Sister Study Breast Cancer Events (any type) [1] and Characteristics: Data Release 10.1 [2]

	N	% of cohort
Total Cases	**4,427**	**8.7**
Breast Cancer Subtype for 1st Diagnosis	N	% of cases
Invasive [3]	3,372	76.2
Ductal only (no lobular component) [4]	2,348	53.0
Lobular only (no ductal component) [4]	296	6.7
Ductal and lobular [4,5]	223	5.0
Inflammatory [6]	8	0.2
Paget's [7]	12	0.3
Phyllodes [8]	5	0.1
Other/unknown if ductal or lobular [9]	480	10.8
In situ	1,015	22.9
Ductal only (no lobular component) [10]	858	19.4
Lobular only (no ductal component) [10]	124	2.8
Ductal and lobular [5,10]	16	0.4
Paget's [7]	5	0.1
Other/unknown if ductal or lobular [11]	12	0.3
Invasiveness unknown [12]	40	0.9
Stage at 1st Breast Cancer Diagnosis [13]		
0	918	23.7
I	1,981	51.1
II	758	19.6
III	198	5.1
IV	22	0.6
Unknown, missing, cannot stage	*550*	
Estrogen Receptor (ER) Status for 1st Breast Cancer Diagnosis [13]		
Positive	3,133	84.7
Negative	557	15.1
Borderline	7	0.2
Unknown, missing, not performed	*730*	
Progesterone Receptor (PR) Status for 1st Breast Cancer Diagnosis [13]		
Positive	2,600	72.6
Negative	972	27.2
Borderline	7	0.2
Unknown, missing, not performed	*848*	
HER2 Amplification Status for 1st Breast Cancer Diagnosis [13]		
Positive	397	12.9
Negative	2,635	85.7
Borderline	41	1.3
Unknown, missing, not performed	*1,354*	
Combined ER, PR, HER2 Status [13]		
Triple negative (ER-, PR-, and HER2-)	304	8.5
Not triple negative	3,279	91.5
Unknown, missing, not performed	*844*	

Characteristic	N (total = 4,427)	% of cases
Diagnosis Confirmation [14]		
Medical	3,635	82.1
NDI Plus or death certificate only	45	1.0
None	747	16.9
Race/Ethnicity [15]		
Non-Hispanic White	3,785	85.5
Non-Hispanic Black	361	8.2
Hispanic	167	3.8
Other	113	2.6
Missing	*1*	
Age (in years) at 1st Breast Cancer Diagnosis [16]		
Under 50	424	9.6
50-64	1,993	45.0
65 and older	2,010	45.4
Menopausal Status at 1st Breast Cancer Diagnosis		
Pre-menopausal	686	15.5
Post-menopausal	3,740	84.5
Unknown	*1*	

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 10.1 includes events from follow-up phases that closed **on or before 10/12/2020**, excluding all data from participants who have withdrawn their data from the study (n=4). 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 10.1 breast cancer events (any type) exclude a) through c) below:

    a) Breast cancers diagnosed prior to participant's baseline date. [Total (n=115): Invasive ductal (n=32); invasive lobular (n=1), invasive ductal and lobular (n=1), invasive - unknown if ductal or lobular (n=5), DCIS (n=20), LCIS (n=54), and *in situ* - unknown if ductal or lobular (n=2)];

    b) Breast cancers diagnosed with unknown timing relative to baseline [Total (n=25): Invasive - unknown if ductal or lobular (n=10), DCIS (n=6), LCIS (n=8), and *in situ* ductal and lobular (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=1)

[4] Invasive ductal only, Invasive lobular only, and invasive 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=7), Ductal and Lobular (n=1), Lobular (n=0)

[7] Paget's - Invasive: Ductal (n=11), 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=475), 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=1)

[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=10) and 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=23); b) Ductal but unknown if invasive or *in situ* (n=8); c) Lobular but unknown if invasive or *in situ* (n=1); d) Neither ductal nor lobular but not otherwise specified and unknown if invasive or *in situ* (n=6); and e) Per MR, unable to determine if DCIS or atypical ductal hyperplasia (n=2)

[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 only; 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=33), age was imputed as the midpoint of the range of plausible ages at diagnosis