# Candidate Information and Goals for Career Development

I am a postdoctoral fellow at the National Institute of Environmental Health Sciences within the Biostatistics and Computational Biology branch. **Broadly speaking, my career goal is to transition into an independent research career focusing on the role of co-occurring modifiable factors to prevent disease unique to women transitioning into midlife past their reproductive years.** The sections that follow demonstrate how my background, including work experience and training, have supported my career goals, led me to the research proposed here, and informed my next steps outlined in the training objectives.

## Candidate background

1. **Prior graduate and applied research**

My masters-level academic degrees in population health, demography, and biostatistics converged into my position as a biostatistician at the University of North Carolina, Department of Psychiatry for more than eight years. In this role, I collaborated with a diverse group of researchers ranging from undergraduate students to faculty in psychiatry and psychology who were writing manuscripts based on both clinical and observational data. This opportunity allowed me to advance my statistical analysis skills in structural equation modeling, learn all aspects of manuscript writing, and observe principal investigators engage in successful grant writing. In applying advanced analytic methods to data from population health studies, these rich experiences motivated my long term goal to establish myself as an independent investigator focusing on public health outcomes.

1. **Dissertation research (University of North Carolina)**

In a decision to further my career goals to become an independent researcher in public health, I enrolled in the doctoral program in Epidemiology at the University of North Carolina, Chapel Hill. My coursework in epidemiology enabled me to study methods focusing on the occurrence of disease as a means to better understand and identify causes of disease. During my training in cardiovascular and genetic epidemiology, I developed a dissertation under the mentorship of Dr. Kari E. North that characterized early infant child growth in a cohort of Chilean infants and its impact on lipid levels in adolescence. My aims, similar to my postdoctoral work described below, centered on a set of biomarker measures – in this case lipid biomarkers related to cardiovascular disease risk. To fund my dissertation work, I obtained an external two-year American Heart Association predoctoral fellowship award (16PRE29200008). When determining the extent to which associations exist between infant growth and lipid outcomes, I furthered my knowledge of longitudinal methods that I had first learned as a statistician working in the UNC Department of Psychiatry, including nonlinear mixed effects models and latent class growth mixture modeling, to characterize growth as an exposure. I used an initial paper I independently developed examining the best measures to characterize infant growth1 to inform my approach in assessing a longitudinal measure of anthropometric measures as an exposure. Within my three aims I was able to: 1) characterize determinants of infant growth applying advanced longitudinal analytic methods;2 2) assess the association between infant growth including latent growth patterns and lipid levels (under review at AJE); and 3) determine if infant growth functions as an effect modifier of candidate genetic variants associated with lipid levels.

1. **Postdoctoral studies (National Institute of Environmental Health Studies)**

Extending my interest in longitudinal exposures and health outcomes in epidemiological research, I have strengthened my experience in time-to-event models with a focus on breast cancer incidence in a large contemporary U.S.-wide study. At the start of my postdoctoral studies I studied familial correlation of age of onset in sisters3 with implications for underlying early life exposures and genetic factors. More recent work continues my focus on biomarkers within the Sister Study, examining serum iron biomarkers and their: a) association with breast cancer incidence;4 b) association with common lifestyle predictors; and c) and correspondence with toenail measures. My work with the Sister Study data sources has strengthened my understanding of the unique and promising aspects of this rich and well-characterized longitudinal data source, preparing me to conduct the proposed research. Research spanning both my dissertation work and postdoctoral fellowship enabled me to conceptualize research problems for both longitudinal exposures within a life course perspective and time-to-event data as it will be applied to a sample of postmenopausal women.

My goals are to apply the knowledge I have gained in methodological and epidemiological research areas to the study of lifestyle exposures and their relationship with cancer and all-cause mortality as women move into their post-reproductive years. To do so I will require further training in aging research, lifestyle exposures, as well as joint analysis models that are part of nascent research that combines both longitudinal exposures and time to event models.

## Career goals and objectives

**Building research that focuses on modifiable factors and changing risk of disease over age and time focused on women’s post-reproductive years within epidemiology is a long-term goal of mine with the ultimate purpose to prevent breast cancer cases and premature mortality.**

Following my work experience and training in statistical and epidemiological methods focusing on biomarkers and risk of disease, I am planning a new direction in research that is aligned with my long-term career goals. Certain lifestyle characteristics figure strongly during the postmenopausal years in women’s risk of breast cancer and number as some of the top ten risk factors for mortality and chronic disease, emphasizing the ability to cross over into study of other outcomes such as mortality. My training objectives and mentoring plan are designed to enhance my knowledge of these lifestyle factors and support the successful completion of my three aims to understand how modifiable factors of lifestyle co-occur, change over time, and relate to disease risk. Once I establish that knowledge, I plan to launch my R00 independent research phase in which I will first investigate associations between the well-defined lifestyle factor changes and breast cancer incidence and all-cause mortality. After that step, I will investigate the role that lifestyle may influence the genetic association with breast cancer risk and all-cause mortality. By focusing on lifestyle factors specific to the postmenopausal age range, the proposed research work would focus on these factors as they operate through a life course perspective. Results from this work could inform areas of lifestyle change that could improve the health of adults as they age – both goals in the NIA strategic plan.5 I plan to capitalize on the knowledge I have gained of factors associated with breast cancer risk and their dependency on certain age and time periods. At the same time, I will rely on my training plan during the K99 phase to extend my research in a direction spanning advanced structural equation modeling, joint analysis, and lifestyle exposures that will set me on a path independent from that of my mentor.

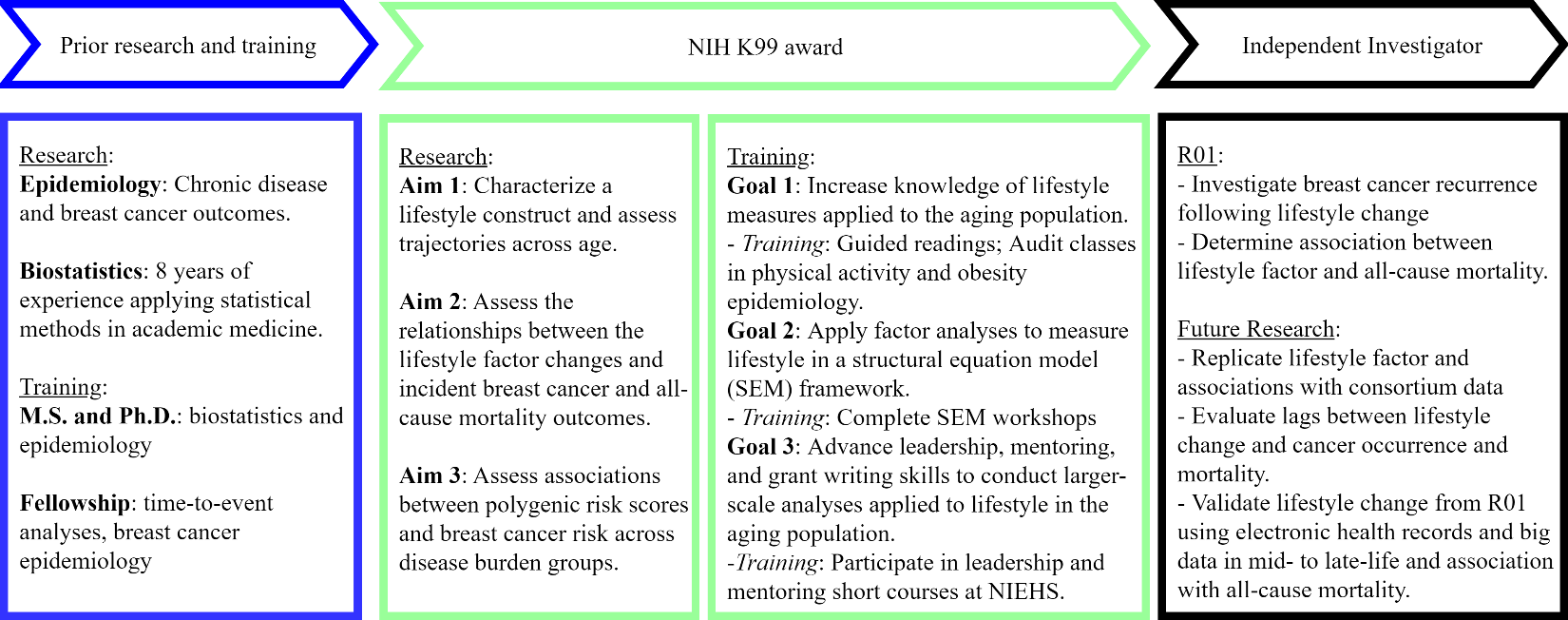


Figure 1: Career timeline

## Career development and training

To best attain my outlined career goals and address the three proposed research aims (Figure 1), I will take advantage of the training opportunities described below to establish myself as an independent investigator studying modifiable lifestyle change following midlife to prevent disease.

### Mentoring team and collaborators

I have assembled a mentoring team that will provide excellent support and guidance as I embark on a new research path involving common modifiable factors and outcomes such as breast cancer in a group of postmenopausal women (Table 1). My primary mentor, **Dr. Clarice R. Weinberg**, a pioneer in the field of epidemiologic methods and a principal investigator for the Sister Study, will continue guiding me as I follow my K99 training period during the first two years. As a postdoctoral fellow in her lab, I meet with her on a weekly basis to plan, develop and write manuscripts related to work based on Sister Study samples. My mentor and co-mentor, **Dr. Clarice Weinberg** and **Dr. Dale Sandler**, are experts in breast cancer epidemiology as well as possessing an extensive and accomplished history of mentorship. Being the principal investigators for the Sister Study, they will guide me regarding lifestyle change within this ongoing contemporary cohort. As I launch into the independent investigator path, I will also draw on the expertise of my mentoring committee, each member fitting within distinct areas of research in which I will need guidance. I will rely on **Dr. Shanshan Zhao’s** expertise in time-to-event modeling as I conduct joint model analyses to understand the relationship between longitudinal factors and risk in a sample of postmenopausal women. As an expert in structural equation models, **Dr. Nisha Gottfredson’s** mentorship will be essential to conduct best practices when capturing the lifestyle information through factor analysis. **Dr. Mary Beth Terry** will play a crucial role as I conduct epidemiologic research related to lifestyle and genetic factors following menopause and their relationship with breast cancer incidence. As I learn from and collaborate with these mentors, I will be able to work independently, but I will also embrace the value in their guided perspectives that can set my career trajectory on an effective and productive course. This mentoring plan will include continuing my weekly meetings with my mentor and having as-needed individual correspondence in the form of email and meetings with my mentorship team (Table 2) as I develop, analyze the data, and write the manuscripts for these proposed projects. Guidance from my mentorship team will be paramount as I search for faculty positions and start my own research group.

Table 1: Mentorship Team

| **Name** | **Associated Specific Aims / Training Goals** | **Position** | **Proposed Role** | **Expertise** |
| --- | --- | --- | --- | --- |
| Clarice R. Weinberg, Ph.D. | Research: 1, 2, 3; Training: 1, 3, 4 | Principal Investigator, Biostatistics and Computational Biology Branch, NIEHS | Primary mentor | Breast cancer, methods and genetic epidemiology |
| Dale P. Sandler, Ph.D. | Research: 1, 2, 3; Training: 1, 3, 4 | Principal Investigator, Epidemiology Branch, NIEHS | Secondary co-mentor | Lifecourse and breast cancer epidemiology |
| Shanshan Zhao, Ph.D. | Research: 1, 2, 3; Training: 2 | Principal Investigator, Biostatistics and Computational Biology Branch, NIEHS | Advisor, statistical methodology | Biostatistics, time-to- event analyses |
| Mary Beth Terry, Ph.D. | Research: 1, 2, 3; Training: 1, 3 | Professor, Epidemiology, Columbia Mailman School of Public Health | Advisor, lifestyle and genetic epidemiology | Genetic and cancer epidemiology |
| Nisha Gottfredson, Ph.D. | Research: 1, 2, 3; Training: 2 | Assistant Professor, Department of Health Behavior, UNC | Advisor, statistical methodology | Factor analysis, longitudinal methods |
|  |  |  |  |  |

### Training objectives

My training objectives in the first two years of the proposed award will draw on the rich interdisciplinary resources available in the Research Triangle Park area and online offerings from across the United States. These activities will include attending seminars, auditing classes, and individual guided readings through mentoring activities (Table 2).

Table 2: Training Timeline

| **Milestones/Benchmarks** | **K99** | | **R00** | | |
| --- | --- | --- | --- | --- | --- |
| **Year 1** | **Year2** | **Year1** | **Year2** | **Year3** |
| **Mentoring Meetings** |  |  |  |  |  |
| Weekly meetings with primary mentor | x (5%) | x (5%) |  |  |  |
| Bi-annual meeting with mentorship committee |  | x (1%) | x (1%) | x (1%) | x (1%) |
| Individual meetings and/or communication on an as-needed basis regarding unanticipated analytic and/or subject matter problems | x (2%) | x (5%) | x (2%) | x (1%) | x (1%) |
| **Research** |  |  |  |  |  |
| Statistical analyses of lifestyle factor and its longitudinal change (Aim 1) | x (55%) | x (25%) |  |  |  |
| Draft and submit manuscript (Aim 1) | x (20%) | x (20%) |  |  |  |
| Statistical analyses of lifestyle factor and breast cancer risk |  | x (10%) | x (40%) | x (25%) |  |
| Draft and submit manuscript (Aim 2) |  |  | x (30%) | x (20%) |  |
| Statistical analyses of polygenic risk scores and lifestyle risk burden (Aim 3) |  |  |  | x (30%) | x (40%) |
| Draft and submit manuscript (Aim 3) |  |  |  |  | x (50%) |
| **Coursework** |  |  |  |  |  |
| Audit "Physical activity epidemiology and public health (EPID 810)" | x (10%) |  |  |  |  |
| Audit "Obesity Epidemiology (EPID 814)" |  | x (10%) |  |  |  |
| **Seminars, workshops, journal clubs** |  |  |  |  |  |
| UNC Bowles Center for Alcohol Studies Spring Seminar Series | x (1%) | x (1%) |  |  |  |
| NIEHS reproductive journal club | x (1%) | x (1%) |  |  |  |
| UNC Odum Institute short course: Introduction to structural equation models | x (2%) |  |  |  |  |
| American Society on Aging summer short course: Managing Health & Chronic Conditions in Older Adults | x (2%) |  |  |  |  |
| **National meetings** |  |  |  |  |  |
| Attend 1-2 meetings per year including SER, ASPO, AACR, and ASHG | x (2%) | x (2%) | x (2%) | x (2%) | x (3%) |
| **Faculty job search and grant writing** |  |  |  |  |  |
| Conduct academic faculty job search |  | x (15%) |  |  |  |
| R01 idea development |  | x (5%) | x (25%) | x (20%) |  |
| R01 submission |  |  |  | x (1%) | x (5%) |

**Training objective 1: Increase knowledge of lifestyle measures relevant to the aging population.**

I will audit physical activity (year 1) and obesity epidemiology (year 2) classes. I will attend the spring seminar series at UNC Bowles Center for Alcohol Studies (<https://www.med.unc.edu/alcohol/spring-2019-seminar-series/>) (years 1 and 2). To gain knowledge of modifiable lifestyle factors in the aging population, I will attend a summer short course, “Managing Health & Chronic Conditions in Older Adults,” offered by American Society on Aging and the USC School of Gerontology (Summer, year 1).

**Training objective 2: Apply factor analyses to measure lifestyle in a structural equation model (SEM) framework.**

Training will include completion of a SEM summer short course offered by Inter-university Consortium for Political and Social Research (ICPSR) in collaboration with the Odum Institute (year 1). During the first and second years I will consult with Dr. Gottfredson regarding the lifestyle factor analyses during the analyses for aim 1, and I will follow any guidelines for directed readings related to my work.

**Training objective 3: Present research at nationally representative conferences.**

This objective will serve multiple goals of: 1) networking as I conduct my search for a faculty position, 2) interacting with experts and leaders who can offer new perspectives and opinions that I can use to improve the proposed work and inform my R01 application, and 3) bringing awareness of the completed scientific work in this proposal.

**Training objective 4: Advance leadership, mentoring, and grant writing skills to conduct larger-scale analyses applied to lifestyle in the aging population.**

I will participate in leadership and mentoring short courses at NIEHS, which include the “Management Bootcamp” offered by the NIEHS Office of Intramural Training and Education (OITE) to learn management concepts independent of the research environment but necessary to develop constructive leadership skills and expand my work by leading a research lab. I will also take advantage of the grant-writing workshops and seminars offered by the OITE. Furthermore, my mentoring team, all of whom have successfully written large-scale grants, will provide individual advice when I start the R01 application process (years 4-5).

### Plans for transition to independence

I can take these learning steps towards these goals to develop an independent research career trajectory as an epidemiologist leveraging advanced methods to focus on modifiable exposures in postmenopausal women as they relate to cancer and mortality outcomes. Pursuing this work will build on work accomplished by my primary mentor but also diverge in a new direction of lifestyle factors, separate from environmental factors, as an exposure in the domain of breast cancer incidence and all-cause mortality.

# Research Strategy

## Background and Significance

Common modifiable health risk factors in the United States include body fatness, exercise, and alcohol use, which also are leading causes of mortality and are shared with other health outcomes such as breast cancer. Despite the importance of these factors, less than 35 percent of NIH-funded prevention research studies measured these modifiable lifestyle factors and less than four percent of all research projects consider more than one leading risk factor contributing to all-cause mortality.6 Many of the leading risk factors mentioned above co-occur in individuals, and their study as an aggregate measure would create a more efficient use of measures of health and well-being as well as capturing a meaningful measure of correlated risk factors. This approach will also follow building research that points towards patterns of multiple lifestyle factors as creating a favorable environment for cancer rather than single causes.7 Developing this unique approach to understand leading modifiable lifestyle factors also aligns with the NIA strategic plan5 to improve the understanding of both: 1) individual effects on aging through a life course perspective and 2) factors that can improve the health of adults as they age with an eye towards informing intervention policies.

Certain lifestyle risk factors for postmenopausal women have a strong body of evidence supporting associations with breast cancer incidence, including body fatness,8–11 physical activity,12–15 and alcohol use.16–19 At a minimum, risk-attributable fraction estimates indicate that one out of five postmenopausal breast cancer cases could be eliminated following modification of lifestyle, with body fatness, alcohol consumption, and physical activity contributing to this estimate.20–22 Importantly, these lifestyle risk factors that are related to breast cancer are also among the top ten attributable causes of all-cause mortality.23

The lifestyle risk factors mentioned above, such as alcohol use, also happen to have the strongest relationship with risk during the postmenopausal period. By focusing on the postmenopausal time, we can target risk factors that are life course-sensitive. Body fatness is another risk factor with a distinct positive association with breast cancer after menopause. Taking advantage of the propensity for change in risk by lifestyle around that time can provide meaningful and actionable knowledge for interventions. Also, leveraging characteristics of a lifestyle factor, with a reduced dimension from several lifestyle indicators into one measure that incorporates the correlations between lifestyle indicators, could be a powerful tool to assess lifestyle change over time and to determine what patterns currently exist in a large U.S.-wide population of women entering the menopausal phase of their lives.

Studies of lifestyle change exist for women diagnosed with breast cancer, but no study to date has addressed this question prospectively in women without a diagnosis of breast cancer. This research gap creates a unique opportunity to better characterize the common risk predictors and examine change across the life course, considered two of the six critical areas of research by the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR).7 Longitudinal changes in a combined lifestyle factor index and its joint association with breast cancer incidence and all-cause mortality can capture the association between lifestyle change and its association with risk of disease and death.24

Considerable research supports both the role of modifiable lifestyle factors, as noted above, and the genetic underpinnings of breast cancer risk. Combining these two exposures through GXE analyses, forms our third research aim. Recent research considering GXE does not provide evidence of multiplicative interactions between genetic and certain lifestyle factors,25–27 but may point towards additive interactions related to breast cancer.28 Our approach in aim 3 to quantify breast cancer risk concentrations29–31 allows us to specify additive interactions and determine if groups bearing larger burdens of risk have stronger associations between genetic variants and breast cancer risk compared to the lowest risk groups. Considering this is an active area of research, it is important to fill these research gaps to determine if well-defined lifestyle factors modify the genetic associations with disease, which would be a novel contribution to this field of research. If so, lifestyle interventions could be personalized based on genetic susceptibility to breast cancer.

## Innovation

### **Characterize lifestyle change over time**

Lifestyle is commonly assessed on a cross-sectional basis, one variable at a time, to study associations with breast cancer risk. We will use data from a large prospective contemporary breast cancer cohort with four follow-up times, which allows us to assess longitudinal change. The factor analysis approach is novel in lifestyle research, and we can leverage correlated lifestyle indicators and summarize them in one measure.

### **Assess change in lifestyle in tandem with risk of breast cancer**

Following implementation of the first aim, we will simultaneously evaluate the association between longitudinal change in the lifestyle factor, considered the exposure, and breast cancer risk with time-to-event data. Joint modeling of longitudinal change of an exposure and time to event data is a recently developed statistical application that has not been used in the context of associations between lifestyle and breast cancer risk and its use can provide a more granular picture of lifestyle change and cancer risk.

### **Use novel definition of risk associated with lifestyle groups to evaluate modification by established genetic underpinnings of breast cancer risk**

Gene-environment analyses commonly assess the change in genetic risk across continuous or categorical measures of an exposure — lifestyle being one example. As an alternative, we plan to define the exposure in terms of its risk concentration using a common measure from economics that is just finding its way into the field of public health: the Lorenz curve. Assessing the exposure in this manner, common in fields such as economics, is less common in public health and has not been used to assess modifications of polygenic associations with breast cancer risk. Our aim is to use this novel application to determine if certain lifestyle groups with the highest burden of risk have stronger polygenic associations with breast cancer risk compared to groups with the lowest burden of risk. We also can use this knowledge to understand how shifting the burden of risk can affect public health interventions.

## Approach

### Overall research design

#### Participants and Setting

**Sister Study cohort**: The proposed study is part of a contemporary prospective cohort of 50,884 women ages 35-74 years of age from 2003 to 2009 who have not been diagnosed with breast cancer upon entry but have a previously diagnosed sister. I plan on using my K99 award to assess a factor-analysis-based healthy lifestyle index based on each of four Sister Study follow-up questionnaires. We will choose healthy lifestyle indicators according to the World Cancer Research Fund and the American Institute for Cancer Research and the availability of repeated measures over the four follow-up surveys. The lifestyle factor will include indicators of body fatness, physical activity, and alcohol use. In addition to determining the relationship between the lifestyle factor index and breast cancer risk, this work will enable the R01 component to determine if certain lifestyle groups with the largest burden of lifestyle-based risk display stronger genetic associations with disease than groups carrying the lowest burden.

#### Data collection

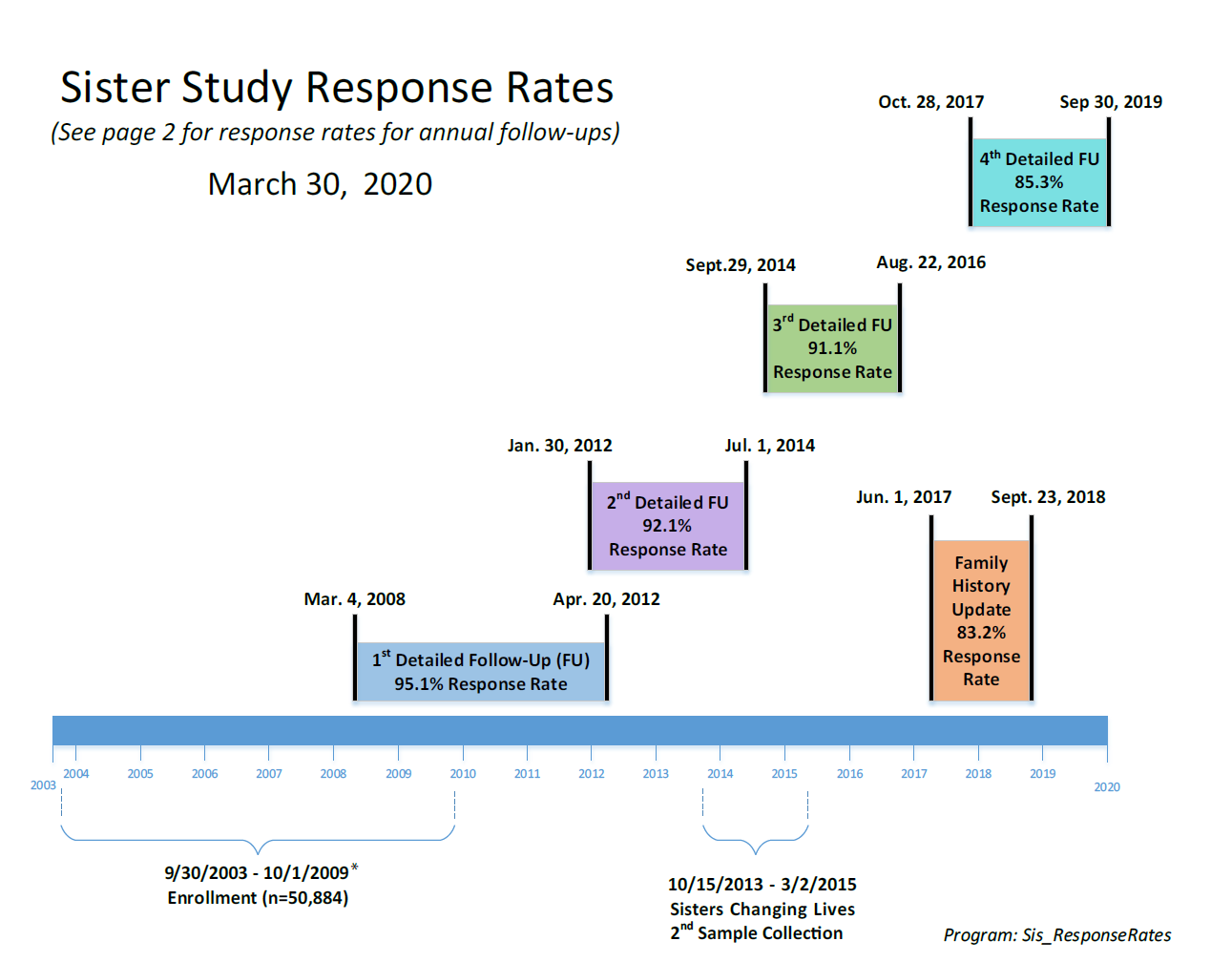
The most recent Sister Study follow-up survey was between 2017-2019, the fourth follow-up, after three bi-annual follow-up surveys approximately 2-3 years apart following the enrollment period between 2003-2009. Participation in these surveys ranged from 95% in the first follow-up survey to 85.3% in the most recent fourth follow-up (Figure 2). Aims 1-3 will use the lifestyle measures from all four follow-up surveys in the subset of postmenopausal women. Data management is centrally managed by contract through DHL Corporation and Westat with a data release system that has been in place since 2003. We expect this reliable and structured data management system — including thorough data cleaning and handling — to continue throughout our study period, and we consider this efficiency a strength of our proposed research.

Figure 2: Follow-up information for the Sister Study.

### Specific Aim 1: Characterize a lifestyle construct from indicators in a contemporary cohort of women and assess trajectories across age.

#### Introduction

Lifestyle is a frequent focus of health research but research on lifestyle as a composite and inter-related measure and its change over time is needed. Our *objective* for this aim is to characterize several lifestyle indicators as a single index and determine trajectories of these groups over a ten-year follow-up period during mid-life to late-life. To attain this aim, we will test the *working hypothesis* that correlated lifestyle factors associated with breast cancer risk will be defined as a single index, which will be consistently defined as age increases. Our *approach* to test this hypothesis will employ confirmatory factor analysis (CFA)32,33 to estimate a single factor and multi-level modeling to assess the factor change over age-time in a contemporary cohort exceeding 50,000 women. The *rationale* for this aim is to establish a well-characterized lifestyle factor reflecting the correlations of lifestyle indicators and its change over a ten-year age span. Upon completion of aim 1, we *expect* to better understand if there are distinct patterns of lifestyle change. Our findings from aim 1 will assist the assessment of associations between lifestyle change and breast cancer risk in aim 2.

#### Methods

**Statistical analysis**

We will use confirmatory factor analysis to identify a lifestyle construct from the four variables representing evidence-based lifestyle components associated with breast cancer incidence and all-cause mortality, also called indicators: BMI, alcohol use, physical activity, and smoking. We will evaluate this factor by racial/ethnic groups to evaluate similarity across groups, also known as group invariance. If the factor loadings are similar across groups and time then we will pool the groups, otherwise further analyses will be stratified by racial/ethnic groups. Following assessment of factors at each of the four follow-up surveys, we will estimate longitudinal trajectories of these factors in the group of postmenopausal women using a ‘Curve-of-Factors Model’ (CFM)34 (Figure 3).

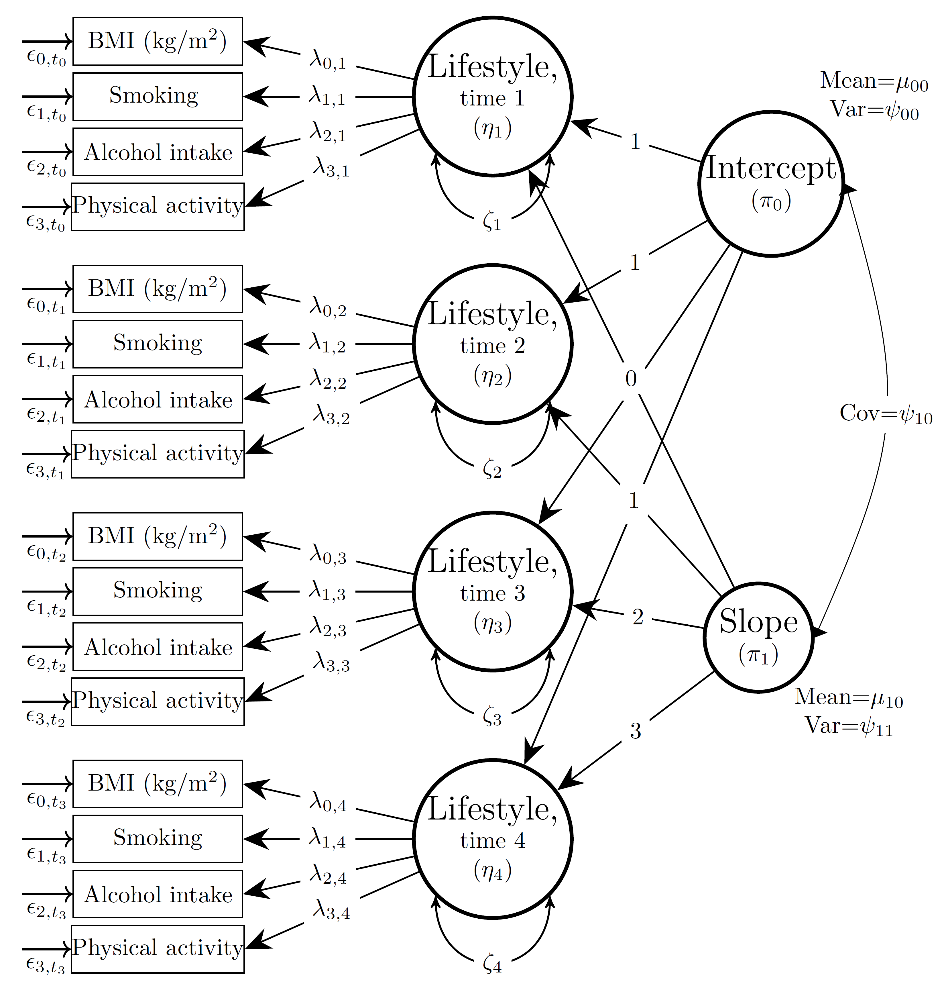


Figure 3: Curve-of-Factors Model for three key lifestyle indicators in postmenopausal women across four follow-up times.

#### Expected Results/Outcomes

The main outcome is a lifestyle factor with three indicators, including body fatness, physical activity, and alcohol use. Estimates include cross-sectional measures of the lifestyle factor and trajectories over the four follow-up times. We expect the lifestyle indicators to have strong loadings to support the factor, and we expect the factor to remain equivalent over time to allow the assessment of longitudinal trajectories. We also expect there to be multiple trajectories of lifestyle change in this cohort of postmenopausal women.

#### Potential problems and alternative strategies

The lifestyle factor is a composite of various indicators, and it may not have a similar structure over time and/or across racial ethnic groups, also known as measurement equivalence or invariance. If this measurement invariance assumption is violated then we cannot compare the lifestyle factor across time or across racial/ethnic groups. Instead, some solutions include looking at each of the lifestyle indicators individually to assess their change over time or stratifying the lifestyle construct by racial/ethnic groups.

### Specific Aim 2: Assess the relationships between the lifestyle factor changes and incident breast cancer and all-cause mortality outcomes.

#### Introduction

Lifestyle changes in midlife may set the stage for higher risk of breast cancer incidence and mortality yet no lifestyle measure exists as one comprehensive exposure. Our *objective* is to assess the change of a lifestyle factor over time and assess the association between these changes and breast cancer risk and all-cause mortality. To achieve this objective, we will test the *working hypothesis* that groups with improving lifestyle trajectories are at lower risk for breast cancer incidence and all-cause mortality. Our *approach* to test this hypothesis will be to identify correlated lifestyle characteristics associated with a “healthy lifestyle” factor followed by joint modeling of longitudinal and time to event methods to capture associations between lifestyle change and breast cancer risk and all-cause mortality. The *rationale* for this aim is to expand upon the existing, mostly cross-sectional, knowledge of lifestyle and breast cancer to add evidence regarding the role of healthy lifestyle change in reducing breast cancer incidence, necessary to understand how to best target prevention efforts. Upon completion of aim 2, we *expect* to identify levels of lifestyle change associated with breast cancer incidence and all-cause mortality, which can inform prevention efforts.

#### Methods

*Statistical analysis*

Following characterization of a lifestyle factor and its change over both age and time in aim 1, we will simultaneously assess longitudinal trajectories of the factor, a Curve-of-Factors model, over the four time points and time-to-event analyses via joint latent class models to assess the association between lifestyle and breast cancer and all-cause mortality risk. The joint latent variable growth-survival analysis entails simultaneously specifying an intercept and slope from the lifestyle factor longitudinal model (Figure 4) and a time-to-event model for incident outcomes (Figure 4). Estimates from these joint regression models will yield estimates of the association between a lifestyle factor as a time-dependent covariate and breast cancer incidence. Advantages of these models include the ability to accommodate simultaneous changes in exposure and risk over age-time, to better capture the role of lifestyle change in risk.

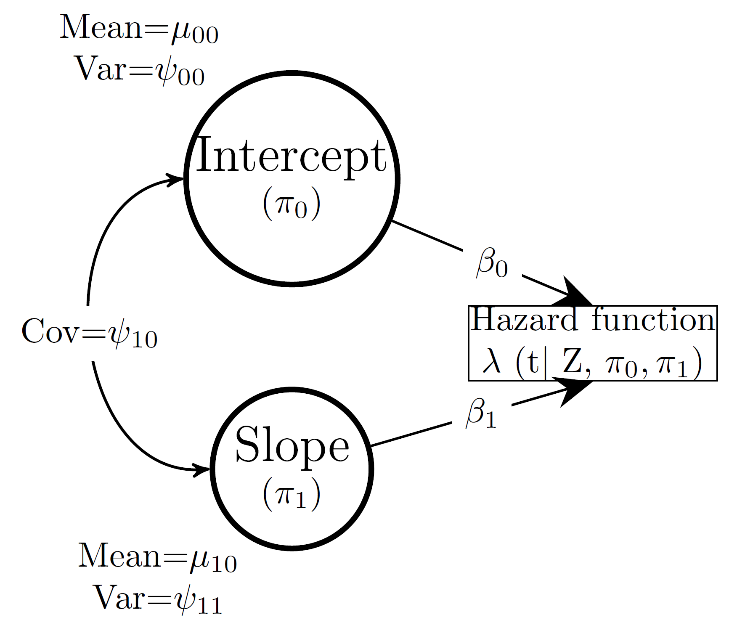


Figure 4: Model for joint analysis of factors and time to event models.

#### Expected Results/Outcomes

The expected outcomes are estimates of associations between longitudinal lifestyle factor patterns of change and breast cancer incidence and all cause mortality.

#### Potential problems and alternative strategies

As mentioned in section , we may not have factors that remain consistently defined over time. If that scenario occurs then we will treat each of the lifestyle variables related to postmenopausal breast cancer separately. Although this alternate analysis will no longer address co-occuring lifestyle variables, which we consider a strength of this proposed study, the analyses will still yield valuable information regarding simultaneous change over time with breast cancer risk for each of the lifestyle variables. This information does not exist in the literature and will still address the research gap. Also, different lifestyle factor specification over time will also constitute valuable information that motivate future research to determine predictors of these changes. For example, the association between exercise and body fatness may attenuate over time, leading to question regarding what environmental conditions may influence this change. This problem that may occur with the proposed research strategy have solutions, and the conditions that motivate this change are also of interest in research.

### Specific Aim 3: Assess associations between polygenic risk scores and breast cancer risk for lifestyle groups with highest disease burden compared to groups with lowest burden.

#### Introduction

Lifestyle and underlying genetic factors play an important role in cancer incidence yet the role of lifestyle as a modifier of the gene-breast cancer risk association remains under active investigation. The *objective* of this aim is to assess whether the joint effect of genetics and postmenopausal lifestyle characteristics is additive for breast cancer risk. To achieve our objective, we will test our *hypothesis* that women with adverse lifestyle characteristics and expected to have a disproportionately higher disease burden will also demonstrate stronger genetic associations with cancer incidence on an absolute scale compared to women with a favorable lifestyle and lower burden. Our *approach* to test this hypothesis will be in two steps: 1) to find high/low risk burden groups through risk inequality estimates determined by Lorenz curves of breast cancer risk concentration conditional on lifestyle factors, and 2) to determine additive effect modification of the genetic association with breast cancer risk by the high/low risk burden groups in time-to-event regression models. The *rationale* for this aim is to better understand how risk inequality can affect the impact of genetic variants on breast cancer risk and to use this information to identify subpopulations that would benefit most from prevention efforts. Once we accomplish our aim, we *expect* to identify groups with multiple, co-occurring adverse lifestyle characteristics that have both higher breast cancer risk and display stronger genetic associations with breast cancer risk, supporting a role of risk inequality in gene-environment associations.

#### Methods

*Statistical analysis*

To assess effect modification of the genetic association with breast cancer incidence by lifestyle factors, we will create a risk inequality covariate and a polygenic risk score (PRS) to use in time-to-event regression models. To characterize risk inequality in breast cancer burden, we will use Lorenz curves, a method first used in the field of economics but of increasing use in epidemiology and public health,29,30,35,36 to estimate the concentration of absolute breast cancer risk conditional on lifestyle factors. Parametric time-to-event models enable the estimation of absolute breast cancer risk at attained ages for each individual according to their lifestyle factor values. At pre-specified attained ages, we can estimate absolute breast cancer risk predicted from the lifestyle factor values then plot the cumulative number of risks versus individual cumulative absolute risk values ordered from lowest to highest risk to create a Lorenz curve (Figure 5). Should the curve follow a 1:1 diagonal line, the estimated risk would be evenly distributed across individuals. Conversely, a Lorenz curve deviating from a diagonal line indicates a disproportionate distribution of risk. In the example shown in Figure 5, 14% of people in this sample carry 25% of the absolute risk. Using that Lorenz curve, we will identify people who belong to upper and lower proportions of the sample with the highest and lowest concentration of risk to serve as the index and referent groups of a risk inequality variable, respectively. For example, we can then identify, characterize, and compare those individuals in the highest carrying the highest and lowest deciles of risk burden. This risk inequality approach also allows us to identify groups of lifestyle factors with a disproportionate amount of breast cancer incidence. In the second step, the breast cancer polygenic risk score and risk inequality variable will then be covariates in subsequent time-to-event models to determine the presence of effect modification of the genetic associations due to uneven distribution of breast cancer risk.

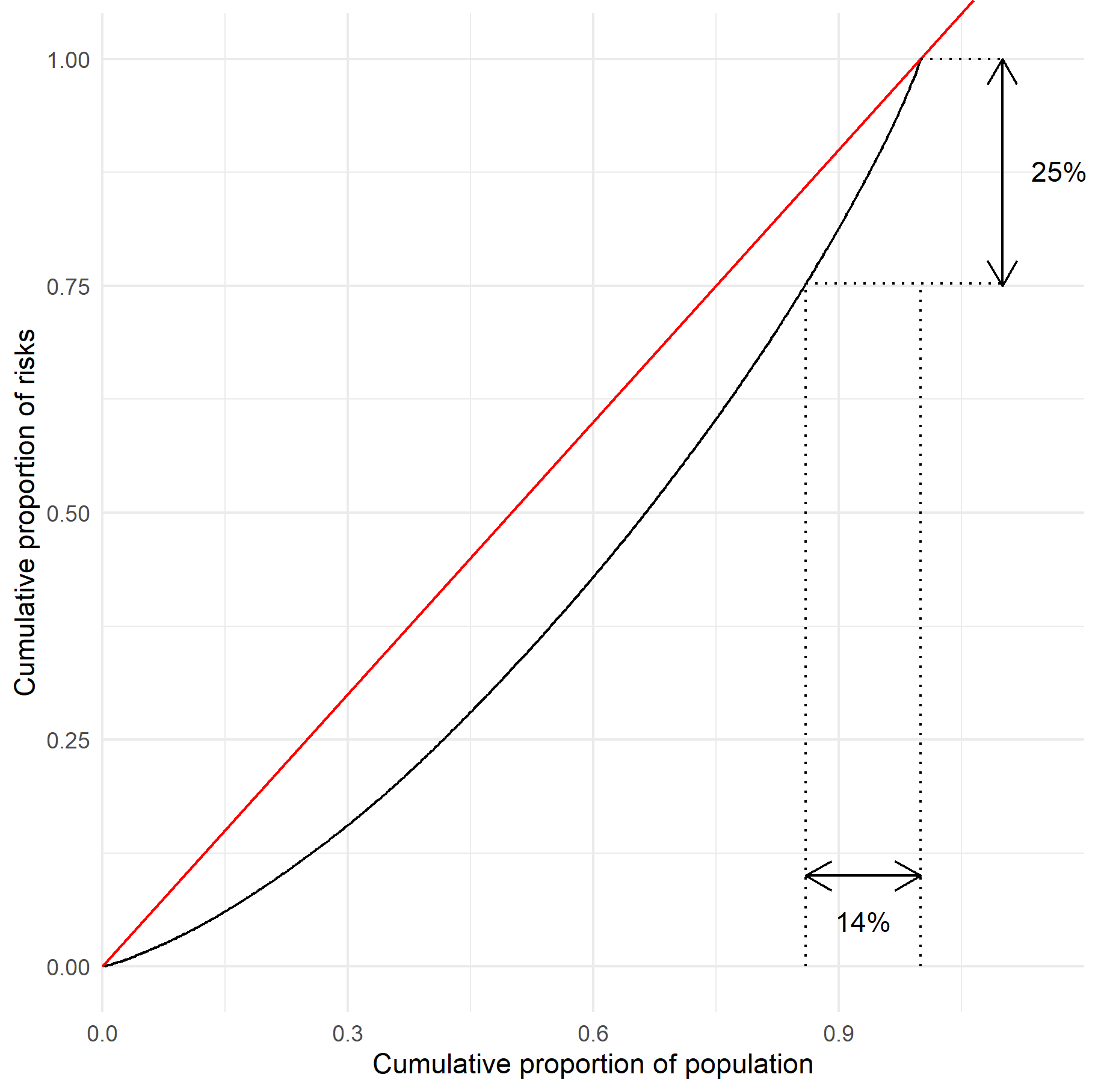


Figure 5: Sample Lorenz curve for lifestyle factors and risk.

#### Expected Results/Outcomes

The expected outcome will be estimates of differences in associations between polygenic risk scores and breast cancer incidence by groups of lifestyle characteristics carrying largest and smallest burdens of risk according to risk inequality measures.

#### Potential Problems and alternative strategies

In this aim we are relying on the distribution of breast cancer risk to be unequal to an extent that powers our comparisons of the genetic associations with breast cancer risk. This approach may not work if the concentrations of risk are evenly distributed, a situation we do not anticipate occurring given evidence for modifiable factors for breast cancer risk in this absolute risk reduction framework.30 However, if this scenario occurs then we will consider separating out the risk factors and assessing independently, under the assumption that both the risk factor and polygenic risk score are related to the incidence outcome.

If more than five percent of observations for combinations of the modifiable risk factors are missing, then we plan to incorporate multiple imputation methods37,38 to account for the missing data and allow us to proceed with planned analyses.

### Preliminary data

In our exploratory analyses with self-reported data from 32,534 Sister Study participants who were postmenopausal at study entry, we found evidence to support a lifestyle factor at study entry. The factor loadings from the lifestyle factor (Figure 6), with three key lifestyle indicators, represent the association between a one unit increase in the standardized factor and that particular indicator. Fit indices are favorable with a Comparative Fit Index of 0.94 and root mean square error of 0.078. We will assume similar favorable model fit for the three subsequent time periods.

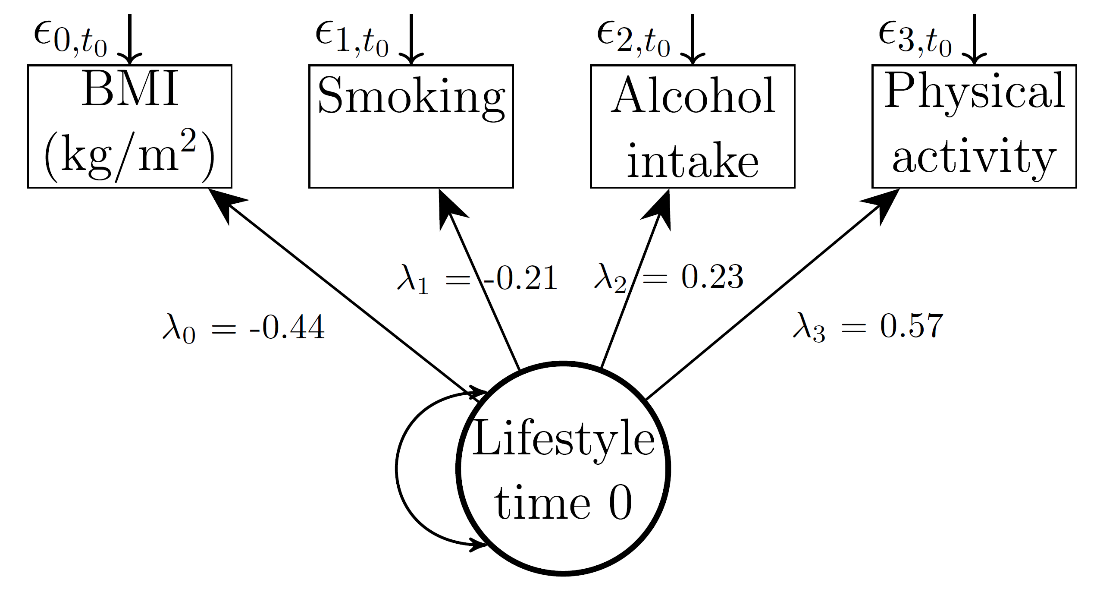


Figure 6: Factor analysis demonstration for three key lifestyle indicators in postmenopausal women at study entry.

#### Preliminary studies

Certain lifestyle characteristics are associated with postmenopausal breast cancer incidence in a large body of literature, including body fatness,8 alcohol use,16 physical activity,12 and diet.39 These important indicators have also been separately studied in relation to breast cancer risk in our proposed sample from the Sister Study. Measures of associations between lifestyle factors at study entry and breast cancer risk in the Sister Study include positive associations between body fatness indicators40,41 and alcohol use.42 Inverse associations exist between physical activity43 and a healthy diet44,45 with little evidence of an association with tobacco use.46 In aim 2 we plan to use a lifestyle factor characterized by elements with the strongest body of evidence during *postmenopause* according to the World Cancer Research Fund,47 including body fatness, physical activity, and alcohol use indicators, to determine how changes in this factor over time impact risk of breast cancer.

### Sample size and statistical power

This study includes the subset of women in the Sister Study who are postmenopausal at study entry (n>30,000). Aim 1 centers on a structural equation model describing lifestyle factors in this group of women and their change over four time points. Using Monte Carlo simulations (n=1,000) with Mplus software48 to estimate the power, we found that sample sizes of 500, 1,000, and 5,000 yielded statistical power of >0.95, to detect a slope of 0.25 (se=0.50) for four evenly spaced lifestyle factors with three continuous normally distributed indicators and an alpha level of 0.05. For aim 2, we used the same specifications from aim 1 in simulations to estimate the power to detect an association between the slope of lifestyle factor change over time and breast cancer incidence assuming a baseline hazard of 300 cases per 100,000 person-years and a hazard ratio of 1.2 (se=0.04). We found a power of 0.30, 0.56, and 0.99 at sample sizes of 500, 1,000, and 5,000, respectively. For aim 3, we have a different analytical approach that compares absolute hazard differences for a one standard deviation change in the polygenic risk score for the lifestyle group with the upper decile of risk burden as demonstrated in 5 from the analytic design corresponding to the third aim compared to a group with the lowest decile of the risk burden was 0.53, 0.99 and 1 for 10%, 25%, and 50% increases, respectively, at an alpha level of 0.05 and sample size of 2,000 cases. Under these assumptions, we believe the sample size for these three aims are sufficiently powered to detect the proposed coefficient values with at least a power of 0.8 and alpha level of 0.05. These assumptions would also hold for mortality outcomes, which are more numerous than incidence breast cancer cases in this sample of women mostly over fifty at study enrollment. Importantly, the large sample size available through the Sister Study that allows us the ability to detect an effect remains even if we have to subset our sample by racial/ethnic groups with participant counts as low as 500 for the first aim and 5,000 in the second aim.

### Timeline and future directions

The timeline for the sequence of work necessary to complete the three aims is shown in Table 7. The estimates of association between lifestyle trajectories with breast cancer incidence and all-cause mortality from Aims 1 and 2 will be the first type of analysis of this kind, and form a basis for replication – consortial data being an ideal data source. Considering these trajectories continue beyond the first ten years of observation we observed in these aims, future analyses can also include longer follow-up times upon completion of the three aims to accommodate time lags between lifestyle and breast cancer incidence. Considering breast cancer recurrence and associations with lifestyle change is another important area of research that would be a natural extension of the proposed research. The broad applicability of these common risk factors, outcomes, and methods offers great potential for other types of future research and avenues for independent investigations. Two other promising downstream projects include: 1) change in lifestyle and associations with breast cancer incidence and all-cause mortality following major life events in mid- to late life including menopause, retirement and bereavement, and 2) measures of epigenetic activity following lifestyle changes and its impact on breast cancer incidence and all-cause mortality stemming from work in Aim 3.

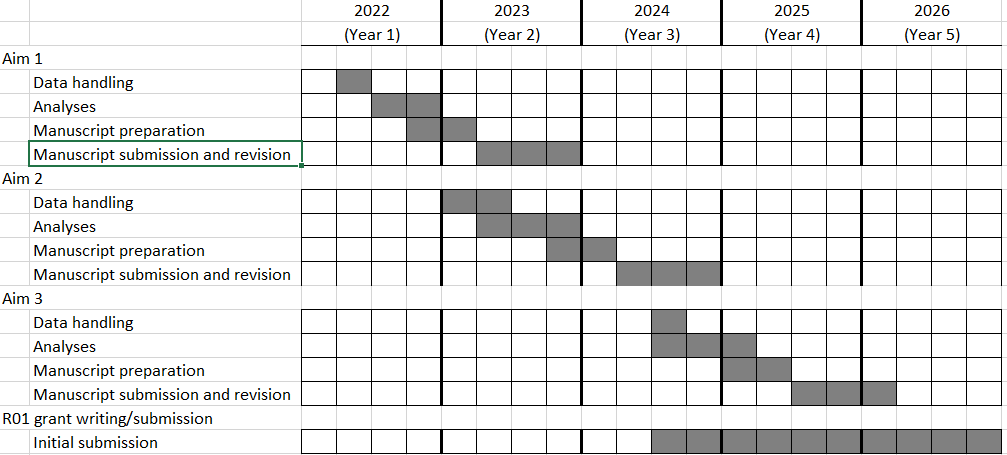


Figure 7: Timeline

# Specific Aims

Body size, alcohol use, physical activity, diet, and tobacco use count among the leading risk factors of mortality in the United States, yet less than ten percent of NIH-funded prevention research projects examine more than one of these indicators. These lifestyle risk factors are also associated with breast cancer, the second most common cancer for women in the United States, who experience a 1 in 8 risk of being diagnosed across their lifetime. What remains unclear is how correlated individual lifestyle indicators comprise a co-existing entity, i.e. a healthy lifestyle factor, and if change in this healthy lifestyle construct is associated with breast cancer incidence. Furthermore, it is unknown if people with extreme lifestyle measures carry a disproportionate breast cancer burden — also characterized as risk inequality. Defining risk inequality measures that involve modifiable behaviors remains an important and underdeveloped area of research in public health and health-promoting lifestyle constructs can also interact with genetic risk factors. With better knowledge of this association, we could do better at identifying subgroups providing optimal targets for breast cancer prevention.

To better understand the role of healthy lifestyle trajectories as they relate to breast cancer incidence and all-cause mortality and their interplay with genetic factors, we plan a three-pronged analytic approach using data from a contemporary large U.S. cohort of women. First, we will use factor analysis to estimate a healthy lifestyle construct and determine if this construct varies across age and racial/ethnic groups. Second, we will assess patterns of change over time in this construct that could influence breast cancer risk and all-cause mortality, through joint modeling of longitudinal effects using time-to-event models. Lastly, we will use risk inequality methods to characterize the concentration of disease burden across the range of lifestyle indicators and use this inequality assessment to determine if it modifies genetic risk of breast cancer through gene-environment models. Postmenopausal women with an unfavorable lifestyle profile and carry a disproportionate disease burden may demonstrate stronger genetic associations with breast cancer risk than those in lower disease-burden groups. Characterizing a constellation of lifestyle indicators as a unified construct and understanding how their changes over time, not just static baseline values, impact cancer and all-cause mortality outcomes can help clarify carcinogenesis and also highlight critical areas of intervention in midlife, providing important insight into avenues of prevention.

**Aim 1**: Characterize a lifestyle construct from indicators in a contemporary cohort of women and assess trajectories across age. *Hypothesis 1*: Correlated lifestyle indicators will fit within a factor and will be consistently defined as age increases.

**Aim 2**: Assess the relationships between the lifestyle factor changes and incident breast cancer and all-cause mortality outcomes. *Hypothesis 2*: Groups with improving lifestyle trajectories are at lower risk for breast cancer and all-cause mortality.

**Aim 3**: Assess associations between polygenic risk scores and breast cancer risk for lifestyle groups with highest disease burden compared to groups with lowest burden. *Hypothesis 3*: Groups carrying a disproportionately higher lifestyle-based disease burden will display stronger genetic associations with cancer incidence compared to those with a lower burden.

Evidence from this research can provide knowledge in an understudied area of breast cancer and mortality prevention to help understand how groups of related modifiable lifestyle indicators: 1) form a construct, 2) change over time and influence breast cancer incidence, and 3) modify genetic risk. This knowledge can help target the most impactful domains within lifestyle indicators to reduce the growing burden of disease that occurs with age in the population of postmenopausal women.

# References

1. Von Holle A, North KE, Tao R, Gahagan S. The perils of standardizing infant weight to assess weight change differences across exposure groups. *Annals of Epidemiology*. 2018;28(8):515-520. doi:[10.1016/j.annepidem.2018.04.006](https://doi.org/10.1016/j.annepidem.2018.04.006)

2. Von Holle A, North KE, Gahagan S, et al. Sociodemographic predictors of early postnatal growth: Evidence from a Chilean infancy cohort. *BMJ Open*. 2020;10(6):e033695. doi:[10.1136/bmjopen-2019-033695](https://doi.org/10.1136/bmjopen-2019-033695)

3. Von Holle A, O’Brien KM, Sandler DP, Weinberg CR. Evidence for familial clustering in breast cancer age of onset. *International Journal of Epidemiology*. November 2020. doi:[10.1093/ije/dyaa201](https://doi.org/10.1093/ije/dyaa201)

4. Von Holle A, O’Brien KM, Sandler DP, Janicek R, Weinberg CR. Association Between Serum Iron Biomarkers and Breast Cancer. *Cancer Epidemiol Biomarkers Prev*. 2021;30(2):422-425. doi:[10.1158/1055-9965.EPI-20-0715](https://doi.org/10.1158/1055-9965.EPI-20-0715)

5. *The National Institute on Aging: Strategic Directions for Research, 2020-2025*.; 2020:49. <https://www.nia.nih.gov/sites/default/files/2020-05/nia-strategic-directions-2020-2025.pdf>.

6. Vargas AJ, Schully SD, Villani J, Caballero LG, Murray DM. Assessment of Prevention Research Measuring Leading Risk Factors and Causes of Mortality and Disability Supported by the US National Institutes of Health. *JAMA Netw Open*. 2019;2(11):e1914718-e1914718. doi:[10.1001/jamanetworkopen.2019.14718](https://doi.org/10.1001/jamanetworkopen.2019.14718)

7. *Future Research Directions*. World Cancer Research Fund/American Institute for Cancer Research; 2018:31. [dietandcancerreport.org](https://dietandcancerreport.org).

8. *Body Fatness and Weight Gain and the Risk of Cancer*. World Cancer Research Fund/American Institute for Cancer Research; 2018:142. [dietandcancerreport.org](https://dietandcancerreport.org).

9. Wada K, Nagata C, Tamakoshi A, et al. Body mass index and breast cancer risk in Japan: A pooled analysis of eight population-based cohort studies. *Ann Oncol*. 2014;25(2):519-524. doi:[10.1093/annonc/mdt542](https://doi.org/10.1093/annonc/mdt542)

10. Schonfeld SJ, Pfeiffer RM, Lacey JV, et al. Hormone-related Risk Factors and Postmenopausal Breast Cancer Among Nulliparous Versus Parous Women: An Aggregated Study. *Am J Epidemiol*. 2011;173(5):509-517. doi:[10.1093/aje/kwq404](https://doi.org/10.1093/aje/kwq404)

11. van den Brandt PA, Spiegelman D, Yaun S-S, et al. Pooled Analysis of Prospective Cohort Studies on Height, Weight, and Breast Cancer Risk. *American Journal of Epidemiology*. 2000;152(6):514-527. doi:[10.1093/aje/152.6.514](https://doi.org/10.1093/aje/152.6.514)

12. *Physical Activity and the Risk of Cancer*. World Cancer Research Fund/American Institute for Cancer Research; 2018:64. [dietandcancerreport.org](https://dietandcancerreport.org).

13. Steindorf K, Ritte R, Eomois P-P, et al. Physical activity and risk of breast cancer overall and by hormone receptor status: The European prospective investigation into cancer and nutrition. *Int J Cancer*. 2013;132(7):1667-1678. doi:[10.1002/ijc.27778](https://doi.org/10.1002/ijc.27778)

14. Howard RA, Leitzmann MF, Linet MS, Freedman DM. Physical Activity and Breast Cancer Risk among Pre- and Postmenopausal Women in the U.S. Radiologic Technologists Cohort. *Cancer Causes Control*. 2009;20(3):323-333. doi:[10.1007/s10552-008-9246-2](https://doi.org/10.1007/s10552-008-9246-2)

15. Suzuki R, Iwasaki M, Yamamoto S, et al. Leisure-time physical activity and breast cancer risk defined by estrogen and progesterone receptor status–the Japan Public Health Center-based Prospective Study. *Prev Med*. 2011 Mar-Apr;52(3-4):227-233. doi:[10.1016/j.ypmed.2011.01.016](https://doi.org/10.1016/j.ypmed.2011.01.016)

16. *Alcoholic Drinks and the Risk of Cancer*. World Cancer Research Fund/American Institute for Cancer Research; 2018:85. [dietandcancerreport.org](https://dietandcancerreport.org).

17. Fagherazzi G, Vilier A, Boutron-Ruault M-C, Mesrine S, Clavel-Chapelon F. Alcohol consumption and breast cancer risk subtypes in the E3N-EPIC cohort. *Eur J Cancer Prev*. 2015;24(3):209-214. doi:[10.1097/CEJ.0000000000000031](https://doi.org/10.1097/CEJ.0000000000000031)

18. Brinton LA, Smith L, Gierach GL, et al. Breast cancer risk in older women: Results from the NIH-AARP Diet and Health Study. *Cancer Causes Control*. 2014;25(7):843-857. doi:[10.1007/s10552-014-0385-3](https://doi.org/10.1007/s10552-014-0385-3)

19. Falk RT, Maas P, Schairer C, et al. Alcohol and risk of breast cancer in postmenopausal women: An analysis of etiological heterogeneity by multiple tumor characteristics. *Am J Epidemiol*. 2014;180(7):705-717. doi:[10.1093/aje/kwu189](https://doi.org/10.1093/aje/kwu189)

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

21. Arriaga ME, Vajdic CM, Canfell K, et al. The preventable burden of breast cancers for premenopausal and postmenopausal women in Australia: A pooled cohort study. *International Journal of Cancer*. 2019;145(9):2383-2394. doi:[10.1002/ijc.32231](https://doi.org/10.1002/ijc.32231)

22. Masala G, Bendinelli B, Assedi M, et al. Up to one-third of breast cancer cases in post-menopausal Mediterranean women might be avoided by modifying lifestyle habits: The EPIC Italy study. *Breast Cancer Res Treat*. 2017;161(2):311-320. doi:[10.1007/s10549-016-4047-x](https://doi.org/10.1007/s10549-016-4047-x)

23. Mokdad AH, Ballestros K, Echko M, et al. The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States. *JAMA*. 2018;319(14):1444-1472. doi:[10.1001/jama.2018.0158](https://doi.org/10.1001/jama.2018.0158)

24. Pan D, Wei Y, Song X. Joint analysis of mixed types of outcomes with latent variables. *Statistics in Medicine*. 2021;40(5):1272-1284. doi:[10.1002/sim.8840](https://doi.org/10.1002/sim.8840)

25. Shi M, O’Brien KM, Weinberg CR. Interactions between a Polygenic Risk Score and Non-genetic Risk Factors in Young-Onset Breast Cancer. *Scientific Reports*. 2020;10(1, 1):3242. doi:[10.1038/s41598-020-60032-3](https://doi.org/10.1038/s41598-020-60032-3)

26. Kapoor PM, Lindström S, Behrens S, et al. Assessment of interactions between 205 breast cancer susceptibility loci and 13 established risk factors in relation to breast cancer risk, in the Breast Cancer Association Consortium. *Int J Epidemiol*. 2020;49(1):216-232. doi:[10.1093/ije/dyz193](https://doi.org/10.1093/ije/dyz193)

27. Rudolph A, Chang-Claude J, Schmidt MK. Gene–environment interaction and risk of breast cancer. *Br J Cancer*. 2016;114(2):125-133. doi:[10.1038/bjc.2015.439](https://doi.org/10.1038/bjc.2015.439)

28. Terry MB. Commentary: No multiplicative GXE interactions for breast cancer risk: Have we reached a verdict or is the jury still out? *Int J Epidemiol*. 2020;49(1):231-232. doi:[10.1093/ije/dyz258](https://doi.org/10.1093/ije/dyz258)

29. Mauguen A, Begg CB. Using the Lorenz Curve to Characterize Risk Predictiveness and Etiologic Heterogeneity: *Epidemiology*. 2016;27(4):531-537. doi:[10.1097/EDE.0000000000000499](https://doi.org/10.1097/EDE.0000000000000499)

30. Petracci E, Decarli A, Schairer C, et al. Risk Factor Modification and Projections of Absolute Breast Cancer Risk. *J Natl Cancer Inst*. 2011;103(13):1037-1048. doi:[10.1093/jnci/djr172](https://doi.org/10.1093/jnci/djr172)

31. Gail MH. Personalized estimates of breast cancer risk in clinical practice and public health. *Statistics in Medicine*. 2011;30(10):1090-1104. doi:[10.1002/sim.4187](https://doi.org/10.1002/sim.4187)

32. Brown TA. *Confirmatory Factor Analysis for Applied Research*. Second edition. New York ; London: The Guilford Press; 2015.

33. Bollen KA. *Structural Equations with Latent Variables*.; 1989. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=839033>. Accessed September 20, 2016.

34. Wickrama KAS, Lee TK, O’Neal CW, Lorenz FO. *Higher-Order Growth Curves and Mixture Modeling with Mplus: A Practical Guide*. New York: Routledge, Taylor & Francis Group; 2016.

35. Stensrud MJ, Valberg M. Inequality in genetic cancer risk suggests bad genes rather than bad luck. *Nature Communications*. 2017;8(1, 1):1-8. doi:[10.1038/s41467-017-01284-y](https://doi.org/10.1038/s41467-017-01284-y)

36. Valberg M, Stensrud MJ, Aalen OO. The surprising implications of familial association in disease risk. *BMC Public Health*. 2018;18(1). doi:[10.1186/s12889-018-5033-5](https://doi.org/10.1186/s12889-018-5033-5)

37. Buuren S van, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(3):1-67. <http://www.jstatsoft.org/v45/i03/>.

38. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*. 2011;20(1):40-49. doi:[10.1002/mpr.329](https://doi.org/10.1002/mpr.329)

39. *Wholegrains, Vegetables and Fruit and the Risk of Cancer*. World Cancer Research Fund/American Institute for Cancer Research; 2018:140. [dietandcancerreport.org](https://dietandcancerreport.org).

40. Park Y-MM, White AJ, Nichols HB, O’Brien KM, Weinberg CR, Sandler DP. The association between metabolic health, obesity phenotype and the risk of breast cancer. *International Journal of Cancer*. 2017;140(12):2657-2666. doi:[10.1002/ijc.30684](https://doi.org/10.1002/ijc.30684)

41. White AJ, Nichols HB, Bradshaw PT, Sandler DP. Overall and central adiposity and breast cancer risk in the sister study. *Cancer*. 2015;121(20):3700-3708. doi:[10.1002/cncr.29552](https://doi.org/10.1002/cncr.29552)

42. White AJ, DeRoo LA, Weinberg CR, Sandler DP. Lifetime Alcohol Intake, Binge Drinking Behaviors, and Breast Cancer Risk. *Am J Epidemiol*. September 2017. doi:[10.1093/aje/kwx118](https://doi.org/10.1093/aje/kwx118)

43. Niehoff NM, Nichols HB, Zhao S, White AJ, Sandler DP. Adult Physical Activity and Breast Cancer Risk in Women with a Family History of Breast Cancer. *Cancer Epidemiol Biomarkers Prev*. 2019;28(1):51-58. doi:[10.1158/1055-9965.EPI-18-0674](https://doi.org/10.1158/1055-9965.EPI-18-0674)

44. Petimar J, Park Y-MM, Smith-Warner SA, Fung TT, Sandler DP. Dietary index scores and invasive breast cancer risk among women with a family history of breast cancer. *Am J Clin Nutr*. 2019;109(5):1393-1401. doi:[10.1093/ajcn/nqy392](https://doi.org/10.1093/ajcn/nqy392)

45. Park Y-MM, Steck SE, Fung TT, et al. Higher diet-dependent acid load is associated with risk of breast cancer: Findings from the sister study. *Int J Cancer*. 2019;144(8):1834-1843. doi:[10.1002/ijc.31889](https://doi.org/10.1002/ijc.31889)

46. White AJ, D’Aloisio AA, Nichols HB, DeRoo LA, Sandler DP. Breast cancer and exposure to tobacco smoke during potential windows of susceptibility. *Cancer Causes Control*. 2017;28(7):667-675. doi:[10.1007/s10552-017-0903-1](https://doi.org/10.1007/s10552-017-0903-1)

47. Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *J Nutr*. 2020;150(4):663-671. doi:[10.1093/jn/nxz268](https://doi.org/10.1093/jn/nxz268)

48. Muthén LK, Muthén BO. *Mplus User’s Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 1998–2017.