

The below data dictionary is for the EA1141 analysis dataset: “ea1141_bcsc_risk_model.sas7bdat”.

This dataset contains one record per participant registered, and is unique by SUBJECT_DE. There are a total of **11** contained in this dataset.

Variable name	Variable definition	Code Table/Values ‘Num’ signifies numeric variable ‘Char’ signifies character/text variable Otherwise, variables are numeric with coded values as indicated
SUBJECT_DE	De-identified case number (<u>numeric</u>).	Num
BCSC_PRIORHISTORYBC	Prior to study enrollment, is there a history of breast cancer (lumpectomy or mastectomy) for the patient? [taken from the variable PREV_HX_BC_PREENROLL from the Patient History form] (<u>numeric, code table</u>).	.M=Missing 0=No 1=Yes 99=Unknown
BCSC_AGE	Participant age at study enrollment [taken from the AGE derived variable from the derived “EA1141_baseline_demog_derived.sas7bdat” dataset. Refer to the data dictionary for that dataset.] (<u>numeric</u>).	Num
BCSC_RACEETHN	Participant race/ethnicity as required for the BCSC risk calculator [taken from the RACE and ETHNICITY derived variables from the “EA1141_baseline_demog_derived.sas7bdat” dataset. Refer to the data dictionary for that dataset.] (<u>numeric, code table</u>). <i>NOTE 1: Refer to the BCSC website and the (https://tools.bcsc-scc.org/BC5yearRisk/intro.htm), and the %bcscrisk_v2() SAS macro (BCSC source code available from the website). Same comment applies to below variables.</i> <i>NOTE 2: Case (5075836) reported two records: one “Black or African American”, the other “Not reported”. Given this, RACE was set to 9</i>	1=White, non-Hispanic 2=Black, non-Hispanic 3=Asian, Native Hawaiian or Pacific Islander 5=American Indian or Alaskan Native 8=Hispanic 9=Other, Mixed (2+ races), or Unknown (treated the same as White, non-Hispanic)
BCSC_FIRSTDEGREE	Family history of breast cancer in a 1 st degree relative (mother, sister, or daughter) [taken from FAM_HX_BRST_CAN_YN from the Genetic and Medical History form, and from FAM_MEM_DX and CK_DX from the Family History of Breast Cancer form.] (<u>numeric, code table</u>).	0=No 1=Yes 9=Missing (no adjustment for missing family history of cancer)

BCSC_FIRSTDEGREEOVARIAN	<p>Family history of ovarian cancer in a 1st degree relative (mother, sister, or daughter) [taken from FAM_HX_OV_CAN_YN from the Genetic and Medical History form, and from FAM_MEM_OVAR_DX and CK_DX_OVAR from the Family History of Ovarian Cancer form.] (numeric, code table).</p> <p>NOTE: This variable is <u>NOT</u> part of the BCSC risk model. However, it was requested by the investigator to summarize patient risk, and so I include it in this dataset to keep variables together.</p>	<p>0=No 1=Yes 9=Missing</p>
BCSC_PBB	<p>History of a breast biopsy (core biopsy, excisional biopsy, or fine-needle aspiration) with benign disease, if known [taken from the derived variables PRIOR_BENIGN_BIOP, PRIOR_BIOP_ADH, PRIOR_BIOP_ALH, and PRIOR_BIOP_LCIS from the derived “EA1141_baseline_demog_derived.sas7bdat” dataset. Refer to the data dictionary for that dataset.] (numeric, code table).</p> <p>NOTE: The Patient History form only captured the number of prior benign breast biopsies, and whether the patient had a prior biopsy with ADH, ALH or LCIS (the last 3 as separate elements). Thus, if the form marked that the patient had at least 1 prior benign biopsy, but none of ADH, ALH or LCIS elements were marked, this was coded 3: Prior biopsy but diagnosis unknown. If instead LCIS was marked this was coded 7:LCIS; if LCIS was not marked, but ADH or ALH was marked, this was coded 6:Proliferative with atypia.</p>	<p>1=Unknown if prior biopsy (self-report missing & nothing in path) (no adjustment for missing result of biopsy) 2=No prior biopsy 3=Prior biopsy but diagnosis unknown 4=Non-proliferative 5=Proliferative without atypia 6=Proliferative with atypia (ADH, ALH) 7=LCIS</p>
BCSC_DENSITY	<p>BIRADS breast density [taken from the derived variable PRIOR_BREAST_DENSITY from the derived “EA1141_baseline_demog_derived.sas7bdat” dataset, and the derived variable TOMO_BREAST_DENSITY_YR0 from the EA1141_year0_screening_derived.sas7bdat dataset.. Refer to the data dictionary for that dataset.] (numeric, code table).</p> <p>NOTE: If the year 0 Tomo was performed, breast density from the year 0 Tomo was used (TOMO_BREAST_DENSITY_YR0); otherwise, breast density from the most recent prior mammogram was used (PRIOR_BREAST_DENSITY).</p>	<p>1=Almost entirely fat 2=Scattered fibroglandular densities 3=Heterogeneously dense 4=Extremely dense 9=Missing (risk not calculated)</p>
BCSC_REASON	<p>Reason BCSC risk score could not be calculated. (numeric, code table).</p>	<p>1=Age > 74 2=Previous diagnosis of breast cancer .N=Risk score was computed</p>

BCSC_5YEAR	<p>5-year risk of developing invasive breast cancer, as calculated by the BCSC risk model. (<u>numeric</u>).</p> <p>NOTE 1: This is output from the %bcscrisk_v2() SAS macro. Refer to the source code from the BCSC.</p> <p>NOTE 2: If BCSC_REASON in(1,2) then BCSC_5YEAR=.N.</p>	<p>Num</p> <p>.N=Risk score could not be computed</p>
BCSC_10YEAR	<p>10-year risk of developing invasive breast cancer, as calculated by the BCSC risk model. (<u>numeric</u>).</p> <p>NOTE 1: This is output from the %bcscrisk_v2() SAS macro. Refer to the source code from the BCSC.</p> <p>NOTE 2: If BCSC_REASON in(1,2) then BCSC_10YEAR=.N.</p>	<p>Num</p> <p>.N=Risk score could not be computed</p>