Automating Cancer Stage Calculations: Structured Data Capture & Web Services

Mary E Edgerton, MD, PhD
Associate Professor of Pathology, Breast and Pathology Informatics Sections
UT MD Anderson Cancer Center
Vice-Chair, PERT Committee
College of American Pathologists

Richard Moldwin, MD, PhD
Lead Physician Informaticist
Structured Data Team
College of American Pathologists
Disclosure

In the past 12 months, the authors and presenters have not had any significant financial interest or other relationship with the manufacturers of the products or providers of the services that will be discussed in my presentation.
Learning Objectives

- After participating in this session the learner should be better able to:
  - Understand the need for automated cancer stage calculations
  - Understand the interaction between the CAP eCC templates, the SDC information model, the CDC web service and the AJCC API for staging rules
Outline

• Intro to Breast Cancer Staging using the AJCC 8th edition system
• Intro to the AJCC API, and the CDC Staging Web Service and dll.
• Intro to Integrating the Healthcare Enterprise (IHE) Structured Data Capture (SDC) information model
• The Breast Cancer Staging Calculator Pilot
What is Cancer Staging?

- A scoring system for severity of disease that helps to predict the probability of survival over time
- Used to include or exclude patients from various therapies and clinical trials
- Used as a basis for patient comparison in cancer registries and for clinical trial analysis.
- The AJCC staging system evolves over time and primarily uses **TNM** status (Tumor, Nodes, Metastases), tumor grade and a variety of biomarkers.
- Cancer Stage is calculated differently based on tumor site, histology, and sometimes based on molecular findings
- Analysis of Cancer Registry data plays a crucial role in creating the Stage definitions for each tumor type.
Stage Groups Predict Survival

Data taken from National Cancer DataBank (ACOS)
Two Types of Stage Calculations

- Clinical (pre-surgical)
- Pathological (post-surgical, no neoadjuvant therapy)
$T = \text{Tumor (extent); clin vs. path}$

How is this data stored?
\[ N = \text{(Regional Lymph) Nodes; clin vs. path} \]
M = Metastases; (clin vs path)

• Tumor that spreads to elsewhere in the body

How is this data stored?
That was TNM

- Next we briefly cover Tumor Grade and Molecular Biomarkers
### Grade

#### Histologic Grade (Nottingham Histologic Score) (Note F)
- **No residual invasive carcinoma**

#### Glandular (Accum) / Tubular Differentiation
- Score 1 (> 75% of tumor area forming glandular / tubular structures)
- Score 2 (10% to 75% of tumor area forming glandular / tubular structures)
- Score 3 (< 10% of tumor area forming glandular / tubular structures)
- Only micrometastases present (not graded)
- Score cannot be determined

#### Nuclear Pheomorphism
- Score 1 (Nuclei small with little increase in size in comparison with normal breast epithelial cells, regular outlines, uniform nuclear chromatin, little variation in size)
- Score 2 (Cells larger than normal with open vesicular nuclei, visible nucleoli, and moderate variability in both size and shape)
- Score 3 (Vesicular nuclei, often with prominent nucleoli, exhibiting marked variation in size and shape, occasionally very large and bizarre forms)
- Only micrometastases present (not graded)
- Score cannot be determined

#### Mitotic Rate
- **Note: See Table 1 in CAP Protocol**
- Score 1 (< 5 mitoses per mm²)
- Score 2 (6-1 mitoses per mm²)
- Score 3 (> 10 mitoses per mm²)
- Only micrometastases present (not graded)
- Score cannot be determined

#### Number of Mitoses per 10 High-Power Fields

#### Diameter of Microscope Field in Millimeters (mm)

#### Overall Grade
- Grade 1 (scores of 3, 4 or 5)
- Grade 2 (scores of 6 or 7)
- Grade 3 (scores of 0 or 9)
- Only micrometastases present (not graded)
- Score cannot be determined

**Note: The grade corresponds to the largest area of invasion. If there are smaller foci of invasion of a different grade, this information should be included under “Additional Pathologic Findings.”**

---


---

**How is this data stored?**

---

Copyright 2018 | College of American Pathologists
Biomarkers

• Estrogen Receptor (ER)
• Progesterone Receptor (PR)
• HER2 by Immunohistochemistry (IHC)
• HER2 by In Situ Hybridization (ISH)
• Oncotype Dx Recurrence Score (RS)
  • Oncotype Dx® is a genomic test based on the assessment of 21 genes; the result is the outcome of a mathematical formula of the weighted expression of each gene combined into a single score. It is measured and reported by RT-PCR, with recurrence score of < 11 the most pertinent cutoff value.14 Oncotype Dx® is required only for assigning prognostic stage group to patients with T1–2 N0 M0, ER-positive, HER2-negative cancers. AJCC Level of Evidence: I
  • Staging data does not factor in the 2018 TAILORx study, which moves the RS cutoff to ≤ 25.
**Principles of Cancer Staging**

- Basic workflow of breast cancer staging:
  - The Clinical Staging timeframe (c)
    - Initial diagnosis (e.g., confirmed by biopsy) (c stage)
    - Biomarkers often obtained on biopsy tissue during the clinical timeframe
    - Neoadjuvant (pre-surgical) therapy (yp for TNM, but no p Stage group (i.e. I-IV)
    - Patient might not receive surgery due to preference or poor health (c stage [or yc stage if they receive medical treatment])
    - A Clinical Stage may be calculated at this point, using the data collected before surgery.
    - The clinical staging window ends with definitive surgery (replaced to p-stage), or at tumor progression if within 4 months, or 4 months after diagnosis, at which time no new data can be added.
Principles of Cancer Staging

- The Pathological Staging **timeframe (p)**
  - Begins with the definitive surgery intended to remove as much of the tumor as possible.
  - Tumor tissue receives extensive analysis through anatomic and molecular pathology approaches. A pTNM, grade and biomarker result assessment is generated from this work.
  - Generally ends with the surgical procedure, although **additional findings may accumulate from clinical workup up until 4 mo. after diagnosis or until tumor progression:**
    - Clinical workup may include: examination of lymph nodes, metastases, lab tests, radiology, etc.
pTNM vs. Pathological Stage

- When can you use pTNM to calculate a Stage versus stopping at TNM measurements

<table>
<thead>
<tr>
<th>Initial Treatment Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Surgical (use pTNM and p stage after surgery)</td>
</tr>
<tr>
<td>☐ Neoadjuvant (use ypTNM after surgery to record residual tumor, no p stage)</td>
</tr>
<tr>
<td>☐ Non-surgical (may use ycTNM after treatment, no p stage)</td>
</tr>
<tr>
<td>☐ No treatment (no yp or yc for TNM, no p stage)</td>
</tr>
</tbody>
</table>
Principles of Cancer Staging

• Synthesis of Data
  • After the required clinical and pathology timeframe data is received, it is assembled by the treating oncologist (aka the managing physician). Clinically-assessed data may be used to supplement the pTNM values.
  • If clinical assessment data (i.e., not pathologically confirmed, such as a radiology report) is used to supplement increase the pT or pN value(s), staging becomes relatively unreliable and therefore a pathological stage is not supposed to be calculated. This rule is fuzzy and sometimes disregarded. Some people call this kind of stage a “working stage” rather than a “prognostic stage.” However, clinically-obtained data for assessment of metastases (cM) may be used for prognostic stage calculation.
  • The staging rules can be subtly different for different tumor types.

• Calculation of Stage
  • In the past, staging was mostly based on anatomical findings, and oncologists consulted short tables to determine the stage. Today, with biomarkers and more complex rules, it is much harder to stage correctly...
AJCC Staging Table and API

• The Invasive Breast Cancer Staging table is 6 pages long with 70 additional pages of Breast cancer staging information in the AJCC manual.

• The rules are relatively complex, and are very specific for invasive breast cancer.

• The 6 page staging table underestimates the complexity. When the full rule set is placed in a fully-expanded table for computer consumption...
### Example: Staging Table for Path Stage

<table>
<thead>
<tr>
<th>Code</th>
<th>T</th>
<th>N</th>
<th>M</th>
<th>Oncotype</th>
<th>Grade</th>
<th>HER2</th>
<th>ER</th>
<th>PR</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1A</td>
<td>pT0</td>
<td>pN1mi</td>
<td>cM0</td>
<td>N/A</td>
<td>G1</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>3</td>
<td>1A</td>
<td>pT0</td>
<td>pN1mi</td>
<td>cM0</td>
<td>N/A</td>
<td>G1</td>
<td>Positive</td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>36215</td>
<td>4</td>
<td>cT4d</td>
<td>cN3c</td>
<td>pM1</td>
<td>N/A</td>
<td>G1</td>
<td>Equivocal</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>36216</td>
<td>4</td>
<td>cT4d</td>
<td>cN3c</td>
<td>pM1</td>
<td>N/A</td>
<td>G2</td>
<td>Equivocal</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>36217</td>
<td>4</td>
<td>cT4d</td>
<td>cN3c</td>
<td>pM1</td>
<td>N/A</td>
<td>G3</td>
<td>Equivocal</td>
<td>Negative</td>
<td>Negative</td>
</tr>
</tbody>
</table>

XML Row #: 630,213

```
<Tvalue>cT4d</Tvalue>
<Nvalue>cN3c</Nvalue>
<Mvalue>pM1</Mvalue>
<N/A
<stagetablevalue outputclass="oncotypedx">N/A</stagetablevalue>
<stagetablevalue outputclass="AJCCfactor-G">G3</stagetablevalue>
<stagetablevalue outputclass="AJCCfactor-HER2">Equivocal</stagetablevalue>
<stagetablevalue outputclass="AJCCfactor-ER">Negative</stagetablevalue>
<stagetablevalue outputclass="AJCCfactor-PR">Negative</stagetablevalue>
<stagegroup>IV</stagegroup>
```
The AJCC API Rule Set

• The rules in the API download for breast cancer staging contain > 600,000 lines of XML
• When converted to an Excel spreadsheet, there are over 36,216 rows for pathological staging and 20,663 rows for clinical staging
• Any combination of factors that does not fit into one of these rows – technically cannot be staged by an automated calculator. However, a stage based on the AJCC manual table can be used to “estimate” an unofficial stage. At this time, we are working through the implications of this discrepancy and trying to determine whether registry data can be used to extend the stage-ability of more cases.
• This is an onerous task for human to slog through while caring for cancer patients. How can we automate the process of Stage calculation?
The Staging Calculator

SDC Forms
- CAP eCC forms
- Add staging form parts

CDC Web Service
- Extract AJCC data
- Wrap in web service

AJCC API
- Staging table for Breast Ca
Behind the Scenes

• Extract all staging rules from the AJCC API and wrap them in a Staging dll (C with a C# wrapper).
  – Anyone who wants to automatically calculate an AJCC stage must contract with AJCC for use of the API. (Managing physicians however can manually determine stage using the AJCC tables in the 8th ed. book)

• Wrap the CDC Staging dll inside a web service that accepts all the input parameters to calculate stage from the outside. Maintain the AJCC API in parallel in order to accept updates and errata from AJCC. Create services for IHE Form Manager and Form Receiver to select and receive the SDC forms.

• Construct an XSLT transform to turn the SDC XML into HTML/JS web pages with full CDS functionality to support Breast Ca Staging through the CDC web service.

• Use the SDC-based web pages to enter the data and stage patients...
<?xml version="1.0" encoding="us-ascii"?>

<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
                    version="1.0" xmlns:sr="http://www.cap.org/pert/2003/010"
                    xmlns:xsl="urn:uuid:orchard:2016">
  <xsl:output encoding="us-ascii" method="html" doctype-public="-//W3C//DTD XHTML 1.0 Transitional//EN"
              doctype-system="http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"/>

  <xsl:variable name="show-toc" select="false"/>
  <xsl:variable name="debug" select="false"/>

  <xsl:template match="/">

    <xsl:variable name="required" select="string(//Header/Property[@type='webposting_date meta']/@val)"/>

    <html>
      <head>
        <title><xsl:value-of select="//Header/@title"/></title>

        <link rel="stylesheet" href="sdctemplate.css" type="text/css" />

        <script type="text/javascript"
            src="https://ajax.googleapis.com/ajax/libs/jquery/1.12.0/jquery.min.js"></script>
        <!--script type="text/javascript" src="sdctemplate.js"></script-->  

        <script type="text/javascript">
          <![CDATA[
            var xmlDoc;
            var repeatIndex = 0; // used to generate unique ids, names in repeated elements
            $(document).ready(function () {
              jQuery.support.cors = true; // not sure if needed because cors setting is on the server

              // support toggle
              
              $("#collapsible").click(function()
                {
                  $(this).siblings().toggle();
                  $(this).toggleClass("HeaderGroup collapsed");
                });
          ]]
        </script>
      </head>

      ...
IHE ITI-34/35 Transaction Diagram

Figure X.4.2.2.2-1: Capture and Submit Pre-Authorization – Process Flow diagram
CAP eCC Forms for Breast Cancer Staging

• Start with CAP eCC forms (IHE SDC XML format) for Invasive Breast Anatomic Pathology and Breast Biomarkers
  • The form should be part of the normal pathology workflow, and saved or transmitted to an EHR
• Add a new SDC form on top for capturing the clinical timeframe cTNM and Treatment Plan (e.g., for neoadjuvant therapy etc.)
  • This form must be saved or transmitted to the EHR.
• Add a third SDC form at the end for the oncologists staging assessment.
  • This form should be automatically populated with the results from the previous clinical and pathology forms. The oncologist may override any value with values from other reports.
  • The form should indicate which required staging data elements are missing
  • The form should automatically calculate clinical and pathological stage and be saved/transmitted to the EHR
Breast Cancer Staging Calculator

SaveState : New
GenericHeaderText : Breast Cancer Staging Calculator
Category : Breast
OfficialName : + Breast Cancer Staging Calculator
TemplateID : 359.1000043
CAP_Required : false
ShortName : Breast.Ca.Stg.Calc
ReleaseStatus : CTP6

ONCOLOGIST: Initial Clinical Staging Assessment

PATHOLOGY: Anatomic and Biomarker Reports

ONCOLOGIST: Staging
3 Levels of Form Submission

ONCOLOGIST: Initial Clinical Staging Assessment
- Primary Tumor (Invasive Carcinoma) (cT)
- Regional Lymph Nodes (cN)
- Distant Metastasis (cM)
  - Initial Non-Surgical Therapy
Submit Initial Clinical Assessment

PATHOLOGY: Anatomic and Biomarker Reports
- Anatomic Pathology - Invasive Breast Cancer
- Biomarkers - Invasive Breast Cancer
Submit Pathology Report

ONCOLOGIST: Staging
- Staging Variable Completion
Submit Form

PROGNOSTIC STAGE RESULTS
Breast Cancer Staging Calculator

SaveState : New
GenericHeaderText : Breast Cancer Staging Calculator
Category : Breast
OfficialName : + Breast Cancer Staging Calculator
TemplateID : 359.1000043
CAP_Required : false
ShortName : Breast.Ca.Stg.Calc
RclascStatus : CTP9

ONCOLOGIST: Initial Clinical Staging Assessment

PATHOLOGY: Anatomic and Biomarker Reports

ONCOLOGIST: Staging
<table>
<thead>
<tr>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Clinical Assessment Report</td>
</tr>
<tr>
<td>Primary Tumor (Invasive Carcinoma) (cT)</td>
</tr>
<tr>
<td>cT1b: Tumor greater than 5 mm but less than or equal to 10 mm in greatest dimension</td>
</tr>
<tr>
<td>Regional Lymph Nodes (cN)</td>
</tr>
<tr>
<td>cN2b: Metastases only in ipsilateral internal mammary nodes in the absence of axillary lymph node metastases</td>
</tr>
<tr>
<td>Distant Metastasis (cM)</td>
</tr>
<tr>
<td>cM0: No clinical or radiographic evidence of distant metastases</td>
</tr>
<tr>
<td>Initial Treatment Approach</td>
</tr>
<tr>
<td>Surgical (use pTNM and p stage after surgery)</td>
</tr>
<tr>
<td>Report</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Pathology Report</td>
</tr>
<tr>
<td>Primary Tumor (Invasive Carcinoma) (cT)</td>
</tr>
<tr>
<td>cT1b: Tumor greater than 5 mm but less than or equal to 10 mm in greatest dimension</td>
</tr>
<tr>
<td>Regional Lymph Nodes (cN)</td>
</tr>
<tr>
<td>cN2b: Metastases only in ipsilateral internal mammary nodes in the absence of axillary lymph node metastases</td>
</tr>
<tr>
<td>Distant Metastasis (cM)</td>
</tr>
<tr>
<td>cM0: No clinical or radiographic evidence of distant metastases</td>
</tr>
<tr>
<td>Initial Treatment Approach</td>
</tr>
<tr>
<td>Surgical (use pTNM and p stage after surgery)</td>
</tr>
<tr>
<td>Overall Grade</td>
</tr>
<tr>
<td>Grade 2 (scores of 6 or 7)</td>
</tr>
<tr>
<td>Primary Tumor (Invasive Carcinoma) (pT)</td>
</tr>
<tr>
<td>pT1b: Tumor &gt; 5 mm but &lt;= 10 mm in greatest dimension</td>
</tr>
<tr>
<td>Category (pN)</td>
</tr>
<tr>
<td>pN2a: Metastases in 4-9 axillary lymph nodes (at least one tumor deposit larger than 2.0 mm)</td>
</tr>
</tbody>
</table>
### Staging Section

- **ONCOLOGIST: Initial Clinical Staging Assessment**
- **PATHOLOGY: Anatomic and Biomarker Reports**
- **ONCOLOGIST: Staging**

**Report**

<table>
<thead>
<tr>
<th>PT</th>
<th>pT1b: Tumor &gt; 5 mm but ≤ 10 mm in greatest dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNM</td>
<td>pN2a: Metastases in 4-8 axillary lymph nodes (at least one tumor deposit larger than 2.0 mm)</td>
</tr>
<tr>
<td>PM</td>
<td>cM0: No clinical or radiographic evidence of distant metastases</td>
</tr>
</tbody>
</table>

**Variables Used for Stage Calculation**

- TreatmentPlan
- cT
- cN
- pT
- pN
- pM
- E1R
- PgR
- HER2
- OncDrx
- Grade

**Clinical Stage, AJCC 8th Ed.**

- IIA
- III

**Pathological Stage, AJCC 8th Ed.**

- IIA
- III
# ONCOLOGIST: Staging

## Staging Variable Completion
- **Overall Grade**

## Biomarker Summary
- **Estrogen Receptor (ER) Results**
- **Progesterone Receptor (PgR) Status**
- **HER2 Evaluation**
- **Oncotype Dx Recurrence Score**

## Pathological Staging Categories (pTNM, AJCC 8th Edition)
- **Primary Tumor (Invasive Carcinoma) (pT)**
- **Regional Lymph Nodes (pN)**
- **Distant Metastasis (pM)**

## PROGNOSTIC STAGE RESULTS
- **Variables Used for Stage Calculation**
- **Clinical Stage, AJCC 8th Ed.**
- **Pathological Stage, AJCC 8th Ed.**
- **Calculate Stage**

Submit Form
PROGNOSTIC STAGE RESULTS

**Variables Used for Stage Calculation**
- Treatment Plan
- cT
- cN
- cM
- pT
- pN
- pM
- ER
- PgR
- HER2
- OncDx
- Grade

**Clinical Stage, AJCC 8th Ed.**
- 0
- IA
- IB
- IIA
- IIB
- IIIA
- IIIB
- IIIC
- IV
- Not Staged

**Pathological Stage, AJCC 8th Ed.**
- 0
- IA
- IB
- IIA
- IIB
- IIIA
- IIIB
- IIIC
- IV
- Not Staged

[Submit Form]
## Staging Report (partial view)

<table>
<thead>
<tr>
<th>Report</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pT1b</td>
<td>Tumor &gt; 5 mm but &lt;= 10 mm in greatest dimension</td>
</tr>
<tr>
<td>Regional Lymph Nodes (pN)</td>
<td>pN2a: Metastases in 4-9 axillary lymph nodes (at least one tumor deposit larger than 2.0 mm)</td>
</tr>
<tr>
<td>Distant Metastasis (pM)</td>
<td>cM0: No clinical or radiographic evidence of distant metastases</td>
</tr>
<tr>
<td>Variables Used for Stage Calculation</td>
<td>TreatmentPlan, cT, cN, cM, pT, pN, pM, ER, PgR, HER2, OncDx, Grade</td>
</tr>
<tr>
<td>Clinical Stage, AJCC 8th Ed.</td>
<td>IIA</td>
</tr>
<tr>
<td>Pathological Stage, AJCC 8th Ed.</td>
<td>IB</td>
</tr>
</tbody>
</table>
Learn More


- **AJCC:** [https://cancerstaging.org/Pages/default.aspx](https://cancerstaging.org/Pages/default.aspx), [https://ajcc.3scale.net](https://ajcc.3scale.net)


- **CAP eCC:** [https://www.cap.org/cancerprotocols](https://www.cap.org/cancerprotocols), [https://www.cap.org/capecc](https://www.cap.org/capecc)
Automating Cancer Stage Calculations using Structured Data Capture and Web Services

CAP:
Keren Hulkower

AJCC:
Laura Meyer Vega
Martin Madera

CDC:
Jennifer Seiffert

Joseph D. Rogers
Sanjeev Baral
Shailendra Bajracharya
Email Rich Moldwin at: rmoldwin@cap.org

Or
Mary Edgerton at medgerton@mdanderson.org