

## **Pathology Report Abstract Data Dictionary**

**SAS dataset = Path\_Abstract\_07212023.sas7bdat**

**N=2998**

The data in this file is one record per subject. Data are derived from multiple pathology reports. Only selected variables are included.

For subjects with multiple first primary tumors, we take the positive/highest value from each tumor.

### **Note:**

For all variables from each pathology report, please refer to SAS dataset "**Path\_Abstract\_Rpt\_07212023.sas7bdat**" and codebook "**CodeBook Exp Path rept w ERS data entry rev 05 03 2012.xls**".

Variable	Description	Comment
STUDYID	CBCS study ID	
ER_STS	ER status Positive Negative Weak Positive / Borderline Indeterminate	If percent staining is available, cut point for positivity: 0 = negative 1-10 = weak positive/borderline >10 = positive  If percent staining is not available, obtain ER status indicated in record.  For tumors with multiple pathology reports: 1. If any positive = positive 2. If one is negative and one is weak positive /borderline (>2%-10%) = weak positive/borderline 3. If one is negative and one is weak positive /borderline (1 or 2 %) = indeterminate
ER_IHC_PERCENT_CELLS_POS	ER IHC percent cell positive (text)	Values could be in ranges, for example: 10-25, <5, >90, etc.

ER_PERCENT_POS	ER IHC % cells positive (numeric value)	<p>Recorded from ER_IHC_PERCENT_CELLS_POS to numeric value.</p> <p>Melissa Troester suggested to recode the data values that are in ranges this way:</p> <ul style="list-style-type: none"> <li>• If data was recorded in a range, take the median value. For example, "11 to 50" is recoded as <math>(11 + 50)/2 = 30.5</math>.</li> <li>• If data was recorded as "&gt;X" or "&lt;X", add/subtract 0.1 to the value. For example, "&gt;10" is recoded to 10.1, and "&lt;5" is recoded to 4.9. There is one exception, "&lt;1" is recoded to 0, to make things consistent with the ER status definition.</li> </ul>
ER_STAINING_INTENISTY	<p>ER average nuclear staining intensity (recoded)</p> <p>0  Weak  Weak to Moderate  Intermediate / Moderate  Moderate to Strong  Strong</p>	<p>Recorded from ER_AVG_NUCLEAR_STAIN_INTENISTY.</p> <p>The original staining variable has these values (0, 1+, 1+ to 2+, 2+, 2+ to 3+, 3+, weak, intermediate/moderate, strong).</p> <p>Dr. Joseph Geradts said the numbers and terms are redundant:</p> <p>1+ = weak  1+ to 2+ = weak to moderate  2+ = moderate  2+ to 3+ = moderate to strong  3+ = strong</p>

PR_STS	PR status Positive Negative Weak Positive / Borderline Indeterminate	If percent staining is available, cut point for positivity: 0 = negative 1-10 = weak positive/borderline >10 = positive  If percent staining is not available, obtain PR status indicated in record.  For tumors with multiple pathology reports: 1. If any positive = positive 2. If one is negative and one is weak positive /borderline (>2%-10%) = weak positive/borderline 3. If one is negative and one is weak positive /borderline (1 or 2 %) = indeterminate
PR_IHC_PERCENT_CELLS_POS	PR IHC percent cell positive (text)	Values could be in ranges, for example: 10-25, <5, >90, etc.
PR_PERCENT_POS	PR IHC % cells positive (numeric value)	Recoded from PR_IHC_PERCENT_CELLS_POS to numeric value.  Melissa Troester suggested to recode the data values that are in ranges this way: <ul style="list-style-type: none"> <li>• If data was recorded in a range, take the median value. For example, "11 to 50" is recoded as <math>(11 + 50)/2 = 30.5</math>.</li> <li>• If data was recorded as "&gt;X" or "&lt;X", add/subtract 0.1 to the value. For example, "&gt;10" is recoded to 10.1, and "&lt;5" is recoded to 4.9. There is one exception, "&lt;1" is recoded to 0, to make things consistent with the PR status definition.</li> </ul>

PR_STAINING_INTENISTY	PR average nuclear staining intensity (recoded) 0 Weak Weak to Moderate Intermediate / Moderate Moderate to Strong Strong	Recoded from PR_AVG_NUCLEAR_STAIN_INTENISTY. The original staining variable has these values (0, 1+, 1+ to 2+, 2+, 2+ to 3+, 3+, weak, intermediate/moderate, strong).  Dr. Joseph Geradts said the numbers and terms are redundant: 1+ = weak 1+ to 2+ = weak to moderate 2+ = moderate 2+ to 3+ = moderate to strong 3+ = strong
HER2_STATUS	HER2 status from IHC/FISH Positive Negative Borderline	Derived from IHC and/or FISH assay.  <ol style="list-style-type: none"> <li>1. If either IHC or Fish = positive, then HER2 = positive.</li> <li>2. If both IHC and Fish are borderline, then HER2 = borderline.</li> <li>3. If one is negative and one is borderline (no data) = negative.</li> <li>4. If one has value and the other is missing, HER2=the non-missing value.</li> </ol> For tumors with multiple pathology reports, adopt the same strategy listed above.

<b>Variable</b>	<b>Description</b>	<b>Comment</b>
LYMPHOVASCULAR_INVASION	Lymphovascular Invasion 1 = Present 2 = Absent 3 = Indeterminate/equivocal 4 = Not stated	For tumors with multiple pathology reports: 1. If any value of any path for the same tumor is Present = Present 2. If one is Absent and one is Indeterminate = Absent 3. If one is Not stated and one is Indeterminate = Indeterminate
ONCO_DX_ASSAY_INCLUDED	Oncotype Dx assay included? N = No Y = Yes	When there are multiple pathology reports with Oncotype Dx assay, use data from the report with the highest ONCO_DX_RECUR_SCORE.
ONCO_DX_ER_SCORE	Oncotype Dx ER score 0 – 100	
ONCO_DX_ER_STS	Oncotype Dx ER status Negative Positive	Negative = if ONCO_DX_ER_SCORE < 6.5  Positive = if ONCO_DX_ER_SCORE ≥ 6.5
ONCO_DX_PR_SCORE	Oncotype Dx PR score 0 – 100	
ONCO_DX_PR_STS	Oncotype Dx PR status Negative Positive	Negative = if ONCO_DX_PR_SCORE < 5.5  Positive = if ONCO_DX_PR_SCORE ≥ 5.5

ONCO_DX_HER2_SCORE	Oncotype Dx Her2 score 0 – 100	
ONCO_DX_HER2_STS	Oncotype Dx Her2 status Negative Equivocal Positive	Negative = if ONCO_DX_HER2_SCORE < 10.7 Equivocal = if $10.7 \leq \text{ONCO\_DX\_HER2\_SCORE} \leq 11.4$ Positive = if ONCO_DX_HER2_SCORE $\geq 11.5$
ONCO_DX_RECUR_SCORE	Oncotype Dx recurrence score 0 – 100	
ONCO_DX_HOSPITAL	Hospital code on the path report with Oncotype Dx scores	
ONCO_DX_PHYSICIAN	Surgeon/Physician code listed on the path report with Oncotype Dx scores.	