ARIC Manuscript Proposal #4144

PC Reviewed: 4/14/26 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title:
Discovering proteins associated with malignant breast cancer incidence in post-menopausal women in the prospective ARIC study

b. Abbreviated Title (Length 26 characters):
Breast cancer, an ARIC study

2. Writing Group:
Writing group members: https://welch.jhmi.edu/
Maria Eleni Syleouni, Joe Coresh, David Couper, Kala Visvanathan, Anna Prizment, Corinne Joshu, Kenneth R Butler, Jiayun Lu, Sabine Rohrmann, Elizabeth Platz
(This is not the final authorship order)

Amendment 4/30/2024: We will add co-authors, including Marc Gunter (Imperial College) from the EPIC Cohort for their work to look up the top hits in their cohort and include those results in this ARIC paper.

Amendment 3/13/2026: Add master's student Manaal Azad to address new aim of the association between the top hit, protein LEG1 homolog, and breast cancer mortality and breast cancer case-fatality, as well as all-cause mortality in breast cancer survivors.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. **_MES____ [please confirm with your initials electronically or in writing]**

First author:
Name: Maria Eleni Syleouni

Address: Epidemiology, Biostatistics and Prevention Institute
University of Zurich
Hirschengraben 84
Zurich 8001, Switzerland
E-mail: maria-eleni.syleouni@uzh.ch

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Elizabeth Platz

Address: Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St., Room E6132
Baltimore, MD 21205
Phone: 410-614-9674
E-mail : eplatz1@jhu.edu

3. Timeline:
Expect to complete this work by October 2023 [published: Syleouni ME, Joshu CE, Coresh J, Gunter MJ, Butler KR, Couper DJ, Guevara M, Lu J, Prizment A, Riboli E, Ru M, Mahamat Saleh Y, Smith-Byrne K, Soori M, Visvanathan K, Burk VA, Wang Z, Chatterjee N, Rohrmann S, Platz EA. Discovering plasma proteins associated with breast cancer incidence in postmenopausal women in the Atherosclerosis Risk in Communities study. J Natl Cancer Inst. 2025 Oct 1;117(10):2044-2052. doi: 10.1093/jnci/djaf170. Erratum in: J Natl Cancer Inst. 2025 Dec 1;117(12):2700. doi: 10.1093/jnci/djaf284. PMID: 40623012; PMCID: PMC12505125]. Amendment 2 11/5/2025: June 30, 2026. Amendment 3: August 2027.

4. Rationale:
Breast cancer is the leading cancer for women worldwide[1], and a woman living in the US has a lifetime risk of developing breast cancer at 12.4%[2]. In the literature, identified predisposing risk factors for breast cancer in women are age, late age of menopause, pregnancy characteristics, hormonal contraceptive use, postmenopausal hormone therapy, genetic factors, benign breast disorders, obesity and overweight, alcohol consumption, smoking, diet, and others, whereas protective risk factors include full-term pregnancy, breast feeding, physical activity [3].

Detecting breast cancer at an early stage coupled with treating it appropriately is crucial, as it can affect the survival of the individuals[4]. Breast cancer screening tests aim to detect a cancer at an early stage. For women at population typical risk, mammography is recommended for screening at age 40 or 50 depending on guidelines. For women at higher risk (e.g., genetic risk) breast magnetic resonance imaging is recommended. However, screening guidelines from international organizations like the Centers for Disease Control and Prevention (CDC) or the European Commission Initiative on Breast Cancer (ECIBC) differ based on individual characteristics, such as age, and depend on the trade-off between potential benefits and harm[5,6]. For women at population typical risk, the use of less invasive tools for early detection of breast cancer with high accuracy and tools to risk stratify to better target existing screening tools would be preferred from both clinical and public health perspectives.

To better target screening and chemoprevention to women across the span of breast cancer risk, numerous breast cancer models have been developed to assist patients and providers in decision-making. These models predict risk for primary breast cancer based on the reproductive and other characteristics and family history of the woman at risk. Well-known examples include the Gail model[7], the Rosner and Colditz model[8] and the Tyrer Cuzick model[9]. Although these models have been externally validated on different populations[10], their prediction ability overall could be improved[11,12] and differences in the accuracy of prediction between populations (e.g., among racial groups[13–15]) minimized.

Plasma samples are easy to acquire, require minimal invasion and are feasible to do at scale providing rich analysis results. Especially with automation of proteomic workflows, multiple samples can be analyzed with minimized sample to sample variation[16]. The availability of proteomic data for large numbers of at-risk individuals in epidemiologic research enhances the ability of investigators to identify potential biomarkers [17]. In the ARIC cohort study, SomaLogic provides proteomic assays of 5000 proteins from a single blood draw, which was performed at three time points (Visits 2, 3, and 5), allowing for the inspection of major biological changes as reflected in plasma over time[18]. ARIC proteomic data have already been used to identify proteins

associated with dementia risk[19], kidney function decline[20], and atrial fibrillation[21]. In breast cancer, proteomics can provide valuable information for diagnosis, prognosis, and predictive biomarkers, as well as identify different breast cancer subtypes[22].

In this study, we will use the proteomics and breast cancer risk factor data from the ARIC cohort study to develop a time-to-event analysis model for malignant breast cancer incidence in post-

menopausal women without a cancer diagnosis. Malignant breast cancer was identified in ARIC with the following ICD codes: ICD-O-1: 174.0-174.9; ICD-O-2: 50.0-50.9; ICD-O-3: 50.0-50.9; ICD-8 & 9: 174.0-174.9; ICD-10: 50.0-50.9. We aim to identify potential protein biomarkers of breast cancer incidence that are independent of the known and purported risk factors for breast cancer. Being able to identify proteins associated with higher risk for breast cancer with high accuracy through minimal intervention, could introduce personalized, risk stratified breast cancer screening, change the course of disease detection and management, or establish different routine breast cancer screening tests, consequently affecting breast cancer mortality or recurrence.

Amendment 4/30/2024:
To confirm the top hit proteins from this discovery study for breast cancer incidence, the EPIC team has agreed to perform a look up in their cohort. They performed a case-cohort study of breast cancer and used the 7K SomaLogic panel. In return, we will provide them with estimates for their top hit proteins for breast cancer incidence for their discovery paper. Only estimates (HRs, 95% CIs, counts) will be exchanged.

A second group from the University Medical Center Groningen, Netherlands, used a mouse model of human breast cancer (orthotopic) to identify proteins that are linked with tumor growth. We will provide them with estimates for the association with breast cancer mortality in women at risk for breast cancer and case-fatality in women with a breast cancer diagnosis for any of their top hit mouse proteins that overlap with the human SomaLogic panel. We will also provide them with estimates for any proteins that overlap between the nominally statistically significant proteins from this discovery for breast cancer incidence. These estimates may be included in their paper reporting on their mouse experiments. Update 3/13/2026: Note that this group decided that the human situation did not sufficiently align with their mouse model, and so did not include ARIC results in their work.

Amendment 11/5/2026:
We identified protein LEG1 homolog as positively associated with post-menopausal breast cancer after FDR correction and confirmed this association in the EPIC cohort. We published that paper in the *Journal of the National Cancer Institute*. Subsequently, when the data were available, we analyzed this association in the UK Biobank and the findings were consistent (deposited in MedRxiv). Very little is known about this protein, other than it is evolutionarily conserved and that rare variants are associated with circulating levels. Thus, we now seek an amendment to characterize correlates of this protein and determine the within-person variation over time in ARIC.

Amendment 3/13/2026:
We continue to explore the import of protein LEG 1 homology and breast cancer. Now, we proposed to investigate its plasma levels in relation to 3 outcomes: breast cancer mortality (Visit 2 levels in women without breast cancer at the start of follow up); death from breast cancer, death from any cause, and death from any cause other than breast cancer (pre-diagnostic levels in women with breast cancer), and death from any cause other than breast cancer in longer term cancer survivors (Visit 5, post-diagnostic levels in women with breast cancer).

5. Main Hypothesis/Study Questions:
The main study question is whether we can uncover specific plasma proteins independent of

other known individual characteristics associated with higher risk of breast cancer in post-menopausal females. We will thus develop a risk prediction model for breast cancer, including a minimum set of proteins additional to other known characteristics.

Amendment 11/5/2025:

1. Characterize correlates of protein LEG1 homolog: demographic, anthropometric, reproductive, medical, and lifestyle factors; plasma proteins measured in the SomaScan panel and conventionally measured protein indicators of kidney and liver damage and inflammation; and plasma estradiol and testosterone
2. Determine the within-person variation over time using SomaScan data from Visits 2, 3, and 5

Amendment 3/13/2026: We hypothesize that protein LEG1 homolog provides breast-cancer specific risk information, but not risk information about death from other causes.

1. Among women without breast cancer at Visit 2, higher plasma protein LEG1 homolog levels are positively associated with breast cancer mortality.
2. Among women with breast cancer, higher pre-diagnostic (closest measurement before the diagnosis) plasma protein LEG1 homolog levels are positively associated with breast cancer-specific mortality, but not death from any cause or death from any cause other than breast cancer. These analyses will be run with and without adjustment for known prognostic markers.
3. Among longer-term breast cancer survivors (e.g., ≥5 years before Visit 5), plasma protein LEG1 homolog levels are not associated with death from any cause other than breast cancer.

6. Design and analysis (study design, inclusion/exclusion, outcome, and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study Design:

We will conduct a prospective study using ARIC cohort data of post-menopausal women, with primary outcome the incidence of a first primary breast cancer event.

Amendment 11/5/2025: We will conduct a cross-sectional analysis at Visit 2 to investigate correlates and a longitudinal analysis to investigate within-person variation over time. We will include post-menopausal women without a history of cancer, and separately, men.

Amendment 3/13/2026:

Hx 1: We will use the same cohort as in the original analysis; prospective cohort of women without a history of cancer.

Hx 2: We will restrict to breast cancer cases who have proteomics data before their diagnosis (closest before of Visit 2, 3, or 5); prospective cohort of women with breast cancer.

Hx 3: We will restrict breast cancer cases to those diagnosed at least 5 years before Visit 5; prospective cohort of longer term breast cancer survivors.

Variables of interest:

Proteomics data

Proteomics data of about 5000 proteins from visit 2, will be included in our analysis as potential

risk factors.

Amendment 11/5/2025: the cross-sectional analysis will use Visit 2 data, and the longitudinal analysis will use data from Visits 2, 3, and 5.

Amendment 3/13/2026: We will use the same covariates as in the original analysis plus fact of death, cause of death, follow-up time to death, stage at breast cancer diagnosis, and receptor status (when available).

Breast Cancer Variables
We will use the adjudicated post-menopausal breast cancer (first primary) case file. Amendment 4/30/2024: In that same file, we will use post-menopausal breast cancer mortality and case-fatality (including stage at diagnosis, which we will adjust for).

Other Variables
We will adjust for known and suspected risk factors for breast cancer, including age, race*field center, education, age at menarche, parity, age at first birth, total number of months of breast feeding, use of hormone replacement therapy, menstrual history, pregnancies, past use of birth control pills, family history of breast cancer, alcohol use, measures of adiposity and physical activity. We will also consider whether to adjust for access to and uptake of care. Specific items are reported below and their description is provided in Table 1.

Visit 1:

- *reproductive history: A1, A2, B11, B15*

Visit 2:

- *anthropometry: Weight, C4a, C4b*

- *Personal and Medical History:*
  A Medical care: A3, A5g-A5l
  C Reproductive History: C14-C23, C25-C28, C39, C40, C42
  D Smoking: D44-D49
  F occupation: F9, F70
- *physical activity: B10, B11, B12*

Visit 3:

- *personal and family medical history: 13, 13a, 14, 14a, 14b, 15, 15a, 16,17, 17a, 18, 18a, 19, 20*

Study population- inclusion and exclusion criteria:

We define visit 2 as our baseline visit since this is the first visit with proteomics data available. In our analysis, we will include any post-menopausal woman of black or white race who was cancer-free during the baseline visit 2 and did not have any previous cancer history. We will exclude any woman of other/unknown race. From our analysis, we will censor (at their date of diagnosis) any women who during follow-up had a cancer event that was not a malignant breast cancer. Additionally, we will exclude women who at visit 2 had a history of any cancer, including breast cancer. Given that the exact times of death or loss to follow-up are available, women lost to follow-up or who died will be considered right censored at the corresponding timepoint. Finally, participants missing proteomic data or covariate information will be excluded from the analysis. In our analysis, with visit 2 as the baseline, we expect to have 6043 at-risk women, of which 524 (8.7%) subsequently were diagnosed with a first primary breast cancer.[21]

Data Analysis Plan:

The analysis is comprised of 4 steps.

Step 1-Descriptive statistics and non-proteomics variable selection.

We will first identify the women who had a first primary breast cancer event and those who did not, and we will provide descriptive statistics of baseline (Visit 2) non-proteomic covariates.

Step 2- Model 1 using only baseline (Visit 2) characteristics.

We will develop multivariate adjusted models using the baseline non-proteomic variables. One strategy is to pre-specify the inclusion of the known and suspected breast cancer risk factors, but this will result in a large number of covariates for this discovery study, and these factors may not be associated with protein levels (i.e., not confounders). Another approach is to use stepwise selection to decide which covariates will be included in the final model. As a first step we will include in the model the covariates for age, race*field center, education, age at menarche, parity, age at first birth, total number of months of breast feeding, use of hormone replacement therapy, menstrual history, pregnancies, past use of birth control pills, family history of breast cancer, alcohol use, measures of adiposity and physical activity. Ultimately, we will confirm whether the known and suspected breast cancer risk factors are associated with the discovered proteins or not. The covariates of this model will be used in step 4 on the final model.

Amendment 3/13/2026: We will used Cox proportional hazards regression to address each hypothesis for protein LEG1 homolog. We will run 3 models: 1) minimally adjusted, (e.g., age, race,

education) 2) additionally adjusted including for breast-cancer related factors (e.g., BMI, reproductive factors, smoking), 3) further adjusted for prognostic factors (i.e., stage at diagnosis and receptor status).

Step 3: Data dimensionality reduction. To identify whether the protein is associated with breast cancer incidence and should be included in the model, we will conduct univariate analysis for each protein and the outcome adjusting for age and race. As a next step, in case of high dimensionality, i.e., large number of proteins to be considered for the final model, we will use Principal Component Analysis (PCA). PCA is used to reduce both the high dimensionality as well as the highly correlated proteins, while preserving the maximum variance[23]. The identified proteins will be included in the final model in step 4.

Step 4: Final model.
We will test for the proportional hazards assumption, and we will develop a multivariate Cox proportional hazards model with proteomics data selected in step 3 and the selected non-proteomic covariates used in model 1 from step 2. From this step we will only keep the proteins that are still associated with breast cancer risk when adjusting for the selected set of non-proteomics data from step 2.

Amendment 11/5/2025:
We will estimate Spearman correlation coefficients for pairwise comparisons and the intraclass correlation coefficient across all 3 visits with proteomic data. We will run linear regression models to identify factors associated with protein LEG1 homolog levels at Visit 2.

Study limitations and challenges:

This study has a limitation arising from multivariate data. More specifically compared to the number of covariates, the number of participants that will be included in the analysis is relatively small. Also, in the study most women (~75%) included are of white race, such that the results might not be transferable to different ethnic populations. From the results of the study, we will not be able to infer causality but mainly associations of the proteins and breast cancer risk. A challenge in our study is the differentiation between proteins whose levels are influenced from an existing but not yet detected breast cancer and proteins that inform the development of a future breast cancer. To identify the proteins that inform the development of a future breast cancer, we plan an additional analysis where the start follow-up will be at visit 3, but we will use the protein and covariate values from visit 2 excluding follow-up time visit 2 and visit 3, i.e., a lagged analysis (Figure 1). Those proteins that remain associated with risk are less likely those that mark an existing but not yet detected breast cancer.

Figure 1 Baseline visit and start of follow-up for the lagged analysis.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes _X_ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ____ Yes ____ No

(The file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? ____ Yes _X___ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"? ____ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/aricproposals/dtSearch.html

_X_____ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

#3482 Plasma Proteins and All-Cause Mortality in Cancer Survivors in ARIC
#3909 Identification of Plasma Proteins for the early detection of cancer in ARIC
#4011 Protein patterns as predictors of cancer risk [this one is on lung cancer]

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _X___ Yes ____ No

11.b. If yes, is the proposal
___ A. primarily the result of an ancillary study (list number* ____
2017.27 Proteomic longitudinal ARIC study: SOMAscan of multiple visits
2011.07 Enhancing ARIC Infrastructure to Yield a New Cancer Epidemiology Cohort
1995.04 Cancer Study
_____)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____)

*ancillary studies are listed by number https://sites.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload**

manuscripts to PubMed Central whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

Table 1 List of item codes by visit

Visit 1	**Visit 2**	**Visit 3**
Reproductive History https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/RHXA.pdf	**Health and medical history (Reproductive History)**	**Reproductive History** https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/RHXB.pdf
Menstrual History and Pregnancies (A)	*Reproductive History (C)*	*Reproductive History (A)*
1. Approximately how old were you when your menstrual periods started?	14. Did the participant have menstrual periods within 2 years prior to visit 1?	~~1. Did the participant have menstrual periods within 2 years prior to her last visit?~~
2. How many times have you been pregnant?	15. Have you had any menstrual periods or bleeding during the past 2 years?	~~2. Have you had any menstrual periods or bleeding during the past 2 years?~~
~~3. How many live-born children have you had?~~	16. In what month was your last menstrual period or bleeding?	~~3. In what month was your last menstrual period or bleeding?~~
~~4. Have you had any menstrual periods during the past 2 years?~~		~~4. Was this a natural period, or was it due to the use of hormones, or to some other cause?~~
~~5. In what month was your last menstrual period?~~	17. In the past two years how many periods did you miss?	~~5. In the past two years how many periods did you miss?~~
~~6. In the past 2 years how many periods did you miss?~~	18. Have you reached menopause?	~~6. Have you reached menopause?~~
~~7. Have you reached menopause?~~	19. At approximately what age did menopause begin?	~~7. At approximately what age did menopause begin?~~
~~8. At approximately what age did menopause begin?~~	20. Was your menopause natural or the result of surgery or radiation?	~~8. Was your menopause natural or the result of surgery or radiation?~~
~~9. Was your menopause natural or the result of surgery or radiation?~~	21. Are you having hot flashes?	~~9. Are you having hot flashes?~~
~~10. Are you having hot flashes?~~	22. Since your first ARIC exam have you taken or used any female hormone pills, dermal patches, shots or implants?	
~~11.~~	23. Please give the names of all female hormones	

~~12.~~	25.At what age did you start taking this hormone for the first time?	
~~13.~~	26. Are you currently taking this hormone?	
~~14.~~	27. At what age did you stop taking this hormone?	
~~15.~~	28. For how long all together since your first ARIC exam have you taken this hormone?	
~~16.~~	39. Have you had surgery to have your uterus or ovaries removed?	
~~17.~~	40. Has your uterus been removed?	
~~18.~~	42. Have you had either one or both ovaries removed?	
Birth Control Pills (B)	**Anthropometry** https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/ANTC.pdf	**Anthropometry** https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/ANTC.pdf
11. Have you ever taken birth control pills?	*Weight*	~~*Height and Weight (A)*~~
~~12. At what age did you start taking them for the first time?~~	Weight (to the nearest lb)	~~1. Standing height~~
~~13. Are you currently taking them?~~		~~2. Weight~~
~~14. At what age did you stop taking them?~~	*Body Size (C)*	~~*Body Size (B)*~~
15. For how many years altogether have you used birth control pills?	4. Girths	~~3. Girths~~
	a) Waist	~~a) Waist~~
	b) Hip	~~b) Hip~~
		Personal and family medical history https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/AMHA_0.pdf
		13. Have you ever been pregnant?
		a) How old were you when you first became pregnant?
		14. Have you ever given birth?
		a) How old were you when your first child was born?
		b) What is the total number of months, adding together all children, that you breast fed?
		15. Have you ever had a mammogram, an x-ray like

		examination of your breast?
		a) In what year was your last mammogram?
		16. Have you ever had a breast biopsy, to remove and examine a small piece of breast tissue?
		17. Have you ever had a mastectomy or lumpectomy to remove part or all of the breast?
		a) Which breast (s)?
		18. Have you ever had breast cancer?
		a) Which breasts?
		19. Has your mother, a full sister, or a child had cancer in both breasts?
		20. Have you ever had chemotherapy or radiation treatment for any kind of cancer?
	Health and medical history	**Personal History** https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/PHXA.pdf
	Medical care (A)	*Medical care (A)*
		~~3. How often do you have a routine physical examination, that is, not for particular illness, but for a general check up?~~
	3. Do you have health insurance, such as Medicare or a medical plan, such as an HMO, which pays part of a hospital, doctor's of surgeon's bill?	~~4. Do you have health insurance, Medicaid, Medicare or a medical plan, such as an HMO, which pays part of a hospital, doctor's or surgeon's bill?~~
		~~5. Do you have: a. prepaid insurance/b. medicare/ c. Medicaid/ d. other~~
	5g. Has a doctor ever said you have cancer?	~~8o. Has a doctor ever said you have cancer?~~
	5h. Can you tell me the part of the body the cancer was located?	~~8p. Can you tell me the part of the body the cancer was located?~~
	5i. And the date it was diagnosed?	~~8q. And the date it was diagnosed?~~
	5j. Have you had another cancer?	~~8r. Have you had another cancer?~~
	5k. Can you tell me in what part of the body the cancer was located?	~~8s. Can you tell me in what part of the body the cancer was located?~~
	5l. And the date it was diagnosed?	8t. ~~And the date it was diagnosed?~~

	Smoking (D)	*Smoking (D)*
	44. Have you ever smoked cigarettes?	~~25. Have you ever smoked cigarettes?~~
	45. Do you now smoke cigarettes?	~~27. Do you now smoke cigarettes?~~
	46. When did you smoke your last cigarette?	~~28. When did you smoke your last cigarette?~~
	47. Prior to quitting, how many cigarettes did you usually smoke per day?	~~29. Prior to quitting, how many cigarettes did you usually smoke per day?~~
	48. How many cigarettes do you smoke per day now?	~~30. How many cigarettes do you smoke per day now?~~
	49. Do you inhale the cigarette smoke?	~~31. Do you inhale the cigarette smoke?~~
	Alcohol (E)	*Alcohol (F)*
	58. Do you presently drink alcoholic beverages?	~~40. Do you presently drink alcoholic beverages?~~
	59. Have you ever consumed alcoholic beverages?	~~41. Have you ever consumed alcoholic beverages?~~
	60. Approximately how many years ago did you stop drinking?	~~42. Approximately how many years ago did you stop drinking?~~
	61. For how many years did you consume alcoholic beverages?	~~43. For how many years did you consume alcoholic beverages?~~
	62. In the past which type of alcoholic beverages did you ordinarily drink?	
	63. What was the usual number of drinks you had per week before you stopped drinking alcoholic beverages?	
	64. How many glasses of wine do you usually have per week?	~~44a. How many glasses of wine do you usually have per week?~~
		~~44b. How many days in a week do you usually drink wine?~~
	65. How many bottles or cans of beer do you usually have per week~~?~~	~~45a. How many glasses, bottles or cans of beer do you usually have per week?~~
		~~45b. How many days in a week do you usually drink beer?~~
	66. How many drinks of hard liquor do you usually have per week?	~~46a. How many drinks of hard liquor do you usually have per week?~~
		~~46b. How many days in a week do you usually drink hard liquor?~~
		~~48. For how many years have you consumed alcoholic beverages?~~
		~~49. Thinking about your entire time you consumed alcoholic beverages, how many glasses of wine did~~

		~~you usually have per week?~~
		~~50. Thinking about your entire time you consumed alcoholic beverages, how many glasses, cans or bottles of beer did you usually have per week?~~
		~~51. Thinking about your entire time you consumed alcoholic beverages, how many glasses of hard liquor did you usually have per week?~~
	Occupation (F)	*Occupation (G)*
	9. Please tell me the letter of the one which best described your current occupation	~~54. Please tell me the letter of the one which best described your current occupation~~
	70. Was the participant retired at visit 1?	~~60. Which of these income groups represents your total combined family income for the past 12 months?~~
	Physical activity https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/PLAA.pdf	**Physical activity/ respiratory symptoms** https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/RPAC.pdf
	Sports (B)	~~*Sports (B)*~~
		~~8. Do you exercise or play sports?~~
	10. Which sport or exercise do you do most frequently?	~~9. Which sport or exercise do you do most frequently?~~
	11. How many hours a week do you do this activity?	~~10. How many hours a week do you do this activity?~~
	12. How many months a year do you do this activity?	~~11. How many months a year do you do this activity?~~

References

1 Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A *et al.* Cancer statistics for the year 2020: An overview. *Int J Cancer* 2021; **149**: 778–789.

2 DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin* 2017; **67**: 439–448.

3 Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer: Targets and Therapy. 2019; **11**: 151–164.
4 Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: Systematic review and meta-analysis to update the 2009 U.S. Preventive services task force recommendation. Ann Intern Med. 2016; **164**: 244–255.
5 Centers for Disease Control and Prevention (CDC). Breast Cancer- Basic Information. https://www.cdc.gov/cancer/breast/basic_info/screening.html. 2021.
6 Ren W, Chen M, Qiao Y, Zhao F. Global guidelines for breast cancer screening: A systematic review. *The Breast* 2022; **64**: 85–99.
7 Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C *et al.* Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. https://academic.oup.com/jnci/article/81/24/1879/1019887.
8 Rosner B, Colditz GA. Nurses' Health Study: Log-Incidence Mathematical Model of Breast Cancer Incidence. https://academic.oup.com/jnci/article/88/6/359/1004933.
9 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004; **23**: 1111–1130.
10 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A. Risk prediction models of breast cancer: A systematic review of model performances. Breast Cancer Res Treat. 2012; **133**: 1–10.
11 Rosner BA, Colditz GA, Hankinson SE, Sullivan-Halley J, Lacey J v., Bernstein L. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. *Breast Cancer Res Treat* 2013; **142**: 187–202.
12 Wang X, Huang Y, Li L, Dai H, Song F, Chen K. Assessment of performance of the Gail model for predicting breast cancer risk: A systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Research* 2018; **20**. doi:10.1186/s13058-018-0947-5.
13 Adams-Campbell LL, Makambi KH, Frederick WAI, Gaskins M, Dewitty RL, McCaskill-Stevens W. Breast cancer risk assessments comparing Gail and CARE models in African-American women. *Breast J* 2009; **15 Suppl 1**: S72–S75.
14 Wang X, Huang Y, Li L, Dai H, Song F, Chen K. Assessment of performance of the Gail model for predicting breast cancer risk: A systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Research* 2018; **20**. doi:10.1186/s13058-018-0947-5.
15 Kurian AW, Hughes E, Simmons T, Bernhisel R, Probst B, Meek S *et al.* Performance of the IBIS/Tyrer-Cuzick model of breast cancer risk by race and ethnicity in the Women's Health Initiative. *Cancer* 2021; **127**: 3742–3750.
16 King CD, Kapp KL, Arul AB, Choi MJ, Robinson RAS. Advancements in automation for plasma proteomics sample preparation. *Mol Omics* 2022. doi:10.1039/D2MO00122E.
17 Mukesh Verma. *Proteomics and Cancer Epidemiology. In: Verma, M. (eds) Cancer Epidemiology. Methods in Molecular Biology*. Humana Press doi:https://doi.org/10.1007/978-1-59745-416-2_10.
18 SomaLogic. Applying proteomics to one of the world's longest-running heart studies. .
19 Walker KA, Chen J, Zhang J, Fornage M, Yang Y, Zhou L *et al.* Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nat Aging* 2021; **1**: 473–489.
20 Grams ME, Surapaneni A, Chen J, Zhou L, Yu Z, Dutta D *et al.* Proteins associated with risk of kidney function decline in the general population. *Journal of the American Society of Nephrology* 2021; **32**: 2291–2302.

21 Norby FL, Tang W, Pankow JS, Lutsey PL, Alonso A, Steffan B *et al.* Proteomics and Risk of Atrial Fibrillation in Older Adults (From the Atherosclerosis Risk in Communities [ARIC] Study). *American Journal of Cardiology* 2021; **161**: 42–50.
22 Neagu A-N, Whitham D, Buonanno E, Jenkins A, Alexa-Stratulat T, Tamba BI *et al.* Proteomics and its applications in breast cancer. 2021www.ajcr.us/.
23 Jollife IT, Cadima J. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016; **374**. doi:10.1098/rsta.2015.0202.