SF 424 R&R and PHS-398 Specific
Table of Contents

SF 424 Cover Page 2

Research & Related Other Project Information..... 3

Project Summary/Abstract..... 4

Project Narrative 5

Facilities and Other Resources 6

PHS 398 Cover Page Supplement.....10

PHS 398 Research Plan 11

Specific Aims 12

Research Strategy 13

Protection of Human Subjects.....22

Inclusion of Women and Minorities.....24

Planned Enrollment Report 25

Inclusion of Children 26

Data Sharing Plan.....27

PI: JOHN, ESTHER M	Title: Risk factors for breast cancer subtypes in racial/ethnic minorities	
	FOA: PAR12-039	
	FOA Title: SMALL GRANTS PROGRAM FOR CANCER EPIDEMIOLOGY (R03)	
	Organization: CANCER PREVENTION INSTIT OF CALIFORNIA	
Senior/Key Personnel:	*Organization:*	*Role Category:*
Esther John Ph.D.	Cancer Prevention Institute of California	PD/PI

RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved? ☒ Yes ☐ No

1.a If YES to Human Subjects

Is the Project Exempt from Federal regulations? ☐ Yes ☒ No

If NO, is the IRB review pending? ☒ Yes ☐ No

2. Are Vertebrate Animals Used? ☐ Yes ☒ No

3. Is proprietary/privileged information included in the application? ☐ Yes ☒ No

4a. Does this project have an actual or potential impact on the environment? ☐ Yes ☒ No

5. Is the research performance site designated, or eligible to be designated, as a historic place? ☐ Yes ☒ No

6. Does this project involve activities outside of the United States or partnerships with international collaborators? ☐ Yes ☒ No

PROJECT SUMMARY/ABSTRACT

Racial/ethnic differences in breast cancer incidence have been well documented, yet the reasons underlying these differences are only partially understood. Relatively few studies have been conducted in racial/ethnic minority populations and some have produced findings on risk factors that are different from those reported for non-Hispanic White women. Furthermore, most studies in minority populations have assessed risk factors for breast cancer overall, despite growing evidence that breast cancer is a heterogeneous disease, with risk factors that differ for breast cancer subtypes defined by estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2) status. Most currently known breast cancer risk factors apply to hormone receptor positive (ER+ and/or PR+) or Luminal A (ER+ and/or PR+, HER2) tumors. Few studies have focused on risk factors for the less common subtypes, such as hormone receptor negative (ER-PR-), triple negative (ER-PR-HER2-), or HER2 over-expressing (ER-PR-HER2+) tumors. Reports on risk factors for the less common subtypes were, with a few exceptions, based on small case numbers and included primarily non-Hispanic white women. Given that breast cancer subtypes are not equally distributed across racial/ethnic groups, with a higher incidence of aggressive subtypes in minority populations, it is important that etiologic studies in minority populations assess risk factors in relation to specific breast cancer subtypes. To address this significant gap in knowledge, we will pool existing interview and cancer registry data for 9,000 breast cancer cases and 7,855 controls who participated in five population-based studies. Participants cover a broad age range (18-79 years) and 71% of cases and 66% of controls are racial/ethnic minorities (Hispanics, Asian Americans, African Americans, non-Hispanic whites). In Aim 1, we will evaluate risk factors for breast cancer subtypes and assess heterogeneity by age and menopausal status. Using polytomous logistic regression, we will assess associations of subtypes with a broad set of risk factors, including family history of breast cancer, hormonal factors and modifiable lifestyle factors. In Aim 2, we will assess whether risk factors for the main subtypes differ across racial/ethnic groups. Leveraging existing data, this study will provide much needed information about risk factors for the less common breast cancer subtypes and will allow direct comparison of subtype-specific risk factors across multiple racial/ethnic groups. Such information will inform the development of preventive strategies that are directly relevant for specific racial/ethnic groups.

PROJECT NARRATIVE

We will analyze existing data from five population-based studies of breast cancer to investigate risk factors for breast cancer subtypes defined by ER, PR, and HER2 status. We will also evaluate whether risk factors for specific breast cancer subtypes differ across racial/ethnic groups.

FACILITIES AND OTHER RESOURCES

Environment – Contribution to Success:

Cancer Prevention Institute of California (CPIC): CPIC offers a wide range of epidemiologic expertise to support Dr. John in the planning and implementation of the proposed case-control analyses. Dr. John, the PI of this subcontract, has over 25 years of experience in conducting epidemiologic studies focused on racial/ethnic minority populations, including Hispanics, African Americans, and Asian Americans. During her 20-year tenure at CPIC, her studies of breast and prostate cancer have recruited of over 13,000 study participants. Dr. John's research program is managed and coordinated by a highly experienced Senior Program Manager who oversees the office and field staff who support on-going studies. Dr. John has many years of experience collaborating with other scientists, both nationally and internationally, and therefore has access to a wide network of expertise that could benefit the proposed study. Dr. John also has access to other intellectual resources to support this application, including access to the Stanford Cancer Institute cores and other Stanford resources, including biostatistics support, GIS and SAS software and the full complement of scientific journals and books from the medical and other libraries. The physical location of the CPIC offices, in the heart of the San Francisco Bay Area, also conveniently places it within easy access to other research organizations, including Stanford University, UC Berkeley, UC San Francisco, and Kaiser Division of Research. The offices themselves are easily accessible via local and air transportation and can easily be used for hosting meetings for up to 150 attendees.

All of these aspects of the CPIC environment and its resources make the CPIC team the ideal partner in this current application.

Institutional Commitment to ESI [include only if PI is an Early Stage Investigator]:
Not applicable.

Facilities:

Laboratory:
Not applicable. No biospecimens will be transferred from CPIC for the proposed study.

Clinical:
CPIC operates the Greater Bay Area Cancer Registry (GBACR), a partner in the NCI-supported Surveillance, Epidemiology, and End Results (SEER) program and the state-supported and mandated California Cancer Registry (CCR). The GBACR, which covers nine Bay Area counties that are rich in population and geographic diversity, collects data on over 27,000 new cancer cases and 10,000 cancer-associated deaths. It also serves as a critical resource of population-based data for epidemiologic and clinical studies. Dr. John's epidemiologic studies conducted in the San Francisco Bay Area have relied on the GBACR to identify newly diagnosed cancer cases. Thus, tumor characteristics for the breast cancer cases and survival data to be included in Dr. in the proposed study will derive from the GBACR.

The resources from the GBACR are critically important to carry out the proposed research.

Animal:
Not applicable.

Computer:
Dr. John will rely on CPIC's computer resources to perform the statistical analyses and communicate with the investigator team (Dr. Wang and Ms. Sangaramoorthy).

CPIC's Information Technology resources consist of a multi-platform communications infrastructure based on:

- Microsoft file & print operations
- Microsoft Authentication services – Active Directory for all user accounts and rights/permissions structures
- SMTP messaging (Exchange)
- FTP (additional file transfer option)
- Cisco networking and security systems:
  - Firewall / Network access system
  - Network core switching
  - VPN remote connectivity solution
  - VOIP Telephony solution
- Symantec/Veritas backup and restore systems
- Trend Micro anti-virus security solutions
- Approx 180 computer systems (combination of desktops and laptops) with additional loaner, test or conference systems
- Approx 40 network servers providing both the services listed above and the foundational systems listed below
- Two desktop publishing workstations

In addition to the communications infrastructure, appropriate foundational systems and support services are provided for all activities required by this study including:

- Dedicated GIS storage/processing system (specific and vertically focused storage and GIS analysis platform, input/output and processing platform, with capabilities of execution, management and analysis of these large sets of data combinations)
- ESRI ArcInfo license
- Access to ESRI ArcGIS products through Stanford
- Application services (e.g., Database, TomCat, IIS) for the GBACR and other similar programs
- Development services (e.g., VB, HTML, XML, SAS)
- FTP (File Transfer Protocol) services for large or secure file transfers with other organizations
- Optical imaging using a 60 platter optical jukebox for storage and retrieval functions
- SAS working platforms
- Secured internet or Extranet connectivity and storage capabilities using SSL, IPSEC or other secure network transport and storage methods
- Print services, including 17 networked printers available to all users / staff (including 2 high speed color printer/copiers), and 30+ dedicated desktop printers
- Tele-conference, video-conference, and Skype conference capabilities

Thus, CPIC has all of the necessary computer hardware and software equipment and associated support to successfully carry out the tasks to which they have committed as part of this research.

Office:
CPIC has 2 office locations, one in Fremont, CA and another in Berkeley, CA.

Fremont Office: The organization occupies 33,598 square feet at the Fremont, CA site. The space includes numerous furnished private offices and cubicles, four conference rooms, a large training room with a moveable wall which enables it to be transformed into a large open meeting space (150 person capacity), a full kitchen and break room and two large copy/storage rooms. Administration, IS/IT, research programs, the Surveillance, Epidemiology, and End Results (SEER) program, the regional office of the California Cancer Registry (CCR) program, and other education/information programs share this space. The Fremont office is a short drive across the Bay from Stanford University and a short distance from the Kaiser Division of Research (DOR). Two large-volume photocopiers are available at the Fremont site for large volume rapid copying and collating. Four smaller photocopy machines are also available. Six facsimile machines are also available. CPIC scientists also have access to office and conference spaces at Stanford. Dr. John has office space in this location.

Berkeley Office: The organization occupies 3,328 square feet in Berkeley, CA. This space includes several offices, a small computer room and kitchen, and two conference rooms. IS/IT resources include Windows-based, portable and non-portable PC microcomputers with mid- to large-size data storage, management and statistical analyses capabilities as well as record linkage, mapping, scanning, graphics, desktop publishing, Internet accessibility and electronic-mail capabilities. The office's PCs are connected to a wide-area network facilitating the communication, research and data exchange capabilities among the staff at the Berkeley office, and externally with the Fremont office and beyond. The Berkeley office also maintains a full complement of printing devices, including high quality, black/white color copier/printer/scanner, a black/white laser printer, a color ink-jet printer and a large format, color ink-jet plotter capable of producing E-size output. The Berkeley office is located within a short walking distance from the University of California, Berkeley, and is 2 blocks from the Berkeley School of Public Health. It is a short driving distance across the Bay from the University of California, San Francisco (UCSF) and its various medical campuses.

Both CPIC office locations are conveniently located within the San Francisco Bay Area, and are conducive to collaborations with neighboring institutions, including Berkeley, UCSF, Stanford, and Kaiser DOR, as well as investigators elsewhere in California or in the US. With easy access from three area airports, both CPIC office locations also house ample space for hosting small- and large-scale meetings, and can easily host investigator meetings.

Other Resources:

Medical literature: Through its partnership with the Stanford Cancer Institute, CPIC scientists are Cancer Institute members and have access to the core resources available through the University. This includes access to a full array of scientific journals through the Stanford University Libraries. Services available include inter-library loan, reference assistance and journal access, and a pull and copy service (for a small fee). These services are available through a number of sources including on-line, email and in-person access. CPIC scientists also have access to other core resources through the Stanford Cancer Institute, including access to faculty in the Biostatistics core.

Internet hosting resources: CPIC maintains a comprehensive website that describes our mission, our investigators and our research efforts within communities/populations throughout

California, the U.S. and internationally. This website specifically contains a detailed profile of each investigator, all of the research studies and projects as well as a complete list of publications of all investigators. Most of these publications are also linked to PubMed abstracts so that others can view the details of the work conducted by the organization. Additionally, the website has the capabilities to upload education and training materials, certain research instruments (i.e., questionnaires) and other tools which can be shared with other researchers outside the organization.

Administrative support: Dr. John has a part-time administrative assistant to provide administrative support for the proposed study, as needed.

Regulatory compliance support: Dr. Robert McLaughlin, provides investigators with study-specific guidance and support to ensure streamlined IRB review and ongoing compliance with state and federal regulations.

Existing collaborations: Dr. John collaborates with a wide network of scientists who have expertise in epidemiology, molecular epidemiology, biostatistics, genetics, surveillance research, behavioral research, laboratory sciences, and clinical sciences.

Thus, Dr. John has access to an array of physical and intellectual resources that optimize her ability to support the development and implementation of the proposed study.

Major Equipment: List the most important equipment items already available for this project, noting the location and pertinent capabilities of each.
Not applicable.

PHS 398 Cover Page Supplement

OMB Number: 0925-0001

1. Project Director / Principal Investigator (PD/PI)

Prefix:	Dr.
First Name:	Esther
Middle Name:	M.
Last Name:	John
Suffix:	Ph.D.

2. Human Subjects

Clinical Trial?	☒ No	☐ Yes
Agency-Defined Phase III Clinical Trial?*	☐ No	☐ Yes

5. Human Embryonic Stem Cells

* Does the proposed project involve human embryonic stem cells?	☒ No	☐ Yes

PHS 398 Research Plan

OMB Number: 0925-0001

Please attach applicable sections of the research plan, below.

1. Introduction to Application (for RESUBMISSION or REVISION only)	
2. Specific Aims	1242-Specific_Aims.pdf
3. Research Strategy*	1243-Research_Strategy.pdf
4. Progress Report Publication List	
Human Subjects Sections	
5. Protection of Human Subjects	1244-Protection_of_Human_Subjectcs.pdf
6. Inclusion of Women and Minorities	1245-Inclusion_of_Women_and_Minorities.pdf
7. Inclusion of Children	1246-Inclusion_of_Children.pdf
Other Research Plan Sections	
8. Vertebrate Animals	
9. Select Agent Research	
10. Multiple PD/PI Leadership Plan	
11. Consortium/Contractual Agreements	1247-Consortium_Contractual_Arrangements.pdf
12. Letters of Support	1248-Letters_of_Support.pdf
13. Resource Sharing Plan(s)	1249-Data_Sharing_Plan.pdf
Appendix (if applicable)	
14. Appendix	

1.SPECIFIC AIMS

Racial/ethnic differences in breast cancer (BC) incidence and mortality are well-documented[1,2], yet it remains unresolved whether BC risk factors differ by race/ethnicity, In 2007-2011, U.S. incidence rates (per 100,000) were highest in non-Hispanic whites (NHWs) (134.1), followed by African Americans (122.8), Hispanics (95.1), Asian Americans/Pacific Islanders (93.6), and American Indians/Alaska Natives (79.3)[3]. Racial/ethnic differences in incidence may be due to racial/ethnic differences in the prevalence of risk factors, in the magnitude of associations with risk factors, or in risk factors yet to be identified[4-7]. BC studies in racial/ethnic minorities have led to conclusions that associations with certain factors observed in NHW women may not hold for minorities. For example, high body mass index (BMI), a well-established risk factor for postmenopausal BC in NHWs, was not associated with risk in African American[8] or Hispanic[9] women. Similarly, high parity, a protective factor well-documented in NHW women[10], was not associated with reduced risk in African Americans[11]. In a recent pooled analysis, we have shown, however, that body size is indeed associated with BC risk in Hispanic women when considering BC subtype and other modifying factors[12,13], and others have reported inverse associations with parity in African American women, but only for those diagnosed with estrogen receptor positive (ER+) BC. These findings highlight the importance of considering subtypes in studies of BC etiology, as there is growing evidence that risk factors vary by BC subtypes[14-18]. In particular, it appears that some of the traditionally known BC risk factors may not apply to the less common, but more aggressive subtypes such as triple negative BC [TNBC; negative for ER, progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2)] or basal-like tumors, or HER2 over-expressing (HER2+) tumors[19-22]. However, risk factors for the less common BC subtypes (i.e., TNBC, HER2+) remain to be identified since most prior studies included only small numbers of cases with these less common subtypes and findings are inconsistent. Furthermore, most investigations of BC subtypes, either in single studies or pooled datasets[14-16,20,23,24], were conducted in NHW women. There is an urgent, yet unmet need to learn about risk factors for BC subtypes in racial/ethnic minority populations. We cannot simply assume that risk factors for BC subtypes are the same across racial/ethnic groups.

There continues to be a significant lack of comprehensive studies of BC etiology in racial/ethnic minority populations. Findings for racial/ethnic minorities often derive from studies with small sample sizes which are difficult to interpret given imprecise risk estimates, some studies include a single racial/ethnic minority group only (e.g., African Americans[25,26]), and most studies in minorities are too small to perform subtype-specific analyses. Given that BC subtypes do not occur at equal frequency across racial/ethnic groups[27-31], it is critical to compare risk factors across racial/ethnic groups for specific BC subtypes rather than all BCs combined. We hypothesize that associations with known risk factors are similar across racial/ethnic groups when considering BC subtypes.

We will test this hypothesis in a large and diverse dataset by pooling interview and other data from 5 population-based studies that included large numbers of racial/ethnic minorities: the San Francisco Bay Area Breast Cancer Study (SFBCS, dx 1995-2002)[32,33]; the Northern California Breast Cancer Family Registry (NC-BCFR, dx 1995-2009)[34,35]; the Triple Negative Breast Cancer Study (TNBCS, dx 2007-2009); the 4-Corners Breast Cancer Study (4-CBCS, dx 1999-2004)[9], and the Asian American Breast Cancer Study (AABCS, dx 1995-2001 and 2003-2006)[36]. Together, these studies comprise 9,000 BC cases and 7,855 population controls. BC subtype will be defined using immunohistochemical (IHC)-based tumor markers as surrogates for molecular BC subtypes[37,38]. Tumor markers available in cancer registry databases include

ER, PR, and HER2 (HER2 status is available for cases diagnosed since 2000). We will address the following specific aims:

1. Assess heterogeneity in breast cancer risk factors (family history, hormonal, and lifestyle factors) by breast cancer subtypes defined by joint ER/PR status (ER+PR+, ER+PR-, ER-PR+, ER-PR-) or joint ER/PR/HER2 status (Luminal A, Luminal B, TNBC, HER2+).

2. For the main breast cancer subtypes (ER+PR+, ER-PR-, Luminal A, TNBC), assess risk factors separately for each racial/ethnic group (Hispanic, African American, Asian American and NHW).

This proposal efficiently leverages uniquely diverse data resources (interview data and cancer registry-based data on tumor characteristics) for all four major U.S. racial/ethnic groups, making excellent use of existing data at a relatively low cost. The analyses will produce much needed information on etiologic factors for specific BC subtypes that will allow direct comparisons across multiple racial/ethnic groups. If differences in risk factors emerge across racial/ethnic groups, such findings will identify areas for more in-depth scientific investigation, as well as opportunities for tailored primary prevention strategies that are directly relevant for specific racial/ethnic groups.

2. RESEARCH STRATEGY

2.1. SIGNIFICANCE AND IMPACT

Several observations suggest that BC subtypes are biologically distinct: risk factors for specific BC subtypes emerging from studies in NHWs are distinct, molecular characteristics of specific subtypes are unique, and prognosis is markedly different for different subtypes. Therefore, studying specific subtypes holds greater promise of understanding etiologic factors influencing BC risk, or identifying novel risk factors, particularly for the less common subtypes[39]. Our understanding of BC etiology in racial/ethnic minority populations will likely improve if subtypes are taken into account. Our investigation will focus on the role of a broad set of risk factors, including potentially modifiable lifestyle factors. Given that so few risk factors have been identified for the less common subtypes like TNBC and HER2+, a better understanding of the role of lifestyle factors could shape the development and implementation of preventive strategies for these aggressive subtypes. As many as 26% of African Americans and 17% of Hispanics diagnosed with BC have TNBC, compared to 12% in NHWs and 10% in Asian Americans[40,41], with an even higher prevalence in women diagnosed at age <40 years (36% in African Americans, 27% in Hispanics, 16% in Asian Americans vs. 22% in NHWs)[41]. HER2+ tumors are also more common in minority populations than NHWs (6-8% vs. 4%)[31], particularly in certain Asian American subgroups[41]. Furthermore, TNBC and HER2+ tumors have a much worse prognosis than other subtypes[38,42,43], and racial/ethnic differences in BC subtype distributions likely contribute to the well-known survival disparities[44]. Therefore, it is critical to identify risk factors for BC subtypes, including modifiable lifestyle factors, so that research can be directly translated into preventive strategies and clinical interventions. Such work is particularly needed for TNBC and HER2+ tumors for which the etiology remains largely elusive; most of the currently known BC risk factors apply to hormone receptor positive BCs. Importantly, subtype-specific risk interventions may need to be tailored to specific racial/ethnic groups if the risk factors for specific subtypes indeed differ by race/ethnicity.

2.2. INNOVATION

Given our large pooled dataset, the proposed analysis will produce more robust data on risk factors for BC subtypes, especially the less common subtypes like TNBC and HER2+ tumors. It will produce novel data on subtype-specific risk factors for BC in Hispanics, African Americans, and Asian Americans, evaluating a broad set of risk factors (demographics, BC family history, hormonal and lifestyle factors). This study will significantly add to the sparse data on BC subtype-specific risk factors in minority populations. Importantly, this will be the first large-scale study to generate data on etiologic factors of BC subtypes for the 4 major U.S. racial/ethnic groups, and not just for a single racial/ethnic minority group as in some consortia[26]. There is a particularly urgent need for a better understanding of BC etiology in Hispanics, a growing U.S. population that has been studied the least[5,33,45-48] among the major racial/ethnic groups.

2.3. APPROACH

Overview. We will harmonize and pool interview and cancer registry data from 5 population-based studies of BC that represent the racial/ethnic diversity of the U.S. population and California in particular. We will leverage existing data to evaluate heterogeneity in risk factors by BC subtypes, as well as subtype-specific heterogeneity by age and menopausal status (Aim 1). For the major BC subtypes, we will also assess heterogeneity in risk factors by race/ethnicity (Aim 2). We will focus on a comprehensive set of risk factors (demographics, BC family history, hormonal and lifestyle factors) that have been associated with BC risk overall. By pooling data from several U.S. population-based BC studies, we will have a dataset that is probably the largest and most racially/ethnically diverse sample of cases and controls to investigate subtype-specific etiologic factors.

Investigator Team. The proposed analysis will be carried out by an experienced team of investigators, including Drs. John, Slattery and Wu, who are the Principal Investigators of the studies being pooled for the collaborative analyses, and Dr. Ingles, a biostatistician, who has collaborated with Dr. John since the late 1990s. The team has worked on BC etiology for many years, with a special focus on racial/ethnic minority populations and a publication record of collaboration on other projects and pooled analyses.

Study Population and Data Collection. The pooled analysis will use existing data from 5 large population-based BC studies (SFBCS, NC-BCFR, TNBCS, 4-CBCS, AABCS; see below) that represent large investments by the NCI, the Department of Defense, and the California Breast Cancer Research Program. Although each individual study is relatively large and focused on one or more racial/ethnic minority populations, the individual sample sizes are too small to investigate risk factors for the less common BC subtypes (i.e., ER-PR-, TNBC, HER2+ tumors) or racial/ethnic differences in subtype-specific risk factors. Each of these studies collected data by in-person interview using similar structured questionnaires, addressing a comprehensive set of factors known or suspected to be associated with BC risk. To maximize high quality data collection, the studies translated questionnaires into the appropriate languages (Spanish, Chinese, Tagalog) for non-English speaking participants. Thus, the study participants are not limited to English-speaking Hispanics and Asian Americans, as is the case in some studies. The large number of exposure variables that are common across the 5 studies will facilitate the harmonization and pooling of interview data. Furthermore, the PIs have engaged in other pooled analyses, thereby demonstrating that data harmonization is feasible between the 5 studies.

The San Francisco Bay Area Breast Cancer Study (SFBCS). SFBCS (PI: E. John), a population-based case-control study[32,33], identified invasive BC cases aged 35-79 yrs through the Greater Bay Area Cancer Registry, including Hispanics (dx 4/1995-4/2002), African Americans (dx 4/1995-4/1999), and NHWs (10% random sample, dx 4/1995-4/1999). Controls were identified through random-digit dialing (RDD), frequency-matched on race/ethnicity and 5-year age group. Interview data are available for 2,258 cases (1,119 Hispanics, 543 African Americans, 596 NHWs) and 2,706 controls (1,462 Hispanics, 598 African Americans, 646 NHWs).

The Northern California Breast Cancer Family Registry (NC-BCFR). NC-BCFR (PI: E. John), a population-based prospective BC family study focused on racial/ethnic minorities[34,35], identified BC cases aged 18-64 yrs through the Greater Bay Area Cancer Registry, including Hispanics and African Americans (dx 1/1995-12/2009); Chinese, Filipinas and Japanese (dx 1/1995-4/2002); and NHWs and other Asians (dx 1/1995-9/1998). Eligibility criteria included: dx at age <35 yrs, personal history of ovarian or childhood cancer, personal history of bilateral BC with a first dx at age <50 yrs, or a first-degree family history of breast or ovarian cancer. Cases not meeting these criteria were randomly sampled (2.5% of NHWs, 33% of other race/ethnicities). Interview data are available for 3,268 invasive cases, 75% of whom are racial/ethnic minorities. Controls were identified through RDD (case:control ratio of 2:1; frequency-matched on race/ethnicity and 5-year age group). Data from the NC-BCFR and SFBCS have been harmonized for other joint analyses[49,50].

Triple Negative Breast Cancer Study (TNBCS). TNBCS (PI: E. John), an ancillary study within the NC-BCFR that used the same protocols and questionnaires, enrolled TNBC cases <65 yrs (dx 1/2007-12/2009). Interview data are available for 215 TNBC cases.

The 4-Corners Breast Cancer Study (4-CBCS). 4-CBCS (PIs: M. Slattery, K. Baumgartner, A. Giuliano, T. Byers), a population-based case-control study conducted in Arizona, Colorado, New Mexico and Utah[9], identified Hispanic and NHW cases aged 25-79 yrs (dx 10/1999-5/2004) through state-wide cancer registries; controls were selected from the target populations of cases, frequency-matched on race/ethnicity and 5-year age group. Interview data are available for 1,936 cases (668 Hispanics, 1,269 NHWs) and 2,526 controls (924 Hispanics, 1,602 NHWs). 4-CBCS and SFBCS are part of the Breast Cancer Health Disparities Study[51] (PI: M. Slattery) which harmonized data between the two studies (see Dr. Slattery's letter of support).

Asian American Breast Cancer Study (AABCS). AABCS (PI: A. Wu), a population-based case-control study conducted in Los Angeles county, identified Asian American cases aged 25-74 yrs (dx 1995-2001 and 2003-2006) through the Los Angeles cancer registry; controls were identified through neighborhood block walking, and frequency-matched on ethnicity and 5-year age group. Interview data are available for 1,847 cases and 2,002 controls. Cases from AABCS and SFBCS are part of the California Breast Cancer Survivorship Consortium (PI: A. Wu) which examines the determinants of survival disparities across racial/ethnic groups[52]. The data have been harmonized between the two studies [52,53] (see Dr. Wu's letter of support).

Available Data for Pooled Analysis. The 5 studies described above collected a comprehensive set of factors known or suspected to be associated with BC risk overall. Interview Data: Common data items include: demographic background (race/ethnicity, ethnic subgroup, country of birth, English language acculturation, education), 1st-degree BC family history, menstrual factors (age at menarche, menopausal status, age at natural or surgical menopause), reproductive factors (nulliparity, age at first full-term pregnancy, number of full-term pregnancies, time between menarche and first pregnancy, time since last full-term

pregnancy), breast-feeding, hormone use (oral contraceptives, menopausal hormones), body size (height, current BMI, young adult-BMI, weight gain, abdominal adiposity), and lifestyle factors (alcohol consumption, cigarette smoking, recreational physical activity). Tumor Characteristics: For cases, information on histology, grade, tumor size, lymph nodes, ER, PR, and HER2 (for cases diagnosed since 2000) has been obtained from the cancer registry in each study.

Data Harmonization. Through prior pooled case-control and survival analyses, we have demonstrated that data harmonization between the 5 studies is feasible (for example, between SFBCS & NC-BCFR[49,50], SFBCS & 4-CBCS[13,54], SFBCS & AABCS[53]). The proposed study benefits from harmonization procedures that have already been developed for a number of variables of interest, including variables that may be more challenging to harmonize (e.g., physical activity[49]). We will generate analytic variables according to common definitions that will be checked for outliers or unreasonable values. Our data pooling will not be limited to a few BC risk factors, as is often the case when data are pooled from many studies[24]. We will create a rich database with a comprehensive set of BC risk factors (demographics, BC family history, hormonal and lifestyle factors).

Sample Size of Pooled Dataset. The pooled dataset will comprise cases with a first primary invasive BC diagnosed from 1995-2009 in California or in the 4-Corners region, and population controls selected from the same geographic areas where the cases were diagnosed. Case-control analyses will be based on 9,000 cases (71% minorities), classified by subtype[22,38,43,55], and 7,855 controls (66% minorities) (Table 1). Race/ethnicity is based on self-report in all studies.

Table 1: Pooled dataset from SFBCS, NC-BCFR, TNBCS, 4-CBCS, and AABCS

		dx 1995-2009				dx 2000-2009			
	All Subjects	ER+PR+	ER+PR-	ER-PR+	ER-PR-	Luminal A[1]	Luminal B[2]	TNBC[3]	HER2+[4]
Hispanics									
Cases	2,593	1,320	239	56	552	499	135	156	77
Controls	2,460								
African Americans									
Cases	1,203	535	126	39	309	210	64	108	37
Controls	671								
Asian Americans									
Cases	2,613	1,413	253	204	456	629	199	136	108
Controls	2,091								
NHWs									
Cases	2,591	1,380	225	50	510	36	9	199	9
Controls	2,633								
Total									
Cases	9,000	4,648	843	349	1,827	1,374	407	599	231
Controls	7,855								

[1] Luminal A: ER+ or PR+, HER2-; [2] Luminal B: ER+ or PR+, HER+; [3] TNBC: ER-PR-HER2-; [4] HER2+: ER-PR-HER2+.

SPECIFIC AIM 1: assess heterogeneity in breast cancer risk factors (family history, hormonal, and lifestyle factors) by breast cancer subtypes defined by joint ER/PR status (ER+PR+, ER+PR-, ER-PR+, ER-PR-) or joint ER/PR/HER2 status (Luminal A, Luminal B, TNBC, HER2+).

Justification and Significance. There is increasing epidemiologic evidence that some risk factors for BC overall (e.g., nulliparity, parity, age at first birth, age at menarche, oral

contraceptive use) are not associated with certain subtypes such as TNBC or basal-like tumors. Findings, however, are not consistent[19-21,23,55-58], likely due to the small numbers of TNBC/basal-like cases included in most studies to date, ranging from less than 200[20-22,55,59,60] to less than 400[23,39 19,56,57,61,62]. Only recent studies included larger numbers of TNBCs[26,58,63,64]. Similarly, risk factor analyses for HER2+ tumors are limited by small numbers and results have not been consistent[20,24,55,56,59,62,63,65]. Furthermore, most of these reports on risk factors for TNBC and HER+, except for the AMBER consortium[26], included primarily NHW women. Large collaborative studies, like the Breast Cancer Association Consortium (BCAC) which reported on risk factors for ER-PR- BC[24], are often limited in their ability to harmonize data across many studies that used different questionnaires, and therefore focus the pooled analysis on a limited number of risk factors that are common to all studies. Thus, thorough and comprehensive analyses of risk factors for the less common subtypes like ER-PR-, TNBC or HER2+ tumors are lacking, yet clearly needed, particularly given the limited number of effective treatments and higher mortality for these more aggressive subtypes[38,66,67]. Our proposed pooled analysis will include sizeable numbers of ER-PR- tumors (n=1,827), TNBCs (n=599), and HER2+ tumors (n=231) and will address a comprehensive set of risk factors (demographics, BC family history, hormonal and lifestyle factors) that have been associated with BC risk overall. Thus, this pooled analysis will provide new insights into the etiology of BC subtypes, particularly the less common subtypes, at a relatively low cost by leveraging existing resources.

Feasibility. As part of prior collaborative projects[13,49,50,53,54], we have demonstrated the feasibility of harmonizing the interview data for selected variables across the 5 studies included in the pooling, and we have determined that sample sizes for the main subtypes are adequate for analyses by race/ethnicity (Table 1).

Research Design and Analysis Plan. In Aim 1, we will assess tumor characteristics and risk factors across BC subtypes and evaluate associations with BC subtypes performing case-case and case-control analyses.

***Exposures*.** We will evaluate associations of BC subtypes with 3 sets of risk factors (hereafter referred to as exposures): Demographics and family history, including ethnic subgroup, country of birth, English language acculturation, education, and 1st-degree BC family history; Hormonal factors, including menstrual factors (age at menarche, menopausal status, age at menopause), and reproductive factors (nulliparity, number of full-term pregnancies, age at first full-term pregnancy, time interval between menarche and first pregnancy, time since last full-term pregnancy); and Modifiable lifestyle factors, including breastfeeding, oral contraceptive use, menopausal hormone therapy use, body size, recreational physical activity, alcohol consumption, and cigarette smoking. Most of these factors have not been thoroughly evaluated in relation to BC subtypes, and because those that are modifiable have direct relevance for primary prevention, a better understanding of their association with specific subtypes is a high priority.

BC Subtypes. We will evaluate associations with the above risk factors for subtypes defined by ER, PR, and HER2 available from cancer registry records. For cases diagnosed from 1995-2009, subtype will be defined by joint ER/PR status: ER+PR+ (n=4,648), ER+PR- (n=843), ER-PR+ (n=349), and ER-PR- (n=1,827). Among BCs with known hormone receptor (HR) status, ~20% are ER-PR- tumors and they include the most aggressive tumors[20]. In our pooled dataset that is enriched with racial/ethnic minorities, ER-PR- BCs account for 24% of tumors with known HR status. For cases diagnosed from 2000-2009, subtypes will be classified according to joint ER/PR/HER2 status and we will distinguish the following subtypes[22,38,43]: Luminal A (ER+ or PR+, HER-; n=1,374), the subtype with the best prognosis; Luminal B (ER+

or PR+, HER+; n=407) a subtype with less favorable outcomes than Luminal A tumors; and triple negative (TNBC; ER-PR-HER2-; n=599) and HER2 over-expressing (HER2+; ER-PR-HER2+; n=231), the two subtypes with the poorest prognosis. A comparison of ER/PR/HER2 data in the SEER cancer registry vs. classification by a single expert pathology laboratory found substantial agreement[68]. We recognize that because this pooled analysis will make efficient use of existing data only, HER2 status will not be available for all BC cases. Similarly, tumor marker data (e.g., CK5 or CK5/6 or EGFR) for more refined subtype classification (e.g., basal subtype) are not available from routine cancer registry records. However, de novo collection of these tumor marker data would be very expensive and not feasible, as tissue sample collection for subtyping and medical record release would require new consent and retrospective tissue collection is challenging for cases diagnosed a long time ago. Furthermore, some studies found no systematic differences in patient and tumor characteristics between cases with and without HR data[23,57]. We also plan to assess differences between the two sets of cases (those with and without tumor marker data). Despite these limitations, our proposed analyses defining subtypes by ER/PR/HER2 are of significant value, as minorities continue to be excluded or under-represented in studies of subtype-specific BC etiology.

Characteristics of BC subtypes. Using cancer registry data on tumor characteristics, we will assess univariate differences in subtype distribution by race/ethnicity, age at diagnosis (<40, 40-49, ≥50 yrs), stage at diagnosis (AJCC stage I-IV), tumor grade (well, moderate, poor differentiation), tumor histology (ductal, lobular, mixed/other), tumor size (<2, 2-4, ≥5 cm), and node involvement (negative, positive). Similarly, we will evaluate differences in subtype distribution by questionnaire-based characteristics and exposures. Chi-square tests will be used to assess differences in frequencies of categorical variables and t-tests will be used to assess differences in means of continuous variables.

Subtype-specific associations. The primary goal of aim 1 is to assess heterogeneity in breast cancer risk factors by BC subtypes. We will perform 2 sets of analyses: *case-case analyses*[69] and *case-control analyses*.

In *case-case analyses*, we aim to understand heterogeneity among cases by comparing the distribution of exposures across BC subtypes, using the most common subtype as the referent group. For cases diagnosed from 1995-2009, we will compare ER+ vs. ER-, PR+ vs. PR-, and ER+PR+ (the most common subtype) vs. ER+PR-, ER-PR+, and ER-PR-. For cases diagnosed from 2000-2009 classified by joint ER/PR/HER2 status, we will compare Luminal A (the most common subtype) vs. Luminal B, TNBC, and HER2+. We will use standard and polytomous unconditional logistic regression models to estimate odds ratios (ORs) 95% confidence intervals (CIs), and p-values for associations between exposures and BC subtypes. Outcome (dependent) variables will be BC subtypes defined by ER only, PR only, joint ER/PR, or joint ER/PR/HER2; explanatory variables will be the exposures of interest, including age at diagnosis (<40, 40-54, ≥55), 1st-degree BC family history (yes, no), age at menarche (≤12, 13, 14, ≥15 yrs), parity (nulliparous, parous), age at first full-term pregnancy (<20, 20-24, 25-29, 30-34, ≥35 yrs), number of full-term pregnancies (1,2,3,4, ≥5), time since last full-term pregnancy (<3, 3-4, 5-9, ≥10 yrs), time interval between menarche and age at first full-term pregnancy (<5, 5-9, 10-14, 15-19, ≥20 yrs), menopausal status (pre-, postmenopausal), type of menopause (natural, surgical), age at menopause (<45, 45-49, 50-54, ≥55 yrs), lifetime breast-feeding (0, <12, 12-23, ≥24 months), duration of oral contraceptive (OC) use (0, <5, 5-9, ≥10 yrs), age at first OC use (<20, 20-24, 25-29, ≥30 yrs), menopausal hormone use (never, past, current), current alcohol consumption (0, <10, 10-19, ≥20 grams/day), smoking status (never, past, current), number of cigarettes smoked (0, <10, 10-19, ≥20/day), recreational physical activity (quartiles), current BMI (<25, 25-29, ≥30 kg/m2), young-adult BMI (quartiles), adult weight gain (quartiles), waist

circumference (quartiles), waist-to-hip ratio (quartiles), and waist-to-height ratio (quartiles). Alternate categorizations of variables may be used, depending on the actual distribution of exposures in the pooled dataset. Associations with subtype will be evaluated for each exposure, adjusting for race/ethnicity, age at diagnosis (continuous), education, and study. Linear trends in associations will be assessed across ordinal values of categorical variables. Exposure variables that are significantly associated with BC subtype will be included in the final multivariate logistic regression models. These case-case comparisons will be useful to uncover etiologic heterogeneity[69]. Importantly, these analyses will cover a comprehensive set of exposures in a large dataset of 9,000 cases.

In *case-control analyses*, we will further investigate the etiology of BC subtypes by comparing separately cases with a particular BC subtype to a common referent group of controls. We will calculate ORs, 95% CIs, and P values to estimate relative risks of developing a particular BC subtype associated with the exposures of interest. We will use conventional and polytomous unconditional logistic regression models with case-control status (each BC subtype vs. controls) as outcome variables and each exposure as explanatory variables. Outcome and explanatory variables will be defined as described above. Regression models will be adjusted for race/ethnicity, age, education, and study. Linear trends in associations will be assessed across ordinal values of categorical variables. Two-sided P values will be used for test of trend and tests of heterogeneity (for polytomous regression). Heterogeneity tests will be performed by testing for equality of regression coefficients across subtype. Exposures found to be significantly associated with a particular subtype will be included in the fully adjusted multivariate logistic regression models. Because two studies (NC-BCFR, TNBCS) recruited cases based on BC family history, cases from those studies will be excluded from analysis of family history. Additionally, since associations with BC risk overall differ between pre- and postmenopausal women for some factors (e.g., obesity[70] or reproductive factors[71]), we will also assess differences in subtype-specific risk factors by menopausal status (pre- vs. postmenopausal). Finally, because TNBCs and HER2+ tumors are more frequent among younger women[19,41], we will also assess differences in subtype-specific risk factors by age (<40 vs. ≥40 yrs). Few studies have investigated risk factors for BC subtypes in younger women[21,59]. To formally evaluate differences in OR estimates for specific subtypes by age and menopausal status, we will include interaction terms by age and menopausal status in the logistic regression models.

Sample size and statistical power. As shown in Table 1, for all women combined, case counts will range from 4,648 (ER+PR+ cases) to 231 (HER+ cases) for the most common and least common subtype, respectively. We estimated power for case-case analyses since the goal of Aim 1 is to assess heterogeneity in exposure associations by subtype. Classifying subtypes by joint ER/PR and assuming a 15% exposure prevalence (EXPV), we will have 80% power to detect ORs as low as 1.23 for ER-PR- and 1.32 for ER+PR- relative to ER+PR+. For analysis of subtypes classified by joint ER/PR/HER2, assuming a 25% EXPV, we will have 80% power to detect ORs as low as 1.36 for TNBC and 1.54 for HER2+ relative to Luminal A. For case-case analyses by age (451 ER+PR+, 308 ER-PR-, 105 TNBC with dx at age <40 yrs), assuming a 15% EXPV, we will have 80% power to detect ORs as low as 1.70 for ER-PR- relative to ER+PR+. For TNBC, assuming a 25% EXPV, we will have 80% power to detect ORs as low as 2.19 relative to Luminal A (or 1.91 relative to ER+PR+). For case-case analyses by menopausal status (1745 ER+PR+, 766 ER-ER-, 248 TNBC who are premenopausal), assuming a 15% EXPV, we will have 80% power to detect ORs as low as 1.38 for ER-PR- relative to ER+PR+. For TNBC, assuming a 25% EXPV, we can detect ORs as low as 1.60 relative to Luminal A (or 1.51 relative to ER+PR+). Thus, the pooled dataset of 9,000 BC cases will have excellent power to detect heterogeneity in exposures by BC subtype.

SPECIFIC AIM 2: For the main BC subtypes (ER+PR+, ER-PR-, Luminal A, TNBC), assess risk factors separately for each racial/ethnic group (Hispanic, African American, Asian American and NHW).

Significance and Justification. U.S. racial/ethnic minority populations continue to be understudied in epidemiologic BC studies. Data are particularly sparse for Hispanics and Asian Americans, two large and growing immigrant populations. Findings on BC risk factors for specific racial/ethnic groups are inconsistent, and because of limited sample size, most prior studies in minorities have examined risk factors for all BCs combined. We hypothesize that the failure to consider BC subtypes may have contributed to inconsistent results in minority populations, given evidence that specific subtypes are not equally distributed across minority populations. To date, most investigations of risk factors for specific BC subtypes have included NHW women only. Only recently have studies begun to examine risk factors for specific BC subtypes in racial/ethnic minorities[9,57,60,72-74], but most studies individually have sample sizes that are too small. More systematic analyses in large studies or pooled datasets, as we aim to do in this proposal, are clearly needed in order to directly compare risk factors for BC subtypes across multiple racial/ethnic groups. A better understanding of BC subtype-specific risk factors in each major U.S. racial/ethnic groups is critical for the tailoring, if warranted, of possible risk-reducing preventive measures. ER-PR- tumors and TNBC will be a special focus of the planned analysis. ER-PR- tumors account for ~20% of all tumors; TNBCs account for 25% and 17% of BCs in African Americans and Hispanics, respectively[40]. Given the lack of effective treatments for TNBCs[66], the disproportionate burden of this disease subtype likely contributes to the higher BC mortality in African Americans.

Research Design and Analysis. In Aim 2, we will assess risk factors for specific BC subtypes in the 4 racial/ethnic groups (Hispanic, African American, Asian American, NHW) represented in the pooled dataset. Our goal is to learn whether there are risk factors for subtypes that are unique to certain racial/ethnic groups.

Analysis approach: We will use the same approach as described above for Aim 1, but stratify the case-case and case-control analyses by race/ethnicity. Exposures and BC subtypes will be defined as described above for Aim 1. Given the more limited sample size when stratifying on race/ethnicity, we will focus the analyses of Aim 2 on the main subtypes with sufficient numbers (ER+PR+, ER-PR-, Luminal A, TNBC). In case-control analyses, we will estimate subtype-specific risk associated with the exposures of interest, separately in each racial/ethnic group. To formally test for heterogeneity in associations by race/ethnicity, we will include interaction terms by race/ethnicity in the logistic regression models. Grouping Asian Americans into a single category is not optimal as they may differ in distributions of risk factors and BC subtypes[41]. Therefore, we will perform exploratory analyses for Chinese (1116 cases, 935 controls) and Filipinas (900 cases, 618 controls), the two largest Asian subgroups in the pooled dataset.

Sample size and statistical power: We estimated power for case-control analyses, since the goal of Aim 2 is to assess subtype-specific risk factors in each racial/ethnic group. For ER+PR+ BC (n=1320 in Hispanics, 535 in African Americans, 1413 in Asian Americans) and assuming a 15% exposure prevalence (EXPV), we will 80% power to detect ORs as low as 1.29, 1.53, and 1.29, respectively. For ER-PR- BC (n=552, 309, 456, respectively), we will have 80% power to detect ORs as low as 1.41, 1.64, and 1.46, respectively. For TNBC (156, 108, 136, respectively), assuming a 25% EXPV, we will have 80% power to detect ORs as low as 1.64 and 1.85, and 1.70 respectively.

Expected Results and Impact. To our knowledge, this will be the first comprehensive analysis that will assess whether differences exist between racial/ethnic groups in risk factors for specific BC subtypes defined by ER/PR/HER2. We expect that 1) among all women combined, there will be heterogeneity across subtypes for some risk factors, but not others. Hormonally related risk factors (e.g., parity) will be limited to hormone receptor (HR) positive tumors, whereas for HR negative tumors, TNBC and HER2+, the risk factors will be different from the well-established risk factors for BC overall. 2) Risk factors for a given subtype will be similar across racial/ethnic groups. A better understanding of risk factors (particularly modifiable lifestyle factors) for the more aggressive BC subtypes with worse prognosis (e.g., ER-PR-, TNBC, HER2+), which are more common in younger minority populations, will be important in terms of primary prevention. For example, the finding of an inverse association between breast-feeding and TNBC is directly translatable into primary prevention strategies. Our results will also provide important directions for future studies of racial/ethnic differences in risk factors, such as genetic susceptibility and gene-environment interactions.

2.4. TIMELINE AND MILESTONES. Months 1-8: Obtain IRB approval at all participating institutions, establish inventory of data available in the 5 studies, identify common data items, assemble data to be harmonized from the 5 studies, harmonize data according to common definitions and create a common set of analytic variables. Months 9-15: Perform logistic regression analyses for Aim 1, and prepare manuscripts describing the study findings related to Aim 1. Months 16-24: Perform logistic regression analyses for Aim 2, and prepare manuscripts describing the study findings related to Aim 2.

PROTECTION OF HUMAN SUBJECTS

Human Subjects Involvement, Characteristics, and Design. The proposed study involves secondary analysis of existing interview data that were provided by breast cancer patients (aged 18 years or older at diagnosis) and women of similar age without a personal history of breast. Participating women completed an interview (in-person or by telephone) for several breast cancer studies: The San Francisco Bay Area Breast Cancer Study (SFBCS, PI: Esther John), the Northern California Breast Cancer Family Registry (NC-BCFR, PI: Esther John), the Triple Negative Breast Cancer Study (TNBCS, PI: Esther John), the 4-Corners Breast Cancer Study (4-CBCS, PI: Marty Slattery), and the Asian American Breast Cancer Study (AABCS, PI: Anna Wu). We will combine the interview data provided by 16,855 women (9,000 women with breast cancer, 7,855 women without breast cancer).

Participating breast cancer cases and women without breast cancer resided in the target area of each of these studies: the Greater San Francisco Bay area (SFBCS, NC-BCFR, TNBCS), Los Angeles county (AABCS), and the four corners area (Utah, Colorado, New Mexico, Arizona; 4-CBCS). They include women from all racial/ethnic populations. Because the studies were focused on studying breast cancer in racial/ethnic minorities, 69% of the participating women are from racial/ethnic minority populations (71% of breast cancer cases, 67% of controls).

Sources of materials. The objective of this data analysis project is two-fold: a) to assess risk factors profiles for specific breast cancer subtypes defined by estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2); and b) to assess whether the risk factors profiles for specific subtypes are the same across racial/ethnic groups. To do so, we will combine and harmonize existing interview data from the above listed studies. The study does not involve subject contact, thus no new data will be collected directly from human subjects.

Written informed consent. All study participants have provided written informed consent when they were enrolled in the above listed studies. The written consent forms are on file at the institution of the PI for each study, and each participating woman was given a signed copy for her own records.

Contacting subjects. This study does not involve any patient contact, only existing data will be analyzed.

Potential risks for subjects. Data to be analyzed come from data files that do not contain personal identifiers such as names or addresses. All data are identified by study identification number only. We therefore expect risks (e.g., breach of confidentiality) to study participants to be extremely low.

Procedures for protecting against or minimizing potential risks. Existing data that will be used in this analysis project are identified by study ID numbers. Completed questionnaires may be consulted if necessary for data cleaning. Questionnaires are stored securely at each institution where the studies originated. Questionnaires and data files do not contain any personal identifying information, such as names or addresses. All data files with the pooled data from the above listed studies will be stored on the network at the Cancer Prevention Institute of California (CPIC) where the PI, Esther John, is located. Files will be accessed only by selected study staff via password-secured computers. All study staff involved in the data analysis will be trained to follow established CPIC confidentiality procedures and will sign confidentiality agreements. No preliminary or final results will be released or published with identifying

information and all epidemiologic data will be presented as statistical summaries only. These procedures are well established at CPIC and have been used in many previous epidemiologic studies.

Discussion of why risks are reasonable in relation to anticipated benefits to subjects. There are no direct benefits for subjects participating in this research study. However, it is our hope that there is the indirect benefit of contributing to potentially valuable research that can be applied to a broader population and be useful for informing intervention strategies to eliminate disparities in breast cancer.

Inclusion of women. This study includes women diagnosed with breast cancer and women without a personal history of breast cancer.

Inclusion of minorities. This study includes all racial/ethnic minority groups, as well as non-Hispanic Whites. Racial/ethnic minorities account for 69% of subjects included in the proposed analyses. The parent studies were specifically focused on studying breast cancer etiology in racial/ethnic minority populations.

Inclusion of children: This study will not involve any children, only women aged 18 years or older were enrolled in the parent studies.

Monitoring for data quality at the Cancer Prevention Institute of California (CPIC). CPIC has a general protocol for data and safety monitoring. Each applicable research project will monitor data for quality.

Monitoring for adverse events & reporting requirements. The IRB requires investigators to report adverse events. This requirement states: "All [CPIC] investigators conducting research with human subjects must report adverse events to the IRB within five working days of the date of occurrence or of the investigator's knowledge that an adverse event has occurred. For this purpose, an adverse event is defined as 'an undesirable and unintended, although not necessarily unexpected, result of therapy or other intervention.' In non-medical research an adverse event can consist of an undesirable and unintended consequence of or reaction to procedures. Any possible change to the risk/benefit ratio must be evaluated, and recruitment and consent procedures may need modification. Subject complaints should be reported." (CPIC IRB Manual page 33). Since this study does not involve direct patient contact, we do not expect any adverse events to happen. If any were to occur (e.g., breach of confidentiality), they would be reported immediately by the Principal Investigator to the IRB administrative staff for appropriate review by the IRB.

Data and safety monitoring specific to current study. The proposed study, while considered to be "human clinical research," is not a clinical trial, behavioral clinical trial, or intervention. It involves the analysis of personal information previously collected via interview with breast cancer cases and women without breast cancer, and data obtained from the cancer registry. Although no adverse events are anticipated, should one occur, we will follow the procedures outlined above.

INCLUSION OF WOMEN AND MINORITES

The proposed study will include 16,855 women, including 11,631 (69%) from racial/ethnic minority populations.

The racial/ethnic composition will be 30% white Hispanic, 31% non-Hispanic white, 11% African-American, and 28% Asian-American.

OMB Number: 0925-0002

Planned Enrollment Report

Study Title: Risk factors for breast cancer subtypes in racial/ethnic minorities

Domestic/Foreign: Domestic

Comments:

Racial Categories	Ethnic Categories				Total
	Not Hispanic or Latino		Hispanic or Latino		
	Female	Male	Female	Male	
American Indian/Alaska Native	0	0	0	0	**0**
Asian	4704	0	0	0	**4704**
Native Hawaiian or Other Pacific Islander	0	0	0	0	**0**
Black or African American	1874	0	0	0	**1874**
White	5224	0	5053	0	**10277**
More than One Race	0	0	0	0	**0**
Total	**11802**	**0**	**5053**	**0**	**16855**

Study 1 of 1

INCLUSION OF CHILDREN

The proposed study includes women ages 18 years or older. No children are involved.

DATA SHARING PLAN

No new data will be generated by the proposed research. The proposed statistical analyses of existing data will be obtained from several studies: 1) the Northern California Breast Cancer Family Registry (NC-BCFR); 2) Triple Negative Breast Cancer Study (TNBCS); the San Francisco Bay Area Breast Cancer Study (SFBCS); the 4-Corners Breast Cancer Study (4-CBCS); and the Asian American Breast Cancer Study (AABCS).

Existing data from the NC-BCFR and TNBCS have been available to qualified researchers since the inception of the Breast Cancer Family Registry (BCFR) in 1995. Information on how to apply for these data is available on the BCFR website: http://www.bcfamilyregistry.org/

For use of existing data from the SFBCS (PI: esther.john@cpic.org), 4-CBCSC (PI: marty.slattery@hsc.utah.edu), and AABCS (PI: annawu@usc.edu), interested collaborators may contact the PIs of each study.