer diagnosis. Nutrition and cancer. 2015; 67:442-8.
127. Kline K, Yu W, Sanders BG. Vitamin E and breast cancer. J Nutr. 2004; 134:3458S-62S.
128. Ho SC, Yeo W, Goggins W, Kwok C, Cheng A, Chong M, et al. Pre-diagnosis and early post-diagnosis dietary soy isoflavone intake and survival outcomes: A prospective cohort study of early stage breast cancer survivors. Cancer Treat Res Commun. 2021; 27:100350.

129. Qiu S, Jiang C. Soy and isoflavones consumption and breast cancer survival and recurrence: a systematic review and meta-analysis. Eur J Nutr. 2019; 58:3079-90.
130. Conroy SM, Maskarinec G, Park SY, Wilkens LR, Henderson BE, Kolonel LN. The effects of soy consumption before diagnosis on breast cancer survival: the Multiethnic Cohort Study. Nutrition and cancer. 2013; 65:527-37.
131. Chi F, Wu R, Zeng YC, Xing R, Liu Y, Xu ZG. Post-diagnosis soy food intake and breast cancer survival: a meta-analysis of cohort studies. Asian Pac J Cancer Prev. 2013; 14:2407-12.
132. Nechuta SJ, Caan BJ, Chen WY, Lu W, Chen Z, Kwan ML, et al. Soy food intake after diagnosis of breast cancer and survival: an in-depth analysis of combined evidence from cohort studies of US and Chinese women. The American journal of clinical nutrition. 2012; 96:123-32.
133. Agriculture. USDoHaHSaUSDo. 2015 – 2020 Dietary Guidelines for Americans. 2015 [cited 2021 07/20/2021]; Available from: https://health.gov/our-work/food-nutrition/previous-dietary-guidelines/2015.
134. Drouin-Chartier JP, Chen S, Li Y, Schwab AL, Stampfer MJ, Sacks FM, et al. Egg consumption and risk of cardiovascular disease: three large prospective US cohort studies, systematic review, and updated meta-analysis. BMJ. 2020; 368:m513.
135. Hu FB, Stampfer MJ, Rimm EB, Manson JE, Ascherio A, Colditz GA, et al. A prospective study of egg consumption and risk of cardiovascular disease in men and women. JAMA. 1999; 281:1387-94.
136. Gonzalez-Ortiz A, Galindo-Hernandez O, Hernandez-Acevedo GN, Hurtado-Ureta G, Garcia-Gonzalez V. Impact of cholesterol-pathways on breast cancer development, a metabolic landscape. J Cancer. 2021; 12:4307-21.
137. Li C, Yang L, Zhang D, Jiang W. Systematic review and meta-analysis suggest that dietary cholesterol intake increases risk of breast cancer. Nutr Res. 2016; 36:627-35.
138. Llaverias G, Danilo C, Mercier I, Daumer K, Capozza F, Williams TM, et al. Role of cholesterol in the development and progression of breast cancer. Am J Pathol. 2011; 178:402-12.
139. Survivors of breast and other cancers: World Cancer Research Fund/American Institute for Cancer Research2018.
140. Siervo M, Lara J, Chowdhury S, Ashor A, Oggioni C, Mathers JC. Effects of the Dietary Approach to Stop Hypertension (DASH) diet on cardiovascular risk factors: a systematic review and meta-analysis. The British journal of nutrition. 2015; 113:1-15.