Visualization and Quantification of the Association Between Breast Cancer and Cholesterol in the All of Us Research Program

Cancer Informatics
Volume 22: 1–10
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351221144132

Jianglin Feng[1], Esteban Astiazaran-Symonds[2] and Jason H Karnes[1,3]

[1]Department of Pharmacy Practice and Science, College of Pharmacy, University of Arizona, Tucson, AZ, USA. [2]Department of Medicine, College of Medicine-Tucson, University of Arizona, Tucson, AZ, USA. [3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA.

ABSTRACT: Epidemiologic evidence for the association of cholesterol and breast cancer is inconsistent. Several factors may contribute to this inconsistency, including limited sample sizes, confounding effects of antihyperlipidemic treatment, age, and body mass index, and the assumption that the association follows a simple linear function. Here, we aimed to address these factors by combining visualization and quantification a large-scale contemporary electronic health record database (the All of Us Research Program). We find clear visual and quantitative evidence that breast cancer is strongly, positively, and near-linearly associated with total cholesterol and low-density lipoprotein cholesterol, but not associated with triglycerides. The association of breast cancer with high-density lipoprotein cholesterol was non-linear and age dependent. Standardized odds ratios were 2.12 (95% confidence interval 1.9-2.48), $P=5.6\times10^{-31}$ for total cholesterol; 1.99 (1.75-2.26), $P=2.6\times10^{-26}$ for low-density lipoprotein cholesterol; 1.69 (1.3-2.2), $P=9.0\times10^{-5}$ for high-density lipoprotein cholesterol at age <56; and 0.65 (0.55-0.78), $P=1.2\times10^{-6}$ for high-density lipoprotein cholesterol at age $\geqslant 56$. The inclusion of the lipid levels measured after antihyperlipidemic treatment in the analysis results in erroneous associations. We demonstrate that the use of the logistic regression without inspecting risk variable linearity and accounting for confounding effects may lead to inconsistent results.

KEYWORDS: Breast cancer, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, antihyperlipidemic treatment, electronic medical records

RECEIVED: July 29, 2022. **ACCEPTED:** November 21, 2022.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work and the University of Arizona - Banner Health All of Us Research Program is supported by OT2OD026549 from the National Institutes of Health's Office of the Director. JHK is supported by the NIH's National Heart, Lung, and Blood Institute (NHLBI, K01HL143137, R01 HL158686, R01 HL156993). The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Jason H Karnes, Department of Pharmacy Practice and Science, University of Arizona College of Pharmacy, 1295 N Martin Avenue, Tucson, AZ 85721, USA. Email: karnes@arizona.edu

Introduction

Breast cancer (BRCA) is the most commonly occurring cancer in females. The association of BRCA with circulating lipids, including low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), total cholesterol (TC) and triglycerides (TG), has been studied repeatedly. However, the results are largely inconsistent. For example, positive association for TC,[1] LDL,[2-4] HDL,[2,5] and TG,[6] null association for TC,[7,8] LDL,[9,10] and HDL,[11,12] and negative association for TC,[13] LDL,[14,15] and HDL[13,16,17,18] have all been reported. Several factors may account for these inconsistent results, including limited sample sizes, the confounding effects of antihyperlipidemic treatment, age, and body-mass index (BMI).[19-25] It is not clear from prior studies how lipids, age, and BMI interact to shape BRCA risk. In addition, logistic regression analysis for a continuous exposure variable assumes that the underlying relationship follows a linear function, which has rarely been validated.

It is generally assumed that a positive association means a monotonic increase of the BRCA risk with the increase of the lipid value over the full spectrum of the lipid value, a negative association means a monotonic decrease of the risk with the increase of the lipid value, and a null association means no clear change of the risk with the change of the lipid value. However, such a detailed relationship over the full spectrum of the lipid values has not been validated. To acquire a clear, reliable understanding of the relationship between BRCA risk and lipid values, an approach that can directly assess the BRCA risk at many different lipid levels, representing the full spectrum of the lipid variable, is needed. Further, to understand the confounding effect of a variable such as age on the relationship of BRCA risk and a lipid, a map can be created to visualize the changes of the risk at two-dimensional coordinates (eg, [LDL, age]).

The All of Us research program is a national-wide effort to collect health-related information in 1 million US residents.[26,27] Electronic health records (EHR) of the current release (2020, v4) includes 315 297 participants. In this study, we aimed to provide a complete assessment of association of BRCA risk across the full spectrum of TC, LDL, HDL, and TG values

Table 1. Characteristics of female breast cancer cohorts from All of Us research program.

VARIABLES	COHORT 1		COHORT 2
	CASE N = 3209	CONTROL N = 31 650	CASE N = 3899
TC (mg/dL)	199.7 ± 40.2	185.3 ± 34.5	188.3 ± 41.0
LDL (mg/dL)	117.6 ± 36.4	107.3 ± 29.8	106 ± 37.5
HDL (mg/dL)	60.5 ± 18.0	58.0 ± 16.8	61.1 ± 17.7
TG (mg/dL)	114.2 ± 56.1	107.8 ± 50.7	116.4 ± 55.3
Age (year)	57.1 ± 11.1	46.2 ± 14.6	58.1 ± 11.1
BMI (kg/m^2)	27.8 ± 7.4	29.0 ± 8.2	27.9 ± 7.6
White (%)	2103 (65.5)	16843 (53.2)	2587 (66.4)
Black (%)	483 (15.1)	5740 (18.1)	614 (15.7)
Hispanic (%)	419 (13.1)	6799 (21.5)	450 (11.5)
Asian (%)	80 (2.5)	918 (2.9)	93 (2.4)

Participants for cohort 1 have lipid measurements prior to lipid treatment. Cases in cohort 2 are mixed with lipid treated and untreated BRCA patients, while the controls are the same as that of cohort 1. Cohort 2 is used to examine the effect of lipid-treatment on the association of breast cancer with cholesterol. Participants with unknown race-ethnicity are not listed. Data are expressed as mean ± SD, or n (%).

while accounting for age, BMI and race/ethnicity using data from the All of Us Research Program.

Materials and Methods

Study design and participants

ICD9 and ICD10 codes 174 and C50 are used to identify BRCA cases in All of Us database.[27] There are 6519 BRCA cases, where 6347 females, 86 males and 86 others. Since the number of non-female BRCA patients was small and BRCA mechanisms for male and female may differ to some extent, male and unknown-gender BRCA patients are excluded from our study. BRCA related carcinoma in situ of BRCA (ICD 9 code 233 and ICD10 code D05) is excluded from our analysis. The age for BRCA cases is calculated from the first diagnosis date, while for controls it is calculated from the date of the lipid measurement. Participant BMI is calculated from weight and height that are measured at the time of physical measurement assessment for all participants. For race/ethnicity, the common categories White, Black, Hispanic, and Asian are used.

To find the intact association of BRCA with cholesterol and to examine the influence of lipid-treatment on the association, we selected 2 cohorts (Table 1) that make use of all available participants for 2 different situations. In cohort 1, cases are female BRCA patients who have all variate and covariates data (TC, LDL, HDL, TG, BMI, age, and race/ethnicity), where TC, LDL, HDL, and TG values that are measured before any statin treatment, including atorvastatin, cerivastatin, fluvastatin, lovastatin, pitavastatin, pravastatin, rosuvastatin, and simvastatin, and are most close to the diagnosis date of BRCA are chosen for analysis. Controls were female participants who have all covariates data but do not have any record of BRCA, carcinoma in situ of BRCA, or statin treatment. So, no influence of lipid-treatment on the association is expected from cohort 1. In Cohort 2, no restriction of lipid treatment is applied for each case, therefore, it includes cases whose lipid values are measured after lipid treatment. The last record of the multiple measurements of TC, LDL, HDL or TG is chosen without considering the date of drug treatment. Cohort 2 uses the same controls as cohort 1.

Visualization of the association

In order to estimate the shape of the BRCA association with each *continuous variable* without assuming a linear relationship, we used prevalence (risk) curves and spline regression curves.[28] A risk curve directly accesses the BRCA prevalence P at a series of intervals for a continuous variable x, while a spline regression curve shows the logit of P (log [P/(1-P)]) at a series of knots of x. For a perfect linear association, the logit curve is a line, while the risk curve is a sigmoid function which has a linear region around the center. To visually estimate the confounding effect between 2 continuous variables, we extend the risk curve to a two-dimensional map that assesses the risk variation over 2 coordinates.

To visualize the association of BRCA with *continuous* variables with above approaches, we take a wide value range for each variable and then divide the range into fine-scale bins to reveal the detailed relationship. Here, a bin is defined as a half-closed interval which is labeled by the end value, for example, the bin of age 50 with a bin-size of 5 years is (45, 50). In this study, the number of intervals or nodes for each variable is set

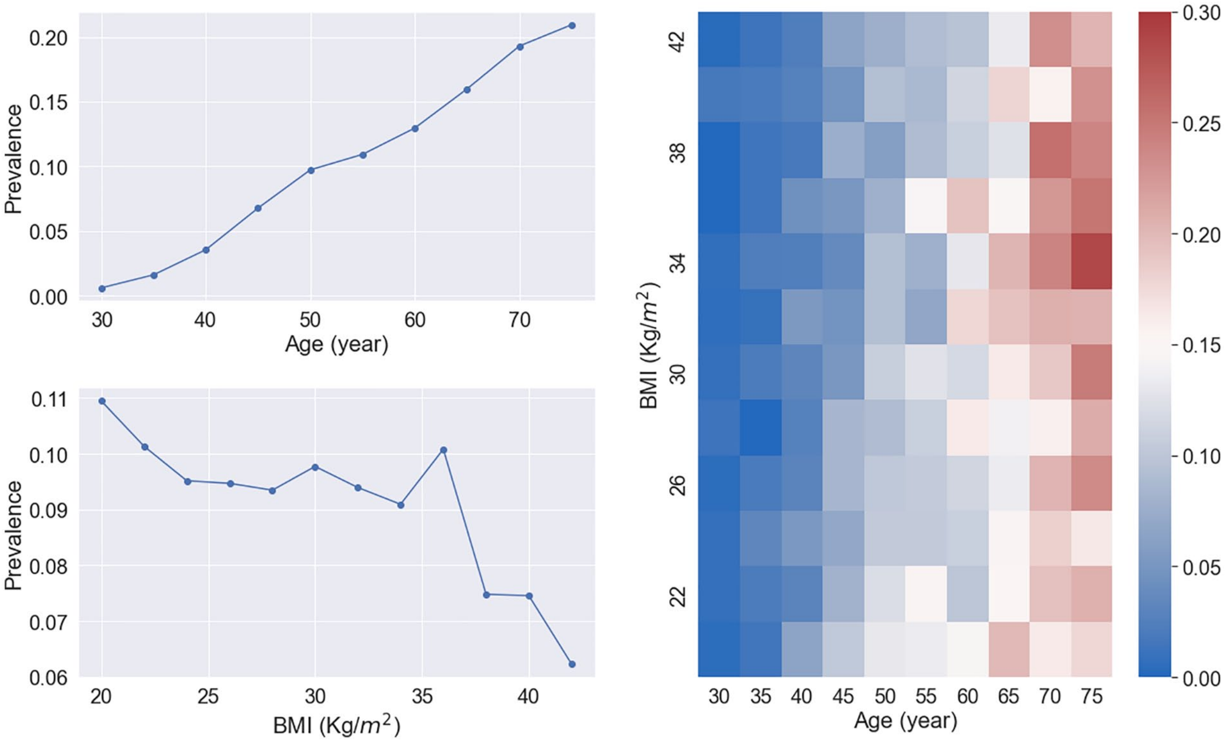

Figure 1. BRCA risk curves of age and BMI (left) and the risk map of age and BMI (right) for cohort 1. Age is positively and strongly associated with BRCA and the association of BMI and BRCA is confounded by Age. The color scale-bar value for the risk map represents breast cancer prevalence.

to at least 10. Values outside the range are put into their nearest bins. The value ranges are 100-300 mg/dL for TC, 40-220 mg/dL for LDL, 30-100 mg/dL for HDL, 50-250 mg/dL for TG, 30-75 years for age, and 22-40 kg/m^2 for BMI. These continuous variables are divided into fine-scale bins (interval size 5-20 mg/dL for lipids, 5 years for age and 2 kg/m^2 for BMI), then BRCA risk (number of BRCA cases over number of participants) at each bin is calculated. This results in a one-dimensional curve for univariate analysis and a two-dimensional map for bivariate analysis.

Quantitation of the association

Logistic regression is used to quantitate the relationship between BRCA with each variable and combination of variables. Lipid values measured after statin treatment are used to evaluate the treatment effect on the BRCA association. For multivariable logistic regression, age, BMI, and race/ethnicity are included as covariates. Our strategy for multivariable logistic regression was to include only non-linear terms that are significant and validated as well by the risk-curve and risk-map. Because of the co-linearity of variable TC with LDL and HDL, and variable White with Black, Hispanic, and Asian, 2 separate multivariable logistic regressions were carried out for the final models: one with variables LDL, HDL, TG, age, BMI, and White, and the other with variables TC, age, BMI, Black, Hispanic, and Asian. The results for age and BMI are the average over the 2 regressions.

The *continuous* variables TC, LDL, HDL, TG, BMI and age either have different units or have the same unit but different ranges, so the usual logistic regression coefficient and odds ratio (OR) cannot be directly compared to rank their association strengths with BRCA. To aid in the direct interpretation between effect sizes of different lipid values, we calculated the standardized regression coefficients by first multiplying the usual coefficients with standard deviations of the independent variables and then dividing by the standard deviation of the dependent variable.[29] The standardized odds ratio can then be calculated from the standardized coefficient. Python module *statsmodels* was used for logistic regression in this work and the Jupyter notebook source code can be accessed as All of Us registered users.

Results

Visualization of associations between BRCA and cholesterol measures

The visualization is applied to *continuous* variables TC, LDL, HDL, TG, age, and BMI. Figure 1 shows the risk curves of BRCA versus age and BRCA versus BMI, and 2D risk map of BRCA versus [age, BMI]. The age curve reveals a near-linear and positive relationship between BRCA and age, and the BMI curve shows an overall negative association between BRCA and BMI with apparent fluctuations. The *risk variability* (highest risk – lowest risk) of a curve may reflect the underlying association strength. The risk variability is 0.2 for the age

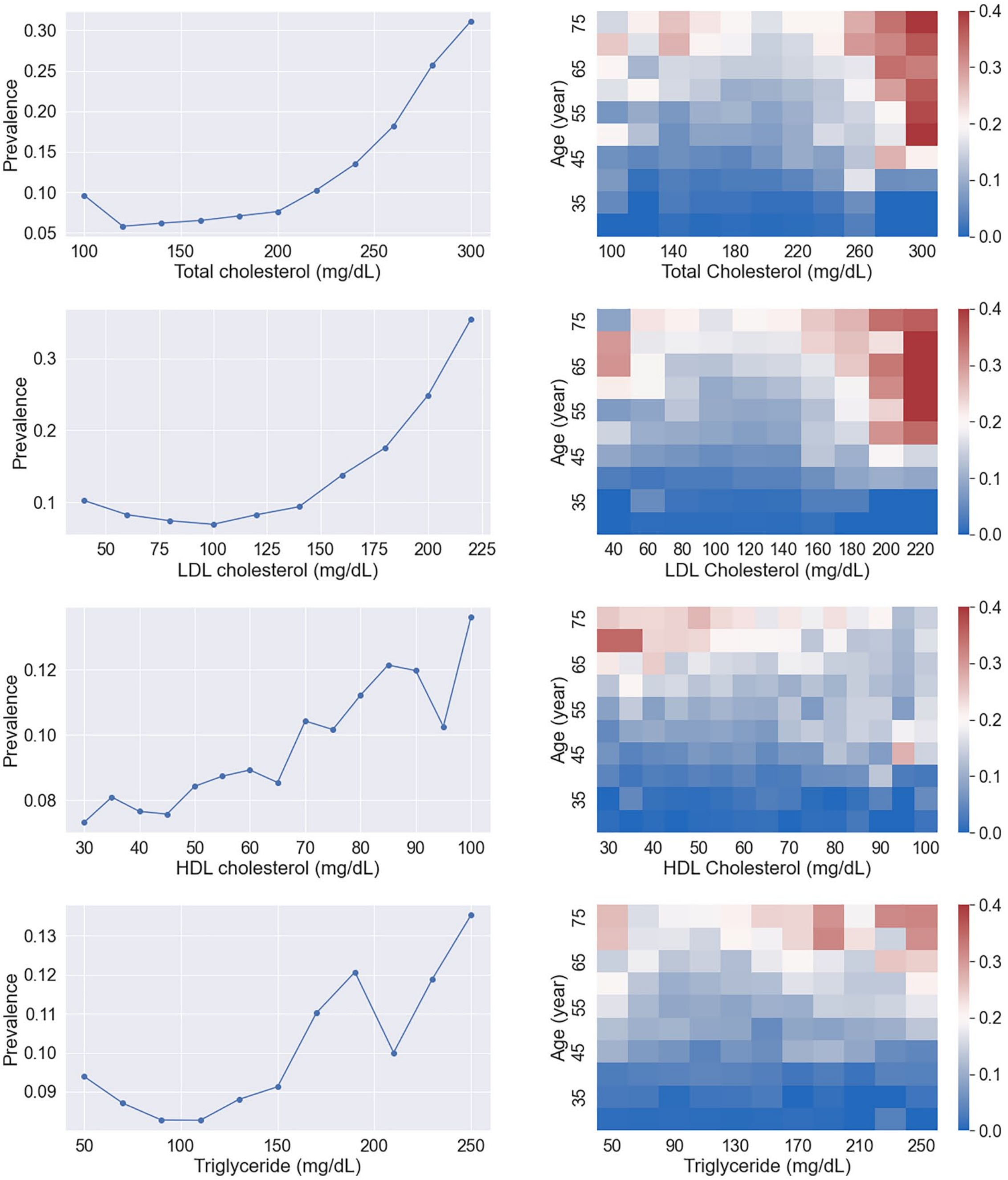

Figure 2. BRCA risk curves (left) of lipids and their risk maps (right) with Age for cohort 1. The overall associations of TC and LDL with BRCA are positive, strong and their curve shape is close to the sigmoid function. The association of HDL with BRCA is confounded by Age. Triglyceride is weakly associated with BRCA. The color scale-bar value for the risk map represents breast cancer prevalence.

curve and 0.05 for the BMI curve, indicating that the association of BRCA with age is much stronger than that of BRCA with BMI. Although the overall BRCA-BMI association is negative, the highest risk region in [age, BMI] map is located at high BMI (3438) and high age (6575). As age is the dominating covariate, only 2D risk maps for lipid with age are shown (Figure 2). The risk curve of BRCA with TC is very similar to that with LDL: strong, positive, smooth, and approximated as Sigmoid functions, although there is a small rise at the lower end. The risk variability is 0.24 for TC and 0.29 for LDL, and both associations are much strong than BMI.

The HDL curve shows an overall positive association between HDL and BRCA. The risk variability is 0.07, suggesting that HDL association is weaker than TC or LDL association. Although the HDL curve shows a positive and near-linear association, the [HDL, age] map shows that the main high-risk region is located at low HDL and high age, indicating a strong confounding effect of age on HDL. By careful examination of this map, a reverse association of HDL with BRCA for high age (⩾60) is discernible. The TG curve has an irregular shape with a risk variability of 0.05. The overall association appears positive but very weak. Spline-regression curves

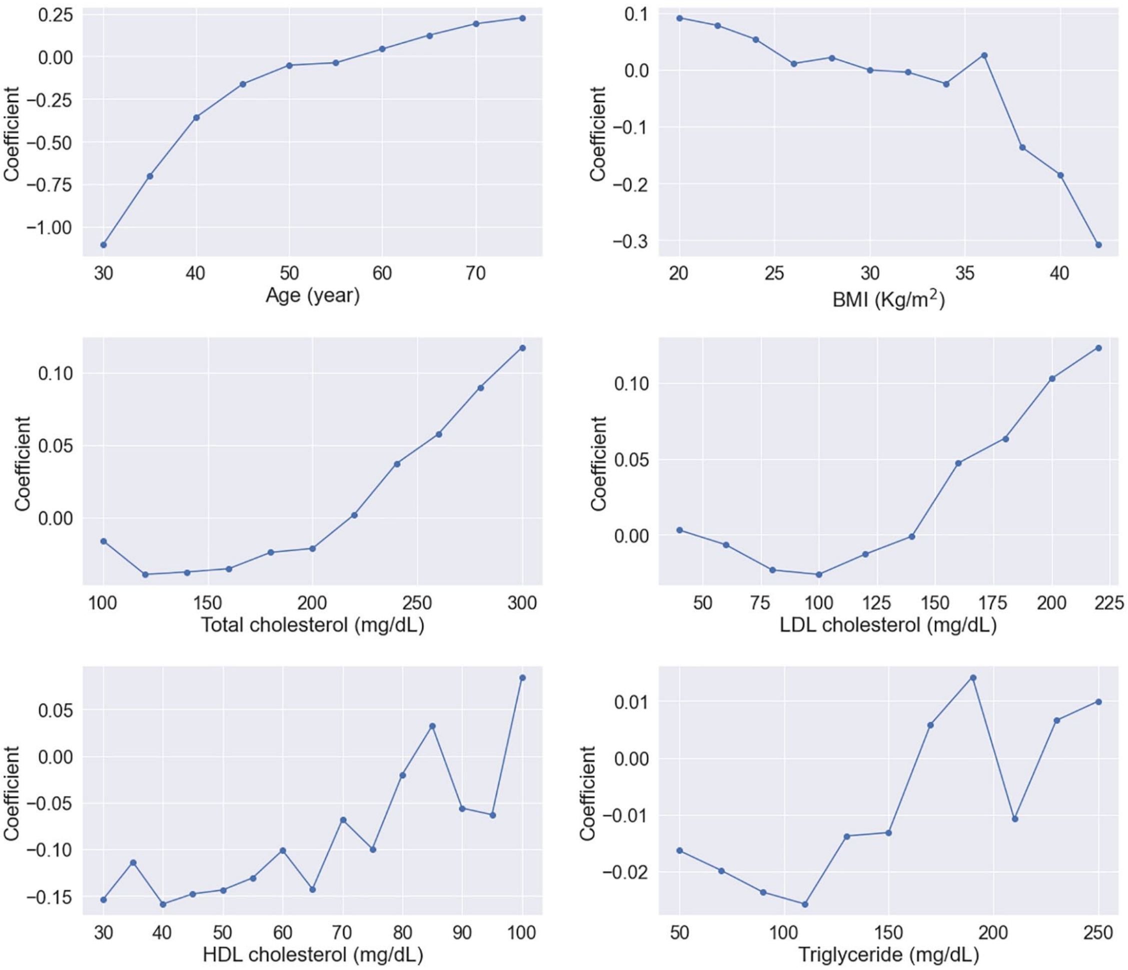

Figure 3. Spline regression curves for Age, BMI and lipids for cohort 1. These curves are similar to the risk curves shown in Figures 1 and 2. The association coefficient of triglyceride with BRCA is much smaller than those of TC, LDL, and HDL, suggesting a null association between TG and BRCA.

showed similar relationships (Figure 3). The association coefficient of triglyceride with BRCA is much smaller than those of TC, LDL and HDL, suggesting a null association between TG and BRCA.

Figure 4 shows the risk curve and maps of TC, LDL, HDL, and TG for cohort 2. These risk curves and maps for TC and LDL are clearly altered when compared to the non-statin treated case in Figure 2: the shape is significantly deviated from the sigmoid/linear function. Significant reverse associations are present for TC and LDL on the lower side. This can be seen more clearly from the [TC, age] and [LDL, age] maps, where the main high-risk region is on the top-left corner. On the contrary, the shapes of HDL and TG risk curves are more linear than their non-treated curves and maps, and the associations appear enhanced (risk variability is 0.09 for HDL and 0.08 for TG).

Quantitation of associations between BRCA and cholesterol measures

The above visual inspection for cohort 1 suggests that variables TC, LDL, and TG can be treated as linear, but for HDL a non-linear cross term HDL × age should be added to the multivariable logistic regression because of the strong confounding effect between HDL and age (Table 2). Coefficients of HDL and HDL × age can be combined as 0.0484(1-age/55.8) HDL, suggesting a positive HDL association for age $<$ 56 and a negative association for age $\geqslant$ 56. At age ~56, the BRCA risk is nearly constant over the full range of HDL (Figure 2). If the cross-term is not included in the model, that is, if a linear HDL relationship is assumed, the logistic regression gives a coefficient of −0.0009 with *P*-value of .49 (listed as *HDL** in Table 2), which would suggest a null association for HDL.

By splitting cohort 1 into 2 sub-cohorts at Age = 56, we can quantitate the 2 sub-cohorts separately. Table 3 lists the quantitation results, which suggest that HDL and BMI are age-dependent, while LDL, TG and race/ethnicity are not age-dependent. The standardized odds ratio is 1.69 for the positive HDL association and 0.653 for the negative association; the pre-menopausal group shows a strong BMI negative association (standardized OR 0.58, $P = 8.9 \times 10^{-5}$), while the post-menopausal group shows a nearly null BMI association (standardized OR 0.954, $P = .53$). The association of age with BRCA decreases for the post-menopausal

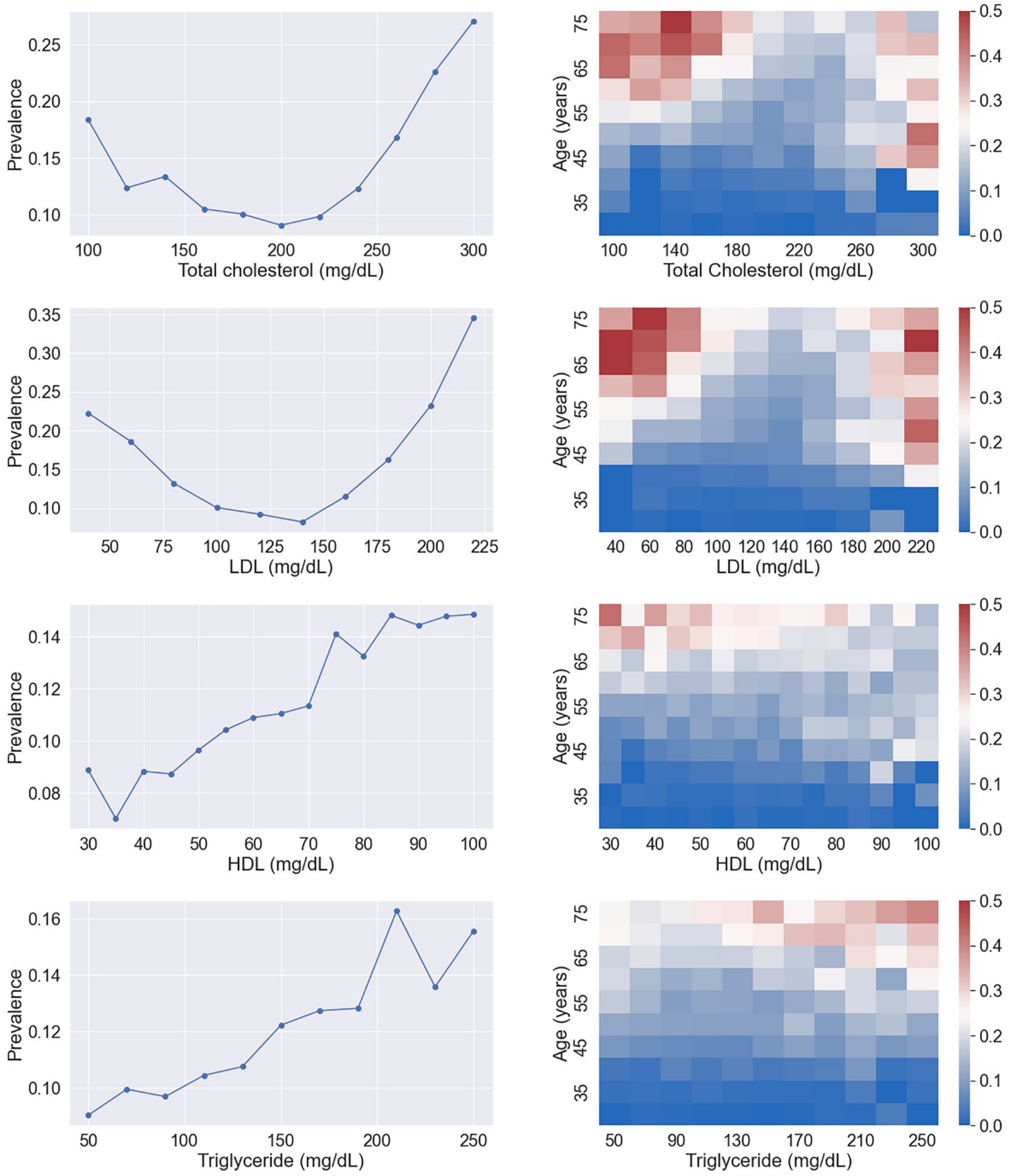

Figure 4. Visualization of the influence of lipid treatment on the association of BRCA with TC, LDL, and HDL for cohort 2. The color scale-bar value for the risk map represents breast cancer prevalence.

group (from standardized OR of 33.9 to 1.48). According to the standardized odds ratio in Tables 2 and 3 we can rank the association strengths of the *continuous* variables with BRCA: for positive association, age (overall 13.8 from Age* in Table 2) >TC (2.12, Table 2) ~LDL (1.99, Table 2) >HDL (pre-menopausal 1.69, Table 3) >TG (1.28, Table 2); for negative associations, BMI (pre-menopausal 0.58, Table 3) is stronger than HDL (post-menopausal 0.65, Table 3). The strong associations of BRCA with variable age, TC and LDL are also reflected by the large differences between the case mean and control mean (Table 1).

The risk curves and maps in Figure 4 were created using cohort 2 and reveal that among lipids TC and LDL are affected mostly by statin treatment and their associations are significantly non-linear, so quadratic term TC^2 and LDL^2 should be included in the multivariable logistic regression. Variable TC and its square term TC^2 can be combined as $1.0028 \times 10^{-4}(TC-207.4)^2$ (Table 4). This suggests that statin treatment leads to 2 strong opposite associations that are separated at 207 mg/dL: a negative association at lower TC, a positive association at higher TC and an optimal TC value 207 mg/dL at which the risk with TC is the minimum. Similarly, LDL and LDL^2 can be expressed as $1.74 \times 10^{-4}(LDL-129.1)^2$, suggesting 2 opposite associations separated at 129 mg/dL. *TC** and *LDL** in Table 4 are logistic regression results assuming linear relationship for TC and LDL. The coefficient of –0.0032 for TC

Table 2. Multivariable logistic regression results for cohort 1.

VARIABLES	COEFFICIENT (95% CI)	OR (95% CI)	STANDARDIZED OR (95% CI)	*P*-VALUE
TC	0.0063 (0.0053, 0.0074)	1.006 (1.005, 1.007)	2.117 (1.9, 2.48)	5.6×10^{-31}
LDL	0.0065 (0.0053, 0.0077)	1.0065 (1.0053, 1.0077)	1.989 (1.752, 2.258)	2.6×10^{-26}
HDL	0.0484 (0.0385, 0.0583)	1.0496 (1.0393, 1.06)	17.162 (9.6, 30.676)	8.6×10^{-22}
HDL × Age	−0.000868 (−0.001, −0.0007)	0.999 (0.999, 1.0)	0.0175 (0.0079, 0.0387)	2.3×10^{-23}
HDL*	−0.0009 (−0.0035, 0.0017)	0.999 (0.997, 1.002)	0.947 (0.813, 1.103)	.49
TG	0.0014 (0.0005, 0.0022)	1.001 (1.0005, 1.002)	1.276 (1.1, 1.479)	.0012
BMI	−0.0066 (−0.011, −0.002)	0.993 (0.989, 0.998)	0.842 (0.745, 0.952)	5.9×10^{-3}
Age	0.103 (0.0923, 0.1137)	1.1085 (1.097, 1.112)	187.98 (109.0, 324.2)	4.0 × 10-79
Age*	0.052 (0.049, 0.055)	1.053 (1.05, 1.056)	13.827 (11.871, 16.1)	8.9 × 10-250
White	0.2217 (0.1412, 0.3023)	1.248 (1.152, 1.353)	1.465 (1.275, 1.684)	6.9×10^{-8}
Black	−0.116 (−0.224, −0.008)	0.89 (0.799, 0.992)	0.857 (0.743, 0.99)	.036
Hispanic	−0.303 (−0.415, −0.191)	0.738 (0.66, 0.826)	0.654 (0.559, 0.765)	1.1×10^{-7}
Asian	−0.007 (−0.249, 0.234)	0.993 (0.78, 1.26)	0.996 (0.866, 1.144)	.95

The odds ratio (OR) for continuous variables is for 1 unit increase (mg/dL for lipid, year for age and kg/m^2 for BMI), and the Standardized OR is unit normalized. The coefficients of HDL and the cross-term HDL × age together can be expressed as 0.0484(1-age/55.8)HDL, which suggests a positive HDL association for age < 56 and a negative association for age >= 56. HDL* and age* are logistic regression results assuming linear relationship (without the cross-term HDL × age).

Table 3. Logistic regression results with 2 age sub-groups of Cohort 1, subgroup with age<56 (~pre-menopausal group), and subgroup with age ⩾ 56 (~post-menopausal group).

VARIABLES	COEFFICIENT (95% CI)		ODDS RATIO (95% CI)		STANDARDIZED OR (95% CI)		*P*-VALUE	
	AGE < 56	AGE ⩾ 56	AGE < 56	AGE ⩾ 56	AGE < 56	AGE ⩾ 56	AGE < 56	AGE ⩾ 56
TC	0.0075 (0.0058, 0.0091)	0.0044 (0.0029, 0.0058))	1.0075 (1.0058 1.0092)	1.0044 (1.0029, 1.0058)	2.97 (2.333, 3.782)	1.518 (1.324, 1.741)	9.9×10^{-19}	2.2×10^{-9}
LDL	0.0063 (0.0045 0.0082)	0.0056 (0.004, 0.0072))	1.0064 (1.0045, 1.0082)	1.0056 (1.004, 1.007)	2.246 (1.773, 2.847)	1.607 (1.404, 1.838)	5.2×10^{-14}	4.9×10^{-12}
HDL	0.0077 (0.0038, 0.0115)	−0.0088 (−0.0123, −0.0052)	1.0077 (1.0038, 1.0116)	0.991 (0.988, 0.995)	1.69 (1.3, 2.198)	0.653 (0.55, 0.776)	9.0×10^{-5}	1.2×10^{-6}
TG	0.0014 (0.0002, 0.0026)	0.0011 (0.0, 0.0023)	1.0014 (1.0002, 1.0026)	1.0011 (1.0, 1.0023)	1.377 (1.053, 1.801)	1.161 (0.995, 1.355)	.019	0.057
Age	0.0792 (0.0722, 0.0863)	0.0237 (0.0156, 0.0318)	1.0825 (1.0748, 1.09)	1.024 (1.016, 1.032)	33.91 (24.77, 46.44)	1.478 (1.294, 1.689)	5.3×10^{-7}	8.9×10^{-9}
BMI	−0.0149 (−0.0224, −0.0075)	−0.0024 (−0.0098, 0.005)	0.985 (0.978, 0.993)	0.998 (0.99, 1.005)	0.58 (0.441, 0.762)	0.954 (0.825, 1.103)	8.9×10^{-5}	0.53
White	0.2524 (0.1379, 0.367)	0.2415 (0.1279, 0.355)	1.287 (1.148, 1.443)	1.273 (1.136, 1.426)	1.722 (1.346, 2.204)	1.355 (1.174, 1.562)	1.6×10^{-5}	3.1×10^{-5}
Black	−0.1674 (−0.322, −0.0127)	−0.1464 (−0.298, 0.0058)	0.8458 (0.725, 0.987)	0.864 (0.742, 1.059)	0.7531 (0.579, 0.988)	0.869 (0.752, 1.006)	.034	0.059
Hispanic	−0.3613 (−0.5111, −0.2115)	−0.2808 (−0.4496, −0.112)	0.6967 (0.5998, 0.8094)	0.7552 (0.638, 0.894)	0.512 (0.388, 0.676)	0.781 (0.673 0.906)	2.3×10^{-6}	1.1×10^{-3}
Asian	−0.0989 (−0.4168, 0.2189)	0.0569 (−0.3167, 0.4305)	0.906 (0.659, 1.245)	1.0585 (0.729, 1.538)	0.926 (0.724, 1.185)	1.02 (0.894, 1.165)	.54	0.77

Table 4. Multivariable logistic regression results for cohort 2 that is affected by lipid treatment.

VARIABLES	COEFFICIENT (95% CI)	OR (95% CI)	STANDARDIZED OR (95% CI)	*P*-VALUE
TC	−0.0416 (−0.0471, −0.0361)	0.959 (0.954, 0.965)	0.0091 (0.0049, 0.017)	2.2×10^{-49}
TC^2	$1.0028\text{x}10^{-4}$ ($8.6\text{x}10^{-5}$, $1.1\text{x}10^{-4}$)	1.0001 (1.0, 1.00011)	76.79 (41.66, 141.5)	5.4×10^{-44}
TC*	−0.0032 (−0.0042, −0.0022)	0.997 (0.996, 0.998)	0.695 (0.621, 0.778)	2.7×10^{-1}
LDL	−0.0449 (−0.0492, −0.0406)	0.956 (0.952, 0.96)	0.0121 (0.0079, 0.0185)	1.6 × 10-92
LDL^2	$1.74\text{x}10^{-4}$ (0.00016, 0.00019)	1.000174 (1.0001,1.0002)	51.767 (34.26, 78.21)	2.3 × 10-78
LDL*	−0.0054 (−0.0065, −0.0043)	0.995 (0.993, 0.996)	0.5876 (0.5259, 0.6566)	5.9×10^{-21}
HDL	0.0041 (0.0017, 0.0064)	1.004 (1.002, 1.006)	1.246 (1.099 1.413)	5.8×10^{-4}
TG	0.0035 (0.0028, 0.0043)	1.0036 (1.0028, 1.0043)	1.79 (1.584, 2.023)	1.1×10^{-2}
Age	0.0585 (0.0557, 0.0613)	1.06 (1.0573, 1.063)	15.86 (13.913, 18.1)	<10-300
BMI	−0.0062 (−0.011, −0.0014)	0.994 (0.989, 0.999)	0.849 (0.749, 0.964)	.01
White	0.2861 (0.2106, 0.3616)	1.3313 (1.234, 1.436)	1.577 (1.399, 1.779)	1.1×10^{-13}
Black	−0.181 (−0.2798, −0.0823)	0.834 (0.756, 0.92)	0.801 (0.71, 0.904)	3.2×10^{-4}
Hispanic	−0.4644 (−0.5727, −0.356)	0.629 (0.564, 0.7)	0.549 (0.478, 0.632)	4.3×10^{-17}
Asian	−0.0447 (−0.272, 0.183)	0.9563 (0.762, 1.2)	0.976 (0.865, 1.102)	.07

Variable units, mg/dL for lipid, year for age and kg/m^2 for BMI. Variable TC and its square term TC^2 can be expressed as $1.0028 \times 10^{-4}(\text{TC-}207.4)^2$, LDL and LDL^2 can be expressed as $1.74 \times 10^{-4}(\text{LDL-}129.1)^2$, suggesting that TC and LDL have 2 opposite associations with BRCA separated at 207 mg/dL and 129 mg/dL respectively. TC* and LDL* are logistic regression results assuming linear relationship for TC and LDL, which lead to an overall negative association for both.

and −0.0054 for LDL would indicate that the overall associations are both negative.

Discussion

The association of BRCA with lipid measures has been studied extensively, but with inconsistent results. Here we utilize a large, contemporary cohort to visualize and quantify the association of BRCA with lipid values and other co-variates. Our analysis suggests that utilization of logistic regression that assumes a simple linear between BRCA and lipids leads to erroneous conclusions due to difference in disease risk across the full spectrum of lipid values. Using one-dimensional risk curves and two-dimensional risk maps, we visually inspected the linearity of associations between BRCA and cholesterol, and the confounding effects between cholesterol values, age, and body mass index. We find that the associations of BRCA with TC and LDL are strong, positive, and near-linear, the association with HDL is non-linear and age-dependent, and the association with TGs is very weak or null. The size of the All of Us cohort also enabled a high resolution of variable relationships. Limitations of our analysis included a lack of consideration of other laboratory data such as estrogen levels and a lack of consideration of non-statin antihyperlipidemic treatment.

Conclusion

In summary, we visualized and quantified the association of BRCA with lipid measures and other co-variates. We find that the associations of BRCA with TC and LDL are strong, positive, and near-linear, the association with HDL is non-linear and age-dependent, and the association with TGs is very weak or null. We explored effects of statin treatment on the associations and demonstrated that the inclusion of treated lipid values can significantly alter the underlying associations. Our study demonstrates that the use of the logistic regression without considering BRCA risk across the spectrum of lipid values may lead to inconsistent results.

Author Contributions

Conceptualization and design: JF, JHK, and EAS; Data curation: JF; Formal analysis: JF; Funding acquisition: JHK; Investigation: JF and JHK; Methodology: JF and JHK; Project administration: JHK; Resources: JHK; Software: JF; Validation: JF; Visualization: JF; Roles/Writing – original draft: JF and JHK; Writing – review & editing: JF, JHK, and EAS.

REFERENCES

1. Kitahara CM, Berrington de González A, Freedman ND, et al. Total cholesterol and cancer risk in a large prospective study in Korea. *J Clin Oncol*. 2011; 29:1592-1598.
2. Nowak C, Ärnlöv J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk. *Nat Commun*. 2018;9:3957.
3. Owiredu WK, Donkor S, Addai BW, Amidu N. Serum lipid profile of breast cancer patients. *Pak J Biol Sci*. 2009;12:332-338.
4. Rodrigues Dos Santos C, Fonseca I, Dias S, Mendes de Almeida JC. Plasma level of LDL-cholesterol at diagnosis is a predictor factor of breast tumor progression. *BMC Cancer*. 2014;14:132.

5. Martin LJ, Melnichouk O, Huszti E, et al. Serum lipids, lipoproteins, and risk of breast cancer: a nested case-control study using multiple time points. *J. Natl. Cancer Inst*. 2015;107:djv032.
6. Potischman N, McCulloch CE, Byers T, et al. Associations between breast cancer, plasma triglycerides, and cholesterol. *Nutr Cancer*. 1991;15:205-215.
7. Ni H, Liu H, Gao R. Serum lipids and breast cancer risk: A meta-analysis of prospective cohort studies. *PLoS One*. 2015;10:e0142669.
8. Bosco JL, Palmer JR, Boggs DA, Hatch EE, Rosenberg L. Cardiometabolic factors and breast cancer risk in U.S. Black women. *Breast Cancer Res Treat*. 2012;134:1247-1256.
9. Chandler PD, Song Y, Lin J, et al. Lipid biomarkers and long-term risk of cancer in the women's health study. *Am J Clin Nutr*. 2016;103:1397-1407.
10. Touvier M, Fassier P, His M, et al. Cholesterol and breast cancer risk, a systematic review and meta-analysis of prospective studies. *Br J Nutr*. 2015;114: 347-357.
11. Borgquist S, Butt T, Almgren P, et al. Apolipoproteins, lipids and risk of cancer. *Int. J. Cancer*. 2016;138:2648-2656.
12. His M, Dartois L, Fagherazzi G, et al. Associations between serum lipids and breast cancer incidence and survival in the E3N prospective cohort study. *Cancer Causes Control*. 2017;28:77-88.
13. His M, Zelek L, Deschasaux M, et al. Prospective associations between serum biomarkers of lipid metabolism and overall, breast and prostate cancer risk. *Eur J Epidemiol*. 2014;29:119-132.
14. Llanos AA, Makambi KH, Tucker CA, Wallington SF, Shields PG, Adams-Campbell LL. Cholesterol, lipoproteins, and breast cancer risk in African American women. *Ethn Dis*. 2012;22:281-287.
15. Li X, Liu ZL, Wu YT, et al. Status of lipid and lipoprotein in female breast cancer patients at initial diagnosis and during chemotherapy. *Lipids Health Dis*. 2018;17:91.
16. Michalaki V, Koutroulis G, Syrigos K, Piperi C, Kalofoutis A. Evaluation of serum lipids and high-density lipoprotein subfractions (HDL2, HDL3) in postmenopausal patients with breast cancer. *Mol Cell Biochem*. 2005; 268:19-24.
17. Kucharska-Newton AM, Rosamond WD, Mink PJ, Alberg AJ, Shahar E, Folsom AR. HDL-cholesterol and incidence of breast cancer in the ARIC cohort study. *Ann Epidemiol*. 2008;18:671-677.
18. Furberg AS, Veierod MB, Wilsgaard T, Bernstein L, Thune I. Serum high-density lipoprotein cholesterol, metabolic profile, and breast cancer risk. *J. Natl. Cancer Inst*. 2004;96:1152-1160.
19. Michels KB, Terry KL, Willett WC. Longitudinal study on the role of body size in premenopausal breast cancer. *Arch Intern Med*. 2006;166:2395-2402.
20. Berstad P, Coates RJ, Bernstein L, et al. A case-control study of body mass index and breast cancer risk in white and African-American women. *Cancer Epidemiol Biomarkers Prev*. 2010;19:1532-1544.
21. White AJ, Nichols HB, Bradshaw PT, Sandler DP. Overall and central adiposity and breast cancer risk in the sister study. *Cancer*. 2015;121:3700-3708.
22. van den Brandt PA, Spiegelman D, Yaun SS, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol*. 2000; 152:514-527.
23. Li CI, Malone KE, Daling JR. Interactions between body mass index and hormone therapy and postmenopausal breast cancer risk (United States). *Cancer Causes Control*. 2006;17:695-703.
24. Neuhouser ML, Aragaki AK, Prentice RL, et al. Overweight, obesity, and postmenopausal invasive breast cancer risk. *JAMA Oncol*. 2015;1:611-621.
25. Sebastiani F, Cortesi L, Sant M, et al. Increased incidence of breast cancer in postmenopausal women with high body mass index at the modena screening program. *J Breast Cancer*. 2016;19:283-291.
26. Denny JC, Rutter JL, Goldstein DB, et al. The "All of us" research program. *N Engl J Med*. 2019;381:668-676.
27. Karnes JH, Arora A, Feng J, et al. Racial, ethnic, and gender differences in obesity and body fat distribution, an all of us research program demonstration project. *PLoS One*. 2021;16:e0255583.
28. Harrell F. *Regression Modeling Strategies*. Springer Science & Business Media; 2015.
29. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized regression coefficients, a further critique and review of some alternatives. *Epidemiology*. 1991;2:387-392.