

**Description of SUMP75_ANALYSIS file for
NSABP B-35 Primary Results Paper
Description updated 4/13/2017**

Variable Name	Definition
Patient Identification Number (PATIENTID)	
Participant Race (RACE)	1=White 2=Black 3=Pacific Islander 4=Asian 5=Native American/Alaskan 6=Multiracial 7=Unknown (Note: If participant selected more than one race, she was categorized as Multiracial. The following individual race variables can be used to determine further detail of the Multiracial group.)
Participant Race is White (RACE_WHITE)	0=No 1=Yes
Participant Race is Black (RACE_BLACK)	0=No 1=Yes
Participant Race is Pacific Islander (RACE_PACIFIC_ISLANDER)	0=No 1=Yes
Participant Race is Asian (RACE_ASIAN)	0=No 1=Yes
Participant Race is Native American/Alaskan (RACE_NATIVE_AMERICAN)	0=No 1=Yes
Participant Race is Unknown (RACE_UNKNOWN)	0=No 1=Yes
Participant Ethnicity (HISPANIC)	0=Non-Hispanic 1=Hispanic or Latino 99=Unknown
Age at Randomization (AGE)	Difference between date of birth and date of randomization, in integer years
Participant Weight at Entry (WT)	Tenths of kilograms
Participant Height at Entry (HT)	Centimeters
Hysterectomy prior to Entry (HYST)	0=No 1=Yes Note: This is an underestimate because it only includes women that had a hysterectomy and no bilateral oophorectomy. So it does not count those who had both.

Variable Name	Definition
Treatment Group (TRT)	1=Tamoxifen 2=Anastrozole
Comedo Necrosis (COMEDO)	0=Absent 1=Present (Blank/missing values are unknown)
Tumor Palpable (TUMORP)	0=No 1=Yes (Blank/missing values are unknown)
Tumor Evident on Mammogram (EVMAMM)	0=No 1=Yes (Blank/missing values are unknown)
Pathological Tumor Size (PPTS)	Tenths of centimeters (Blank/missing values are unknown)
Vital Status (DEATH)	0=Alive 1=Dead
Underlying Cause of Death (COD)	ICD8 CODES (Blank/missing if patient is alive; 9999 indicates unknown cause of death)
Time to Last Contact (TLC)	Difference between date of randomization and date of last contact, regardless of the source of contact. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month.
Time to Last Physical Contact (TLPC)	Difference between date of randomization and date of last physical contact. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month.
Any Breast Cancer Event (BCFI)	0 = None 1 = Invasive IBTR 2 = Non-Invasive IBTR 3 = IBTR Not Otherwise Specified 4 = Other Local/Regional Recurrence 5 = Distant Recurrence 6 = Invasive Contralateral Breast 7 = Non-Invasive Contralateral Breast 8 = Contralateral Breast Not Otherwise Specified 9 = Breast Second Primary (Not Ipsilateral Recurrence Or Contralateral) If more than one event was diagnosed simultaneously, the preference for assignment is as follows: CBT, DIST, IBTR, LOCREG, BREAST 2nd PRIM
Any Breast Cancer Event Indicator (BC)	0=No breast cancer 1=Breast cancer
Time to Breast Cancer Event (TBC)	Difference between date of randomization and date of the first breast cancer event; else difference between date of randomization and date of last physical contact or death. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month.

Variable Name	Definition
Disease-Free Survival Indicator (DFS)	0=No disease-free survival event 1=Event (includes any breast cancer, non-breast second primary, and death)
Time to First Disease-Free Survival Event (TFE2)	Difference between date of randomization and date of the first disease-free survival event; else difference between date of randomization and date of last physical contact. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month.
Non-breast Second Primary as a First Event (NB2P)	0=No non-breast second primary cancer 1=Non-breast second primary cancer
Uterine Cancer (UTERINE2P)	0=None 1=Uterine Cancer
Time to Uterine Cancer (TUTERINE2P)	Difference between date of randomization and date of diagnosis for uterine cancer; if no event, this is the time to last physical contact. If patient had a hysterectomy pre-entry then this is set to 0. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month.
Site of Non-breast Second Primary (SITOP2)	ICD8 code for site of second primary cancer (Blank/missing if no second primary; 9999 indicates unknown site of second primary)
Description of Site of Non-breast Second Primary (SITOP2DESCR)	Text description of site of second primary cancer (Blank/missing if no second primary)
First Osteoporotic Fracture (FRAC)	0=None 1=Hip, wrist, or spine fracture
Time to First Osteoporotic Fracture (TFRAC)	Difference between date of randomization and date of fracture. If no event, this is the time to last contact, TLC. If FRAC=1 and the date of fracture is unknown, then TFRAC is missing. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month.
Date of Form file Download (DFUP)	Date of download of form files
Date of Summary File (DREP)	DREP is the cutoff date for events included in this summary file.
Body Mass Index (BMI)	Calculated value as weight in kilograms divided by height in meters-squared.
Thromboembolic Events (THROM)	Maximum grade reported per patient. Missing means no adverse event information available. 0=Grade 0/1 (none/superficial thrombosis) 2=Grade 2 (deep-vein thrombosis) 3=Grade 3 (uncomplicated pulmonary embolism) 4=Grade 4 (life-threatening pulmonary embolism) 5=Grade 5 (death)

Variable Name	Definition
Arthralgia (ARTHRL)	Maximum grade reported per patient. Missing means no adverse event information available. 0=Grade 0/1 (none/mild pain) 2=Grade 2 (moderate pain) 3=Grade 3 (severe pain) 4=Grade 4 (disabling)
Myalgia (MYALG)	Maximum grade reported per patient. Missing means no adverse event information available. 0=Grade 0/1 (none/mild pain) 2=Grade 2 (moderate pain) 3=Grade 3 (severe pain) 4=Grade 4 (disabling)
Overall Toxicity (OVERG)	Maximum grade reported per patient. Missing means no adverse event information available. 0=Grade 0/1 2=Grade 2 3=Grade 3 4=Grade 4 5=Grade 5 (death)
Stopped treatment early for reasons other than disease progression or death (OFF)	0=No 1=Yes
Time to off treatment (TOFF)	Difference between date of first treatment and the last date the patient went off therapy. Time interval is expressed as the number of months and is calculated to the nearest 10 th of a month. A value of 0 indicates patient never started treatment.
Treatment Status (TRTSTATUS)	1 = Treatment completed 2 = Toxicity/side effects/complications 3 = Patient withdrawal or refusal after beginning trt 4 = Patient withdrawal or refusal prior to beginning trt 5 = Disease progression, relapse during trt 6 = Disease progression, relapse prior to beginning trt 7 = Death after beginning protocol therapy 8 = Death prior to beginning protocol therapy 9 = Alternative therapy 10 = Patient off-treatment for other complicating disease 11 = Consent Withdrawal 12 = Stopped Therapy b/c site closed & not reassigned 13 = Lost to Follow-up 98 = Other