# Cancer Trial Eligibility Annotation Schema (v. 1.4)

Translational Biomedical Informatics Center at MUSC

November 29, 2018

**Introduction:** This document explains and provides guidelines for annotation of cancer trial eligibility criteria that are relevant for the Clinical Trial Eligibility project. It also serves to maintain consistency across all annotators. The natural language processing tool will learn from the expert annotations and then automatically extract this information from electronic health records of various types. Annotations will be done to help localize spans of text that contain evidence for a particular criterion and the conclusions to be drawn from said evidence. The annotation will be done using an annotation tool available online: WebAnno.

**Conventions applied to classes and slots below:** In the provided examples, highlighted text provides examples of what to annotate. Annotate the full span of text containing the measurement or value of interest and the label associated with that value. Some of these measurements are reports of historical facts. For those annotations, you should select the *Historical* checkbox. The value that you set an attribute to should reflect the status at the time the measurement is taken, even if that value differs from the current status. Some of these mentions are actually generic uses of term and do not specifically denote the presence or absence of a condition for the patient. For those annotations, you should select the *Generic* checkbox. Some of these measurements represent uncertain, hypothetical, or future possible facts. For those annotations, you should select the *Uncertain* checkbox. Some of the measurements will be undefined, unreported, or ambiguous. For those annotations, you should leave the value blank (if numerical) or choose *Indeterminant* from the drop-down. Beside the highlighted span is an information box showing the attribute(s) to set for the given span. Labels and measurement values may appear in other formats not documented below but nontheless should be annotated.

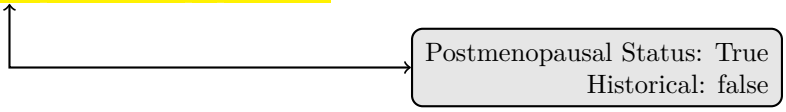
1. Postmenopausal Status

   **Values:** *True*, *False*, *Indeterminant*

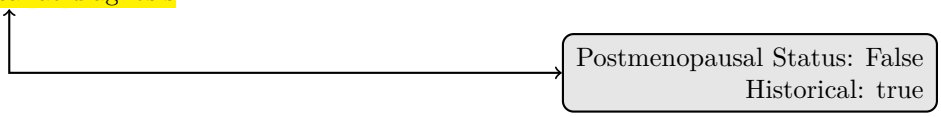
   Menopausal status is sometimes conveyed as a simple categorical attribute. Menopausal status can be given '*at diagnosis*' in a report far removed from the original diagnosis. These status values should be treated as *Historical*. In these historical cases, the value that you set *Postmenopausal Status* to should reflect the status at that point in time, regardless of the current status. When the last menstrual period (LMP) is given in date format, cancer trial eligibility requires that the last menstrual period be more than 2 years ago.

   Examples:

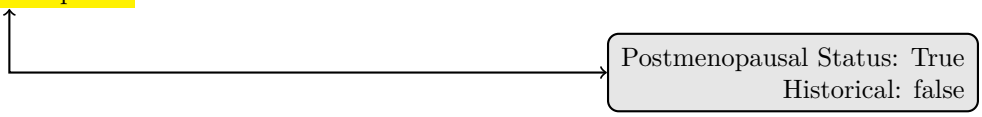
   (1) MENOPAUSE STATUS: Post menopausal at diagnosis



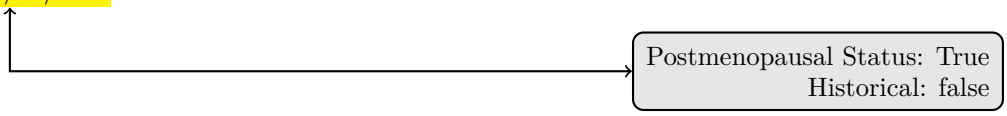
(2) Pre-menopausal at diagnosis



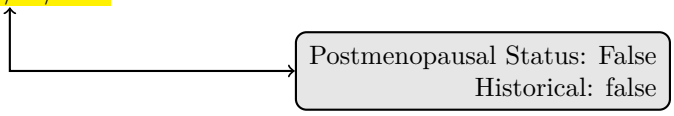

(3) now post-menopausal




(4) LMP 01/01/1992




(5) Report Date: 11/27/2017 … LMP 11/01/2017

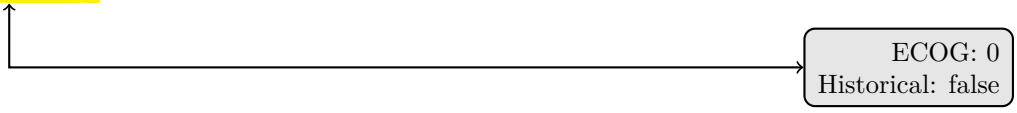



2. ECOG Scale of Performance Status[1]

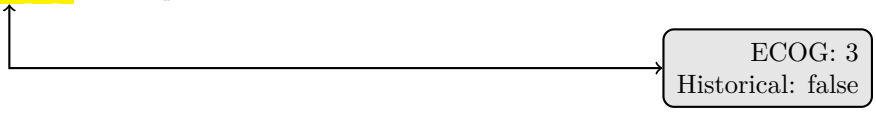
**Values:** *0–5*

Examples:

(6) ECOG status 0

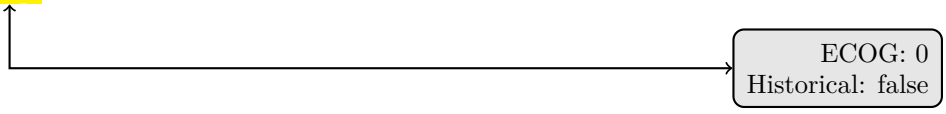



(7) Currently her ECOG is 3 due in part to …




(8) ECOG is not always spelled out and instead surfaces as some variant of ‘performance status‘:

SpO2 99% PS: 0




[1] http://ecog-acrin.org/resources/ecog-performance-status

3. Biomarker Status

   Each of the three receptors (*ER*, *PR*, and *HER2*) have four attributes of interest:

   (a) ER, PR, and HER2 Status

   **Values:** *Positive*, *Negative*, *Indeterminant*

   The hormone receptor status is sometimes described in categorical terms. If an explicit categorical *Positive* or *Negative* is not given, then you should consider the attribute *Indeterminant*.

   (b) ER, PR, and HER2 Positivity

   **ER and PR Values:** *0–100*
   **HER2 Values:** *0–3*

   The hormone receptor positivity is a percentage (ranging from zero to 100 for ER and PR and ranging from zero to three for HER2). When an explicity number is not provided in the text span, leave this attribute empty.

   (c) Allred Status

   **Values:** *Positive*, *Negative*, *Indeterminant*

   The Allred status is sometimes described in categorical terms. If an explicit categorical *Positive* or *Negative* is not given, then you should consider the attribute *Indeterminant*.

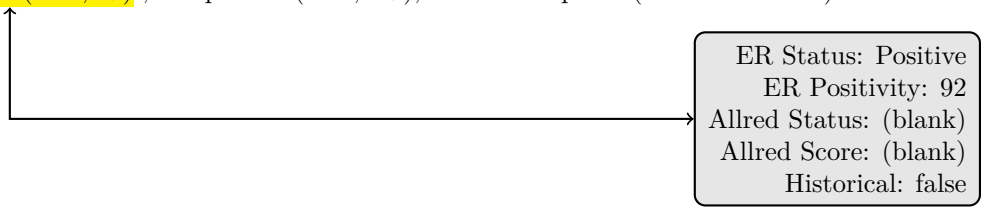
   (d) Allred Score

   **Values:** *0–8*

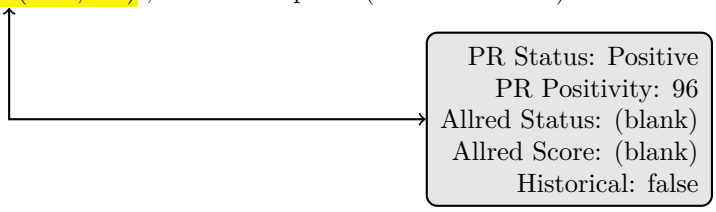
   The Allred score ranges from zero to eight. Leave this attribute empty if only the Allred intensity or proportion score is provided and not the composit Allred score.

   Examples:

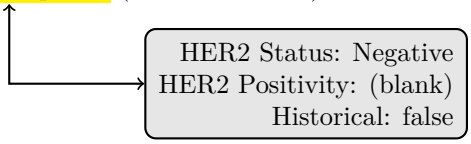
   (9) ER positive(92%,3+) , PR positive(96%, 2+), HER unamplified(ratio 1.01 FISH)

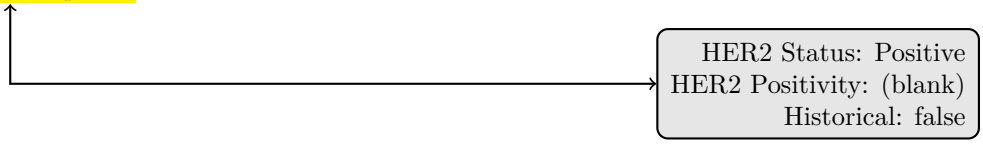
   (10) ER positive(92%,3+), PR positive(96%, 2+) , HER unamplified(ratio 1.01 FISH)

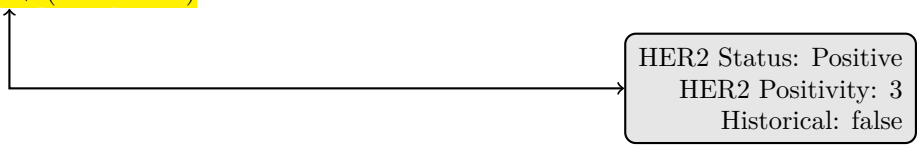
   (11) ER positive(92%,3+), PR positive(96%, 2+), HER unamplified (ratio 1.01 FISH)

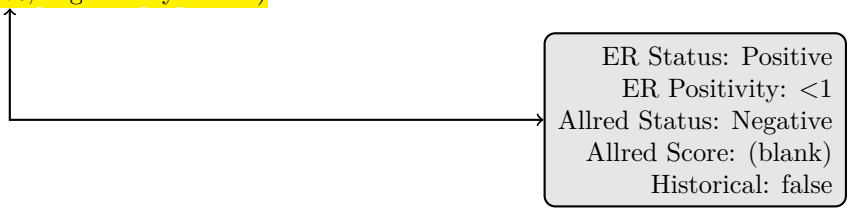
(12) HER-2/neu amplified




(13) o HER2/NEU: 3+ (POSITIVE)




(14) ER focal positive (<1%; negative by Allred)




(15) Positivity and negativity do not need to spelled out. They can also be indicated by symbols:

ER+ GAPA+


4. Breast Cancer Staging[2]

   (a) Timing
   **Values:** *Clinical*, *Pathological*
   The staging values can be determined clinically or pathologically. Leave this attribute empty if no explicit timing information is given.

   (b) Primary Tumor (T)
   **Values:** *T0*, *T1*, *T2*, *T3*, *T4*

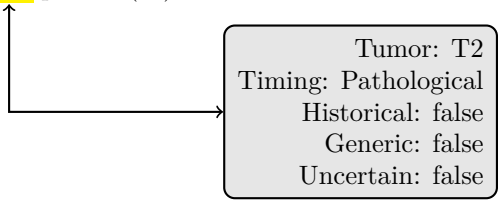
   (c) Regional Lymph Nodes (N)
   **Values:** *N0*, *N1*, *N2*, *N3*

   (d) Distant Metastases (M)
   **Values:** *M0*, *M1*

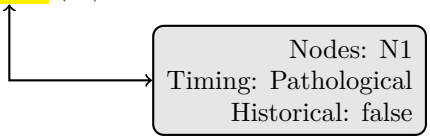
   Examples:

[2] https://cancerstaging.org/references-tools/quickreferences/Documents/BreastMedium.pdf

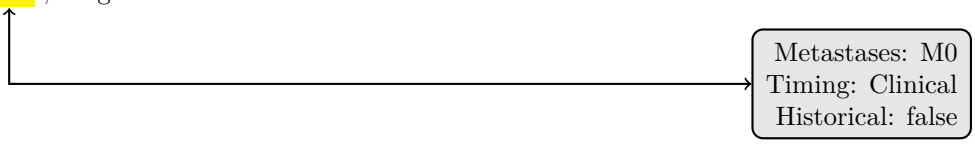
(16) Invasive Ductal carcinoma with associated DCIS, pT2 pN1mi (sn)




(17) Invasive Ductal carcinoma with associated DCIS, pT2 pN1mi (sn)



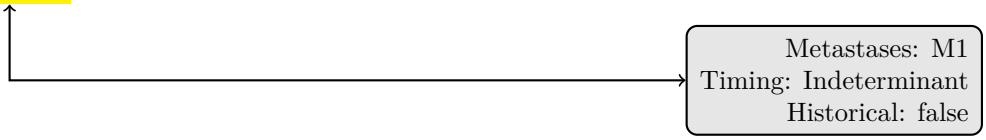

(18) cT0 cN1 cM0 , Stage IIA




(19) likely T3 N1


(20) In most instances, we require an explicit mention of staging. However, for *M1* it is enough to see a mention of *metastatic* to qualify as *M1* staging:

with metastatic adenocarcinoma




(21) Timing information is sometimes not included as part of the staging word. In these instances, we will treat the "clinical stage" as a context clue and not directly annotate it:

clinical stage T2 N0