The Premenopausal Breast Cancer Collaboration: A pooling project of studies participating in the National Cancer Institute Cohort Consortium

Formatted for Cancer Epidemiology Biomarkers and Prevention


*Co-first authors
†Co-senior authors

Affiliations (Institution, Country):
1 University of North Carolina Gillings School of Global Public Health, USA
2 Division of Genetics and Epidemiology, The Institute of Cancer Research, UK
3 Karolinska Institutet, MEB, Sweden; Harvard T.H. Chan School of Public Health, USA; University of Oslo Institute of Health and Society, Norway
4 Fondazione IRCCS Istituto Nazionale dei Tumori, Italy
5 Centre for Research in Epidemiology and Population Health (CESP), France.
6 Beckman Research Institute of City of Hope, USA
7 Slone Epidemiology Center at Boston University, USA
8 Vanderbilt University School of Medicine, USA
9 University of Pittsburgh Graduate School of Public Health and University of Pittsburgh Cancer Institute, USA
10 NYU School of Medicine, USA
11 National Cancer Institute, National Institutes of Health, USA
12 International Agency for Research on Cancer, France
13 Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School, USA; Harvard T.H. Chan School of Public Health, USA
14 Cancer Council Victoria, University of Melbourne; Australia
15 University of Tromsø, The Arctic University of Norway, Norway
School of Public Health and Health Sciences, University of Massachusetts, USA
Johns Hopkins Bloomberg School of Public Health, USA
Division of Cancer Epidemiology, DKFZ, Heidelberg, Germany
University of Oxford, UK
Dalla Lana School of Public Health, University of Toronto, Canada
Duke-NUS Medical School, Singapore
Karolinska Institute, Institute of Environmental Medicine, Sweden
UiT (University of Tromsø), Norway
School of Public Health, Imperial College London, UK
CIBER Epidemiología y Salud Pública (CIBERESP), Spain
Aarhus University, Denmark
Radiation Effects Research Foundation, Japan
University Medical Center, Utrecht, Netherlands
Albert Einstein College of Medicine, USA
Umeå University, Sweden
Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School; Harvard T.H. Chan School of Public Health, USA
Hellenic Health Foundation, Greece
Norwegian University of Science and Technology, Norway
Johns Hopkins School of Medicine
Karolinska Institutet, Department of Epidemiology and Biostatistics; Department of Research, Head, Group of Etiological Cancer Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway; Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland; Department of Community Medicine, University of Tromsø, The Arctic University of Norway, Tromsø, Norway
Departments of Nutrition and Epidemiology, Harvard T.H. Chan School of Public Health, USA
Karolinska Institutet, Institute of Environmental Medicine, Sweden
National Institute of Environmental Health Sciences, National Institutes of Health, USA
Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

Key words: Premenopausal breast cancer, epidemiology, prospective, pooling, cohort

Corresponding author:
Hazel B. Nichols
Assistant Professor, Department of Epidemiology
University of North Carolina Gillings School of Global Public Health
2104F McGavran-Greenberg Hall, 135 Dauer Drive
Chapel Hill, NC 27599-7435
Phone: (919) 966-7456
Email: hazel.nichols@unc.edu
Acknowledgements:
We wish to acknowledge all study participants, staff, and participating cancer registries as well as Hoda Anton-Culver, Jianwen Cai, Jessica Clague, Christina Clarke, Dennis Deapen, Niclas Håkansson, Allison Iwan, Diane Kampa, James Lacey, Eunjung Lee, Siew-Hong Low, David Nelson, Susan Neuhausen, Katie O’Brien, Hannah Park, Jerry Reid, Peggy Reynolds, Sophia Wang, Renwei Wang, Jiawei Xu, Jeffrey Yu, and Argyrios Ziogas.

Financial support:
Support for this research comes, in part, from the Avon Foundation (02-2014-080); Breast Cancer Now; The Institute of Cancer Research, London; the United States National Institutes of Health National Institute of Environmental Health Sciences (Z01 ES044005; P30 ES000260) and National Cancer Institute (UM1 CA176726; UM1 CA186107; R01 CA058420; UM1 CA164974; CA077398; CA144034; UM1 CA182876; UM1 CA182876; UM1 CA182934); the United States National Center for Advancing Translational Sciences (KL2-TR001109), the National Program of Cancer Registries of the Centers for Disease Control and Prevention, and the Department of Energy; the Swedish Research Council and Swedish Cancer Foundation; the Japanese Ministry of Health, Labour and Welfare; the Hellenic Health Foundation; Karolinska Institutet Distinguished Professor Award Dnr: 2368/10-221; Cancer Council Victoria and the Australia National Health and Medical Research Council (209057; 396414; 504711); the State of Maryland, the Maryland Cigarette Restitution Fund, and the United Kingdom National Health Service funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

The coordination of the European Prospective Investigation in Cancer (EPIC) is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l’Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom).
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 paper 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.
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 aged 15-39 years worldwide (1). Further, breast cancer among premenopausal women often presents at more advanced stages and has less favorable prognosis than among postmenopausal women (2, 3), and its onset frequently coincides with caregiving responsibilities for children and/or aging parents.

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 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 factors 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 (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 paper 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 2 or more time points (baseline and at least one follow-up, to allow for exposure information and menopausal status to be updated) were 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)(6, 10-28) with data submitted to the Collaborative Group as of June 2016. Participating cohorts are shown in Table
and span North America, Europe, Asia, and Australia. The numbers of female participants from these cohorts aged <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 incident 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 were identified by the Coordinating Centers. A comprehensive data request was
developed to minimize the need for supplemental data requests once analyses were underway. Complete capture of all information across exposures was not required for participation.

After confirming eligibility, cohort-specific questionnaires were reviewed to evaluate availability of variables, and a data request template was sent to cohorts that wished to participate.

Requested exposure data included: 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. Most of these characteristics were collected at enrollment and each follow-up, as available. Breast cancer information included age at diagnosis, stage, grade, histology, and expression of ER, PR, HER2, CK5/6, or EGFR.

Participating studies were asked where possible to recode their own data to fit the data request template to minimize the potential for error in the recoding or understanding of variables in their original form. However, if this was not possible due to programming support constraints or other reasons, data were sent to the Coordinating Centers in their original form with a study-specific contact person identified to address questions from Coordinating Center programmers who reformatted the information to fit the standard definitions in the data request template.

After data transfer agreements were signed between each individual study and the Coordinating Centers, completed datasets were transferred from each individual study to the coordinating centers using secure file transfer protocols. Each cohort submitted their data to one of the two
Coordinating Centers who took responsibility for data transfer and harmonization procedures. By having two data coordinating sites, one located in the United States and the other in the United Kingdom, we were able to minimize time zone differences to facilitate rapid communication, and accommodate studies that were only able to send data to certain locations because of country-specific information governance requirements.

Data harmonization procedures were standardized across Coordinating Centers. Quality control checks were run on each dataset to identify (i) potential data inconsistencies for each questionnaire round (e.g. nulliparous women reporting more than zero births), (ii) inconsistencies between questionnaire rounds (e.g. number of births at follow-up being lower than at baseline questionnaire), and (iii) implausible values. Data checking procedures were automated with a shared program that was run at each Coordinating Center with standardized output. Each cohort was contacted regarding any issues that were identified, and clarifications or updates were incorporated into the study-specific dataset. Where issues could not be resolved, pre-established recoding rules were applied to the data. When study-specific variables could not be recoded to meet the data template formats (e.g. age at exposure was collected in categories but a continuous variable was requested), differences were documented and original data were retained for potential future use. Once the datasets were recoded to the standardized formats, data were merged to create a pooled dataset containing values from all cohorts.

Defining menopausal status

A primary issue for the Collaborative Group analyses is the definition of menopausal status during follow-up and at diagnosis. Menopausal status was contributed by cohorts at each follow-
up round for which it was available. In addition, we requested at least one follow-up round after age 55 or breast cancer diagnosis (if available) to allow menopausal status to be defined retrospectively. In analyses conducted by menopausal status we will explore different lag periods to determine patterns for ‘premenopausal’ or ‘perimenopausal’ breast cancer, as menopause can be a gradual transition.

Statistical approach

Two statistical approaches are being used to analyze the data. We first examine study-specific estimates and a pooled estimate across studies using a random-effects model that weights estimates 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. Cochran’s $Q$ statistic is used to examine statistical heterogeneity between 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 hazard ratios (HR) 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, e.g. 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 (2, 5). 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
(57), breastfeeding practices, family history of breast cancer (48, 58), or varies by breast cancer subtype (55, 56, 59) or other tumor characteristics (60, 61).

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, 62). 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 ~30% (63). However, parous women have a 50-90% increased risk of basal-like or ER/PR-breast cancer overall (56, 64, 65). 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 idea, 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, 58).

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 (66-68). 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 (69). 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, 64). 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 (70). Individual studies have had insufficient statistical power or lacked key information to evaluate these characteristics jointly, making the Collaborative Group an ideal setting 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 a protective effect on breast cancer risk (71-73). This protective effect of premenopausal BMI is already evident with respect to higher adiposity in childhood and adolescence, and appears to be associated with a lower risk of breast cancer at both premenopausal (72, 74-76) and postmenopausal (75-77) ages. Whether further weight gain contributes additional reductions in premenopausal breast cancer risk is not entirely clear (78, 79). A protective effect of adiposity at
premenopausal ages is in contrast to the effect of adiposity at postmenopausal ages, with greater BMI after menopause associated with higher risk of breast cancer, probably through production of oestrogens by aromatase in adipose tissue (80).

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 (81). Similarly, an effect of polycystic ovary syndrome (PCOS) has been proposed, although Nurses’ Health Study II data did not support this (72). 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 (71, 76). The association between adiposity and premenopausal breast cancer has been reported to vary by ethnicity, with strong associations in Caucasian, but not in Asian (82) or African-American (83), women, and associations are possibly stronger for ER+ than ER-premenopausal breast cancer (72). 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 (71).
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 factors 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 (84). However, very few studies (35, 85, 86) have published prospective data for premenopausal breast cancer risk in relation to physical activity, whereas others have published by age at breast cancer (87-89) or menopausal status at study entry (90-93), or have included premenopausal women in their study but did not publish effect estimates for these women separately (94, 95).

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 oestrogen metabolism (96-98). 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.
Additionally, 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 (99). There is also emerging evidence of differential effects of activity by ethnicity, normal weight, parity and family history of breast cancer, but mostly based on data from postmenopausal women (35, 89, 100). It is a limitation, however, that physical activity information is collected in many different ways and is difficult to harmonize (101).

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.

Some limitations and challenges have emerged. Our analyses, as currently constituted, will not address premenopausal breast cancer risk after age 55. Age 55 years is approximately the 88th percentile for age at menopause in the United States (102) and may account for an even higher proportion of the premenopausal lifespan in other countries. 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 identifiers 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 (103). 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.
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 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>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) 0.0-17.2</td>
</tr>
<tr>
<td>Campaign against Cancer and Heart Disease (CLUE II) (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 Épidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale (E3N) (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>719</td>
<td>Self-report and national registry</td>
<td>72,058</td>
<td>5.4 (1.7) 0-9.7</td>
</tr>
<tr>
<td>Helseundersøkelsen i Nord-Trøndelag (HUNT2)(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) 0.25-14.0</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 (104)</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>Study Name</td>
<td>Country</td>
<td>Age Range</td>
<td>Time Period</td>
<td>Data Collection Method</td>
<td>Number Respondents</td>
<td>Data Source</td>
<td>Number Subjects</td>
<td>Confidence Interval</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------------------------------------</td>
<td>--------------------</td>
<td>------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>Radiation Effects Research Foundation Life Span Study (21)</td>
<td>Japan</td>
<td>18-54</td>
<td>1950-1993</td>
<td>Administered or mailed questionnaire</td>
<td>130</td>
<td>City registry</td>
<td>18,420</td>
<td>13.5 (8.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1-36.7</td>
<td></td>
</tr>
<tr>
<td>Singapore Chinese Health Study (22)</td>
<td>Singapore</td>
<td>43-54</td>
<td>1993-1998</td>
<td>Administered questionnaire</td>
<td>134</td>
<td>National cancer registry</td>
<td>16,056</td>
<td>5.3 (3.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3-11.5</td>
<td></td>
</tr>
<tr>
<td>Sister Study (6)</td>
<td>United States</td>
<td>35-54</td>
<td>2003-2009</td>
<td>Telephone and written questionnaire</td>
<td>679</td>
<td>Self-report</td>
<td>24,044</td>
<td>4.7 (2.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1-10.6</td>
<td></td>
</tr>
<tr>
<td>Southern Community Cohort Study (23)</td>
<td>United States</td>
<td>40-54</td>
<td>2002-2009</td>
<td>Administered questionnaire</td>
<td>233</td>
<td>State registry</td>
<td>30,289</td>
<td>5.1 (2.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1-13.3</td>
<td></td>
</tr>
<tr>
<td>Sweden Women’s Lifestyle and Health Study (27)</td>
<td>Sweden</td>
<td>29-49</td>
<td>1991-1992</td>
<td>Mailed questionnaire</td>
<td>1,192</td>
<td>National registry</td>
<td>49,010</td>
<td>14.4 (5.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1-21.1</td>
<td></td>
</tr>
<tr>
<td>Swedish Mammography Cohort (24)</td>
<td>Sweden</td>
<td>38-54</td>
<td>1987-1990</td>
<td>Mailed questionnaire</td>
<td>649</td>
<td>National registry</td>
<td>34,126</td>
<td>8.3 (4.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0-16.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0-22.8</td>
<td></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 Epidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale and Norwegian Women and Cancer Study directly.
Table 2. Breast cancer characteristics among women younger than 55 years across the Premenopausal Breast Cancer Collaborative Group.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Combined N</th>
<th>Total N studies with data available*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total breast cancers diagnosed</td>
<td>21,766</td>
<td>20 (all)</td>
</tr>
<tr>
<td>Age at diagnosis (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>1,245</td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>3,340</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>7,053</td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>10,096</td>
<td></td>
</tr>
<tr>
<td>Extent of disease</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>In situ</td>
<td>3,645</td>
<td></td>
</tr>
<tr>
<td>Invasive</td>
<td>17,364</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>757</td>
<td></td>
</tr>
<tr>
<td>Estrogen receptor status</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Positive</td>
<td>9,583</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>3,182</td>
<td></td>
</tr>
<tr>
<td>Borderline</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>8,949</td>
<td></td>
</tr>
<tr>
<td>Progesterone receptor status</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Positive</td>
<td>7,919</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>3,939</td>
<td></td>
</tr>
<tr>
<td>Borderline</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>9,813</td>
<td></td>
</tr>
<tr>
<td>HER2/neu overexpression</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Positive</td>
<td>1,092</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>4,808</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>15,188</td>
<td></td>
</tr>
</tbody>
</table>

*contributed as of June 2016.
References

58. Albrektsen G, Heuch I, Thoresen S, Kvale G. Family history of breast cancer and short-
Hypothesized role of pregnancy hormones on HER2+ breast tumor development. Breast cancer 
60. Albrektsen G, Heuch I, Thoresen SO. Histological type and grade of breast cancer 
tumors by parity, age at birth, and time since birth: a register-based study in Norway. BMC 
61. Albrektsen G, Heuch I, Thoresen S, Kvale G. Clinical stage of breast cancer by parity, 
age at birth, and time since birth: a progressive effect of pregnancy hormones? Cancer 
62. Innes KE, Byers TE. First pregnancy characteristics and subsequent breast cancer risk 
and risk of breast cancer: a meta-analysis of 8 studies from the Nordic countries. International 
factors and oestrogen/progesterone receptor-negative breast cancer in the Breast Cancer 
66. Johansson AL, Andersson TM, Hsieh CC, Cnattingius S, Lambe M. Increased mortality 
in women with breast cancer detected during pregnancy and different periods postpartum. 
67. Stensheim H, Moller B, van Dijk T, Fossa SD. Cause-specific survival for women 
diagnosed with cancer during pregnancy or lactation: a registry-based cohort study. J Clin 
Oncol. 2009;27:45-51.
Stage at diagnosis and mortality in women with pregnancy-associated breast cancer (PABC). 
mammary gland involution drives progression of ductal carcinoma in situ through collagen and 
Postpartum Remodeling, Lactation, and Breast Cancer Risk: Summary of a National Cancer 
71. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. 
Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. 
72. Michels KB, Terry KL, Willett WC. Longitudinal study on the role of body size in 
breast cancer risk factors in relation to molecular subtypes of breast cancer. Breast cancer 
75. Berkey CS, Frazier AL, Gardner JD, Colditz GA. Adolescence and breast carcinoma 
76. Baer HJ, Tworoger SS, Hankinson SE, Willett WC. Body fatness at young ages and risk 


