

## E2197T5 Clinical Data Description

The data in the spreadsheet are the analysis data for the MetaSite Breast™ score analysis published in

Sparano JA, Gray R, Oktay MH, Entenberg D, Rohan T, Xue X, Donovan M, Peterson M, Shuber A, Hamilton DA, D'Alfonso T, Goldstein LJ, Gertler F, Davidson NE, Condeelis J, Jones J. A metastasis biomarker (MetaSite Breast™ Score) is associated with distant recurrence in hormone receptor-positive, HER2-negative early-stage breast cancer. *NPJ Breast Cancer*. 2017 Nov 8;3:42

The cases used are a 600-patient subset of the 776 patients previously analyzed for Oncotype DX Recurrence Score and expression of other genes (Goldstein et al, Prognostic utility of the 21-gene assay in hormone receptor–positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol* 2008; 26:4063-71. PMID:PMC2654377). Those 776 cases were a stratified case-control sample from the parent E2197 protocol (NCT00003519). Because different proportions of subjects were sampled for recurrences and nonrecurrences within strata, the sampling fractions need to be taken into account to make valid inferences. The variable weightmeta in the spreadsheet gives the inverse sampling fractions within strata and recurrence combinations, which are appropriate weights to use for some types of analyses. (The groups for the weights are defined using the recurrence status at the time the original sample was selected, and the weights have been updated to reflect the actual numbers in the groups for the 600 patients analyzed here.)

Patients on E2197 were only followed for disease recurrence until the first occurrence of a breast cancer event (evidence of recurrent disease or a new breast primary). The variable rfi in the spreadsheet is the time from entry on E2197 to the first breast cancer event (recurrence or new primary breast cancer), or to the date of the last documented free of recurrence (for patients without a recurrence). rfind=1 indicates that the time in rfi is a recurrence event, and drfind=1 indicates that the time in rfi is a distant recurrence. The analysis of Distant Recurrence-Free Interval in the paper used rfi as the time to event/censoring variable and drfind as the event indicator. The analysis of breast cancer specific survival used survtime as the time to event/censoring variable and the indicator of cause=2 as the event indicator.

### Data File Description

Field	Name	Description	Coding
1	blindid	Case ID number	numeric code, up to 7 digits
2	rxarm	Assigned Treatment	A = AT, B = AC
3	race	Patient's race/ethnicity	1=white, 2=Hispanic, 3=black 4=Asian, 6=Native American, 9=unknown/not reported
4	age	Age at randomization on E2197	years (integer); range 25-85
5	meno	Menopausal status at randomization on E2197	1=premenopausal, 2=postmenopausal
6	nndpos	# positive axillary lymph nodes	count (0, 1, 2, 3)
7	tumsz	Maximum diameter of the primary tumor	value in millimeters (integer)
8	tumszg	Tumor size category	Character: '(0,20]' is less than or equal to 20mm, '(20,50]' is greater than 20mm and less than or equal to 50mm, '(50,100]' is greater than 50mm (the max is actually 85mm)

9	surgproc	Primary surgical procedure	1=tumorectomy, 2=mastectomy
10	grade	Histologic grade from Central Review	Character: Well = Low, Moderate = Intermediate, Poor = High
11	CentralERPos	Estrogen Receptor Status based on central evaluation	0 = Negative, 1 = Positive, 2, 3 = Central evaluation not performed; case classified as negative (2) or positive (3) based on local results and/or gene expression
12	CentralPRPos	Progesterone Receptor Status based on central evaluation	0 = Negative, 1 = Positive, 2, 3 = Central evaluation not performed; case classified as negative (2) or positive (3) based on local results and/or gene expression
13	CentralHER2Pos	HER2 Status based on central evaluation	0 = Negative, 1 = Positive, 2, 3 = Central evaluation not performed; case classified as negative (2) or positive (3) based on local results and/or gene expression
14	SubType	Breast Cancer Subtype	1 = ER-neg, PR-neg, HER2-neg 2 = PR-pos or ER-pos, HER2-neg 3 = HER2-pos
15	Recurrence	Oncotype DX Recurrence Score	Numeric 0-100
16	RecG	Grouped Recurrence Score (RS)	1 = RS < 18 2 = RS 18 to 30 3 = RS > 30
17	RecT	Grouped Recurrence Score using TAILORx study cutoffs	'[0-11]' = RS < 11 '[11-25]' = RS 11 to 25 '(25,100]' = RS > 25
18	MetaSite	MetaSite score	Integer 0 to 199
19	metag3	Grouped MetaSite score with tertile cutoffs	'[0,5]' = 0 to 5 '(5,17]' = greater than 5 and less than or equal to 17 '(17,199]' = greater than 17
20	weightmeta	Analysis weights (inverse sampling fractions)	Numeric
21	strat	Strata used for sampling subjects from the E2197 study	Integer 1 = Arm A, ER and PR neg, Node-neg 2 = Arm B, ER and PR neg, Node-neg 3 = Arm A, ER or PR pos, Node-neg 4 = Arm B, ER or PR pos, Node-neg 5 = Arm A, ER and PR neg, Node-pos 6 = Arm B, ER and PR neg, Node-pos 7 = Arm A, ER or PR pos, Node-pos 8 = Arm B, ER or PR pos, Node-pos (ER and PR were defined from local

			evaluation, so do not always agree with CentralERPos and CentralPRPos)
22	recind1	Recurrence indicator from original sample creation	1 = recurrence reported at time of sampling 0 = no recurrence at time of sampling (Note: during the intervening years, 4 of these cases were subsequently determined not to be recurrences)
23	rfi	Recurrence-free interval: time from randomization to first recurrence or to last documented free of recurrence	Numeric (years)
24	rfind	Recurrence indicator	1 = recurrence (rfi is time of recurrence) 0 = no recurrence
25	drfind	Distant recurrence indicator	1 = distant recurrence (rfi is time of distant recurrence) 0 = no distant recurrence
26	survstat	Survival status	1 = dead 0 = alive
27	survtime	Time from randomization to death or date last known alive	Numeric (years)
28	cause	Cause of death	0 = Alive 2 = Breast cancer (includes cases with cause unknown/not reported with prior distant recurrence) 3 = Other