Sister Study Breast Cancer Events (any type) [1] and Characteristics: Data Release 9.0 [2]

	N	% of cohort
Total Cases	**4,150**	**8.2**
Breast Cancer Subtype for 1st Diagnosis	**N**	**% of cases**
Invasive [3]	3,108	74.9
Ductal only (no lobular component) [4]	2,185	52.7
Lobular only (no ductal component) [4]	270	6.5
Ductal and lobular [4,5]	211	5.1
Inflammatory [6]	6	0.1
Paget's [7]	10	0.2
Phyllodes [8]	5	0.1
Other/unknown if ductal or lobular [9]	421	10.1
In situ	993	23.9
Ductal only (no lobular component) [10]	824	19.9
Lobular only (no ductal component) [10]	126	3.0
Ductal and lobular [5,10]	18	0.4
Paget's [7]	5	0.1
Other/unknown if ductal or lobular [11]	20	0.5
Invasiveness unknown [12]	49	1.2
Stage at 1st Breast Cancer Diagnosis [13]		
0	887	24.1
I	1,874	50.9
II	713	19.4
III	182	4.9
IV	23	0.6
Unknown, missing, cannot stage	*471*	
Estrogen Receptor (ER) Status for 1st Breast Cancer Diagnosis [13]		
Positive	2,924	84.7
Negative	523	15.1
Borderline	7	0.2
Unknown, missing, not performed	*696*	
Progesterone Receptor (PR) Status for 1st Breast Cancer Diagnosis [13]		
Positive	2,409	72.5
Negative	906	27.3
Borderline	8	0.2
Unknown, missing, not performed	*827*	
HER2 Amplification Status for 1st Breast Cancer Diagnosis [13]		
Positive	385	13.4
Negative	2,439	85.0
Borderline	44	1.5
Unknown, missing, not performed	*1,282*	
Combined ER, PR, HER2 Status [13]		
Triple negative (ER-, PR-, and HER2-)	294	8.8
Not triple negative	3,065	91.2
Unknown, missing, not performed	*791*	

Characteristic	N (total = 4,150)	% of cases
Diagnosis Confirmation [14]		
Medical	3,293	79.4
NDI Plus or death certificate only	45	1.1
None	812	19.6
Race/Ethnicity [15]		
Non-Hispanic White	3,555	85.7
Non-Hispanic Black	328	7.9
Hispanic	156	3.8
Other	110	2.7
Missing	*1*	
Age (in years) at 1st Breast Cancer Diagnosis [16]		
Under 50	422	10.2
50-64	1,904	45.9
65 and older	1,824	44.0
Menopausal Status at 1st Breast Cancer Diagnosis		
Pre-menopausal	663	16.0
Post-menopausal	3,469	84.0
Unknown	*18*	

Abbreviations: BrCa= Breast cancer; DCIS=Ductal carcinoma *in situ*; LCIS = Lobular carcinoma *in situ*; MR = Medical report (medical record and/or pathology report)

[1] Breast cancer events (any type) are defined by the following ICD10 codes: D05 (includes lobular carcinoma *in situ*, ductal carcinoma *in situ*, other and unspecified *in situ* breast carcinomas), C50 (all invasive breast cancer), and D48.6 (neoplasm of uncertain behavior of breast)

[2] Data release 9.0 includes events from follow-up phases that closed **on or before 9/30/2019**, excluding all data from participants who have withdrawn their data from the study (n=3). To be considered a breast cancer event, breast must be the primary site, and the cancer diagnosis must have occurred after the participant's baseline date. Consequently, Data Release 9.0 breast cancer events (any type) exclude a) through c) below:
   a) Breast cancers diagnosed prior to participant's baseline date. [Total (n=113): Invasive ductal (n=31); invasive lobular (n=1), invasive ductal and lobular (n=1), invasive, unknown if ductal or lobular (n=5), DCIS (n=19), LCIS (n=54), *in situ* unknown if ductal or lobular (n=1), and invasiveness unknown (n=1)];
   b) Breast cancers diagnosed with unknown timing relative to baseline [Total (n=33): Invasive, unknown if ductal or lobular (n=11), DCIS (n=10), LCIS (n=9), *in situ* ductal and lobular (n=1), *in situ* unknown if ductal or lobular (n=1), and invasiveness unknown (n=1)];
   c) Unknown primary cancer diagnoses where breast was unlikely to be the primary site, but the possibility could not be ruled out (n=5)

[3] Includes 'presumed invasive' (i.e. a carcinoma *in situ* with positive nodes) (n=2)

[4] Ductal only, Lobular only, and Ductal and Lobular exclude inflammatory breast cancer and Paget's

[5] Includes BrCa events with a) tumors with mixed ductal and lobular characteristics and/or b) multiple tumors of different histology (e.g. a ductal only tumor and a lobular only tumor)

[6] Inflammatory – Ductal (n=5), Ductal and Lobular (n=1), Lobular (n=0)

[7] Paget's - Invasive: Ductal (n=9), Lobular (n=1); *in situ*: Ductal (n=5), Lobular (n=0)

[8] Phyllodes is neither ductal nor lobular; it is a sarcoma

[9] Includes: a) Invasive but unknown if ductal or lobular (n=415), these are typically self-report without MR confirmation; b) Invasive type "neither ductal nor lobular, not otherwise specified" (n=4), these are typically self-report without MR confirmation; and c) Per MR, presumed invasive - *in situ* tumor, positive nodes (n=2)

[10] *In situ* Ductal only, *in situ* Lobular only, and *in situ* Ductal and Lobular exclude Paget's

[11] Includes: a) *In situ* but unknown if ductal or lobular (n=18); b) *In situ* type "neither ductal nor lobular, not otherwise specified" (n=2); these are typically self-report without MR confirmation

[12] Invasiveness unknown includes: a) Unknown if invasive or *in situ* and unknown if ductal or lobular (n=27); b) Ductal but unknown if invasive or *in situ* (n=10); c) Lobular but unknown if invasive or *in situ* (n=1); d) Neither ductal nor lobular (other) but not otherwise specified and unknown if invasive or *in situ* (n=7); e) Both ductal and lobular but unknown if invasive or *in situ* (n=1); and f) Per MR, unable to determine if DCIS or atypical ductal hyperplasia (n=3)

[13] Priority for stage and ER/PR/HER2 status when diagnosis is reported by multiple sources: Medical record>Breast Cancer Follow-Up Form

[14] Medical confirmation of reported breast cancer is generally from medical records but may only include pathology report; National Death Index (NDI) Plus or death certificate confirmation is specific to cases that do not have medical confirmation

[15] Race/ethnicity: precedence given to Hispanic ethnicity first, then Black race, then Other race, then White

[16] For those missing age at diagnosis (n=29), age was imputed as the midpoint of the range of plausible ages at diagnosis