The Premenopausal Breast Cancer Collaboration: A Pooling Project of Studies Participating in the National Cancer Institute Cohort Consortium


Abstract

Breast cancer is a leading cancer diagnosis among premenopausal women around the world. Unlike rates in postmenopausal women, incidence rates of advanced breast cancer have increased in recent decades for premenopausal women. Progress in identifying contributors to breast cancer risk among premenopausal women has been constrained by the limited numbers of premenopausal breast cancer cases in individual studies and resulting low statistical power to subcategorize exposures or to study specific subtypes. The Premenopausal Breast Cancer Collaborative Group was established to facilitate cohort-based analyses of risk factors for premenopausal breast cancer by pooling individual-level data from studies participating in the United States National Cancer Institute Cohort Consortium. This article describes the Group, including the rationale for its initial aims related to pregnancy, obesity, and physical activity. We also describe the 20 cohort studies with data submitted to the Group by June 2016. The infrastructure developed for this work can be leveraged to support additional investigations.

Cancer Epidemiol Biomarkers Prev; 26(9); 1–10. ©2017 AACR.

University of North Carolina, Gillings School of Global Public Health, Chapel Hill, North Carolina. 2Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, United Kingdom. 3Karolinska Institutet, MEB, University of Oslo Institute of Health and Society, Sweden, Norway. 4Fondazione IRCCS Istituto Nazionale dei Tumori, Italy, France. 5Centre for Research in Epidemiology and Population Health (CESP), France. 6Beckman Research Institute of City of Hope, Monrovia, California. 7Slone Epidemiology Center at Boston University, Boston, Massachusetts. 8Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee. 9University of Pittsburgh Graduate School of Public Health, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania. 10NYU School of Medicine, New York, New York. 11Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland. 12International Agency for Research on Cancer, France. 13Channing Division of Network Medicine, Brigham and Women’s Hospital, Harvard T.H. Chan School of Public Health and Harvard Medical School, Boston, Massachusetts. 14Cancer Council Victoria, University of Melbourne, Melbourne, Australia. 15University of Tromsø, The Arctic University of Norway, Tromsø, Norway. 16Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Toronto, Toronto, Ontario, Canada. 17Duke-NUS Medical School, Singapore. 18Division of Cancer Epidemiology, DKFZ, Heidelberg, Germany. 19University of Oxford, Oxford, United Kingdom. 20Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. 21Duke-NUS Medical School, Singapore. 22Karolinska Institutet, Institute of Environmental Medicine, Stockholm, Sweden. 23UIT (University of Tromsø), Tromsø, Norway. 24School of Public Health, Imperial College London, United Kingdom. 25Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain. 26CIBER Epidemiología y Salud Pública (CIBERESP), Department of Health and Social Sciences, Universidad de Murcia, Murcia, Spain. 27Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark. 28Radiation Effects Research Foundation, Hiroshima, Japan. 29University Medical Center, Utrecht, the Netherlands. 30Albert Einstein College of Medicine, Bronx, New York. 31Umeå University, Umeå, Sweden. 32Hellenic Health Foundation, Greece. 33Norwegian University of Science and Technology, Trondheim, Norway. 34Johns Hopkins School of Medicine, Baltimore, Maryland. 35Karolinska Institutet, Department of Epidemiology and Biostatistics, Oslo, Norway. 36Department of Research, Head, Group of Etiological Cancer Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway. 37Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland. 38Department of Community Medicine, University of Tromsø, The Arctic University of Norway, Tromsø, Norway. 39Department of Nutrition and Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. 40Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Durham, North Carolina. 41Division of Breast Cancer Research, The Institute of Cancer Research, London, United Kingdom.
Introduction

Breast cancer is the most common cancer diagnosed among women worldwide, with an estimated 1.67 million cases diagnosed in 2012, accounting for a quarter of all new cancers in women. Breast cancer is also the most common cancer diagnosed among women ages 15 to 39 years worldwide (1). Furthermore, breast cancer among premenopausal women often presents at more advanced stages (2) and, at the youngest ages, has less favorable prognosis (3) than among postmenopausal women.

Identifying contributors to breast cancer risk in younger women is critical to prevention. In the United States, incidence rates of advanced breast cancer have increased among premenopausal women in recent decades, whereas they have consistently decreased among women ages 60 and older during the same period (4). Accumulating evidence supports etiologic heterogeneity between pre- and postmenopausal breast cancer. Several lifestyle factors, including childbirth (5), obesity (6), and cigarette smoking (7), have been reported to have differential associations with breast cancer risk before and after menopause. Breast cancer subtypes, including those defined by gene expression (8), or clinical markers including estrogen receptor (ER), progesterone receptor (PR), or HER2/neu oncogene expression, have emerged as critical considerations for risk factor associations and are differentially distributed by menopausal status (9). Investigations of breast cancer etiologic heterogeneity require large sample sizes to have sufficient statistical power to account jointly for menopausal status and tumor subtype.

The Premenopausal Breast Cancer Collaborative Group (referred to hereafter as the Collaborative Group) was established to facilitate cohort-based analyses of risk factors for premenopausal breast cancer, both overall and according to tumor characteristics. This article describes the formation of the Collaborative Group, the methods used for ongoing efforts, and provides the rationale for initial analyses related to pregnancy, obesity, and physical activity. The infrastructure developed to address these questions can support future investigations of additional potential risk factors.

Collaborative Group Studies

The National Cancer Institute (NCI) Cohort Consortium was formed to address the need for large-scale collaborations to pool data in cohort studies of cancer and hence to quicken the pace of research (http://epi.grants.cancer.gov/Consortia/cohort.html). The Collaborative Group was initiated within the Cohort Consortium in 2013 by investigators at The Institute of Cancer Research (ICR) in London and the National Institute of Environmental Health Sciences (NIEHS). The ICR and the NIEHS serve as the Data Coordinating Centers.

Eligibility

Prospective cohorts in the Cohort Consortium with at least 100 female breast cancers diagnosed during follow-up before age 55 and data collection at two or more time points (baseline and at least one follow-up, to allow for exposure information and menopausal status to be updated) are eligible to participate.

Participating cohorts

This report describes the 20 cohort studies (counting the European Prospective Investigation into Cancer and Nutrition, which has many cohorts within it, as a single cohort; refs. 6, 10–28) with data submitted to the Collaborative Group as of June 2016. Participating cohorts are shown in Table 1 and span North America, Europe, Asia, and Australia. The number of female participants from these cohorts ages <55 at enrollment ranges from 5,671 (Campaign Against Cancer and Heart Disease to 117,733 (Nurses’ Health Study cohort). The cohorts were initiated as early as 1950 (the Radiation Effects Research Foundation Life Span Study) or as recently as 2003 (Generations and Sister Study cohorts). All cohorts have conducted more than one round of data collection; however, follow-up data are not yet fully available for three cohorts. The number of follow-up rounds for which data have been submitted as of June 1, 2016, ranged from 1 to 16 across cohorts.

Breast cancer ascertainment

To date, data have been received for 1,030,761 women, and include 21,766 incidents invasive or in situ breast cancers diagnosed after study enrollment and before age 55 years (Table 2). Across studies, cancer diagnoses are identified by linkage with city/state/provincial/regional (10, 12, 13, 23, 28–31) or national (11, 12, 14, 24, 26, 32, 33) population-based cancer registries, and/or through self-report followed by medical record review (6, 10, 11, 14, 15, 25, 34, 35). All participating studies established case ascertainment procedures and published findings related to incident breast cancer risk prior to joining the Collaborative Group.

Data exchange and harmonization

After approval by the NCI Cohort Consortium executive committee, the aims of the proposed collaboration were circulated to all Consortium members in 2013. Key exposure, covariate, and outcome information necessary to address the initial analyses and potential confounding or effect modification was identified by the Coordinating Centers. Complete capture of all information across exposures is not required for participation.

After confirming eligibility, a data request template is sent to cohorts that wish to participate. Requested exposure data include: age/year of cohort entry, length of follow-up, demographic characteristics (age, race/ethnicity, education, socioeconomic status), lifestyle factors (physical activity, anthropometric characteristics, alcohol intake, smoking information, mammography use), reproductive history (menarche, menstrual cycle characteristics, gravidity, parity, pregnancy complications, infertility, breastfeeding, hormonal medications, menopausal status), benign breast disease, and family history of breast cancer (Supplementary Table S1). Most characteristics are requested at enrollment and each follow-up, as available. Breast cancer information includes age at diagnosis, stage, grade, histology, and expression

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

H.B. Nichols and M.J. Schoemaker are co-first authors of this article.

D.P. Sandler and A.J. Swerdlow are co-senior authors of this article.

Corresponding Author: Hazel B. Nichols, University of North Carolina Gillings School of Global Public Health, 155 Dauer Drive, Chapel Hill, NC 27599. Phone: 919-966-7456; Fax: 919-962-2089; E-mail: hazel.nichols@unc.edu
doi: 10.1158/1055-9965.EPI-17-0246
©2017 American Association for Cancer Research.
Table 1. Characteristics of women younger than 55 years in cohorts included in the Premenopausal Breast Cancer Collaborative Group

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Location</th>
<th>Ages at enrolment; mean (SD), range</th>
<th>Calendar years of enrollment</th>
<th>Baseline data collection methods</th>
<th>N of data collection rounds</th>
<th>Breast cancer cases (N)</th>
<th>Breast cancer ascertainment sources</th>
<th>Cohort size (women &lt;55 years)</th>
<th>N years of follow-up (SD, range (&lt;55 years))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Women's Health Study (10)</td>
<td>United States</td>
<td>37.1 (8.6) 20–54</td>
<td>1995</td>
<td>Mailed questionnaire</td>
<td>9</td>
<td>1,299</td>
<td>Self-report and state registry</td>
<td>52,543</td>
<td>12.6 (5.6) 0-18.6</td>
</tr>
<tr>
<td>California Teachers Cohort (28)</td>
<td>United States</td>
<td>40.4 (7.4) 22–54</td>
<td>1995-1998</td>
<td>Mailed questionnaire</td>
<td>4</td>
<td>1,185</td>
<td>State registry</td>
<td>47,516</td>
<td>11.6 (5.0) 0-17.2</td>
</tr>
<tr>
<td>Campaign against Cancer and Heart Disease (CLUE II; ref. 13)</td>
<td>United States</td>
<td>39.6 (9.6) 18-54</td>
<td>1989</td>
<td>Administered questionnaire</td>
<td>6</td>
<td>131</td>
<td>State registry</td>
<td>5,671</td>
<td>10.8 (5.4) 0.3-26.0</td>
</tr>
<tr>
<td>Canadian Study of Diet, Lifestyle, and Health (12)</td>
<td>Canada</td>
<td>44.1 (6.9) 23-54</td>
<td>1991-1999</td>
<td>Mailed questionnaire</td>
<td>1</td>
<td>377</td>
<td>Provincial and national registry</td>
<td>1,589</td>
<td>8.1 (4.7) 0-18.6</td>
</tr>
<tr>
<td>European Prospective Investigation into Cancer and Nutrition (14)</td>
<td>Europe</td>
<td>44.2 (8.1) 19-54</td>
<td>1991-2000</td>
<td>Self-reported/ administered questionnaires</td>
<td>1</td>
<td>2,122</td>
<td>Self-report and national/regional registries</td>
<td>150,291</td>
<td>7.5 (4.2) 0-16.6</td>
</tr>
<tr>
<td>Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l’Education Nationale (E3N; ref. 15)</td>
<td>France</td>
<td>46.5 (4.2) 38-54</td>
<td>1989-1991</td>
<td>Mailed questionnaire</td>
<td>8</td>
<td>1,908</td>
<td>Self-report</td>
<td>72,748</td>
<td>8.1 (4.2) 0-16.4</td>
</tr>
<tr>
<td>Generations Study (11)</td>
<td>United Kingdom</td>
<td>39.8 (9.5) 16-54</td>
<td>2003-2012</td>
<td>Mailed questionnaire</td>
<td>2</td>
<td>793</td>
<td>Self-report and national registry</td>
<td>72,058</td>
<td>5.4 (1.7) 0-9.7</td>
</tr>
<tr>
<td>Helseundersøkelser i Nord-Trendelag (HUNT2; ref. 26)</td>
<td>Norway</td>
<td>38.9 (9.7) 20-54</td>
<td>1995-1997</td>
<td>Administered questionnaire</td>
<td>1</td>
<td>209</td>
<td>National cancer registry</td>
<td>20,974</td>
<td>10.2 (4.1) 0.3-14.0</td>
</tr>
<tr>
<td>Melbourne Collaborative Cohort Study (16)</td>
<td>Australia</td>
<td>47.5 (4.4) 31-54</td>
<td>1990-1994</td>
<td>Administered questionnaire</td>
<td>3</td>
<td>227</td>
<td>State registry</td>
<td>12,029</td>
<td>7.3 (4.4) 0-20.1</td>
</tr>
<tr>
<td>New York University Women's Health Study (19, 20)</td>
<td>United States</td>
<td>45.2 (5.5) 31-54</td>
<td>1984-1991</td>
<td>Self-administered questionnaire</td>
<td>6</td>
<td>371</td>
<td>Self-report and state registry</td>
<td>8,757</td>
<td>9.5 (5.5) 0-23.5</td>
</tr>
<tr>
<td>Norwegian Women and Cancer Study (105)</td>
<td>Norway</td>
<td>45.7 (6.0) 31-54</td>
<td>1991-2007</td>
<td>Mailed questionnaire</td>
<td>3</td>
<td>2,124</td>
<td>National registry</td>
<td>117,633</td>
<td>9.0 (5.8) 0.3-20.5</td>
</tr>
<tr>
<td>Nurses' Health Study (17)</td>
<td>United States</td>
<td>42.6 (7.1) 29-54</td>
<td>1976-1978</td>
<td>Mailed questionnaire</td>
<td>16</td>
<td>2,743</td>
<td>Self-report</td>
<td>117,730</td>
<td>12.2 (7.0) 0.1-25.5</td>
</tr>
<tr>
<td>Nurses' Health Study II (18)</td>
<td>United States</td>
<td>34.8 (4.7) 24-44</td>
<td>1989-1990</td>
<td>Mailed questionnaire</td>
<td>12</td>
<td>3,765</td>
<td>Self-report</td>
<td>116,415</td>
<td>18.7 (3.7) 0.1-23.7</td>
</tr>
<tr>
<td>Radiation Effects Research Foundation Life Span Study (21)</td>
<td>Japan</td>
<td>41.3 (8.5) 18-54</td>
<td>1963-1993</td>
<td>Administered or mailed questionnaire</td>
<td>6</td>
<td>130</td>
<td>City registry</td>
<td>18,420</td>
<td>13.5 (8.5) 0.1-36.7</td>
</tr>
<tr>
<td>Singapore Chinese Health Study (22)</td>
<td>Singapore</td>
<td>49.6 (3.0) 43-54</td>
<td>1993-1998</td>
<td>Administered questionnaire</td>
<td>2</td>
<td>134</td>
<td>National cancer registry</td>
<td>16,056</td>
<td>5.3 (3.0) 0.3-11.5</td>
</tr>
<tr>
<td>Sister Study (6)</td>
<td>United States</td>
<td>47.9 (4.9) 35-54</td>
<td>2003-2009</td>
<td>Telephone and written questionnaire</td>
<td>3</td>
<td>679</td>
<td>Self-report</td>
<td>24,044</td>
<td>4.7 (2.5) 0.1-10.6</td>
</tr>
<tr>
<td>Southern Community Cohort Study (23)</td>
<td>United States</td>
<td>47.3 (4.2) 40-54</td>
<td>2002-2009</td>
<td>Administered questionnaire</td>
<td>2</td>
<td>233</td>
<td>State registry</td>
<td>30,289</td>
<td>5.1 (2.4) 0.1-13.3</td>
</tr>
</tbody>
</table>

(Continued on the following page)
Table 1. Characteristics of women younger than 55 years in cohorts included in the Premenopausal Breast Cancer Collaborative Group (Cont’d)

<table>
<thead>
<tr>
<th>Cohort Description</th>
<th>Location</th>
<th>Ages at enrollment; mean (SD), range</th>
<th>Calendar years of enrollment</th>
<th>Baseline data collection methods</th>
<th>N of data collection rounds*</th>
<th>Breast cancer cases (N)</th>
<th>Breast cancer ascertainment sources</th>
<th>Cohort size (women &lt;55 years)</th>
<th>N years of follow-up, mean (SD), range (&lt;55 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden Women's Lifestyle and Health Study (27)</td>
<td>Sweden</td>
<td>39.7 (5.3) 29–49</td>
<td>1987–1990</td>
<td>Mailed questionnaire</td>
<td>2</td>
<td>1192</td>
<td>National registry</td>
<td>49,010</td>
<td>14.4 (5.3) 0.1–21.1</td>
</tr>
<tr>
<td>Swedish Mammography Cohort (24)</td>
<td>Sweden</td>
<td>46.6 (4.3) 38–54</td>
<td>1987–1990</td>
<td>Mailed questionnaire</td>
<td>2</td>
<td>649</td>
<td>National registry</td>
<td>34,126</td>
<td>8.3 (4.3) 0–16.6</td>
</tr>
<tr>
<td>United States Radiologic Technologist Cohort (25)</td>
<td>United States</td>
<td>36.8 (7.3) 22–54</td>
<td>1983–1998</td>
<td>Mailed questionnaire</td>
<td>3</td>
<td>1,570</td>
<td>Self-report</td>
<td>62,862</td>
<td>14.5 (5.6) 0–22.8</td>
</tr>
</tbody>
</table>

*Contributed as of June 2016: includes baseline and each follow-up.

The Canadian Study of Diet, Lifestyle, and Health is the only case-cohort study. The cohort size (N = 1,589) represents the subcohort only.

The European Prospective Study into Cancer and Nutrition (EPIC) dataset does not include the French or Norwegian EPIC sites which contributed from the Etude Epidemiologique auprès de femmes de la Mutuelle Générale de l’Education Nationale and Norwegian Women and Cancer Study directly.

A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis, which was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.

Two statistical approaches are being used to analyze the data. A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In addition, we request at least one follow-up round after age 55 of breast cancer diagnosis (if available). To allow menopausal status to be defined retrospectively, an analysis is conducted on a pre-defined group of women who were enrolled prior to age 55 and who remained in the cohort thereafter. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.

A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In addition, we request at least one follow-up round after age 55 of breast cancer diagnosis (if available). To allow menopausal status to be defined retrospectively, an analysis is conducted on a pre-defined group of women who were enrolled prior to age 55 and who remained in the cohort thereafter. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.

A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In addition, we request at least one follow-up round after age 55 of breast cancer diagnosis (if available). To allow menopausal status to be defined retrospectively, an analysis is conducted on a pre-defined group of women who were enrolled prior to age 55 and who remained in the cohort thereafter. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.

Two statistical approaches are being used to analyze the data. A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In addition, we request at least one follow-up round after age 55 of breast cancer diagnosis (if available). To allow menopausal status to be defined retrospectively, an analysis is conducted on a pre-defined group of women who were enrolled prior to age 55 and who remained in the cohort thereafter. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.

Two statistical approaches are being used to analyze the data. A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In addition, we request at least one follow-up round after age 55 of breast cancer diagnosis (if available). To allow menopausal status to be defined retrospectively, an analysis is conducted on a pre-defined group of women who were enrolled prior to age 55 and who remained in the cohort thereafter. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.

Two statistical approaches are being used to analyze the data. A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In addition, we request at least one follow-up round after age 55 of breast cancer diagnosis (if available). To allow menopausal status to be defined retrospectively, an analysis is conducted on a pre-defined group of women who were enrolled prior to age 55 and who remained in the cohort thereafter. Menopausal status was ascertained by cohorts at each follow-up round for which it was available. In analyses conducted by menopausal status, we will explore differences in lag periods to determine patterns for premenopausal or ‘postmenopausal’ breast cancer, as menopause can be assumed by the inverse of the study-specific variance (36–38). An advantage of this approach is that each study-specific estimate can be derived based on its own available covariates. Cohort-specific Q-statistic is used to examine statistical heterogeneity between studies.
studies by comparing a weighted measure of difference between individual study estimates and the pooled estimate (39, 40). We calculate the I^2 statistic to examine the proportion of variance that is due to between-study heterogeneity rather than chance (41). Potential sources of heterogeneity are investigated.

Maximum flexibility for confounder adjustment and assessment of effect modification can be achieved by pooling individual-level data across cohorts. If homogeneity assumptions are not violated, we pool data into a single dataset to conduct aggregate analyses stratified by study and adjusted for potential confounders that are available in all included studies.

In both approaches, Cox regression models are used to calculate HRs and 95% confidence intervals (CI) for breast cancer (42). Regression models are constructed with age as the time scale such that person-time is accrued from age at cohort entry until breast cancer diagnosis, age at last follow-up, or other exit age, whichever occurred first. Follow-up time is stratified by time-updated exposures obtained from follow-up questionnaires, as appropriate. We test the proportional hazards assumption for exposures of interest, and in case of time-varying associations, for example, an interaction between attained age and the risk factor of interest, we investigate the addition of time-varying covariates in the model. In pooled analyses, potential variation in the association between exposures and breast cancer risk according to tumor subtype are assessed using Cox proportional hazards regression accounting for alternative tumor subtypes as competing risks (43, 44).

Rationale for Initial Aims

**Pregnancy**

A “dual effect” of pregnancy on breast cancer risk has been used to describe the short-term increase in breast cancer risk observed after childbirth followed by a long-term protective effect of parity. This pattern has been reported in epidemiological studies nested within European population registries (45–49) and in other case-control (50–55) and cohort (56) studies. Observational studies have reported 1.25- to 3-fold increases in breast cancer risk for up to 10 years after the last birth (5, 57). The magnitude of the pregnancy-related increase in breast cancer risk varies across studies, and may be influenced by maternal, pregnancy, or post-partum characteristics. Although a period of increased breast cancer risk after childbirth has been reported across several studies, it remains unclear whether this observation is different for, or limited to, specific groups defined by age (5, 50, 51), parity (45, 52, 53), oral contraceptive use (58), breastfeeding practices, family history of breast cancer (48, 59), or varies by breast cancer subtype (55, 56, 60) or other tumor characteristics (61, 62).

Women who have a first birth at an older age may have the greatest initial increase in breast cancer risk, and the longest interval until a protective effect appears (5, 49, 54, 63). Over the last 50 years, more women have postponed childbirth to older ages (5); this trend may have contributed to the increasing advanced-stage breast cancer rates among reproductive-age women. Pregnancy may also have opposite effects on risks of different breast cancer subtypes. For example, without considering menopausal status or subtype, parity reduces overall breast cancer risk by approximately 30% (64). However, parous women have a 50% to 90% increased risk of basal-like or ER−/PR− breast cancer overall (56, 65, 66). Associations for pre- and postmenopausal breast cancer combined often reflect patterns among the majority postmenopausal breast cancer cases. Our study will be well positioned to examine potential variation in the association between recent pregnancy and breast cancer subtype among premenopausal women. Others have proposed that pregnancy-related increases in breast cancer risk may also be affected by the relatively greater influence of genetic predisposition at younger versus older ages at diagnosis (48). In support of this hypothesis, at least two studies have shown stronger associations with recent birth and breast cancer risk among women with a mother or sister who was diagnosed with breast cancer (48, 59).

Theories to explain the transient increased risk of breast cancer after childbirth vary. High levels of estrogen and progesterone and the rapid expansion of breast cells during pregnancy could promote latent initiated tumor cells. However, breast tumors diagnosed postpartum are more often at an advanced stage and are associated with lower survival compared with those diagnosed during pregnancy (67–69). This evidence has led to increased focus on the role of post-partum exposures, including lactational involution (the process that returns the mammary gland to a non-milk producing state), as contributors to a pro-tumorigenic microenvironment that may be favorable for cancer cell migration and metastasis (70). Potential adverse effects of lactational involution on the breast microenvironment must also be reconciled with demonstrated lower risks of specific tumor subtypes among parous women who breastfeed, including ER-negative or basal-like tumors that confer a worse prognosis (56, 65). A better understanding of the factors that contribute to short-term increases in breast cancer risk after pregnancy, including potential variation by age, parity, oral contraceptive use, breastfeeding, family history, or tumor subtype could provide necessary information for refining hypotheses about carcinogenesis in reproductive-age women (71). Individual studies have had insufficient statistical power or have lacked key information to evaluate these
characteristics jointly, making the Collaborative Group an ideal setting in which to advance understanding of pregnancy’s role in premenopausal breast cancer development.

**BMI and other anthropometrics**

There is epidemiological evidence for higher BMI at premenopausal ages having an inverse association with breast cancer risk (72–75). Higher adiposity in childhood and adolescence appears to be associated with a lower risk of breast cancer at both premenopausal (73, 76–78) and postmenopausal (77–79) ages. Whether further weight gain contributes to additional reductions in premenopausal breast cancer risk is not entirely clear (80, 81). A protective effect of adiposity at premenopausal ages is in contrast with the effect of adiposity at postmenopausal ages, with greater BMI after menopause associated with higher risk of breast cancer, probably through production of estrogens by aromatase in adipose tissue (82).

The reason for the protective effect of adiposity at premenopausal ages is unclear, although several hypotheses have been put forward. Fewer ovulatory cycles in heavier women, and consequent lower sex hormone levels, has been suggested as a potential explanation (83). Similarly, an effect of polycystic ovary syndrome (PCOS) has been proposed, although Nurses’ Health Study II data did not support this (73). To find the reasons for the inverse associations with premenopausal adiposity, large study populations are needed to produce stable estimates and to stratify by potentially explanatory factors.

Few published studies have had sufficiently large numbers of premenopausal cases to produce age-specific estimates over a range of ages, or to explore whether risks differ by other explanatory factors or by breast cancer subtype. The few that stratified by established breast cancer risk factors such as parity have so far reported risk estimates to be similar across these factors (72, 78). The association between adiposity and premenopausal breast cancer has been reported to vary by ethnicity, with strong associations in Caucasian, but not in Asian (84) or African-American (85), women, and associations are possibly stronger for ER− than ER+ premenopausal breast cancer (73). It is not clear what level of BMI confers the highest breast cancer risks—one study reported a non-linear association between BMI and risk, with the highest risk around 24 kg/m² (72).

The Collaborative Group, with its large number of cases in the pooled dataset and data on a wide range of risk factors, will be able to clarify the contribution of premenopausal adiposity to breast cancer risk, by examining which subtypes of breast cancer are affected, analyzing associations by exposures such as menstrual factors, and by assessing the effect of changes in adiposity over time.

**Physical activity**

Physical activity is of particular interest in that it constitutes a potentially modifiable risk factor for breast and other cancers. For premenopausal women, the effect of physical activity on reducing breast cancer risk appears to be smaller and less certain than for postmenopausal women (86). However, very few studies (35, 87, 88) have published prospective data for premenopausal breast cancer risk in relation to physical activity, whereas others have published by age at breast cancer (89–91) or menopausal status at study entry (92–95), or have included premenopausal women in their study but did not publish effect estimates for these women separately (96, 97).

The biological mechanisms through which physical activity could exert an effect in premenopausal women is less clear than in postmenopausal women, but might be through an effect on menarche, menstrual dysfunction, cycle length, endogenous hormone levels or estrogen metabolism (98–100). A smaller effect of physical activity in premenopausal than postmenopausal women is possible because, in contrast to postmenopausal women, in whom the protective effect of physical activity on breast cancer risk is partly through its effect on reducing adiposity, adiposity in premenopausal women has a protective effect on breast cancer risk. In addition, the impact of physical activity on hormone levels might be less influential among premenopausal women given their high levels of circulating hormones.

To aid prevention, information is needed on the type, frequency and intensity of exercise required to influence breast cancer risk, as well as the ages and characteristics of women for whom it is most effective. There might be periods of life during which physical activity has a higher impact than others, such as the time-period between menarche and first birth (101). There is also emerging evidence of differential effects of activity by ethnicity, weight, parity and family history of breast cancer, but mostly based on data from postmenopausal women (35, 91, 102). It is a limitation, however, that physical activity information is collected in many different ways and is difficult to harmonize (103).

The Collaborative Group aims to address premenopausal breast cancer risk by frequency, intensity, type, and ages of exercise, within strata defined by factors such as BMI, family history of breast cancer and age at diagnosis, and to explore specific breast cancer subtypes and stages, on a much larger scale than previously. The information gained can be used to advise young women about the extent and type of exercise that can influence their breast cancer risk.

**Opportunities and challenges**

The Collaborative Group is an international collaboration formed to address etiological factors for breast cancer that may be particular to, or differ in, premenopausal or perimenopausal women. By harmonizing a wide range of exposure variables across 20 studies and developing quality assurance and analysis programs, our collaboration is in a position to conduct initial analyses of pregnancy, obesity and physical activity, and to leverage the research infrastructure and established collaboration model for investigations of other risk factors. Our initial aims do not require the use of biospecimens. However, biospecimens have been collected in many of the participating studies, as described in the Cancer Epidemiology Descriptive Cohort Database (available at https://ceded.nci.nih.gov/biospecimen.aspx) and could potentially be incorporated to address future hypotheses.

Some limitations and challenges have emerged. As in many consortia, information from the participating studies in the Collaborative Group was not collected with future pooling efforts in mind and follow-up data are not collected at standardized intervals. Therefore, harmonization efforts must identify common data elements that are collected with minimal levels of measurement error. Identification of these elements can be complicated by questionnaires and codebooks that must be translated to a common language.

Another aspect of working on pooling cohorts that requires planning and forethought is the potential for overlap of
participants between studies, for example, in Scandinavian countries with multiple cohorts that have wide geographic catchment areas. Although the existence of national identities makes it theoretically possible to identify women who may contribute information to more than one study in a country, the logistics for obtaining approval and merging datasets can be prohibitive. Therefore, we have worked with study investigators to identify the individual cohorts within a country with the most relevant information for specific Collaborative Group aims, and to develop strategies for excluding specific geographic regions from one cohort, but not another, where overlap of cohort catchment areas is known to exist.

The value of cancer consortia to address scientific questions efficiently and create new opportunities has become increasingly recognized [104]. Conducting analyses across multiple studies requires ongoing communication and transparency. Our Collaborative Group holds in-person working group meetings in conjunction with the NCI Cohort Consortium annual meeting, as well as regular telephone conferences. These meetings provide a forum to discuss additional hypotheses that can be addressed in the future to maximize the value of the created infrastructure. The Cohort Consortium provides valuable coordinating and communication services and dedicated time and space through the annual meeting; however, other research support for data preparation, ongoing infrastructure development, and dedicated time for collaboration remains a challenge faced across many large-scale projects. Our Collaborative Group and others continue to work to identify and streamline data sharing models to maximize productivity and collaborative opportunity.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We wish to acknowledge all study participants, staff, and participating cancer registries as well as Hoda Anton-Culver, Jianwen Cai, Jessica Clague, Christine Clarke, Dennis Deapen, Nicolas Hakansson, Allison Iwan, Diane Kampa, James Lacy, Eun Jung Lee, Shiew-Hong Low, David Nelson, Susan Neuhouser, Katie O’Brien, Hannah Park, Jerry Reid, Peggy Reynolds, Sophia Wang, Renwei Wang, Mark Weaver, Jiawei Xu, Jeffrey Yu, and Angyrisz Ziegas.

**References**


