A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP1923.

The Sister Study Cohort: Baseline Methods and Participant Characteristics

Dale P. Sandler,[1] M. Elizabeth Hodgson,[2] Sandra L. Deming-Halverson,[2] Paula S. Juras,[1] Aimee A. D'Aloisio,[2] Lourdes M. Suarez,[2] Cynthia A. Kleeberger,[2] David L. Shore,[3] Lisa A. DeRoo,[4] Jack A. Taylor,[1] Clarice R. Weinberg[5] and the Sister Study Research Team

[1]Epidemiology Branch, National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Research Triangle Park, North Carolina, USA
[2]Social & Scientific Systems, Inc., Durham, North Carolina, USA
[3]Westat, Durham, North Carolina, USA
[4]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway
[5]Biostatistics and Computational Biology Branch, NIEHS, NIH, DHHS, Research Triangle Park, North Carolina, USA

Background: The Sister Study was designed to address gaps in the study of environment and breast cancer by taking advantage of more frequent breast cancer diagnoses among women with a sister history of breast cancer and the presumed enrichment of shared environmental and genetic exposures.

Objective: The Sister Study sought a large cohort of women never diagnosed with breast cancer but who had a sister (full or half) diagnosed with breast cancer.

Methods: A multifaceted national effort employed novel strategies to recruit a diverse cohort, and collected biological and environmental samples and extensive data on potential breast cancer risk factors.

Results: The Sister Study enrolled 50,884 U.S. and Puerto Rican women 35–74 y of age (median 56 y). Although the majority were non-Hispanic white, well educated, and economically well off, substantial numbers of harder-to-recruit women also enrolled (race/ethnicity other than non-Hispanic white: 16%; no college degree: 35%; household income <$50,000: 26%). Although all had a biologic sister with breast cancer, 16.5% had average or lower risk of breast cancer according to the Breast Cancer Risk Assessment Tool (Gail score). Most were postmenopausal (66%), parous with a first full-term pregnancy <30 y of age (79%), never-smokers (56%) with body mass indexes (BMIs) of $<29.9\,kg/m^2$ (70%). Few (5%) reported any cancer prior to enrollment.

Conclusions: The Sister Study is a unique cohort designed to efficiently study environmental and genetic risk factors for breast cancer. Extensive exposure data over the life-course and baseline specimens provide important opportunities for studying breast cancer and other health outcomes in women. Collaborations are welcome. https://doi.org/10.1289/EHP1923

Introduction

Breast cancer is the leading (non-skin) cancer in U.S. women, with over 240,000 diagnoses of invasive breast cancer and 40,000 deaths estimated to have occurred in 2016 (SEER-NCI 2016). As the U.S. population ages, and more women enter the decades with the highest breast cancer incidence [median age at diagnosis = 62 y of age (SEER-NCI 2016)], these numbers are expected to rise. Both invasive and *in situ* breast cancer can lead to significant morbidity and health care resource utilization (Feiten et al. 2014; Fontes et al. 2016; Scott et al. 2016; Tian et al. 2013). Known risk factors explain little of the variation in breast cancer risk, and heritability is modest (Ford et al. 1995; Mucci et al. 2016).

Responding to public concerns, we proposed a novel approach to the study of environment and breast cancer. At that time, there were already many large U.S. cohort studies of women's health generally or breast cancer specifically (Belanger et al. 1978; Colditz and Hankinson 2005; Hays et al. 2003; Russell et al. 2001; Women's Health Initiative Study Group 1998). Although these studies collected breast cancer incidence data, many did not collect biological samples from the full cohort and few focused on non-lifestyle environmental factors. Large population-based case–control studies (Gammon et al. 2002) did focus on environmental exposures, collecting both environmental data and biological samples, but retrospective studies are subject to bias in assessing exposures or biologic measurements that might change following breast cancer diagnosis or treatment. Thus, we saw the need for a large prospective study focused on environmental and genetic drivers of breast cancer risk, a necessity for studying a disease with complex etiology and potentially long time course between relevant exposures and clinical disease (Swerdlow et al. 2011; Weinberg et al. 2007). The Sister Study, a prospective study of 50,884 U.S. women who have had at least one sister diagnosed with breast cancer but had no personal history of breast cancer at enrollment, was designed to fill this gap. The study was not designed around one particular *a priori* hypothesis. The primary objective was to create a resource from which to study current and emerging hypotheses regarding environmental and genetic risk factors for breast cancer. By collecting data on a wide range of potential risk factors, including commonly studied and novel exposures, along with environmental and biological samples, we hoped to create a framework from which to think more broadly about environmental causes and gene–environment interactions. The approach was premised on the general paradigm that by studying genetic and environmental factors in a cohort of women with enhanced risk we would be much more likely than in previous studies to identify preventable risk factors for breast cancer (Weinberg et al. 2007). There were a number of specific and broadly defined environmental factors of interest at study initiation, including vitamin D, light at night, hormone replacement therapies, diet, pesticides, solvents, air pollution, personal care products that may contain endocrine disruptors, environmental tobacco smoke, organochlorines, and exposure to medical hyperstimulation of the ovaries (see Table S1). The prospective design allows us to assess these and other

Address correspondence to D.P. Sandler, Epidemiology Branch, NIEHS, P.O. Box 12233, M.D. A3-05; 111 T.W. Alexander Dr., Research Triangle Park, NC 27709-2233 USA. Telephone: (919) 541-4668. Email: Dale.Sandler@nih.gov

Supplemental Material is available online (https://doi.org/10.1289/EHP1923).

The authors declare they have no actual or potential competing financial conflicts.

Received 22 March 2017; Revised 9 October 2017; Accepted 26 October 2017; Published 20 December 2017.

exposures before disease onset, avoiding biases inherent in case–control studies. In addition, it was recognized that hypotheses of interest at the time the study was initiated might no longer be of interest by the time sufficient cases accrued. Thus, the prospective design with ongoing data collection also creates a framework for addressing future hypotheses as science advances over the follow-up period (see Table S1), and for studying health outcomes other than breast cancer.

Assembling and following a cohort of women who have had a sister diagnosed with breast cancer provided two advantages. The sisters of women with breast cancer are at approximately twice the risk of breast cancer as the general female population (Collaborative Group on Hormonal Factors in Breast Cancer 2001), allowing faster case accrual than in similar sized cohorts that are not enriched by family history. The increased risk in sisters could be due to enhanced genetic susceptibility, shared environmental risk factors, or some combination of the two. The prevalence of multiple gene variants related to breast cancer is expected to be higher in a cohort recruited based on sister history (Weinberg et al. 2007). This may confer increased power for finding environmental factors that interact with genetic factors, as demonstrated mathematically by Weinberg et al. (2007) and illustrated in a recent analysis of polycyclic aromatic hydrocarbon (PAH) exposure, familial risk, and breast cancer (Shen et al. 2017). The prevalence of relevant environmental exposures may also be increased to the extent that sisters share similar experiences, further enhancing statistical power (Weinberg et al. 2007).

Second, in-depth information on exposures over the life-course collected at baseline, along with high follow-up rates over time, are critical to the success of any long-term study. Sisters of women diagnosed with breast cancer potentially provide a very motivated and engaged cohort, enhancing retention and allowing us to collect data on commonly studied factors such as reproductive history, hormone use, and diet as well as less well-studied occupational and environmental exposures.

The Sister Study cohort has now matured to the point where there are sufficient cancer and non-cancer outcomes for etiologic studies. The purpose of this paper is to describe the study methods, which may be useful to others planning new cohorts, and to describe the baseline characteristics of Sister Study participants. Future publications will compare cohort participant characteristics to those of women in the National Health and Nutrition Examination Survey (NHANES), a nationally representative sample of United States, as well as provide additional details on the Sister Study biorepository.

Methods

The Sister Study is a long-term prospective cohort of women residing in the United States (including Puerto Rico) who have had a sister diagnosed with breast cancer but did not have breast cancer themselves at enrollment. Interest in risk factors for breast cancer drove design decisions, but the cohort is also appropriate for studies of other cancer and noncancer health outcomes and, through extended follow-up of all participants, for studies of cancer survivors.

The institutional review board (IRB) of the National Institute of Environmental Health Sciences and the Copernicus Group IRB approved the study. All participants provided written consent. Data included in this report come from Sister Study Data Release 5.0.1 (August 2015), unless otherwise noted.

Eligibility

Women residing in the United States, including Puerto Rico, were eligible for the Sister Study if they were 35–74 y of age, had a sister (full or half) diagnosed with primary breast cancer, and had not themselves ever had a diagnosis of ductal carcinoma *in situ* (DCIS) or invasive breast cancer. A history of cancer other than breast was not considered a basis for exclusion. Women with a prophylactic mastectomy were considered eligible given that they are still at risk for breast cancer, albeit very low risk, and preventive surgeries were documented. Special efforts were made to maximize inclusion of typically underrepresented women including nonwhite women, older women, and women of lower socioeconomic status.

Recruitment

Women were recruited in a "Vanguard" run-in phase in selected cities (Phoenix, AZ, Providence, RI, Tampa, FL, and St. Louis, MO) beginning in July 2003. In 2004, recruitment was expanded from the four pilot cities to the four states in which each city was located. Then, following a national press release in October 2004, recruitment was extended to all 50 U.S. states and Puerto Rico. Enrollment of women in most demographic groups ended 1 April 2008, but continued through March 2009 for underrepresented groups (i.e., African American, Latina, Asian American, Native American, less than a college degree, ≥65 y of age). To enroll a diverse group of women with different educational levels, job exposures, and ages, recruitment was multifaceted, ranging from community-based local efforts to nationally endorsed campaigns, as described in detail in Appendix A in the Supplemental Material.

Because there are no lists of women with a sister with breast cancer, recruitment targeted women more broadly, including the general population and breast cancer survivors who could lead us to eligible women. The approaches most often used were *a*) word-of-mouth and flyer distribution through breast cancer support and advocacy groups, women's volunteer organizations, enrolled Sister Study participants and contacts made at local and national women's events; *b*) outreach through hospitals, mammography centers, churches, unions, and trade organizations; and *c*) direct mail, mass emails, and media outlets (television, web, radio, newspapers, and magazines).

Recruitment materials included brochures and flyers in English and Spanish, some with tailored messaging for women in trades, older women, and women of various races and ethnicities. A brief video featuring early study participants and their sisters was also distributed. Giveaways such as fans, pins, notepads, and logo magnets with contact information also helped promote the study. These materials were provided to recruitment staff, volunteers, and organizations, along with study talking points and sample newspaper articles for local media.

A direct mail and email campaign also targeted minorities, seniors, and women in trades. For example, postcards were mailed to African-American women using a purchased list of confirmed addresses, and emails were sent to women receiving *Essence*, *People en español*, *Blacks N LA*, and *Las Comadres*, and to a commercial list of Asian-American women.

We worked with a wide range of breast cancer and minority advocacy organizations to promote the study (see Appendix A in the Supplemental Material). The Sister Study principal investigator (PI) promoted the study in presentations to women's groups, minority health organizations, and other groups, seeking input on design and research questions. Sister Study recruiters distributed study materials at local and national conferences and trade shows (e.g., the National Hair Show in Atlanta, Georgia) and made materials available to local volunteers, partner organizations, hospitals, mammography centers, minority-focused groups, and breast cancer support groups. The *Dr. Susan Love Research Foundation* selected the Sister Study as the first study for which

Table 1. Baseline Sister Study data collection.

Data component	Details
First study contact	Per IRB stipulations, web-screened women were required to call the study to sign up. Data collected: age, race/ethnicity, sister, and personal breast cancer status.
Two-part computer-assisted telephone interview (CATI1 and CATI2)	Required baseline activity. Usually scheduled on different days. For same-day interviews, minimum 15-min break scheduled between CATI1 and CATI2. Data collected in CATI1: age, race/ethnicity, education; income; family size; personal history of cancer, *BRCA1/2* screening; sisters' history of cancer(s) including breast; overall health, screening behavior, breast conditions; environmental exposures and residential history including farm exposures; physical activity; sun exposure; smoking; alcohol use; sleep patterns. Questions on exposures and lifestyle factors addressed childhood and adulthood. Data collected in CATI2: occupational history and exposures including 19 occupation-specific modules; age at menarche; pregnancy and fertility data; hormone use (contraceptives and hormone therapy); medical conditions and medications; current height and weight as well as during 30s and teens and at $\sim$10 y of age.
Biometrics (examiner form)	Part of required home visit. Data collected: height, weight, hip and waist circumference; blood pressure and pulse.
Biologic and environmental specimens	Part of required home visit. Specimens collected by examiner (phlebotomist): blood (saliva for DNA if blood could not be collected). Specimens collected by participant: first morning urine; toenail clippings; dust swabs of home environment.
Past 24 h questionnaire (self-administered; paper)	This questionnaire was included in mailed kit to be filled out for the 24 h just before the examiner visit; completed questionnaire given to home visit examiner (phlebotomist). Data collected: medications, smoking, alcohol, and chemical exposures in the 24 h before blood collection; environmental exposures in the weeks preceding blood collection.
Other questionnaires (self-administered; paper) • Family history • Food frequency (Block 98) • Personal care products	Included in the study kit mailed to participants; completed questionnaires usually given to home visit examiner but could be returned later by participant. Data collected: Family history questionnaire: participants' birth characteristics and mother's pregnancy experiences (*in utero* exposures); cancer history of first-degree relatives and others; noncancer medical conditions in biological family members. Dietary questionnaire: frequency and amount of foods consumed in the last 12 months; meal patterns; complementary and alternative medicines. Personal care products questionnaire: current and childhood (10–13 y of age) use of products such as makeup, moisturizers, other creams and lotions, acne-related products, skin lighteners or tanners, wrinkle-reducing products, talc, douches, hair care products and dyes, nail care products, mouthwash, deodorant, antiperspirant.

the *Love/Avon Army of Women* (https://www.armyofwomen.org/) would help recruit participants.

We used free and paid media as well as celebrity endorsements. We distributed press releases and media kits including talking points, ads, and newsletter text. Brief mentions and feature articles appeared in magazines such as *Woman's Day, Ladies Home Journal, Essence, and People*. Articles in AARP magazines and bulletins (English and Spanish) reached older women. The study PI and participant volunteers appeared live and via remote satellite on local and national morning news programs and conducted radio interviews and media tours. Radio public service announcements in English and Spanish also were distributed. Radio campaigns featured on-air mentions by radio station personalities. Novel approaches included billboards and bus ads in selected cities and a media campaign with Reach Media's *Tom Joyner Morning Show*, a popular national radio show aimed at the African-American community, which included live on-air mentions by Tom Joyner, on-air interviews, and a web campaign on his *Black America Web* site.

Enrollment

Interested women completed a brief eligibility questionnaire via website or telephone call. Women found to be eligible through website screening were asked to confirm interest by making a telephone call to the study. Those who agreed to enroll were mailed study kits containing self-administered questionnaires, consent documents, support materials for the telephone interview and home visit, and supplies and instructions for collecting urine, toenail, and house-dust samples. A follow-up interview confirmed kit receipt, reviewed materials, and explained what the participant could expect at the home visit. Eligible women had to complete a two-part computer-assisted telephone interview (CATI1 and CATI2) and a home visit to be considered enrolled in the cohort. Women who signed up for the study but did not complete required baseline activities are being followed for mortality and cause of death via the National Death Index (NDI), provided they completed at least one telephone interview.

Data collection. An overview of baseline Sister Study data collection is provided in Table 1. This table includes data collection components and their corresponding details.

Questionnaires. Trained interviewers administered the two-part telephone interview (CATI1 and CATI2) in either English or Spanish. The interview, which took about 2 h to complete overall, collected information on breast cancer risk factors, residential history, medical history, lifetime occupational history, reproductive history, socioeconomic status, and other information, including sister history of breast cancer (https://sisterstudy.niehs.nih.gov/english/baseline.htm and Table S1). The questionnaires were longer than those in other cohort studies to allow for collection of information on commonly studied known and potential risk factors as well as to collect data on occupational and environmental exposures that were not being collected in most other prospective studies.

Environmental and occupational exposures of interest included but were not limited to chemicals previously identified as mammary carcinogens or endocrine disruptors (Bennett and Davis 2002; Rudel et al. 2007) and shift work; we asked about history of working in industries and occupations where exposure to these factors was possible as well as exposures at home, such as pesticides, paints, or hobby materials, and gardening. In addition to the time of enrollment, questions focused on periods that may be relevant to breast cancer risk, including *in utero* and childhood exposures, particularly around menarche. Addresses for current, longest adult, and longest childhood residence have been geocoded for linkage with various GIS databases for environmental exposures, such as air pollution, and census data for socioeconomic and neighborhood factors.

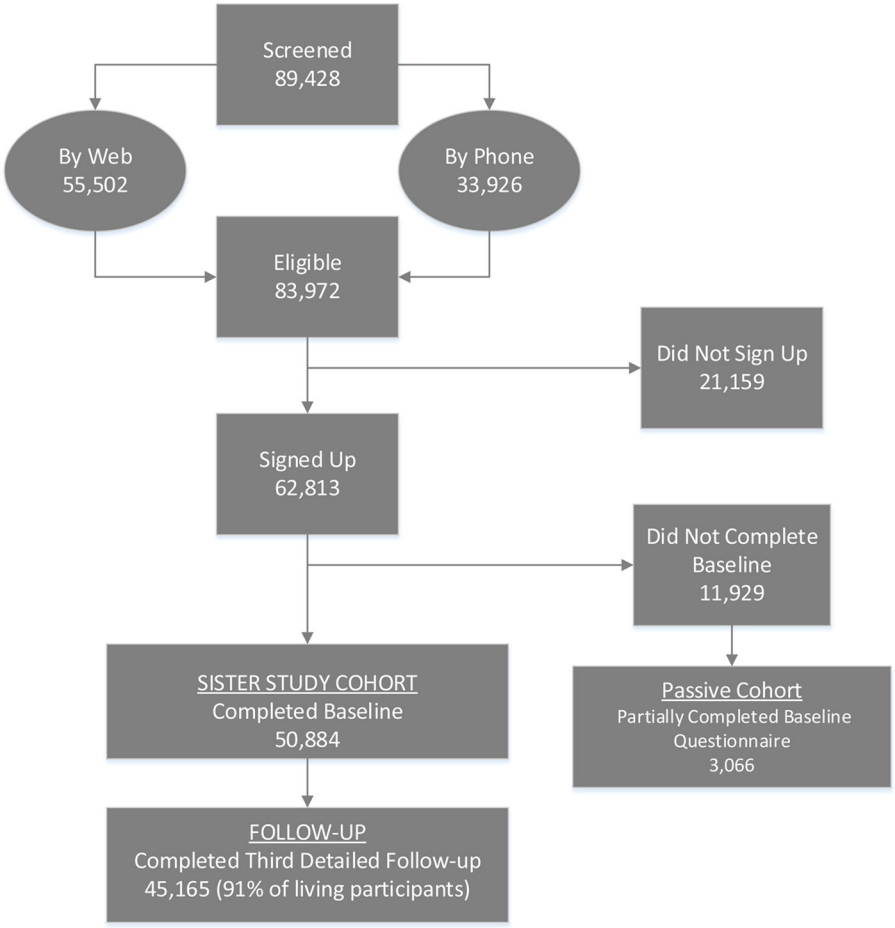

Figure 1. Sister Study cohort enrollment and retention.

Participants completed self-administered questionnaires on diet, personal care products, family history of cancer, and early-life exposures, including the participant's mother's exposures during her pregnancy with the participant. The food frequency questionnaire (Block 98) (Boucher et al. 2006) was supplemented with questions about cooking practices, dietary intake of phytoestrogens, childhood diet, vitamin supplements, and complementary and alternative medicines and practices.

Home visit. During a 45-min visit to participants' homes (or in rare instances another site such as a doctor's office), female examiners from a national in-home phlebotomy service [Examination Management Services, Inc. (EMSI)] collected fasting blood samples, anthropometry data (height, weight, and waist and hip circumference), and blood pressure following standardized study protocols. Ahead of the home visit, participants were sent detailed instructions and materials for all specimen collections: first morning void urine (~60 mL phthalate-free cup), toenail clippings (all; polish-free), and house dust (six alcohol wipes and plastic bag; two wipes each for the tops of three door frames in different rooms). Written consent was obtained prior to collecting biological samples. Participants were asked to record the date of their last menstrual period, the use of medications and hormones, smoking, and alcohol in the 24 h prior to the home visit, and information about the self-collected urine, dust, and toenail samples. Examiners retrieved self-administered questionnaires and participant-collected biological and environmental samples.

Examiner-drawn blood samples totaled ~45 mL and were collected in six Becton Dickinson Vacutainer® tubes. These included two EDTA tubes (one purple top, one metal-free tan top), two red-top serum tubes, and two yellow-top ACD-B tubes. Red-top tubes were centrifuged in the field and serum and clot separated prior to shipping. Serum was transferred to an amber tube to protect from sunlight. Using custom-designed multi-compartment Styrofoam packaging, urine and serum were shipped cold, and whole blood and clot were shipped at ambient temperature to a central laboratory (Social & Scientific Systems, Inc., Durham, NC) for processing and storage. In the event of unsuccessful blood collection, saliva was collected for DNA analyses (Oragene DNA self-collection saliva kit; DNA Genotek, Ottawa, ON, Canada). All samples were barcoded with participant ID prior to shipping.

Specimen processing. Upon receipt at the central laboratory, any evident adverse conditions and examiner errors were documented. Kit contents were scanned and inventoried, and daily reports on examiner performance were fed back to EMSI; 92.6% of kits were received at the central laboratory within 24 h of collection.

Serum was stored in 0.5 mL CryoBioSystem™ (CBS™) straws in liquid nitrogen (LN) vapor phase. Blood clots were stored in −80°C freezers and LN vapor phase. EDTA whole blood was stored in a cryovial, and spotted (60 µL per spot) and stored on two types of dry blood storage cards: a card chemically impregnated to lyse cells and stabilize DNA (Whatman FTA Classic Card) and an untreated card (Whatman 903 Protein Saver Card). Remaining EDTA whole blood was centrifuged and the plasma was stored in 0.5-mL CBS™ straws in LN vapor phase. A 3.0-mL EDTA BD Vacutainer® tube (tan top, metal free) was stored untouched at −20°C for future analysis of metals, trace elements, and environmental contaminants. One ACD-B whole

Table 2. Sociodemographic characteristics of Sister Study participants and incomplete enrollees (passive cohort) at baseline, 2003–2009.

Characteristic	Study participants		Passive cohort	
	n	(%)[a]	*n*	(%)[a]
Participants	50,884	100.0	3,066	100.0
Year of enrollment				
2003	843	1.7	24	0.8
2004	13,297	26.1	569	18.6
2005	7,712	15.2	389	12.7
2006	7,161	14.1	411	13.4
2007	13,717	27.0	970	31.6
2008	6,616	13.0	538	17.6
2009	1,538	3.0	165	5.4
Age at baseline (y)				
35–39	2,100	4.1	189	6.2
40–44	4,479	8.8	435	14.2
45–49	7,703	15.1	606	19.8
50–54	9,817	19.3	657	21.4
55–59	10,109	19.9	547	17.8
60–64	7,803	15.3	333	10.9
65–69	5,824	11.4	190	6.2
70–74	3,049	6.0	109	3.6
Median (range) (y)	55.6 (25.0–76.5)		52.1 (35.1–78.5)	
Race/ethnicity				
Non-Hispanic white	42,558	83.7	1,939	63.4
Non-Hispanic black	4,462	8.8	723	23.6
Hispanic	2,515	4.9	293	9.6
Other	1,334	2.6	106	3.5
Unknown or missing	15		5	
Marital status				
Never married	2,759	5.4	264	8.6
Divorced/separated	7,550	14.8	700	22.9
Widowed	2,564	5.0	167	5.5
Legally married	35,870	70.5	1,748	57.1
Living as married	2,127	4.2	180	5.9
Missing	14		7	
Education				
Less than high school	627	1.2	72	2.4
High school/GED	7,178	14.1	542	17.7
Some college, no degree	9,957	19.6	751	24.6
Associate or technical degree	7,224	14.2	531	17.4
Bachelor's degree	13,714	27.0	672	22.0
Master's degree	10,103	19.9	398	13.0
Doctoral degree	2,069	4.1	92	3.0
Missing	12		8	
Household income				
<$20,000	2,296	4.7	247	8.5
$20,000–49,999	10,284	21.0	775	26.7
$50,000–99,999	19,907	40.7	1,082	37.3
$100,000–200,000	12,868	26.3	625	21.5
>$200,000	3,534	7.2	174	6.0
Missing	1,995		163	
Household Size				
1	8,991	17.7	667	21.9
2	24,228	47.7	1,157	38.1
3	7,216	14.2	497	16.3
4	6,782	13.4	468	15.4
≥5	3,548	7.0	252	8.3
Missing	119		25	
Mean (range)	2.47 (1–20)		2.55 (1–11)	
Household members				
<18 y of age				
0	37,318	73.5	1,940	63.8
1	6,249	12.3	508	16.7
2	5,005	9.9	399	13.1
≥3	2,187	4.3	194	6.4
Missing	125		25	
Mean (range)	0.46 (0–11)		0.65 (1–7)	
U.S. Census region				
Northeast	8,532	16.8	500	16.6
Midwest	13,689	26.9	649	21.5
South	16,743	32.9	1,215	40.2
West	11,010	21.7	607	20.1
Puerto Rico	883	1.7	51	1.7
Missing	27		44	

[a]Total percentages may not always equal 100% due to missing values and rounding.

blood Vacutainer® tube (yellow top) was aliquotted and cryopreserved with 10% DMSO (dimethylsulfoxide) using a freezer that reduces the temperature to −80°C in preprogrammed steps to improve cell viability. Twelve percent of the time, this ACD-B tube was selected for lymphocyte isolation under an alternative protocol, with selection based on an algorithm that oversampled women from a high-risk group based on age of enrollment and the affected sister's age at diagnosis. The buffy coat (lymphocytes) was isolated from the whole blood, washed, resuspended, and stored in LN vapor phase. For urine, a basic chemistry urinalysis (Multistix Pro 10LS reagent strips) was performed immediately upon receipt to measure protein, creatinine, blood, leukocytes, nitrite, glucose, ketone, pH, and specific gravity (Bayer Clinitek 500). Urine was aliquoted into twenty 0.5-mL CBS™ straws, five 1.0-mL vials, and one 3.6-mL vial. The straws were stored in LN vapor phase and vials in −80°C mechanical freezers. Toenails were stored in envelopes at ambient temperature, with the large toenails kept separate from all other toenails. Participants collected dust from three locations in their home using prepackaged alcohol wipes, which were stored at −20°C after receipt. Samples from a single individual were stored across multiple freezers, and extensive quality assurance measures were put in place to track and document conditions for each sample. See https://sisterstudy.niehs.nih.gov/English/specimen.htm for further details.

Follow-up

Participants are contacted each year for either a short (~two pages) annual update questionnaire or a detailed follow-up questionnaire (approximately every 2–3 y; two to three booklets of 20–30 pages

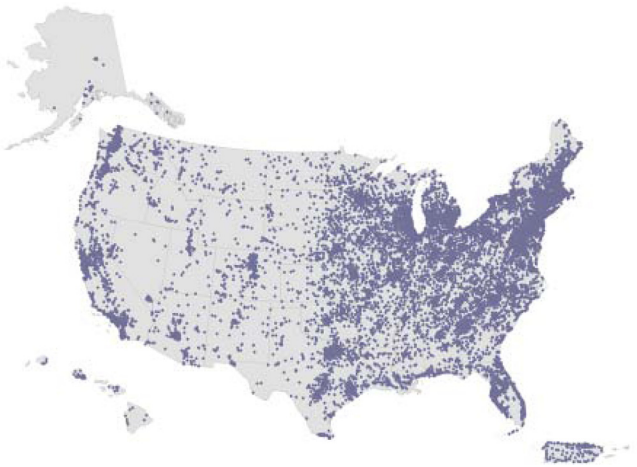

Figure 2. Sister Study participants' residence at enrollment (ArcGIS, version 10.3.1, Esri; U.S. Census map, 2013 Cartographic Boundary File, State for United States, http://www2.census.gov/geo/tiger/GENZ2013/STATE/cb_2013_us_state_500k.zip).

Table 3. Health and lifestyle characteristics of Sister Study participants at baseline, 2003–2009.

Characteristic	*n*	(%)[a]
Participants	50,884	100.0
Smoking status		
Never	28,552	56.1
Current	4,175	8.2
Former	18,141	35.7
Missing	16	
Alcohol status		
Nondrinker (never/former)	9,679	19.1
Light (≤3 drinks per week)	27,615	54.4
Moderate (>3–7 drinks per week)	7,878	15.5
Heavy (>7 drinks per week)	5,625	11.1
Missing	87	
BMI (kg/m^2)[b]		
<18.5	563	1.1
18.5–24.9	18,875	37.1
25.0–29.9	16,151	31.8
30.0–34.9	8,800	17.3
35.0–39.9	3,959	7.8
≥40.0	2,519	5.0
Missing	17	
Median (range)	26.6 (11.5–72.1)	
Personal history of cancer other than breast or NMSC		
No pre-baseline cancer reported	48,238	94.8
Ovarian cancer	167	0.3
Other cancer	2,479	4.9
Missing		

Note: BMI, body mass index; NMSC, non-melanoma skin cancer.
[a]Total percentages may not always equal 100% due to missing values and rounding.
[b]99.3% of BMI based on examiner data; remaining is based on self-reported height and weight.

each). Questionnaires are typically offered on the web first, then on paper, followed by telephone contact and a CATI questionnaire. To ensure maximum retention and response rates, there is a comprehensive prompting protocol for nonresponders at each stage, which can include email, postal mail, and/or telephone, as appropriate to the participant's contact history. Women who are more vulnerable to nonresponse (e.g., women with a history of slow response) are assigned a personal study advocate who regularly reaches out to them to encourage completion of study activities and help with prioritization of tasks.

Follow-up questionnaires include updates on menopausal status and health, including incident breast cancer, updates on exposures, and new exposures of interest. Women reporting breast cancer are asked to provide additional details and permission to retrieve medical records and paraffin-embedded tumor tissue blocks. Pathology reports or authorization to retrieve pathology reports are requested following report of other cancers. In 2014, a second home visit (including a second blood draw) was completed for a subset of 2,461 Sister Study participants (breast cancer cases and a random sample of the cohort).

Results

Response

As shown in Figure 1 and Table S2, nearly 90,000 people completed an eligibility questionnaire through the website (62.1%) or by telephone (37.9%). Very few women who completed a web screener were found to be ineligible for the study during their subsequent enrollment call, whereas 16.0% of those who only telephoned were ineligible. In all, 62,813 women provided verbal consent to join the study. Of the women who signed up, 81.0% ($n = 50,884$) completed required baseline activities and were enrolled; an additional 3,066 who completed some, but not all, of the required baseline activities before the enrollment end date. They are being followed through mortality linkage (and possibly cancer registry linkage) as a "passive cohort."

Most commonly, participants completed at least one telephone interview prior to home exam, however 18.7% of participants had their home visit prior to completing CATI1; 38.0% completed it after CATI1 and before CATI2. The median time between CATI1 and home visit blood draw for all women was 22 d (interquartile range, 10–44 d).

The vast majority of participants also completed all four self-administered questionnaires and provided all biological samples, along with written consent to use their samples (see Table S3). Women in the passive cohort most often completed only CATI1 [although some also completed CATI2 (41%) or provided biospecimens], completing home exams prior to CATI completion. Biological samples from these women were anonymized for pilot studies.

Among participants, 40.2% reported their primary referral source was a sister with breast cancer or some other family member or friend (see Table S4). Print materials such as magazines and newspapers were the next most cited (32.3%). All direct mail and email efforts yielded enrollees; however, success rates varied. Emails sent through a trusted source such as *Essence* or the Susan Love Army of Women were much more successful than an unendorsed email to women from a purchased list (data not shown). Endorsements were also successful, particularly those of TV newswoman Robin Roberts and Luisa Gándara, wife of the Governor of Puerto Rico, resulting in hundreds of African-American enrollees and over a thousand Latina enrollees (data not shown).

Baseline Characteristics

The median age of Sister Study participants was 55.6 y (range, 35.0–76.5 y) at completion of all required enrollment activities (Table 2). They were predominantly non-Hispanic white (83.7%) and married or living as married (74.7%). The vast majority had some college education (84.7%), with just over 50% having a bachelor's degree or higher. Educational attainment was high across all race/ethnicity groups. The fraction that were nonwhite decreased with older age (see Figure S1). Participants were generally well off; approximately two-thirds reported having a total annual household income of between $50,000 and $200,000. Household size was generally small and only a quarter (26.5%) reported children (<18 y of age) in the household at time of enrollment. Residential distribution across the United States and Puerto Rico is shown in Figure 2; women from all 50 U.S. states and the District of Columbia participated.

As seen in Table 2, women in the passive cohort were more likely than full participants to be younger, of a race/ethnicity other than non-Hispanic white, unmarried, have less than a bachelor's degree, and have a household income of less than $50,000; they were also more likely to have children in the household.

Over half of participants never smoked and only 8.2% were current smokers (Table 3). Light/moderate alcohol consumption was common, with only 19.1% reporting no current alcohol consumption. Approximately two-thirds of participants had BMI in the normal or overweight range (37.1% and 31.8%). Although never having been diagnosed with DCIS or breast cancer was a study requirement, 59 women were diagnosed sometime before completing their final enrollment activity and were retained in the cohort. Approximately 5% of Sister Study participants reported having been diagnosed with other cancers prior to enrollment.

Table 4. Reproductive characteristics, screening, and breast conditions among Sister Study participants at baseline, 2003–2009.

Characteristic	*n*	Prevalence (%)[a]
Participants	50,884	100.0
Reproductive		
Age at menarche (y)		
Early (<12)	10,405	20.5
Typical (12–13)	28,525	56.1
Late (≥14)	11,908	23.4
Missing	46	
Median (range)	13.0 (5–40)	
Parity (number of births)		
0 (nulliparous)	9,246	18.2
1	7,442	14.6
2	18,868	37.1
3	10,124	19.9
4	3,482	6.9
≥5	1,687	3.3
Missing	35	
Median (range) (parous only)	2 (1–12)	
Age at first full term pregnancy (y)[b]		
<20	6,137	15.3
20–24	15,349	38.4
25–29	11,251	28.1
30–34	5,204	13.0
≥35	2,081	5.2
Parous, missing age	1,579	
Median (range) (parous only)	24 (10–54)	
Menopausal status[c]		
Premenopausal	17,513	34.4
Postmenopausal	33,338	65.5
Natural	25,209	49.5
Surgical or other (i.e., medical)	7,938	15.6
Unknown	32	0.1
Missing	1	
Ever used hormonal birth control[d]		
Yes	43,121	85.2
Missing	277	
Ever used hormone therapy		
Yes	22,932	45.2
Missing	159	
Screening		
Breast		
Timing of most recent mammogram		
<1 y ago	40,929	80.5
1–2 y ago	7,500	14.7
>2 y ago	1,875	3.9
Never had mammogram	566	1.1
Missing	14	
Non-breast screening		
Most recent physical exam		
<6 mo ago	20,056	41.3
From 6 mo to 1 y ago	19,702	40.6
>1 but <2 y ago	5,328	11.0
2–5 y ago	2,440	5.0
>5 y ago or never	1,030	2.1
Missing	2,328	
Ever had colon-/sigmoidoscopy		
Yes	31,610	62.2
Missing	21	
Pap smear in past 12 mos		
Yes	38,910	76.5
Missing	41	
Breast procedures		
Ever had needle biopsy of the breast		
Yes	10,332	22.8
Missing	5,603	
Ever had any other type of breast biopsy[e]		
Yes	9,316	20.6
Missing	5,611	
Ever had a lumpectomy		
Yes	6,922	13.7
Missing	243	
Ever had prophylactic mastectomy		
Yes	238	0.5
Missing	9	

[a]Total percentages may not always equal 100% due to missing values and rounding.
[b]Among parous women only.
[c]Participant is postmenopausal if no menstrual period in the last 12 mos or had any qualifying medical intervention that caused menstrual periods to cease. Qualifying interventions include both ovaries removed; chemo/radiation that stopped periods; hysterectomy, ablation, or embolization and ≥55 y of age; ovarian suppressing drugs or contraception that eliminated menstrual flow and ≥55 y of age.
[d]Hormonal birth control includes birth control pills or patches, Norplant implants, Depo-Provera injections, IUD containing hormones.
[e]Includes participants who have had a surgical biopsy other than a needle biopsy (e.g., excisional biopsy).

As can be seen in Table 4, most participants had menarche at 12–13 y of age, most were parous (median number of births, 2), completed their first full term pregnancy in their 20s (median age, 24 y), and had used hormonal birth control (85.2%). Most were postmenopausal (65.5%) with 15.6% reporting surgical, medical, or other (nonnatural) types of menopause. Nearly half had used some form of hormone therapy. Prevalence of health screenings was high; virtually all participants had had at least one mammogram and 80.5% reported having had one within the previous year.

There were no restrictions on the number of sisters within a family that could join the Sister Study. We identified 4,318 sibships with more than one sister in the cohort through linkage using birth dates and other familial details, comprising 18.8% of the cohort.

Although the vast majority of participants (95.8%) had at least one full biologic sister who had been diagnosed with breast cancer as of enrollment (Table 5), women with half-sister(s) with breast cancer were also eligible. At enrollment, most participants had a single sister with breast cancer (89.8%); 18.7% had a mother with breast cancer. Over half of participants (57%) had a first-degree female relative (full sister, mother, or daughter) with young onset (<50 y of age at diagnosis) disease. Those families were targeted for a family-based "Two Sister Study" (Fei et al. 2012). Just under 4% had a first-degree family history of ovarian cancer.

Although most women had a Gail score in the high-risk range at enrollment, 16.5% had a 5-y risk score below 1.67%, the National Cancer Institute cutoff for defining high risk. Approximately two-thirds had a lifetime Gail score of <20%, another cut point used to indicate high risk (American Cancer Society 2016; Graubard et al. 2010). We did not ask directly about Ashkenazi Jewish heritage, but 22% reported Eastern European ancestry (data not shown). Few reported testing for *BRCA1* or *BRCA2* (3.1%). Of these, 17.3% ($n=256$) reported being told they had a mutation in a known breast cancer gene.

Cohort Retention and Response Rates

Response rates (n of responses during field period/n assumed alive at start of field period, where n of responses = n of completed questionnaires + n of deceased during field period) for the first three short updates (i.e., annual update questionnaires) were 96.3%, 95.6%, and 94.0%, respectively. Response rates for the first three detailed follow-ups were only slightly lower at 94.9%, 92.1%, and 91.0% (completed August 2016) despite the significantly longer questionnaires (see the Sister Study website for follow-up questionnaires: https://sisterstudy.niehs.nih.gov/English/fu-data.htm). As of July 2017, 1,643 (3.2%) participants are known to be deceased and 1,716 (3.4%) of the 50,884 women enrolled in the Sister Study have withdrawn (i.e., requested no further study contact—including just 2 participants who requested that their

Table 5. Familial risk factors for breast and/or ovarian cancer in Sister Study participants at baseline, 2003–2009.

Characteristic	*n*	Frequency (%)[a]
Participants	50,884	100.0
First-degree female relatives with breast cancer[b]		
0	1,709	3.4
1	36,377	71.5
≥2	12,795	25.1
Missing	3	
First-degree female relatives with young-onset breast cancer[b,c]		
0	21,563	43.0
1	25,910	51.6
≥2	2,703	5.4
Missing	708	
Sisters with breast cancer		
1 half-sister, no full sisters	1,905	3.8
1 full sister, no half-sisters	43,736	86.1
1 sister (unknown if half or full)	28	0.1
2 half-sisters	136	0.3
2 sisters (1 full and 1 half)	275	0.5
2 full sisters	4,096	8.1
2 sisters (1 or both half or unknown)	7	0.01
≥3 sisters (half, full, or unknown)	629	1.2
Missing	72	
Mother had breast cancer		
Yes	9,135	18.7
Missing	1,915	
First-degree relatives with ovarian cancer[b]		
0	48,858	96.1
≥1	1,994	3.9
Missing	32	
Mother had ovarian cancer		
Yes	1,157	2.4
Missing	1,917	
First-degree family history of breast and ovarian cancer		
Yes	1,940	3.8
No	48,912	96.2
Missing	32	
Ever been tested for *BRCA1* or *BRCA2*		
Yes	1,551	3.1
No	49,094	96.9
Missing	239	
Told you have a mutation in one of the breast cancer genes		
Yes	256	17.3
No	1,223	82.7
Missing	72	
Gail score		
5-y absolute risk		
≤1.66%	8,371	16.5
≥1.67%	42,399	83.5
Missing	114	
Median (range)	2.8 (0.2–14.3)	
Lifetime (90 y) absolute risk		
<15%	17,174	33.8
15–<20%	17,107	33.7
≥20%	16,489	32.5
Missing	114	
Median (range)	17.1 (2.4–61.2)	

[a]Total percentages may not always equal 100% due to missing values and rounding.
[b]Does not include half-sisters (second-degree relatives).
[c]Young onset defined as diagnosed at <50 y of age.

data not be included in any new analyses). Thus at least 47,525 (93.4%) women are still actively participating in the cohort.

To date, medical records have been obtained for 81.1% of those reporting an incident breast cancer diagnosis included in Data Release 5 (https://sisterstudy.niehs.nih.gov/English/brca-validation.htm). Using data from an earlier data release (Data Release 4; 82.0% with medical records), we evaluated the validity of a self-reported diagnosis (D'Aloisio et al. 2017). Among those with medical records, the positive predictive value (PPV) was better than 99% for total and invasive cancer. The PPV was also high for self-report of breast cancer subtypes such as ductal cancer (98.8%) and estrogen-receptor positive breast cancer (99.3%) (https://sisterstudy.niehs.nih.gov/English/brca-validation.htm and D'Aloisio et al. 2017). Tumor tissue blocks have been obtained for 1,683 women with incident breast cancer as of Data Release 5.

Discussion

The Sister Study was designed to address concerns that not enough was being done to evaluate the impact of widespread environmental exposure to chemicals on trends in breast cancer incidence. In many ways, the Sister Study design anticipated the subsequent recommendations of commissioned review panels such as the Institute of Medicine (2012) and the Interagency Breast Cancer & Environmental Research Coordinating Committee (2013). We have created a resource to, as these panels suggested, prospectively address questions about genetic factors and environmental exposures during time periods of potentially higher sensitivity to breast cancer induction and/or progression, including gestation, early life, childhood, and the reproductive and perimenopausal years. Furthermore, the extensive and varied data and biospecimens collected is supporting a wide array of mutlidisciplinary research projects.

One of the main difficulties inherent in cohort studies focusing on breast cancer is the need for extensive data collection (reproductive data, known and suspected risk factors, potential confounders and effect measure modifiers as well as biological and environmental samples) while achieving and maintaining high participation and retention rates. The Sister Study approach was to recruit women already disposed to be engaged and motivated by their personal experience with family members' diagnoses to participate in a long-term, detailed data collection effort aimed at better understanding, and possibly preventing, breast cancer. Not coincidentally these same women are, on average, at modestly elevated risk of breast cancer themselves, making them good candidates for a well-powered study, especially given the need to evaluate gene–environment interactions.

Requirements for the Sister Study were substantial—the two-part baseline interview averaged 2 h, and the home visit required as much as another hour of participant time. The number of women contacting the Sister Study (~89,000) is evidence of great interest, and the high proportion of women who signed up that completed all the required baseline activities (81.0%) attests to the level of participant engagement and dedication. Women who did not complete baseline activities were more likely to be nonwhite and younger, with relatively lower income and education levels, and more likely to either live alone or have larger households. Nonetheless, thousands of women with these characteristics did complete the baseline activities.

Great effort was devoted to enhancing the number of nonwhite women participating in the Sister Study; 16.3% of the cohort consider themselves a race/ethnicity other than non-Hispanic white. Although lower than the percentage in the United States, that fraction is higher than in other national cohorts (Hays et al. 2003; VanKim et al. 2017) with the obvious exceptions of the Black Women's Health Study (Russell et al. 2001) and the Multi-Ethnic Cohort (Kolonel et al. 2000), which were designed to target specific racial/ethnic groups. Regardless of the proportion of nonwhite women in the United States and Puerto Rico, as long as the relevant data are collected, these cohort participants can and do serve as the basis for valid conclusions

regarding potential risk and modifying factors within and among differing race/ethnicity groups.

Design Considerations

The Sister Study is a risk-based prospective cohort study (Weinberg et al. 2007). One advantage of this design is the more rapid accrual of incident cases than non–risk-based designs, markedly enhancing power to detect etiologic factors and gene–environment interactions. The increased power is driven, in part, by higher prevalence of potentially relevant environmental risk factors and by increased prevalence of multiple relatively uncommon genetic susceptibility factors, rather than by rare high penetrance genes (Weinberg et al. 2007). Modest enrichment of even unidentified genetic risk factors will also enhance the ability to detect any environmental factors with which they interact (Shen et al. 2017; Weinberg et al. 2007).

Based on age-specific incidence rates of invasive breast cancer from SEER (the National Cancer Institute's Surveillance, Epidemiology, and End Results program), proposed age ranges for recruiting, and an on-average 2-fold increased risk among women with a sister with breast cancer, we estimated that approximately 1,500 cases of invasive breast cancer would accrue in the first 5 y of the Sister Study (300 cases per year). As of Data Release 5 there were 2,163 cases of invasive breast cancer with an average follow-up time in the full cohort of 7.5 y (average number of cases per year = 288). As might be expected in a cohort of women with a sister history of breast cancer, breast screening is common (95% report a mammogram within the 2 y prior to baseline) and most of the invasive cancers (91%) are early stage (Stage 1, 66%; Stage 2, 25%).

At study initiation, there was no consensus on whether lobular carcinoma *in situ* (LCIS) should be considered breast cancer or a risk factor for breast cancer. Consequently, at that time some women diagnosed with LCIS were told they had breast cancer and others were told they did not. In order to avoid confusion to potential participants, LCIS was not considered an exclusion criterion for the Sister Study. Women with LCIS diagnosed prior to enrollment can be handled analytically in a number of ways, depending on study question.

As with other volunteer cohorts, participants in the Sister Study are not entirely representative of the U.S. population, although they do reside in all 50 U.S. states, the District of Columbia, and Puerto Rico. In addition to having a sister with breast cancer, they are generally older and more educated and more likely to be white, relatively economically well off, and healthy. However, population representativeness may not be an appropriate or necessary goal (Rothman et al. 2013). Rather, for a scientific study, as opposed to a study measuring population attributes (e.g., population prevalence of disease or exposures), the goal should be to achieve internal validity and scientific generalizability, not population representativeness (Rothman et al. 2013), something with which it is occasionally conflated. One would expect that whereas exposure prevalence may vary among specific study subgroups, and perhaps influence power to detect a substantive relative risk, the exposure would exert its effect in a mechanistically similar fashion. For the Sister Study, scientific generalizability, as opposed to representativeness, means the ability to draw conclusions about circumstances and mechanisms of disease etiology relevant to women in specific subgroups, as well as (potentially) to women in general. Nonetheless, as with other U.S. cohorts, the degree to which the Sister Study cohort is similar (or dissimilar) to the U.S. population is of interest and may affect interpretation of any findings. Consequently, a manuscript comparing Sister Study participants to women in NHANES, a nationally representative sample of the U.S. population, in terms of classic breast cancer risk factors, lifestyle factors, morbidity, and other factors, is in preparation.

Some might worry that women with a sister history of breast cancer would be overly vulnerable to highly penetrant gene mutations, such as those in *BRCA1* or *BRCA2*. However, such genes are unlikely to dominate breast cancer etiology in the Sister Study cohort (Weinberg et al. 2007). Based on a 2% prevalence of *BRCA*-positive status in breast cancer cases, the prevalence in sisters of breast cancer cases would be about 1%. If the odds ratio is 10 for carriers, one would expect fewer than 8% of the invasive breast cancer cases in the first 5 y of the study to be *BRCA*-positive (Weinberg et al. 2007).

Although any cohort including an element of family history in the selection criteria would be expected to have an elevated risk relative to the general population, we believe it would be a misnomer to characterize the Sister Study as high-risk, a specific term typically defined by cut points. Rather, the Sister Study should be characterized as a cohort with modestly elevated average risk, composed of participants with a wide range of absolute risks. Given the modestly increased risk conferred by a having a sister with breast cancer (approximately 2-fold) and the heterogeneity in risk among those with different family histories (Collaborative Group on Hormonal Factors in Breast Cancer 2001), as well as variation across individuals in established reproductive, lifestyle, and environmental factors, the Sister Study cohort includes participants with a wide range of risk levels. The notable breast cancer risk heterogeneity in the Sister Study is apparent in the distribution of Gail scores. At the lower end of the risk spectrum, 16.5% of the cohort had 5-y Gail scores indicating average or lower risk relative to U.S. women, whereas at the upper end, 32.5% were in a high-risk group (lifetime risk of breast cancer of at least 0.20) (American Cancer Society 2016; Graubard et al. 2010). Only 5.4% had two or more first-degree relatives with young-onset (<50 y of age) breast cancer, whereas three quarters of the Sister Study participants have only a half-sister or single first-degree relative with breast cancer (3.4% and 71.5%, respectively). This argues that the various biologic mechanisms underpinning breast cancer development are well represented in the Sister Study cohort.

Challenges and priorities

One of the first challenges with risk-based sampling is to identify those at elevated risk but without disease. Identifying and connecting with women who have a sister with breast cancer is not straightforward. There are no lists of such women. We considered strategies based on identifying women with breast cancer and recruiting their sisters. Cancer registries proved inefficient and impractical because of the large number of separate cancer registries that would need to be approached, each with its own application and IRB requirements, and the necessity of contacting the potential participants only through the sister with breast cancer rather than directly. Survival bias would also have become a serious issue. Recruiting through cases participating in existing case–control studies was also considered, but few investigators could share needed contact information.

Some studies have restricted recruitment to states with SEER registries to simplify breast cancer identification and case confirmation. However, many of our strategies involved nationwide approaches; restricting enrollment to selected states would have been a disincentive for our many partner organizations. Consequently, we prioritized geographic diversity over the ease of cancer validation.

By focusing on sisters of women with breast cancer, women without sisters are excluded. In fact, women from larger families were likely oversampled because the chance of having a sister with breast cancer rises with the number of sisters in a family.

Because the risk enhancement for women with many sisters but only one with breast cancer will be less than that for women with a single sister, who has breast cancer, it is possible that sibship size could become a confounder in this cohort if relevant risk factors (e.g., parity, age at first birth) are associated with family size. Average family size in the United States was at its highest during the baby boom years (1946–1964) peaking at 3.7 children per woman in 1957, the year that a 50-y-old Sister Study participant enrolled in 2007 would have been born. (CDC 1999)

The universe of eligible women is not known, nor do we know how many women would have known about the opportunity to enroll, but response to recruitment overall was good, with 74.8% of identified eligible women signing up for the study after talking to study staff and learning what would be expected of them. Of note, among women who did the eligibility screener by telephone, 92.1% of eligible women signed up for the study, as opposed to the 65.9% of web-screened eligible women who signed up. Had women been allowed to sign up for the Sister Study online after doing the web screener, rather than needing to subsequently call the Sister Study, it seems likely that a much higher proportion of web-screened women would have enrolled; however, subsequent retention might have been lower.

We received funding for additional outreach to women underrepresented in breast cancer research (nonwhite women, women with lower income and education, and older women). For example, women in the southern region of the United States were heavily recruited and response was good. The proportion of the final cohort from the South Atlantic states (21.6%) was similar to U.S. Census Bureau estimates for 2004 (18.8%) (U.S. Census Bureau, Population Division 2004). Despite focused recruitment efforts in the Mississippi Delta region (of special interest to the Institute of Minority Health and Health Disparities which provided supplemental funding) being hampered by Hurricane Katrina in 2005, overall 32.9% of the cohort are from the U.S. South, similar to U.S. Census estimates for 2004 (36.1%) (U.S. Census Bureau, Population Division 2004). The recruiting push in southern states helped increase the numbers of nonwhite women (mainly African Americans) in the Sister Study. Although black women and Latina women are included in the cohort in sufficient numbers for some stratified analyses, Asian- and Native American women are not. Because family sizes tend to be smaller for some Asian groups (Pew Research Center 2017), requiring a sister with breast cancer likely resulted in a reduced pool of eligible Asian-American women.

Collaborative opportunities

The Sister Study is a rich resource for collaborative research involving scientists inside and outside of the National Institutes of Health. Collaborations may involve use of existing data or samples, and proposals for add-on studies involving new data collection are considered. Because we also collect information about non-breast cancers, and non-cancerous conditions, the Sister Study also offers the opportunity to evaluate environmental exposures with respect to these outcomes. The prospective nature of the study allows for generation and investigation of new hypotheses for a range of outcomes, including studies of cancer survivors. Procedures for requesting access to study data or for proposing add-on or nested substudies can be found on the study website at https://sisterstudy.niehs.nih.gov.

Conclusions

The Sister Study is a unique cohort designed to efficiently study environmental and genetic risk factors for breast cancer. Its risk-based design affords enhanced statistical power to detect gene–environment interactions. Our goal was to create a resource for studying environmental and genetic contributors to breast cancer and other diseases in women. To date more than 90 papers have been published in the peer-reviewed literature using Sister Study data. The Sister Study has provided the platform for seven extramurally funded grants, including three led by extramural collaborators, and is the basis for collaborative research with the Centers for Disease Control and Prevention (CDC) on survivorship and the impact of a breast cancer diagnosis on family members. In addition, the Sister Study participates in many cohort consortia focused on gene discovery, gene–environment interactions, and lifestyle and environmental risk factors for breast cancer and rare outcomes that cannot be studied in a single cohort. High rates of enrollment and participation over time as well as extensive data collected about exposures over the life-course and baseline biospecimens from nearly all women in the cohort provide many opportunities for studying breast cancer and other health outcomes in women.

Acknowledgments

We thank the Sister Study Research Team for their hard work and dedication: P. Armsby, A. Bilhorn-Janssens, H. Carroll, D. Bittner, R. Jesrani, A. Jones, I. Khodosh, C. Scheier, P. Schwingl, J. Ter Maat, E. Revak, T. Young, and F. Yucel. We also thank all former and current Sister Study staff at Social & Scientific Systems, Inc. and at Westat, Inc.; our many volunteer recruiters and spokespersons for the Sister Study; and the dedicated Sister Study participants. We thank W. Arroyave for table preparation and helpful discussion.

The Sister Study was funded by the Intramural Research Program of the NIH, NIEHS (Z01ES044005), with support from the Institute of Minority Health and Health Disparities, NIH.

Reference

American Cancer Society. 2016. http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs.

Belanger CF, Hennekens CH, Rosner B, Speizer FE. 1978. The Nurses' Health Study. Am J Nurs 78(6):1039–1040, PMID: 248266.

Bennett LM, Davis BJ. 2002. Identification of mammary carcinogens in rodent bioassays. Environ Mol Mutagen 39(2–3):150–157, PMID: 11921183.

Boucher B, Cotterchio M, Kreiger N, Nadalin V, Block T, Block G. 2006. Validity and reliability of the Block98 food-frequency questionnaire in a sample of Canadian women. Public Health Nutr 9(1):84–93, PMID: 16480538, https://doi.org/10.1079/PHN2005763.

CDC (Centers for Disease Control and Prevention). 1999. Achievements in Public Health, 1900-1999: Family Planning. MMWR Morb Mortal Wkly Rep 48(47):1073–1080. https://www.cdc.gov/MMWR/preview/mmwrhtml/mm4847a1.htm [accessed 16 November 2017].

Colditz GA, Hankinson SE. 2005. The Nurses' Health Study: lifestyle and health among women. Nat Rev Cancer 5(5):388–396, PMID: 15864280, https://doi.org/10.1038/nrc1608.

Collaborative Group on Hormonal Factors in Breast Cancer. 2001. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 358(9291):1389–1399, PMID: 11705483, https://doi.org/10.1016/S0140-6736(01)06524-2.

D'Aloisio AA, Nichols HB, Hodgson ME, Deming-Halverson SL, Sandler DP. 2017. Validity of self-reported breast cancer characteristics in a nationwide cohort of women with a family history of breast cancer. BMC Cancer 17(1):692. , PMID: 29058598, https://doi.org/10.1186/s12885-017-3686-6.

Fei C, DeRoo LA, Sandler DP, Weinberg CR. 2012. Fertility drugs and young-onset breast cancer: results from the Two Sister Study. J Natl Cancer Inst 104(13):1021–1027, PMID: 22773825, https://doi.org/10.1093/jnci/djs255.

Feiten S, Dünnebacke J, Heymanns J, Köppler H, Thomalla J, van Roye C, et al. 2014. Breast cancer morbidity: questionnaire survey of patients on the long term effects of disease and adjuvant therapy. Dtsch Arztebl Int 111(31–32):537–544, PMID: 25145512, https://doi.org/10.3238/arztebl.2014.0537.

Fontes F, Pereira S, Castro-Lopes JM, Lunet N. 2016. A prospective study on the neurological complications of breast cancer and its treatment: updated analysis three years after cancer diagnosis. Breast 29:31–38, PMID: 27394676, https://doi.org/10.1016/j.breast.2016.06.013.

Ford D, Easton DF, Peto J. 1995. Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. Am J Hum Genet 57(6):1457–1462, PMID: 8533776.

Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, et al. 2002. The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. Breast Cancer Res Treat 74(3):235–254, PMID: 12206514.

Graubard BI, Freedman AN, Gail MH. 2010. Five-year and lifetime risk of breast cancer among U.S. subpopulations: implications for magnetic resonance imaging screening. Cancer. Cancer Epidemiol Biomarkers Prev 19(10):2430–2436, PMID: 20841391, https://doi.org/10.1158/1055-9965.EPI-10-0324.

Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. 2003. The Women's Health Initiative recruitment methods and results. Ann Epidemiol 13(9 suppl):S18–S77, PMID: 14575939.

Institute of Medicine. 2012. Breast Cancer and the Environment: A Life Course Approach. Washington, DC:National Academies Press.

Interagency Breast Cancer & Environmental Research Coordinating Committee. 2013. Breast Cancer and the Environment: Prioritizing Prevention, Durham, NC: National Institute of Environmental Health Sciences, https://www.niehs.nih.gov/about/assets/docs/breast_cancer_and_the_environment_prioritizing_prevention_508.pdf.

Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. 2000. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 151(4):346–357, PMID: 10695593.

Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. 2016. familial risk and heritability of cancer among twins in Nordic countries. JAMA 315(1):68–76, PMID: 26746459, https://doi.org/10.1001/jama.2015.17703.

Pew Research Center. 2017. Analysis of 1986–2014 Current Population Survey June Supplement Data. http://www.pewsocialtrends.org/2015/05/07/childlessness-falls-family-size-grows-among-highly-educated-women/st_2015-05-07_childlessness-12/ [accessed 16 November 2017].

Rothman KJ, Gallacher JE, Hatch EE. 2013. Why representativeness should be avoided. Int J Epidemiol 42(4):1012–1014, PMID: 24062287, https://doi.org/10.1093/ije/dys223.

Rudel RA, Attfield KR, Schifano JN, Brody JG. 2007. Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. Cancer 109(12 suppl):2635–2666, PMID: 17503434, https://doi.org/10.1002/cncr.22653.

Russell C, Palmer JR, Adams-Campbell LL, Rosenberg L. 2001. Follow-up of a large cohort of Black women. Am J Epidemiol 154(9):845–853, PMID: 11682367.

Scott JM, Adams SC, Koelwyn GJ, Jones LW. 2016. Cardiovascular late effects and exercise treatment in breast cancer: current evidence and future directions. Can J Cardiol 32(7):881–890, PMID: 27343744, https://doi.org/10.1016/j.cjca.2016.03.014.

SEER-NCI (National Cancer Institute–Surveillance, Epidemiology, and End Results Program). 2016. Cancer Statistics. http://seer.cancer.gov/statfacts/html/breast.html [accessed 16 November 2017].

Shen J, Liao Y, Hopper JL, Goldberg M, Santella RM, Terry MB. 2017. Dependence of cancer risk from environmental exposures on underlying genetic susceptibility: an illustration with polycyclic aromatic hydrocarbons and breast cancer. Br J Cancer 116(9):1229–1233, PMID: 28350789, https://doi.org/10.1038/bjc.2017.81.

Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, et al. 2011. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. Br J Cancer 105(7):911–917, PMID: 21897394, https://doi.org/10.1038/bjc.2011.337.

Tian Y, Schofield PE, Gough K, Mann GB. 2013. Profile and predictors of long-term morbidity in breast cancer survivors. Ann Surg Oncol 20(11):3453–3460, PMID: 23702642, https://doi.org/10.1245/s10434-013-3004-8.

U.S. Census Bureau, Population Division. 2004. Table 8: Annual Estimates of the Population for the United States, Regions, and Divisions: April 1, 2000 to July 1, 2004 (NST-EST2004-08). https://www2.census.gov/programs-surveys/popest/tables/2000-2004/state/totals/nst-est2004-08.pdf [accessed 16 November 2017].

VanKim NA, Austin SB, Jun HJ, Hu FB, Corliss HL. 2017. Dietary patterns during adulthood among lesbian, bisexual, and heterosexual women in the Nurses' Health Study II. J Acad Nutr Diet 117(3):386–395, PMID: 27889314, https://doi.org/10.1016/j.jand.2016.09.028.

Weinberg CR, Shore DL, Umbach DM, Sandler DP. 2007. Using risk-based sampling to enrich cohorts for endpoints, genes, and exposures. Am J Epidemiol 166(4):447–455, PMID: 17556763, https://doi.org/10.1093/aje/kwm097.

Women's Health Initiative Study Group. 1998. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials 19:61–109, PMID: 9492970.