Circulating Carotenoids and Risk of Breast Cancer: Pooled Analysis of Eight Prospective Studies


Background
Carotenoids, micronutrients in fruits and vegetables, may reduce breast cancer risk. Most, but not all, past studies of circulating carotenoids and breast cancer have found an inverse association with at least one carotenoid, although the specific carotenoid has varied across studies.

Methods
We conducted a pooled analysis of eight cohort studies comprising more than 80% of the world's published prospective data on plasma or serum carotenoids and breast cancer, including 3055 case subjects and 3956 matched control subjects. To account for laboratory differences and examine population differences across studies, we recalibrated participant carotenoid levels to a common standard by reassaying 20 plasma or serum samples from each cohort together at the same laboratory. Using conditional logistic regression, adjusting for several breast cancer risk factors, we calculated relative risks (RRs) and 95% confidence intervals (CIs) using quintiles defined among the control subjects from all studies. All P values are two-sided.

Results
Statistically significant inverse associations with breast cancer were observed for α-carotene (top vs bottom quintile RR = 0.87, 95% CI = 0.71 to 1.05, \( P_{\text{trend}} = 0.04 \), \( \beta \)-carotene (RR = 0.83, 95% CI = 0.70 to 0.98, \( P_{\text{trend}} = 0.02 \)), lutein+zeaxanthin (RR = 0.84, 95% CI = 0.70 to 1.01, \( P_{\text{trend}} = 0.05 \)), lycopene (RR = 0.78, 95% CI = 0.62 to 0.99, \( P_{\text{trend}} = 0.02 \)), and total carotenoids (RR = 0.81, 95% CI = 0.68 to 0.96, \( P_{\text{trend}} = 0.01 \)). \( \beta \)-Cryptoxanthin was not statistically significantly associated with risk. Tests for heterogeneity across studies were not statistically significant. For several carotenoids, associations appeared stronger for estrogen receptor negative (ER−) than for ER+ tumors (eg, \( \beta \)-carotene: ER+: top vs bottom quintile RR = 0.52, 95% CI = 0.36 to 0.77, \( P_{\text{trend}} = 0.001 \); ER−: RR = 0.83, 95% CI = 0.66 to 1.04, \( P_{\text{trend}} = 0.06 \); \( P_{\text{heterogeneity}} = 0.01 \)).

Conclusions
This comprehensive prospective analysis suggests women with higher circulating levels of α-carotene, \( \beta \)-carotene, lutein+zeaxanthin, lycopene, and total carotenoids may be at reduced risk of breast cancer.


Carotenoids, natural pigments ranging from yellow to red, are important for photosynthesis in plants. More than 600 carotenoids have been identified (1), approximately 40 are present in the US diet, 20 of which are measurable in tissue and serum. α-Carotene, \( \beta \)-carotene, \( \beta \)-cryptoxanthin, lutein, zeaxanthin, and lycopene are the most prevalent, comprising 90% of circulating carotenoids (1–3). These micronutrients are suspected to be anticarcinogenic, with possible biologic activities including antioxidation, enhanced gap-junction intercellular communication, immunoenhancement, inhibition of tumorigenesis and malignant transformation, and metabolism to retinoids, which, in turn, contribute to cellular differentiation (4–10). Experimental studies suggest carotenoids inhibit tumor progression and reduce proliferation in both estrogen receptor positive (ER+) and ER− breast cancer cells (11).

Studies of dietary intake, of fruits and vegetables overall and carotenoids specifically, have had mixed results, although a modest inverse association has been suggested in several studies, as summarized in meta-analyses and pooled analyses (12–14). Several (15–19), but not all (20), studies of dietary patterns that include higher fruit and vegetable intake observed lower risk of ER−, but not ER+, breast tumors. Most recently, in pooled analyses of 18 cohort studies, stronger inverse associations were observed for ER− (vs ER+) tumors with intake of fruits and vegetables (S Y Jung, D Spiegelman, L Baglietto, et al., unpublished data) as well as intake of α-carotene, \( \beta \)-carotene, and lutein+zeaxanthin (21).

Examining circulating carotenoid levels overcomes potential weaknesses of dietary data, including recall of past diet (3), inaccuracies of nutrient databases to determine carotenoid content of
specific foods (22), influences of cooking and storage on carotenoid content (3,23), geographic and seasonal variation of foods (22), and individual differences in nutrient absorption. As a result of these limitations, the correlations between intake and plasma carotenoid levels are modest \( r = 0.2–0.4 \) (24,25). Over the last 13 years, 10 prospective studies have investigated the associations between circulating carotenoids and breast cancer risk (26–35), eight of which are included in this analysis (26–33). Most studies observed inverse associations, either suggested or statistically significant, with at least one of five carotenoids (\( \alpha \)-carotene, \( \beta \)-carotene, \( \beta \)-cryptoxanthin, lutein+zeaxanthin, lycopene), although the specific carotenoid and/or subgroup of importance varied among studies.

The collaborating investigators of eight studies conducted a pooled analysis of original data examining associations between circulating carotenoids and breast cancer risk, the results of which are presented here.

**Methods**

A review of the literature at the time this project was proposed revealed five small, early studies (14 to 67 case subjects) (36–40) and six more recent, larger prospective studies (26–31). Investigators of the six larger studies were contacted, and all agreed to collaborate. During the data collection phase, an additional two studies were published and subsequently included in the collaborative project (32,33). The eight prospective studies in this pooled analysis are: Columbia, Missouri (26); Umea, Sweden (27); New York University Women's Health Study (NYUWHS), New York, New York (28); CLUE I and CLUE II, Washington County, Maryland (29); Nurses' Health Study (NHS), United States (30); Women's Health Study (WHS), United States (31); Shanghai Women's Health Study (SWHS), Shanghai, China (32); and Multiethnic Cohort Study (MEC), California and Hawaii (33). Two studies were published after the analyses were underway, and were not included in this pooled analysis (34,35).

Each of the eight studies was a case–control study nested within a cohort, with plasma (27,29–33) or serum (26,28,29) samples collected among initially healthy women who were then followed for subsequent risk of breast cancer. Investigators of the original studies contributed data on circulating carotenoid concentrations, matching factors, covariates, and breast cancer diagnoses. Each study obtained informed consent (or proxy, such as implied consent with return of blood samples and questionnaires in the NHS) and approval from their respective institutional review boards. The pooled analysis was approved by the Committee on the Use of Human Subjects in Research at the Brigham and Women’s Hospital.

**Carotenoid Assays**

Reverse-phase high-performance liquid chromatography was used to measure plasma or serum carotenoids in each of the original studies, but assays were conducted at five different laboratories. To distinguish between-laboratory variation from between-population variation, we conducted a recalibration study. Plasma (27,29–33) or serum (26,28,29) samples from each study that had been assayed for carotenoids in the original analysis were sent to the NHS blood laboratory. Samples were from control subjects for seven of the eight studies; quality control samples were used for the remaining study (MEC). Control subject samples were chosen to represent a range of carotenoid levels, with selection of two control subjects from each study-specific quintile of \( \alpha \)-carotene and total carotenoids. The original data in the NHS showed considerable laboratory drift across four batches run in different years (30); thus, 20 control subjects from each of four batches were sent for recalibration. Samples from seven of eight studies (not SWHS) were realiquoted, labeled to blind the assay laboratory to the study identification, and packaged together in one project with additional quality control samples and 10 “drift pool” samples, which are additional unique quality control samples used to assess drift among batches. SWHS provided samples \(( n = 10 )\) that were assayed in a separate, later batch but with identical quality control and drift pool samples as were sent with the other study samples. Samples were assayed by reverse-phase high-performance liquid chromatography, using the methods described by El-Sohemy et al. (41), at the Micronutrient Analysis Laboratory in the Department of Nutrition at the Harvard School of Public Health. Individual carotenoid measures include all specific isomers (eg, cis+trans); lutein and zeaxanthin are read together, and therefore are a sum of the two individual carotenoids (lutein+zeaxanthin). Coefficients of variation from blinded quality control replicates ranged from 4.6% (lutein+zeaxanthin) to 8.2% (\( \alpha \)-carotene).

**Matching Factor and Covariate Data**

Information was collected on matching factors, measured plasma or serum carotenoid levels, and case–control status. Matching factors varied across cohorts, with cohorts matching on three to seven factors, including age at blood collection; date, time, and fasting status at blood collection; menopausal status; date of last menstrual period and/or phase and day of menstrual cycle in premenopausal women; postmenopausal hormone use; race or ethnicity; study center; smoking status; follow-up time; availability of food frequency questionnaire; use of antibiotics in the last week; number of blood collections within the cohort; and diagnosis of benign breast disease in the previous 2 years. Matches on date of blood collection included \( \pm 1 \) month (CLUE I, CLUE II, NHS, SWHS), \( \pm 2–3 \) months (Umea, Sweden; New York University Women's Health Study), \( \pm 6 \) months (MEC), and \( \pm 1 \) year (Columbia, MO). Although blood collection date was not a matching factor in WHS, case subjects were matched to control subjects on follow-up time (\( \pm 6 \) months), and 80% of case–control pairs had blood collection dates within 10 months (all were within 2 years).

To comprehensively adjust for possible confounders and allow for stratified analyses, we requested information from each study for the case and control subjects on age at menarche, parity, age at first birth, menopausal status, age at menopause, body mass index (BMI), family history of breast cancer, personal history of benign breast disease, smoking status, alcohol use, oral contraceptive use, postmenopausal hormone use, cholesterol level, physical activity, and race/ethnicity. The NHS and WHS cohorts had information available on each of these covariates. Data from each cohort was available for all variables except BMI, personal history of benign breast disease, alcohol use, cholesterol level, and physical activity. A range of one (CLUE II, SWHS) to four (CLUE I) variables was missing from each of the remaining cohorts.

Covariates were coded with as much detail as possible within the restriction of creating common definitions across cohorts.
Multiple cohorts were missing data on physical activity (5 cohorts), alcohol consumption (4 cohorts), and cholesterol levels (2 cohorts).

Case Subjects
From each cohort, we attempted to collect diagnosis information, including date, hormone receptor status, differentiation, size, and nodal involvement. A total of 3055 case subjects were included, matched with 3956 control subjects. Seven cohorts had information on hormone receptor status (n = 1481 ER+ case subjects; n = 417 ER− case subjects). We obtained information on grade from six cohorts (n = 1479 case subjects), information of size from four cohorts (n = 1617 case subjects), and information on nodal involvement from six cohorts (n = 1810 case subjects).

Statistical Analysis
Outliers, based on the original carotenoid values, were detected within each cohort using the extreme Studentized many-deviate approach (42). Outliers were removed for the individual carotenoids, with the sum of outliers across cohorts ranging from 2 to 17 for each carotenoid.

Original circulating carotenoid values were recalibrated to have a distribution comparable with the samples rerun together at the Harvard laboratory. Using linear regression within each cohort, we regressed the rerun values on the original values and used the resulting intercept and slope to predict recalibrated values for all cohort participants. Quintiles and deciles of individual and total carotenoids were defined among control subjects, within and across studies, using both original and recalibrated values, thus resulting in cut points classified in four ways: study-specific original values, study-specific recalibrated values, common original values, and common recalibrated values.

Analyses were conducted in two ways, by using a two-stage approach of pooled relative risks (RRs) and by an aggregated data approach of pooled raw data. In the two-stage approach, natural-log relative risks were calculated from conditional logistic regression models within each cohort and pooled using DerSimonian and Laird random effects models (43), with relative risks weighted by the inverse of their variance. We tested for heterogeneity among studies by the DerSimonian and Laird Q statistic (43). In the aggregated data approach, we pooled the raw data from each cohort and calculated a single, combined relative risk from a conditional logistic regression model (44).

Multivariable adjustment was achieved by considering the most thorough model with the least amount of missing information. Our final, multivariable-adjusted, conditional logistic regression model included the following variables: menopausal status (premenopausal, postmenopausal, unknown); age at menopause (years, continuous); age at menarche (≤12, 13, ≥14 years, missing); parous (yes, no); age at first birth (years, continuous); exogenous hormone use (oral contraceptives or postmenopausal hormones; yes, no, missing); BMI (kg/m2, continuous); current smoking (yes, no, missing); race (Caucasian, African-American, Asian/Pacific Islander, other, missing); personal history of benign breast disease (yes, no, missing); and family history of breast cancer (yes, no, missing). Although some of these variables were matching factors in some studies (eg, menopausal status was a matching factor in five studies), we included them as covariates to ensure we achieved control for possible confounding in the combined analysis. To assess the effect of simplifying variable definitions, we conducted sensitivity analyses within NHS and WHS, cohorts with the most detailed covariate information, comparing detailed vs simplified coding in the multivariable-adjusted models; results were not appreciably different. In addition, we conducted sensitivity analyses including vs excluding physical activity, alcohol consumption, and cholesterol levels within the cohorts that had information on these variables; results generally were similar.

For stratified analyses, we used unconditional logistic regression, additionally adjusting for the following matching factors: age (years), date of blood collection (months), time of blood collection (hours, missing), fasting status (yes, no, missing), and study (12 indicators including CLUE I and II separately as well as analytic subsets within Umea, Sweden (27) and MEC (33)). We conducted analyses stratified by lifestyle factors (menopausal status [premenopausal vs postmenopausal], postmenopausal hormone use [current use, yes vs no], smoking status [current, yes vs no], alcohol consumption [<4 vs ≥4 drinks per week], BMI [<25, 25 to <30, ≥30 kg/m2], and time between blood collection and diagnosis [<2, 2–5, ≥5 years]) and by tumor characteristics (size [<2 vs ≥2 cm], nodal involvement [0 vs ≥1 positive nodes], differentiation [well, moderate, poor], and ER status [ER− vs ER+]).

Tests for trend were conducted by modeling the medians of quintiles (or deciles) as a continuous variable and calculating the Wald statistic. The shape of the dose–response curves and tests for nonlinearity were assessed using restricted cubic spline models with stepwise selection of knots over the range of carotenoid distributions (45). Wald tests for interaction between stratification variables and carotenoids compared the slopes of the quintile medians between groups. To test whether associations differed by tumor characteristics, we used polychotomous logistic regression with multiple endpoints (eg, ER−, ER+, and no breast cancer) (46). We used a likelihood ratio test to compare a model with separate slopes for the carotenoids in each case group with a model with a common slope. Correlations with plasma or serum carotenoid levels were assessed by Spearman rank correlations. All P values were based on two-sided tests and were considered statistically significant if less than or equal to .05. Analyses were conducted using SAS version 9 (SAS Institute, Cary, NC); polychotomous logistic regressions were conducted using STATA version 11.0 (StataCorp, College Station, TX).

Results
Each of the eight studies included in the pooled analysis contributed a range of 103 to 962 breast cancer case subjects (total = 3055) and a range of 115 to 962 control subjects (total = 3956) (Table 1). Mean age at blood collection for case subjects ranged across studies from 51.3 to 66.0 years. Most participants were postmenopausal at blood collection within studies (50%–100%) and overall (67%). Median time between blood collection and diagnosis was 4.3 years (range = 0.8–13.7 years).

Recalibration of original plasma or serum carotenoid values resulted in consistent, high correlations between the original and recalibrated carotenoids values. The regression models of original and rerun carotenoid levels for each carotenoid resulted in
Table 1. Case and control characteristics in pooled analysis of prospective data on circulating carotenoids and breast cancer risk*

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Reference</th>
<th>Subjects No.</th>
<th>Mean age, y</th>
<th>Mean BMI, kg/m²</th>
<th>Mean age menarche, y</th>
<th>Nulliparous, %</th>
<th>Postmenopausal, %</th>
<th>Using PMH, %†</th>
<th>Median time to diagnosis, y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columbia, MO</td>
<td>Dorgan et al., 1998 (26)</td>
<td>Case 105</td>
<td>58.2</td>
<td>26.1</td>
<td>12.9</td>
<td>18.1</td>
<td>77.1</td>
<td>23.5</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 209</td>
<td>56.2</td>
<td>26.8</td>
<td>12.9</td>
<td>14.4</td>
<td>79.4</td>
<td>20.5</td>
<td>—</td>
</tr>
<tr>
<td>Umea, Sweden</td>
<td>Hultén et al., 2001 (27)</td>
<td>Case 201</td>
<td>54.5</td>
<td>25.3</td>
<td>13.3</td>
<td>12.7</td>
<td>66.7</td>
<td>45.5</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 389</td>
<td>54.4</td>
<td>25.7</td>
<td>13.4</td>
<td>9.8</td>
<td>66.6</td>
<td>40.2</td>
<td>—</td>
</tr>
<tr>
<td>CLUE I</td>
<td>Sato et al., 2002 (29)</td>
<td>Case 244</td>
<td>51.3</td>
<td>†</td>
<td>12.7</td>
<td>1.1</td>
<td>57.0</td>
<td>13.0</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 244</td>
<td>51.1</td>
<td>†</td>
<td>12.7</td>
<td>1.3</td>
<td>57.8</td>
<td>14.2</td>
<td>—</td>
</tr>
<tr>
<td>CLUE II</td>
<td>Sato et al., 2002 (29)</td>
<td>Case 115</td>
<td>60.4</td>
<td>26.2</td>
<td>12.6</td>
<td>1.1</td>
<td>79.1</td>
<td>11.0</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 115</td>
<td>60.2</td>
<td>25.4</td>
<td>12.9</td>
<td>1.1</td>
<td>79.1</td>
<td>13.2</td>
<td>—</td>
</tr>
<tr>
<td>NHS</td>
<td>Tamimi et al., 2005 (30)</td>
<td>Case 962</td>
<td>57.2</td>
<td>25.4</td>
<td>12.5</td>
<td>6.4</td>
<td>68.6</td>
<td>52.6</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 962</td>
<td>57.3</td>
<td>25.4</td>
<td>12.6</td>
<td>6.6</td>
<td>68.3</td>
<td>52.7</td>
<td>—</td>
</tr>
<tr>
<td>WHS</td>
<td>Sesso et al., 2005 (31)</td>
<td>Case 508</td>
<td>55.7</td>
<td>25.5</td>
<td>12.3</td>
<td>2.3</td>
<td>62.4</td>
<td>59.3</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 508</td>
<td>55.7</td>
<td>25.9</td>
<td>12.4</td>
<td>2.6</td>
<td>61.0</td>
<td>54.2</td>
<td>—</td>
</tr>
<tr>
<td>NYUWHS</td>
<td>Toniolo et al., 2006 (28)</td>
<td>Case 269</td>
<td>53.4</td>
<td>25.3</td>
<td>12.4</td>
<td>34.6</td>
<td>53.2</td>
<td>0</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 269</td>
<td>53.2</td>
<td>25.4</td>
<td>12.6</td>
<td>29.4</td>
<td>53.2</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>SWHS</td>
<td>Dorjgochoo et al., 2009 (32)</td>
<td>Case 365</td>
<td>53.1</td>
<td>24.3</td>
<td>14.8</td>
<td>4.1</td>
<td>49.6</td>
<td>7.7</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 365</td>
<td>53.1</td>
<td>24.5</td>
<td>15.0</td>
<td>3.5</td>
<td>50.6</td>
<td>3.3</td>
<td>—</td>
</tr>
<tr>
<td>MEC</td>
<td>Epplein et al., 2009 (33)</td>
<td>Case 286</td>
<td>66.0</td>
<td>25.6</td>
<td>12.5</td>
<td>12.2</td>
<td>100.0</td>
<td>34.6</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 535</td>
<td>66.0</td>
<td>25.4</td>
<td>12.6</td>
<td>10.8</td>
<td>100.0</td>
<td>34.6</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>Case 3055</td>
<td>66.5</td>
<td>25.3</td>
<td>12.8</td>
<td>9.0</td>
<td>66.5</td>
<td>37.2</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control 3956</td>
<td>66.7</td>
<td>25.4</td>
<td>13.1</td>
<td>8.1</td>
<td>67.5</td>
<td>33.0</td>
<td>—</td>
</tr>
</tbody>
</table>

* BMI = body mass index; MEC = Multiethnic Cohort Study; NHS = Nurses’ Health Study; NYUWHS = New York University Women’s Health Study; PMH = postmenopausal hormones; SWHS = Shanghai Women’s Health Study; WHS = Women’s Health Study; † = not applicable because controls were not diagnosed with breast cancer.
† Using PMH calculated among postmenopausal women only.
‡ BMI not collected in CLUE I.

relatively high $R^2$ values from individual studies ($\alpha$-carotene = 0.89–0.99; $\beta$-carotene = 0.86–0.99; $\beta$-cryptoxanthin = 0.92–0.99; lutein+zeaxanthin = 0.74–0.98; lycopene = 0.80–0.99) with the exception of WHS ($R^2 = 0.56$) and MEC ($R^2 = 0.18$). There was no obvious explanation for the low lycopene correlations in WHS and MEC. Although three high original values in the WHS appeared to influence the regression, and excluding them increased the $R^2$ to 0.85, recalibration results were similar whether including or excluding these values. Although correlations generally were high, absolute values of original and recalibrated individual carotenoid values differed with recalibration (Figure 1; Supplementary Table 1, available online). In general, differences among studies were not as pronounced with recalibration. For example, the range of median values of $\alpha$-carotene decreased from the original (range = 1.7–12.7 g/dL) to the recalibrated (2.3–8.9 g/dL) values. In most instances, the ranking of medians across studies was maintained with recalibration. The exceptions include $\beta$-carotene values in three studies (Umea, Sweden, from middle to high; Columbia, MO, from high to lowest; MEC, from highest to middle with recalibration), $\beta$-cryptoxanthin and lutein+zeaxanthin in NHS (from low to middle-high with recalibration), and lycopene in WHS (from second lowest to second highest with recalibration). Recalibrated values highlight population differences in carotenoid levels between studies. For example, the 90th percentile of $\alpha$-carotene in SWHS is nearly comparable with the 10th percentile in WHS; the 90th percentiles of $\beta$-cryptoxanthin in CLUE I and CLUE II are comparable with the median in MEC (Supplementary Table 1, available online). In SWHS, the 10th percentile of lutein+zeaxanthin is higher than the medians of all but two cohorts (NHS and New York University Women’s Health Study); in contrast, the median of lycopene in SWHS is lower than the 10th percentiles of all but two cohorts (Columbia, MO, and Umea, Sweden).

We first examined associations between carotenoid levels and breast cancer risk using the two-stage approach, with original carotenoid values and study-specific quintiles (data not shown). Suggestive but non–statistically significant inverse associations were observed for $\alpha$-carotene (top vs bottom quintile RR = 0.84, 95% CI = 0.63 to 1.01, $P_{trend} = .19$); $\beta$-carotene (RR = 0.77, 95% CI = 0.56 to 1.05, $P_{trend} = .16$), and total carotenoids (RR = 0.72, 95% CI = 0.51 to 1.01, $P_{trend} = .07$). However, statistically significant heterogeneity between studies was observed for each of these carotenoids ($P = .03, .004, .002$, respectively). $\beta$-Cryptoxanthin was not associated with breast cancer risk (RR = 0.98, 95% CI = 0.83 to 1.17, $P_{trend} = .67$). Associations for lutein+zeaxanthin (RR = 0.91, 95% CI = 0.77 to 1.07, $P_{trend} = .24$) and lycopene (RR = 0.89, 95% CI = 0.72 to 1.08, $P_{trend} = .04$) were similar to one another. Tests for heterogeneity were not statistically significant for $\beta$-cryptoxanthin, lutein+zeaxanthin, or lycopene.

Using the recalibrated data and common quintiles across cohorts, we conducted analyses with the aggregated data approach (Figure 2). In unadjusted conditional logistic regression models, we observed suggestive inverse associations with some carotenoids and a statistically significant inverse trend for lycopene (top vs bottom quintile RR = 0.80, 95% CI = 0.64 to 1.01, $P_{trend} = .05$). Associations strengthened with multivariable adjustment because of small contributions by several factors (eg, BMI, parity, smoking, menopausal status, and postmenopausal hormone use). Statistically significant inverse
associations were observed for all carotenoids except \( \beta \)-cryptoxanthin (\( \alpha \)-carotene: top vs bottom quintile RR = 0.87, 95% CI = 0.71 to 1.05; \( P \) _{trend} = 0.04; \( \beta \)-carotene: RR = 0.83, 95% CI = 0.70 to 0.98, \( P \) _{trend} = 0.02; \( \beta \)-cryptoxanthin: RR = 0.98, 95% CI = 0.82 to 1.18, \( P \) _{trend} = 0.21; lutein+zeaxanthin: RR = 0.84, 95% CI = 0.70 to 1.01, \( P \) _{trend} = 0.05; lycopene: RR = 0.78, 95% CI = 0.62 to 0.99, \( P \) _{trend} = 0.02; total carotenoids: RR = 0.81, 95% CI = 0.68 to 0.96, \( P \) _{trend} = 0.01). None of the associations was statistically significantly nonlinear in cubic spline models (data not shown). We reran the model assessing the association with lycopene excluding MEC and/or WHS given the low correlations between original and rerun assay levels; results essentially were unchanged (data not shown). Using the two-stage approach with recalibrated data and common quintiles, results generally were similar to the aggregated data approach (data not shown). Tests for heterogeneity among cohorts were no longer statistically significant (data not shown). All subsequent analyses were conducted using the aggregated data approach with recalibrated carotenoid levels and common cut points.

Expanding the quantile analysis to deciles, we generally observed that breast cancer risk continued to decrease in the top decile of carotenoids (\( \alpha \)-carotene: top vs bottom decile RR = 0.79, 95% CI = 0.61 to 1.03, \( P \) _{trend} = 0.02; \( \beta \)-carotene: RR = 0.77, 95% CI = 0.61 to 0.98, \( P \) _{trend} = 0.01; lutein+zeaxanthin: RR = 0.72, 95% CI = 0.56 to 0.92, \( P \) _{trend} = 0.02; total carotenoids: RR = 0.71, 95% CI = 0.56 to 0.91, \( P \) _{trend} = 0.03). The association between lycopene and breast cancer risk was not as consistent across deciles as it was across quintiles (top vs bottom decile RR = 0.89, 95% CI = 0.65 to 1.21, \( P \) _{trend} = 0.11).

When we analyzed the associations between carotenoids and breast cancer risk by ER status, inverse associations were apparent among ER− tumors (Figure 3). Although inverse associations were suggested among ER+ tumors for \( \alpha \)-carotene and \( \beta \)-carotene, multivariable relative risks and tests for trend were not statistically significant (top vs bottom quintile RR = 0.85, 95% CI = 0.65 to 1.12, \( P \) _{trend} = 0.16; RR = 0.85, 95% CI = 0.66 to 1.04, \( P \) _{trend} = 0.06, respectively). However, the associations were statistically significant for both \( \alpha \)-carotene and \( \beta \)-carotene among ER− tumors (top vs bottom quintile RR = 0.61, 95% CI = 0.40 to 0.93, \( P \) _{trend} = 0.04; RR = 0.52, 95% CI = 0.36 to 0.77, \( P \) _{trend} = 0.011, respectively). There was statistically significant heterogeneity between ER− and ER+ tumors for \( \beta \)-carotene (\( P \) = 0.01) but not for \( \alpha \)-carotene (\( P \) = 0.11). Associations for other carotenoids were not statistically significantly different by ER status.

We compared associations between carotenoids and breast cancer risk across categories of tumor size, nodal involvement, and differentiation. No statistically significant differences were observed for tumor size (<2 vs ≥2cm) or nodal involvement (0 vs ≥1 positive nodes) (data not shown). A statistically significant difference by differentiation was observed for lutein+zeaxanthin, with stronger inverse associations with less differentiation (well-differentiated (\( n = 301 \)): top vs bottom quintile RR = 1.35, 95% CI = 0.87 to 2.09, \( P \) _{trend} = 0.28; moderately differentiated (\( n = 664 \)): RR = 0.88, 94% CI = 0.64 to 1.23, \( P \) _{trend} = 0.28; poorly differentiated (\( n = 508 \)): RR = 0.68, 95% CI = 0.46 to 0.99, \( P \) _{trend} = 0.15; \( P \) _{heterogeneity} = 0.04). Differences in the associations by differentiation were not observed for any other carotenoid.

We conducted analyses stratified by several lifestyle factors, including menopausal status, postmenopausal hormone use, smoking status, alcohol consumption, BMI, and time between blood collection and diagnosis. Interactions with most of these factors were not statistically significant. However, several of the associations of carotenoid with breast cancer risk varied by BMI and smoking status (Table 2). Interactions with BMI were fairly consistent across carotenoids, with stronger inverse associations with breast cancer risk observed among leaner (BMI ≤25) women and suggestively or statistically significantly positive associations among obese women.
(BMI = ≥30) (eg, total carotenoids: BMI <25: top vs bottom quintile RR = 0.65, 95% CI = 0.51 to 0.83, \( P_{\text{trend}} = .001 \); BMI ≥30: RR = 1.74, 95% CI = 0.97 to 3.13, \( P_{\text{trend}} = .04 \); \( P_{\text{heterogeneity}} = .01 \)). Interestingly, as has been documented previously (3), BMI was inversely associated with plasma carotenoid levels (total carotenoids \( r = -0.19 \)).

Statistically significant interactions also were observed with current smoking for lutein+zeaxanthin and total carotenoids, with stronger inverse associations among current smokers (Table 2) (eg, total carotenoids: non-smokers: top vs bottom quintile RR = 0.93, 95% CI = 0.77 to 1.11, \( P_{\text{trend}} = .38 \); current smokers: RR = 0.47, 95% CI = 0.30 to 0.73, \( P_{\text{trend}} = .002 \); \( P_{\text{heterogeneity}} = .01 \)). Although interactions with smoking were not statistically significant for the remaining carotenoids, generally similar patterns were observed, with stronger associations among smokers. This pattern persisted among both lean and heavy women (data not shown). Similarly, the interaction with BMI persisted among nonsmokers (data not shown).

Statistically significant interactions with smoking persisted regardless of ER status for lutein+zeaxanthin and total carotenoids (data not shown). Although interactions with BMI were not statistically significant, differences in the carotenoids associations by BMI appeared more striking among ER- tumors than among ER+ tumors (eg, among ER- tumors: \( \alpha \)-carotene BMI <25: RR = 0.34, 95% CI = 0.18 to 0.64; vs BMI ≥25: RR = 1.06, 95% CI = 0.54 to 2.09; among ER+ tumors: \( \alpha \)-carotene BMI <25: RR = 0.80, 95% CI = 0.51 to 1.24; vs BMI ≥25: RR = 1.13, 95% CI = 0.75 to 1.70). Although we observed

Figure 2. Relative risks (RRs) of breast cancer and 95% confidence intervals (CIs) according to quintile of plasma carotenoids (μg/dL), recalibrated data. Diamonds represent relative risks; lines represent 95% confidence intervals. To convert μg/dL to μmol/L, multiply by the following factors: 0.01863 for \( \alpha \)-carotene, \( \beta \)-carotene, and lycopene; 0.01810 for \( \beta \)-cryptoxanthin; 0.01753 for lutein+zeaxanthin. ref = referent.
no association between β-cryptoxanthin and breast cancer risk overall and by ER status, we observed strong inverse associations with ER− tumors among lean women and strong positive associations with ER− tumors among overweight and obese women (BMI <25: top vs. bottom quintile RR = 0.50, 95% CI = 0.28 to 0.88, P trend = .03; BMI ≥25: RR = 2.18, 95% CI = 1.13 to 4.24, P trend = .07; P heterogeneity = .16).

To investigate the role of vitamin supplement use in the observed associations, we repeated the analyses among six cohorts with data on multivitamin and/or supplement use and restricted analyses to nonusers in these cohorts. Overall results among these cohorts (n = 2749 case subjects) were similar to the total pooled results. Restricting to nonvitamin users (n = 1364 case subjects) generally resulted in similar point estimates but non–statistically significant trends, except for lutein+zeaxanthin and lycopene, for which the point estimates were somewhat attenuated as well as non–statistically significant (data not shown).

To examine the potential impact of preclinical disease on our results, we excluded 897 case subjects diagnosed less than or equal to 2 years after blood collection, ranging from 4% (CLUE I) to 69% (Umea, Sweden) of case subjects within individual cohorts. Risk estimates were essentially unchanged for α-carotene, β-carotene, β-cryptoxanthin, and total carotenoids, but somewhat attenuated for lutein+zeaxanthin and lycopene (data not shown).

The levels of most individual carotenoids were statistically significantly correlated with one another, ranging from r equal to 0.14 (P <.001) (lutein+zeaxanthin with α-carotene) to r equal to 0.65 (P <.001) (α-carotene with β-carotene); lutein+zeaxanthin and lycopene were not correlated (r = −0.02, P = .27). When pairs of carotenoids were included in the same statistical model, most associations were attenuated and no longer statistically significant (eg, β-carotene adjusting for α-carotene: top vs bottom quintile RR = 0.87, 95% CI = 0.70 to 1.08, P trend = .19) except when β-cryptoxanthin was the second carotenoid (eg, β-carotene adjusting for β-cryptoxanthin: RR = 0.85, 95% CI = 0.71 to 1.02, P trend = .07). The inverse association between lycopene and breast cancer risk remained statistically significant with adjustment for either β-cryptoxanthin or lutein+zeaxanthin (eg, lycopene: RR = 0.80, 95% CI = 0.63 to 1.02, P trend = .06 adjusting for lutein+zeaxanthin). Point estimates for the association between lycopene and breast cancer were similar to the overall point estimates when adjusting for α-carotene (correlation with lycopene: r = 0.42, P <.001) or β-carotene (r = 0.25, P <.001), but the trends were no longer statistically significant (eg, RR = 0.81, 95% CI = 0.64 to 1.03, P trend = .06 adjusting for α-carotene).

![Figure 3](https://jnci.oxfordjournals.org)
To date, four initial small (N = 14–67 case subjects) (36–40) and 10 larger, more recent (26–35) nested case–control studies of circulating carotenoid levels and breast cancer risk have been published. In most (26–30,32,35), but not all (31,33,34), of the larger studies, statistically significant or suggestive inverse associations were observed, with relative risk less than or equal to 0.7 for at least one carotenoid. Two studies (34,35) were published after this project was underway; thus, the 566 case subjects in those studies are not included in this analysis. In the Women’s Health Initiative (N = 190
In contrast with our results, a suggestive positive association was observed with lycopene among invasive tumors (RR = 0.55, 95% CI = 0.34 to 0.90, $P_{\text{rand}} = .02$) and case subjects diagnosed less than 5 years after blood draw (eg, 1–3 years: RR = 0.42, 95% CI = 0.23 to 0.75, $P_{\text{rand}} = .002$) (35). In contrast with our results, a suggestive positive association was observed with lycopene among invasive tumors (RR = 1.47, 95% CI = 0.98 to 2.22, $P_{\text{rand}} = .06$). In E3N–EPIC (N = 366 case subjects), no statistically significantly associations with breast cancer risk were observed; however, many of the results were largely consistent with our results (eg, total carotenoids: top vs bottom quintile RR = 0.74, 95% CI = 0.47 to 1.16, $P_{\text{rand}} = .38$) (34). Thus, it is likely that our pooled results would not have been substantially altered had we been able to include these two additional studies.

A unique advantage of this analysis was the ability to accommodate true population differences in carotenoid levels between studies by recalibrating previously assayed blood samples to a common standard. Results were fairly consistent and robust regardless of whether we pooled study-specific relative risks or calculated relative risks from pooled recalibrated data. However, there was statistically significant between-study heterogeneity with original data and study-specific quintiles that was not observed with recalibrated data. Given that population differences in individual carotenoid levels were evident with recalibration, heterogeneity using study-specific quintiles was not unexpected.

Another strength of this analysis was the ability to examine associations by ER status. Given that ER− tumors are less common, there was limited statistical power within each individual study to study these tumors. The inverse associations we observed among ER+ tumors highlight carotenoids as one of the first modifiable risk factors for this poor prognosis tumor type. Most well-established breast cancer risk factors are hormonal factors that are more strongly associated with ER+ tumors, such as parity, age at first birth, and postmenopausal BMI (47, 48). Although experimental evidence has shown some carotenoids to inhibit growth of both ER+ and ER− breast cancer cell lines (11), it is possible that an effect of carotenoids on ER+ tumors is masked by the hormone-related associations that dominate as risk factors for ER+ tumors.

Recent studies of dietary intake and breast cancer risk by ER status support our findings. An analysis of the Dietary Approaches to Stop Hypertension diet in the NHS found that a higher intake of fruits and vegetables was inversely associated with ER−, but not ER+ tumors (top vs bottom tertile RR = 0.71, 95% CI = 0.55 to 0.90, $P_{\text{rand}} = .005$) (49). Similar associations with dietary patterns reflecting higher fruit and/or vegetable consumption also were observed in the Melbourne Collaborative Cohort Study (16), the Black Women's Health Study (17, 18), and the NIH–AARP study (19). In contrast, in the E3N–EPIC study, the healthy/Mediterranean dietary pattern was inversely associated with ER+ progesterone receptor negative tumors (top vs bottom quartile RR = 0.65, 95% CI = 0.49 to 0.87, $P_{\text{rand}} = .001$) (20). Most recently, in pooled analyses of 18 cohort studies, intakes of both fruits and vegetables (S Y Jung, D Spiegelman, L Baglietto, et al, unpublished observations) and carotenoids specifically (21) were inversely associated with ER+ tumors. Statistically significant differences in the associations between carotenoids and breast cancer by ER status were observed for α-carotene, β-carotene, β-cryptoxanthin, and lutein–zeaxanthin (eg, β-carotene: ER−: top vs bottom quintile RR = 0.85, 95% CI = 0.77 to 0.93, $P_{\text{rand}} = .002$; ER+: RR = 1.04, 95% CI = 0.98 to 1.09, $P_{\text{rand}} = .24$; $P_{\text{heterogeneity}} = .01$) (21).

Given the potential antioxidant effects of carotenoids, it is possible that women who smoke or consume alcohol, lifestyle factors associated with oxidative stress, may gain more benefit from carotenoids. We observed statistically significantly stronger inverse associations of lutein–zeaxanthin and total carotenoids with breast cancer risk among current smokers, and similar patterns were observed with the other carotenoids. We did not observe statistically significant differences by alcohol intake. Interestingly, although obesity also contributes to oxidative stress (50, 51), we observed interactions with BMI in the opposite direction, with statistically significantly stronger inverse associations with carotenoids among leaner women and statistically significant positive associations among obese women. This finding was unexpected and not easily explained. BMI is inversely associated with plasma carotenoid levels (3), as observed in our data, as well as adipose tissue carotenoid levels (52). However, obese women with high levels of plasma carotenoids likely also have very high adipose tissue levels. Although carotenoid levels in breast adipose tissue have been inversely associated with breast cancer risk (53), very high concentrations of carotenoids have been shown to act as pro-oxidants in animal and in vitro studies, although it is unclear whether such pro-oxidant activities occur in humans (54–57).

There are several proposed mechanisms by which carotenoids may influence carcinogenesis. α-carotene, β-carotene, and β-cryptoxanthin may decrease cancer risk indirectly through their metabolism to retinol, which in turn regulates cell growth, differentiation, and apoptosis via direct and indirect effects on gene expression (10, 58, 59). Carotenoids also may be directly anticarcinogenic by several other mechanisms, including improved gap junction communication, enhanced immune system functioning, or antioxidant scavenging of reactive oxygen species (5–9, 60–62); this may inhibit cellular dysregulation or DNA damage.

There are several limitations to our study. First, despite biologic plausibility that carotenoids are responsible for the observed inverse associations with breast cancer risk, there are alternative explanations for our findings. Other phytochemicals in fruits and vegetables, such as flavonoids and other phenolic compounds, or an interaction among various phytochemicals (63) could be associated with carotenoids and responsible for the observed association. Future studies should explore the associations between additional phytochemical components and breast cancer risk. There also is the possibility of unmeasured confounding. However, associations persisted after adjustment for other health-related behaviors, including physical activity and postmenopausal hormone use, as well as exclusion of people with health-associated behaviors, such as multivitamin use. Second, for this analysis only one blood sample was available from each participant. However, reproducibility of circulating carotenoids over a few years was very good in two of the included cohorts; intraclass correlation coefficients for individual carotenoids measured over a 2 to 3 year period were 0.63 to 0.85 in New York University Women's Health Study (28) and 0.73 to 0.88 in NHS (64). Third, although circulating carotenoids are indirect markers of activity at the breast tissue, plasma levels are
positively correlated with breast tissue levels ($r = 0.44–0.88$) (65). Given the relatively high correlations among individual carotenoids, isolating the independent effects of a specific carotenoid is difficult. However, the inverse association with lycopene persisted with adjustment for other carotenoids, despite its correlations with other carotenoids ($r = 0.25$ with $\beta$-carotene, $r = 0.42$ with $\alpha$-carotene). Finally, although we conducted a large number of analyses, our hypotheses were biologically motivated, and in some cases, have been published previously in the individual studies included in the pooled analysis. However, we interpret the results with caution and note that future studies should be conducted to confirm our findings.

The results of this large pooled analysis suggest that women with higher circulating carotenoid levels are at reduced breast cancer risk. The statistically significant positive associations between circulating carotenoids and risk we observed among overweight and obese women warrant further study. Additional work also is needed to determine if carotenoids are the causal factor in the observed associations. Given the possibility that another bioactive compound is responsible for the observed associations, as well as the uncertainty about the specific carotenoid(s) that are important, use of specific carotenoid supplements is not advised and may indeed be harmful among smokers (66,67). Carotenoids are available in a wide variety of fruits and vegetables common to the US diet, including: carrots for $\alpha$-carotene; sweet potatoes and leafy greens for $\beta$-carotene; citrus fruits for $\beta$-cryptoxanthin; leafy greens for lutein+zeaxanthin; and tomatoes for lycopene (68). A diet high in carotenoid-rich fruits and vegetables offers many health benefits, including a possible reduced risk of breast cancer.

References

63. Liu RH. Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. *Am J Clin Nutr.* 2003;78(3 Suppl):517S–520S.

**Funding**

This work was supported primarily by the National Cancer Institute at the National Institutes of Health (R03 CA128073 to AHE). Funding to individual cohorts included grants from the National Cancer Institute (R01 CA131218, P01 CA87969, R01 CA49449, R01 CA098661, P01 CA106807, P01 CA33619, R37 CA34281, CA-047988, R01 CA106591, R17 CA70867); the National Institute of Environmental Health Sciences Center (ES000260); the National Heart Lung and Blood Institute (HL043851, HL080467, HL099355); contracts from the National Institutes of Health (N01 PC35137, N01 PC35139, N02 CP110866); the Intramural Research Program of the National Institutes of Health; the Swedish Cancer Society; Swedish Scientific Council; Swedish Research Council; Regional Government of Västerbotten, Sweden; and a grant from DSM Nutritional Products, Inc. (formerly Roche Vitamins). SJH was supported in part by a training grant (5 T32 CA090013-5) from the National Institutes of Health.

**Notes**

The study sponsors had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

**Affiliations of authors:** Channing Division of Network Medicine (AHE, RMT, SEH) and Division of Preventive Medicine (JEB, HDS), Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA; Department of Epidemiology (AHE, SJH, RMT, SEH) and Department of Nutrition (SJH, HC), Harvard School of Public Health, Boston, MA; Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD (LAB); Vanderbilt University School of Medicine, Nashville, TN (QD, WZ); Fox Chase Cancer Center, Philadelphia, PA (JFD); University of Hawaii Cancer Center, Honolulu, HI (AAF, MTG, LRW); Shanghai Cancer Institute, Shanghai, China (Y-TG); Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden (GH, KH); George W. Comstock Center for Public Health Research and Prevention, Johns Hopkins University,
Bloomberg School of Public Health, Baltimore, MD (KJH, JH-B); The Prevention and Research Center, Mercy Medical Center, Baltimore, MD (KJH); Division of Health Studies, Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention, Atlanta, GA (ALS); Department of Obstetrics and Gynecology (PT) and Department of Environmental Medicine (AZ-J), New York University Cancer Institute, New York University School of Medicine, New York, NY; Department of Clinical Nutrition, University of Gothenburg, Göteborg, Sweden (AW); Division of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA (SEH).