**Using data from:**

* Molecular Linkage file (main file)
* Diagnosis file (for primary site and stage info)
* Metastatic Disease (to help assign stage)

**Definitions**

* **Melanoma diagnosis** (Diagnosis file): HistologyCode = {list of melanoma codes previously sent}
* **Tumor sequenced** (Molecular Linkage file): Tumor/Germline variable = Tumor

**New variables to create**

* **MelanomaDiagnosisCount:** count for number of melanoma clinical diagnoses for a patient
  + Create using the number of unique [AgeAtDiagnosis and PrimaryDiagnosisSite combinations] for each patient
  + This approach may work best since a few patients have multiple melanomas diagnosed at same age, so it cannot be used on its own.
    - *Example with multiple diagnoses at same age with same stage: ILE2DL0KMW*
* **SequencedTumorCount**: count for number of sequenced tumor samples for a patient
  + Create using the number of unique [DeidSpecimenID and AvatarKey combinations] for each patient
  + This approach may work best since a few patients have multiple sequenced tumors at the same age (or stage, etc), so these variables cannot be used on their own.
    - *Example with multiple tumors sequenced at same age/stage: 59OP5X1AZL*
* **AssignedPrimarySite**: {cutaneous, ocular, mucosal, unknown}
  + Based on the parameters outlined below; primary site variable to be used for the analysis
* **AssignedStage**: {I, II, III, IV}
  + Based on the parameters outlined below; stage variable to be used for the analysis
* **AssignedGroup**: {A, B, C, D} (NEW – just to keep track of these better)

**AssignedPrimarySite (SAME DEFINTIONS AS BEFORE)**

* + IF PrimaryDiagnosisSite contains “skin” OR “ear” OR “eyelid” OR “vulva”, THEN **AssignedPrimarySite** = cutaneous
    - *Vulvar melanoma is included here given that all appear to have been staged as cutaneous (not mucosal) melanoma.*
  + If PrimaryDiagnosisSite contains “choroid” OR “ciliary body” OR “conjunctiva”, then **AssignedPrimarySite** = ocular
  + If PrimaryDiagnosisSite contains “sinus” OR “gum” OR “nasal” OR “urethra” then **AssignedPrimarySite** = mucosal
    - *The list only includes the primary mucosal sites present in this dataset and not all possible mucosal sites for melanoma in general.*
  + If PrimaryDiagnosisSite contains ‘unknown’, then **AssignedPrimarySite** = unknown

**AssignedGroup (SAME DEFINTIONS AS BEFORE)**

* **Group A**: **MelanomaDiagnosisCount** = 1 AND **SequencedTumorCount =** 1 🡪 n = 327
* **Group B**: **MelanomaDiagnosisCount** = 1 AND **SequencedTumorCount >** 1 🡪 n = 19
* **Group C**: **MelanomaDiagnosisCount** > 1 AND **SequencedTumorCount =** 1 🡪 n = 30
* **Group D**: **MelanomaDiagnosisCount** > 1 AND **SequencedTumorCount >** 1 🡪 n=3

**AssignedGroup**

**Group B =** 1 melanoma diagnosis and >1 tumor sequenced 🡪 n=19

CHANGE FROM PRIOR: Do not exclude any, can still use those with WES only for TMB analysis. A few clarifications on the rules also in red text below.

SpecimenSiteofCollection is the correct field

For the record: these two patients had the wrong SpecimenID in my file but the correct SpecimenID in by your code according to the rules below. Patient ID 317K6G9N41 should have SpecimenID = 53LFUMZSW8ACOX5IPUZ45ICJ0. Patient ID U2CUPQJ4T1 should have SpecimenID = HIWF190182C5JJE3FHYR5BIXR

* If RNAseq is available for just one tumor, select the tumor with RNAseq data (even if no WES)
* If RNAseq data is available for > 1 tumors OR if only WES is available for all tumors:
  + If the patient has a tumor with SpecimenSiteofCollection that contains “skin” AND does not also have a tumor with SpecimenSiteofCollection that contains either “lymph node” or “soft tissue”, then select the tumor with SpecimenSiteofCollection that contains “skin”
    - If multiple skin specimens meet this rule, then use the one with earliest Age At Specimen Collection
  + If none of the patient’s tumors have SpecimenSiteofCollection that contains “skin” or “soft tissue”, then select the tumor with SpecimenSiteofCollection that contains “lymph node”
  + If a patient has a tumor with a SpecimenSiteofCollection that contains [“skin” OR “soft tissue”] AND a tumor with a SpecimenSiteofCollection that contains “lymph node”, then select the one with earliest Age At Specimen Collection.
    - If both skin/soft tissue and lymph node collected at same age (including if both are in the Age 90 or older category), then use the specimen with Primary/Met field = “Primary” (usually will be the skin specimen)
    - If none of the specimens have Primary/Met field = “Primary”, then use the lymph node specimen.
  + If the patient does NOT have any tumor with a SpecimenSiteofCollection that contains [“skin” OR “soft tissue” or “lymph node”], then select the tumor with the earliest Age At Specimen Collection

**Group C = >** 1 melanoma diagnosis and 1 tumor sequenced 🡪 n=30

CHANGE FROM PRIOR: Simplified rules with changes in red text to ensure correct diagnosis is selected.

SpecimenSiteofCollection is the correct field

* For specimens with the field Primary/Met = “**Primary**”
  + IF AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for only one of the diagnoses, then use that diagnosis.
    - 2FEX3JHKCW*,* 8BCBHGEO4N*,* FNP53S81KA, L5876HQBT9, MD578977I9, QC7QX0VWAY, VLY2FWYN29
  + IF AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for more than one of the diagnoses OR AgeAtDiagnosis is unknown for at least one of the diagnoses, but the PrimaryDiagnosisSite (from the Diagnosis file) is NOT the same for all diagnoses, then use the diagnosis that has a PrimaryDiagnosisSite = SpecimenSiteOfOrigin
    - ILE2DL0KMW, L2R9RJJ88C
* For specimens with the field Primary/Met = “**Metastatic**”
  + IF SpecimenSiteOfCollection does not contain “lymph node” AND only one diagnosis contains PathGroupStage EQUALS“IV” OR [ClinGroupStage EQUALS“IV” AND PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”]], then use the diagnosis with stage IV
    - 9NOLH4M870, HLIXS3VDZ6, SHTJKKY76C, DEB9M36STN
  + IF SpecimenSiteOfCollection does not contain “lymph node” AND NONE of the diagnoses have {PathGroupStage EQUALS“IV” OR [ClinGroupStage EQUALS “IV” AND PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”]]}, then use the diagnosis with the earliest AgeAtDiagnosis
    - 087FO3NF65, AC2EJBKWJO, BXVDLL792A, F3BE85LAWN, JUDAJ1LHL9, R06W2EUXCM, R2TNVTF684
  + IF SpecimenSiteOfCollection contains “lymph node” AND AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for only one diagnosis, then use that diagnosis within 90 days of specimen collection.
    - 643X8OLYWR, ILKRH6I83A, RAB7UH51TS
  + IF SpecimenSiteOfCollection contains “lymph node” AND AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for more than one diagnosis AND only one diagnosis has PathNStage that does NOT contain [“N0”, “Nx” OR “Unknown/Not Applicable” OR “No TNM applicable for this site/histology combination”], then use that diagnosis (the one with known positive nodes on PathNStage).
    - 39TYSJBNKK, 5BS8L7PCCE, 6RX3G5GV02
  + IF SpecimenSiteOfCollection contains “lymph node” AND AgeAtSpecimenCollection is GREATER THAN 90 days AFTER the AgeAtDiagnosis for all of the diagnoses AND the PrimaryDiagnosisSite (from Diagnosis file) = SpecimenSiteofOrigin for only one of the diagnosis, then use the diagnosis associated with that PrimaryDiagnosisSite
    - 7HU06PZK4Q, KWLPMWV0FM, XPZE95IE7I
  + IF SpecimenSiteOfCollection contains “lymph node” AND AgeAtSpecimenCollection is GREATER THAN 90 days AFTER the AgeAtDiagnosis for all of the diagnoses AND the PrimaryDiagnosisSite (from Diagnosis file) DOES NOT EQUAL SpecimenSiteofOrigin for any diagnosis, then use the diagnosies with earliest age of AgeAtDiagnosis
    - 383CIRVHH2
* **Group D**: >1 melanoma diagnosis and >1 tumor sequenced 🡪 n=3

CHANGE FROM PRIOR: Rules below instead of using rules from Group B and C, to ensure correct tumors and diagnosis pairs are selected.

SpecimenSiteofCollection is the correct field

* + First select tumor:
    - If just one tumor with RNAseq data, then select that tumor.
      * 59OP5X1AZL, FUAZTE7LVQ
    - If > 1 tumor with RNAseq data, then select tumor with earliest Age At Specimen Collection
      * 7HOWLJKDEM
  + Then select diagnosis:
    - If SpecimenSiteOfCollection contains “lymph node”, then select diagnosis with latest Age At Diagnosis
      * 59OP5X1AZL
    - If SpecimenSiteOfCollection contains either “skin” or “soft tissue”, then select diagnosis with earliest Age At Diagnosis
      * 7HOWLJKDEM, FUAZTE7LVQ

**Additional Fields for Stage Assignments**

Field Name: Discrepancy (0/1 or No/Yes)

This is to keep track of all the questionable ones that I reviewed with Dr. Slingluff regardless of whether we made an exception to the rule for staging or kept with the staging assigned by the rule. These will be identified under each rule, but are summarized here since I imagine this will be coded separately from the rules (this field doesn’t impact the rules/exceptions).

* Discrepancy = 1 if the AvatarKey = ["087FO3NF65" or "227RDTKST8" or "2AP9EDU231" or "2X7USSLPJC" or "6DL054517A" or "6HWEJIP63S" or "7YX8AJLMWR" or "87AJ4KITK8" or "9DLKDVIQ2W" or "9EYYI5H9SU" or "9HA9MZCSU2" or "A594OFU98I" or "AC2EJBKWJO" or "APGLZDFLYJ" or "DTUPUJ06B5" or "EKZGU61JTP" or "FUAZTE7LVQ" or "GEM0S42KIH" or "GITAF8OSTV" or "HTKAEZOC7V" or "HZD0O858UJ" or "IALUL9JC9Y" or "ILE2DL0KMW" or "KEHK6YTVAK" or "L2R9RJJ88C" or "MD5OTA3E8A" or "MPHAPLR8K1" or "N5Q9122LTG" or "NXPOH3RBWY" or "QC7QX0VWAY" or "QE43I70DGC" or "QLWU5QNQIB" or "TIZXXXVCV9" or "X9AZUY3R1C" or "XHTXLE3MLC" or "XPZE95IE7I" or "YMC959TA29" or "YRGE6MVYNK" or "Z7CEUA8SAJ"]
* For all other AvatarKey IDs not specified, Discrepancy = 0

Field Name: Possible New Primary (0/1 or No/Yes)

This is to keep track of the ones that may be new primary melanomas but we do not have an additional diagnosis or other information to definitively know this. The staging for these will be carried out like all the others. These will be identified under each rule, but are summarized here since I imagine this will be coded separately from the rules (this field doesn’t impact the rules/exceptions).

* Possible New Primary = 1 if the AvatarKey = ["87AJ4KITK8" or "9DLKDVIQ2W" or "9EYYI5H9SU" or "9HA9MZCSU2" or "A594OFU98I" or "AC2EJBKWJO" or "FUAZTE7LVQ" or "HZD0O858UJ" or "L2R9RJJ88C" or "MD5OTA3E8A" or "NXPOH3RBWY" or "QC7QX0VWAY" or "QE43I70DGC" or "XHTXLE3MLC"]
* For all other AvatarKey IDs not specified, Possible New Primary = 0

These are the exceptions to the staging rules. I imagine this should be coded before the other staging rules in your script, so I’ve named it **Rule#0 EXCEPTION**, but you can put it in your code differently if there is a better way.

**RULE#0 EXCEPTION**

* If AvatarKey = [“227RDTKST8” or “ILE2DL0KMW” or “YRGE6MVYNK” or “EKZGU61JTP”], then AssignedStage = III
* If AvatarKey = [“GEM0S42KIH” or “CG6JRI0XIX” or “WDMTMU4SV2”], then AssignedStage = IV
* This rule applies to 6 patients, all cutaneous; 5 patients in Group A and 1 patient in Group C (ILE2DL0KMW)

**AssignedStage**

**Once AssignedStage field is populated by a rule, it should not be further evaluated by the remainder of the script. This simplifies the rules if the script can be written so that the assignment happens in this order and the patient is no longer evaluated once assigned.**

* For example, patient 50I7H1PID7 has AgeAtSpecimenCollection = “Age 90 or older” and a Path Stage of IV. It will be assigned stage IV by the AGE rule and then removed from further evaluation, so should not then be evaluated/assigned for the PATHIV rule.
* **Simplified rules now apply to all specimens regardless of their primary site (cutaneous, ocular, mucosal, or unknown)**
* **I have provided the counts that meet each rule by primary site and group.**

**RULE #1: AGE**

* IF AgeAtDiagnosis = “Age 90 or older”, then **AssignedStage** = *numerical* value of PathGroupStage OR ClinGroupStage (if PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”])
  + Total: 5 patients, all cutaneous
    - 4 in group A
    - 1 in group B (D42OPM2PLC)

**RULE #2: PATHIV**

* IF PathGroupStage contains “IV”, then **AssignedStage** = “IV”
  + Total: 75 patients
    - Cutaneous: 70 patients
      * 66 in group A; 1 in group B (317K6G9N41); 3 in group C (9NOLH4M870, DEB9M36STN, HLIXS3VDZ6)
    - Ocular: 1 patient, group A
    - Mucosal: 3 patients
      * 2 patients in group A; 1 patient in group C (R06W2EUXCM)
    - Unknown: 1 patient, group B (MYCVCULC8L)

**RULE #3: CLINIV**

* IF ClinGroupStage contains “IV”, THEN **AssignedStage** = “IV”
  + *This rule no longer state requires path staging to be unknown. In all but one case is the path stage also IV or unknown. In that one case, the path staging and metastatic disease file are unclear, and it appears appropriate to classify as stage IV disease.* 
    - Confirm with Slingluff: confirmed to follow rule, keep as stage IV
      * QLWU5QNQIB: Trunk melanoma diagnosed at age 60.35 with path stage III (TxN3M0, 7th) and clinical stage IV (but recorded as TxN3M0 for that as well). Specimen is lymph node (NOS) obtained at age 60.64 (105 days after initial diagnosis). Reported to have “regional” colon and peritoneum mets at unknown age, as well as regional node (NOS) at unknown age, all associated with the trunk diagnosis. Also reported to have a distant skin met on lower extremity, also at unknown age and associated with trunk diagnosis. Rule seems appropriate given unknowns.
        + For this patient ID, set Discrepancy = 1.
  + 27 patients
    - Cutaneous: 24 patients
      * 22 patients in group A; 1 patient in group B (WAYUEWGM1O), 1 patient in group C (SHTJKKY76C)
    - Mucosal: 2 patients, group A
    - Ocular: 1 patient, group A

**RULE #4: METSITE**

**SpecimenSiteofCollection is the correct field**

* IF SpecimenSiteofCollection *does not contain* [“skin” OR “ear” OR “eyelid” OR “vulva”, OR “head” OR “soft tissues” OR “breast” OR “lymph node” OR “parotid” OR “vagina”], THEN **AssignedStage** = “IV”
  + Specimen from a distant metastatic site
    - Confirm with Slingluff: confirmed, rule captures any muscle or chest wall invasion as stage IV, but does not automatically assign breast or other soft tissue specimens as stage IV due to these not all appearing to be stage IV disease in this database
  + This rule applies to all patients regardless of primary site (cutaneous, ocular, mucosal, or unknown). This works for our database since it does not appear that we have any primary mucosal specimens. This rule would need to be edited if used in the future with mucosal melanomas.
  + Confirm with Slingluff the appropriateness of stage IV assignment, confirmed; updated rule to exclude vaginal specimen and move EKZGU61JTP to rule #8 (SKINLESS90D)
    - KEHK6YTVAK: trunk melanoma stage IIIC (T4aN3M0, 8th) at age 63.28. Specimen is anus (NOS), designated as “primary”, obtained at age 63.55. Reported to have distant pelvic nodes at 63.55 associated with trunk melanoma, and then two additional distant mets reported (inguinal nodes, soft tissue of pelvis) associated with trunk diagnosis. No mention of an anal melanoma. Rule seems appropriate to consider this as metastatic site instead of new primary mucosal (in addition to cutaneous primary of trunk); confirmed, follow rule to assign stage IV given previously reported distant mets
      * For this patient ID, set Discrepancy = 1.
  + 37 patients
    - Cutaneous: 35 patients
      * 30 patients in group A; 3 patients in group B (CILTNWAT6B, P5JJX6UI7G, 48J2GNDBEN); 2 patients in group C (F3BE85LAWN, BXVDLL792A)
    - Ocular: 2 patients, both group C (JUDAJ1LHL9, R2TNVTF684)

**RULE #5: PRIORDISTANT**

* IF {combined entry on Metastatic file with: **MetsDzPrimaryDiagnosisSite contains** [“skin” OR “ear” OR “eyelid” OR “vulva” OR “eye” OR “choroid” OR “ciliary body” OR “conjunctiva” OR sinus” OR “gum” OR “nasal” OR “urethra”] AND **MetastaticDiseaseInd = “Yes – Distant”**, AND **AgeAtMetastaticSite <= AgeAtSpecimenCollection+0.005**}, THEN **AssignedStage** = “IV”
  + Patients with recorded distant metastatic disease prior to specimen collection
  + *Must add 0.005 to the AgeAtSpecimenCollection due to only 2 decimal points recorded for age in that file, but 3 decimal points recorded on the metastatic disease file.*
  + This works for our database. This rule would need to be edited if used in the future to reflect the primary diagnosis site keywords applicable to that dataset (e.g., if there are other mucosal sites that should be included).
    - 37 patients
      * Cutaneous: 36 patients
        + 31 patients in group A; 3 patients in group B (PP7QY6B66M, U2CUPQJ4T1, WX0NQIQIXM); 2 patients in group C (087FO3NF65, QC7QX0VWAY)
      * Ocular: 1 patient, group A
    - Count update reflects exceptions listed below as well as the two patients (CG6JRI0XIX, WDMTMU4SV2) that I had erroneously counted by this rule but that weren’t actually captured. CG6JRI0XIX has prior metastatic disease (liver) reported with MetastaticDiseaseInd = “Yes – NOS”. WDMTMU4SV2 has prior metastatic disease (bone) reported with MetastaticDiseaseInd = “Yes – Regional” due to upper extremity melanoma; consider metastasis to bone as stage IV disease.
  + *This rule assumes that the designation of “Distant” (instead of Regional or NOS) in the metastatic disease file is accurate.* In general, this designation seems appropriate for our database, but it is possible that some cases are misclassified.
    - Confirm with Slingluff the appropriateness of stage IV assignment, confirmed details below
      * 087FO3NF65: Trunk IIIC (T3aN2aM0, 7th) at age 61.8. Regional axillary nodes at 62, then “Breast, NOS” met classified as “Distant” at age 63.59. Specimen is the Breast NOS met. Reported to have lung mets with “regional” intrathoracic nodes at age 64.32. Rule seems appropriate in this case; keep as stage IV for distant met.
        + For this patient ID, set Discrepancy = 1 and Exception = 0.
      * ILE2DL0KMW: Two diagnoses at the same time: scalp IIC and trunk IIIC (T4bN2cM0, 8th). Specimen is skin of trunk, collected 57 days after diagnosis, designated as “primary”. Reported to have “distant” axillary nodes associated with the trunk diagnosis (at the same time as the trunk skin specimen collection) and multiple “regional” skin mets associated with the scalp diagnosis. No additional entries to suggest later distant mets. No ICB treatment. Surgery file shows trunk diagnosis with first related biopsy taken at time of specimen collection. This may represent the primary lesion at IIIC diagnosis. Make exception to the rule to assign as stage III.
        + For this patient ID, set Discrepancy = 1. The PRIORDISTANT rule does not apply since this patient will be captured by EXCEPTION rule. This patient is not included in the patient counts for this rule anymore.
      * 227RDTKST8: Trunk IIIC (T3N3cM0, 8th) initial diagnosis. Specimen is lymph node (NOS), designated as “metastatic”, collected 35 days after diagnosis. Reported to have “distant” axillary nodes at same age as diagnosis, then “regional” inguinal nodes at same age as specimen collection. Specimen likely represents an inguinal node. Specimen collected Pre-ICB treatment. No additional entries to suggest later distant mets. Surgery file shows three entries, one at diagnosis and two at age of specimen collection. Perhaps axillary node was found first, then later discovered skin lesion and inguinal nodal basin involvement, but this would still be IIIC disease. Make exception to the rule to assign as stage III.
        + For this patient ID, set Discrepancy = 1. The PRIORDISTANT rule does not apply since this patient will be captured by EXCEPTION rule. This patient is not included in the patient counts for this rule anymore.
      * 9EYYI5H9SU: Lower extremity clinical IIB disease (no path stage) at initial diagnosis at age 43. Reported to have lung mets at age 59 associated with lower extremity diagnosis. Specimen is skin of trunk, designated as “primary”, collected at age 63.46. Reported to later develop pancreas mets at age 63.64, associated with lower extremity diagnosis. Surgery file only contains entries for lower extremity diagnosis, including time points that correspond to age of specimen collection. Trunk lesion could represent new primary, but rule seems appropriate in this case. Keep as stage IV with likely new primary in the setting of metastatic disease.
        + For this patient ID, set Discrepancy = 1.
        + For this patient ID, set Possible New Primary = 1.
      * HTKAEZOC7V: Scalp/neck III (T4bN2bM0, 6th) disease at initial diagnosis at age 70.9. Specimen is axillary node, designated as “metastatic”, at age 82.55. Reported to have regional skin mets at age 71, then “distant” axillary nodes at age 79.7, followed by regional nodes of head/face/neck at age 81. Later lung mets at age 83.9. Specimen collected post-ICB. Rule seems appropriate in this case. Follow rule, keep as stage IV.
        + For this patient ID, set Discrepancy = 1.
      * IALUL9JC9Y: Trunk IIIA (T3aN1aM0, 7th) at age 26.08. Reported to have “distant” nodes of head/face/neck at age 26.12 associated with trunk diagnosis. Then reported to have “regional” lymph node (NOS) associated with trunk diagnosis at age 27.22. Specimen is lymph node (NOS) at age 27.22. No ICB treatment. No additional entries to suggest later distant mets. Given prior “distant” nodes in head/face/neck, considered stage IV. Not sure about how different nodal basins are generally viewed for trunk melanomas. Follow rule, keep as stage IV.
        + For this patient ID, set Discrepancy = 1.
      * QE43I70DGC: Lower extremity IIIB (T3bN1M0, 6th) disease at age 55.115. Reported to have “distant” inguinal nodes at age 55.115 (same time as initial diagnosis). Specimen is a skin of lower extremity obtained at age 63.6, designated as a “primary”. Reported to later have brain mets at age 64. Skin specimen could represent new primary given timing, but could also represent regional recurrence of prior disease. Not sure why inguinal nodes were labelled as distant at initial diagnosis since not reported as stage IV at the time. Surgery file reports two specimens at age 55.115 (likely skin + node), but then another surgery at 55.293 before the specimen that we have at 63.4, and another later surgery at 63.79. Specimen obtained post-ICB treatment. Given brain mets within a year of specimen, rule likely appropriate given limited information. Keep as stage IV with likely new primary in the setting of metastatic disease.
        + For this patient ID, set Discrepancy = 1.
        + For this patient ID, set Possible New Primary = 1.
      * XHTXLE3MLC: Lower extremity IIC disease at age 50.5. Reported distant upper extremity skin met at 53.44, associated with the lower extremity diagnosis. Specimen is skin from lower extremity, designated as “primary”, obtained at age 54.56. No additional entries to suggest additional distant disease. Specimen obtained post-ICB treatment. While specimen could represent new primary, rule seems appropriate given limited information. Keep as stage IV with likely new primary in the setting of metastatic disease.
        + For this patient ID, set Discrepancy = 1.
        + For this patient ID, set Possible New Primary = 1.
      * YMC959TA29: Trunk IIIB (T4aN3cM0, 7th) disease at age 37.35. Specimen is skin (NOS), designated as “metastatic”, obtained at age 37.43 (29 days after initial diagnosis). Reported to have “distant” skin met (NOS) at that time, associated with trunk diagnosis. Specimen collected pre-ICB treatment. Surgery file has first entry at 37.43, followed by two later (age 40) entries. This could represent the skin from the initial IIIB diagnosis, but since listed as “metastatic” and there is a reported “distant” skin (NOS) met, classified as stage IV. Rule seems appropriate given limited information. Follow rule, keep as stage IV.
        + For this patient ID, set Discrepancy = 1.
      * YRGE6MVYNK: Lower extremity IIIA (T2aN1aM0, 8th) disease at age 29.97. Reported “distant” inguinal nodes at age 30.02, associated with lower extremity diagnosis. Specimen is skin from lower extremity, designated as “primary”, obtained at age 30.02 (same age as “distant” inguinal nodes). No additional entries to suggest later distant mets. No ICB treatment. Surgery file has an entry at same age of diagnosis labeled as “unknown site” and another entry at same age of specimen collection, also labeled as “unknown site”, followed by another surgery at age 30.56 that does have lower extremity diagnosis associated with it. This skin specimen likely represents the initial IIIA diagnosis. Make exception to the rule to assign as stage III.
        + For this patient ID, set Discrepancy = 1. The PRIORDISTANT rule does not apply since this patient will be captured by EXCEPTION rule. This patient is not included in the patient counts for this rule anymore.
      * QC7QX0VWAY: Two melanoma diagnoses: trunk IIB at age 68.518 and lower extremity IIB at age 75.08. Lung met reported at age 71 associated with the trunk melanoma. No other mets reported. Specimen is skin from lower extremity, designated as “primary”, obtained at age 75.08 (time of that diagnosis). Specimen obtained pre-ICB treatment. Since the rule does not distinguish by site (trunk, extremity, etc), it counts the lung met as distant disease for the lower extremity melanoma, and thus assigns stage IV. If this lesion represents a new primary, then this is incorrect. Unable to know from information available. Keep as stage IV with likely new primary in the setting of metastatic disease.
        + For this patient ID, set Discrepancy = 1.
        + For this patient ID, set Possible New Primary = 1.
        + This is the patient that had the correct stage (IV) by your script according to the rules, but an error in my file (key) for this patient. The key has been corrected.

**RULE #6: NOMETS**

* IF Metastatic Disease file field **MetastaticDiseaseInd = “No”,** THEN **AssignedStage** = *numerical* value of PathGroupStage OR ClinGroupStage (if PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”])
* No recorded regional or distant metastatic disease. Rule assigns stage at initial diagnosis.
* *Must add 0.005 to the AgeAtSpecimenCollection due to only 2 decimal points recorded for age in that file, but 3 decimal points recorded on the diagnosis file.*
* 27 patients, all cutaneous
  + 21 patients in group A; 1 patient in group B (92S5RV6HJS), 2 patients in group C (5BS8L7PCCE, L5876HQBT9)
* MetastaticDiseaseInd = No (meaning no regional or distant metastatic disease reported).
  + Except for 1 lymph node specimen (5BS8L7PCCE), all that have MetastaticDiseaseInd = No are skin specimens listed as “Primary”. However, the time between age at specimen collection and age at diagnosis ranges from 0 days to 721 days, including about half that were obtained within 30 days of initial diagnosis. About 70% have clinical or pathologic stage II disease at initial diagnosis. While some could represent new primaries, the skin site of specimen collection is the same as the primary skin site of diagnosis for all but one, and in that one case, the specimen site is “skin NOS”. While some of these may represent new primaries, this rule treats all specimens as related to the initial diagnosis given the limited information to know if these are truly new primaries.
* Three patients with specimens obtained > 90 days:
  + - 9DLKDVIQ2W: initial IIB upper extremity diagnosis; specimen is skin of upper extremity obtained 721 days later; likely represents new primary; given limited information, keep as stage II
      * For this patient ID, set Discrepancy = 1
      * For this patient ID, set Possible New Primary = 1
    - MD5OTA3E8A: initial IIC lower extremity diagnosis; specimen is skin of lower extremity obtained 592 days later; likely represents new primary; given limited information, keep as stage II
      * For this patient ID, set Discrepancy = 1
      * For this patient ID, set Possible New Primary = 1
    - MPHAPLR8K1: initial IIIC face melanoma diagnosis; specimen is skin of face obtained 94 days later; keep as stage III assuming locoregional recurrence (not new primary)
      * For this patient ID, set Discrepancy = 1

**RULE#7: NODE**

**SpecimenSiteofCollection is the correct field**

* IF **SpecimenSiteofCollection** contains [“**lymph node**” OR “**parotid**”], THEN **AssignedStage** = III
  + Assigns stage III to node specimens that remain after the above rules have been evaluated.
  + 100 patients
    - 98 cutaneous
      * 87 patients in Group A
      * 1 patient in Group B: WOW011YH6I
        + This patient was appropriately captured by this rule in your script, but I had an error in my file (key) that had the correct stage (III) but incorrect rule called. The key has been corrected.
      * 9 patients in Group C: 383CIRVHH2, 39TYSJBNKK, 643X8OLYWR, 6RX3G5GV02, 7HU06PZK4Q, ILKRH6I83A, KWLPMWV0FM, RAB7UH51TS, XPZE95IE7I
      * 1 patient in Group D: 59OP5X1AZL
    - 1 ocular patient: 8OR7RX5NO5 (group A)
    - 1 mucosal patient: Z7CEUA8SAJ (group A)
* Confirm with Slingluff; confirmed, all to follow rule (stage III), but all should be marked with Discrepancy = 1.
  + N5Q9122LTG: Lower extremity stage III (clinical, no path information) diagnosed age 36. Specimen is lymph node (NOS) obtained at age 55 (~20 years later). Reported to only have “distant” inguinal nodes at unknown age. Surgery file has multiple entries ranging from age 43 – 57.9 associated with this diagnosis (no information on what the specimens were). No additional information to suggest distant mets. Specimen obtained Pre-ICB treatment. Rule seems appropriate given the limited information.
  + 2X7USSLPJC: Lower extremity stage III (clinical T3cN1bM0, no path information) at age 55.13. Specimen is a lymph node (NOS) obtained at age 55.54. Only one entry in metastatic disease file: distant pelvic nodes at age 55.578 (so about 2 weeks after specimen collection). Surgery file has 3 entries (no information on types of specimens): two done at 55.13 (so likely represents skin + node) and then one done at 55.578. Specimen obtained post-ICB treatment (looks like just one dose ipi/nivo since both meds have start and stop ages at 55.197). Rule seems appropriate given information available.
  + 7YX8AJLMWR: Lower extremity stage IIIB (T4bN1aM0, 7th) disease at age 47.29. Specimen is lymph node (NOS) obtained at age 49.19 (about 2 years after initial diagnosis). Metastatic disease file only has two entries for regional node (inguinal) and regional skin met, both at age 47.468. Multiple surgery file entries ranging from age 47.29 at diagnosis to age 50.23 (no specimen information available). Specimen obtained post-ICB treatment (end age 48.995). Rule seems appropriate given limited information.
  + DTUPUJ06B5: Trunk melanoma IIIA (T2N1aM0, 7th) at age 77. Reported regional inguinal nodes at 77.44 (160 days after initial diagnosis). Specimen is lymph node (“intra-abdominal”) at 77.44 (same age as reported regional inguinal nodes). Two entries in surgery file: one at 77 and the other at 77.44. No ICB treatment. Despite being labeled as “intra-abdominal” nodes, no other information to suggest stage IV disease, so rule seems appropriate with assumption that these were the regional inguinal nodes.
  + X9AZUY3R1C: Lower extremity IIIC (T4bN2cM0, 7th) at age 48.67. Regional inguinal nodes reported at that initial diagnosis. Specimen is lymph node (NOS) obtained at 52.7. No ICB treatment. No additional information to suggest other mets. Rule seems appropriate.
  + 2AP9EDU231: Face melanoma III (T3N2cM0, 8th) at age 62.56. Reported regional nodes of face/head/neck at unknown age. Specimen is node (NOS, non-sentinel) at age 69.66. Surgery file with just one entry at 69.658. Specimen obtained pre-ICB treatment. No other information. Rule seems appropriate given limited information.
  + 6HWEJIP63S: Lower extremity III (T4aN2bM0, 7th) at age 52.9. Reported regional nodes (NOS) at unknown age. Also brain mets at unknown age. Specimen is node (NOS, regional) obtained at 55.14. Specimen is pre-ICB treatment. Rule seems appropriate given limited information.
  + GITAF8OSTV: Lower extremity III (T3bN3Mx, 7th) at age 60.29. Regional node (NOS) reported at unknown age. Specimen is node (NOS) at age 60.72. No ICB treatment. Rule seems appropriate given limited information.
  + XPZE95IE7I: Lower extremity III (T3bN2Mx, 7th) at age 62.87. Regional nodes (inguinal, pelvic) both at unknown age. Specimen is pelvic node obtained at 63.99, post-ICB treatment. Rule seems appropriate given limited information.
  + Z7CEUA8SAJ: Urethra melanoma stage III at initial diagnosis at age 77.78. Specimen is a lymph node (NOS, non-sentinel) obtained at age 77.82 (48 days after initial diagnosis). Reported to have distant inguinal nodes at unknown age. No additional entries in the metastatic disease file or surgery file to suggest additional later distant mets. Rule keeps this stage as III and ignores the “distant” classification for the inguinal nodes. Rule seems appropriate in this case.

**At this point, only cutaneous melanomas with skin/soft tissue/other subcutaneous tissue specimens should remain unassigned.**

**RULE#8: SKINLESS90D**

**This rule needs to use MetastaticDiseaseSite from the Metastatic file and has been edited to reflect that.**

* IF **AgeAtSpecimenCollection+0.005 WITHIN 90 days of AgeAtDiagnosis** AND ***does not have any*** {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND **MetastaticDiseaseInd = [“Yes - Regional” OR “Yes – NOS”]** AND **MetastaticDiseaseSite *contains*** [“skin” OR “ear” OR “eyelid” OR “vulva” OR “breast”] AND **[AgeAtMetastaticSite <= AgeAtSpecimenCollection+0.005]**}, THEN **AssignedStage** = *numerical* value of PathGroupStage OR ClinGroupStage (if PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”])
* Skin specimens collected within 90 days of initial diagnosis and no SKIN metastatic sites reported prior to specimen collection. Rule only specifies Regional and NOS for the metastatic sites since all Distant sites have already been assigned with previous rule. Rule assigns path stage (or clinical stage if no path information) at initial diagnosis.
* Rule allows for regional lymph node metastatic sites prior to specimen collection with the assumption that these will have been captured by the initial staging (since these or skin specimens, not node specimens, which have been previously captured by a prior rule), which appears true for this database. This rule will need to be edited if used in the future for other datasets, since it is possible for node metastases to occur within 90 days of diagnosis that may result in re-staging from initial staging.
* 44 patients, all cutaneous (count not edited since one patient was removed for exception, but one patient was added)
  + 33 patients in Group A
    - You noted an issue with patient ID = YQJCZJ5Z6U, but we determined that this issue should be resolved with the correction to the rule to use MetastaticDiseaseSite instead of SpecimenSiteofCollection.
  + 7 patients in Group B (DVJMQLZBJV, EXW43C3IGC, FWQ9OSPM9E, NJ6HY2U9GC, QMSJQAHVG2, ZFCLZOPWSJ, CVKXM5RBNR,)
    - Patient ID WOW011YH6I was erroneously included in this count. This patient is captured by rule NODE.
  + 4 patients in Group C (2FEX3JHKCW, 8BCBHGEO4N, FNP53S81KA, MD578977I9)
* Confirm with Slingluff; confirmed, details below:
  + APGLZDFLYJ: Trunk melanoma IIIA (T4N1aMx, clinical M0; 7th) diagnosed at 53.09. Specimen is skin of trunk (back), designated as “primary”, obtained at age 53.10 (4 days after diagnosis). Metastatic disease file shows a distant lung met (‘overlapping lesion of lung’) at age 53.162 (23 days after specimen collection/diagnosis), but this met is associated with a scalp/neck melanoma. This scalp/neck melanoma is not listed as a separate diagnosis for this patient. Rule seems appropriate despite close distant disease. Keep as stage IV.
    - For this patient ID, set Discrepancy = 1
  + GEM0S42KIH: Lower extremity III (T3aNxMx, 6th; not sure how stage III by this) diagnosed at age 53.20. Specimen is skin of heel, designated as “primary”, obtained at 53.28 (28 days after diagnosis). Reported distant nodes (axilla) and skin met (trunk) at age 53.299 (1 week after specimen collected). Rule assigns as stage III based on the ages/dates, but likely represents missed metastatic disease, particularly since substaging was incomplete at initial diagnosis. Make exception to rule to assign this as stage IV.
    - For this patient ID, set Discrepancy = 1. The SKINLESS90D rule does not apply since this patient will be captured by EXCEPTION rule. This patient is not included in the patient counts for this rule anymore.

**RULE#9: SKINREG**

**This rule needs to use MetastaticDiseaseSite from the Metastatic file and has been edited to reflect that.**

* IF {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND **MetastaticDiseaseInd = [“Yes - Regional” OR “Yes – NOS”]** AND **MetastaticDiseaseSite *contains*** [“skin” OR “ear” OR “eyelid” OR “vulva”, OR “head” OR “soft tissues” OR “breast” OR “lymph node”] AND **[AgeAtMetastaticSite <= AgeAtSpecimenCollection+0.005]**} OR **does not contain any** {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND **MetastaticDiseaseInd = [“Yes - Distant”]** AND **AgeAtMetastaticSite = “Age Unknown/Not Recorded”**}, THEN **AssignedStage** = III
* These are the remaining skin specimens with reported regional skin or nodal metastases prior to specimen collection. It also covers special cases that don’t fit other rules.
  + No distant metastatic sites at unknown age reported (and no regional disease and the only distant disease is after specimen collection): 2 patients (WLYUAJ9AAE, 9HA9MZCSU2) with initial stage III disease that should remain stage III.
* This works for our database. This rule would need to be edited if used for future datasets.
  + Confirm with Slingluff, confirmed to follow rule: This rule primarily captures those that were initial stage II disease, but now likely represent stage III disease. However, it is possible that some represent new primaries, local recurrence, or distant disease not otherwise recorded, but this cannot be definitively determined from available information.
    - 8 patients for which the specimen really may not represent stage III disease (i.e. either new primary, local recurrence, or distant disease): 6DL054517A (could be distant), 87AJ4KITK8 (could be new primary), 9HA9MZCSU2 (could be new primary or distant skin met), A594OFU98I (could be new primary), FUAZTE7LVQ (could be new primary or distant skin met), HZD0O858UJ (could be new primary), L2R9RJJ88C (could represent new primary), NXPOH3RBWY (could be new primary)
    - These 8 patient IDs should have Discrepancy = 1.
    - Additionally these 7 IDs (87AJ4KITK8, 9HA9MZCSU2, A594OFU98I, FUAZTE7LVQ, HZD0O858UJ, L2R9RJJ88C, NXPOH3RBWY) should have Possible New Primary = 1.
  + 16 patients, all cutaneous
    - 12 patients in Group A
    - 2 patients in Group C (L2R9RJJ88C, VLY2FWYN29)
    - 2 patients in Group D (FUAZTE7LVQ, 7HOWLJKDEM)

**RULE#10: SKINUNK**

* IF {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND **MetastaticDiseaseInd = [“Yes - Distant” OR “Yes – NOS”]** AND **AgeAtMetastaticSite = “Age Unknown/Not Recorded”**}, THEN **AssignedStage** = IV
* Remaining patients with skin specimens that have distant disease reported at an unknown time. This rule assumes that these skin specimens represent stage IV disease.
* This works for our database. For generalizability, this would need to be edited if used in the future with other datasets.
* 7 patients, all cutaneous
  + 6 patients in Group A
  + 1 patient in Group C (AC2EJBKWJO)
* Confirm with Slingluff; confirmed, follow rule to assign stage IV
  + Two patients not only have unknown age of distant mets, but also unknown age of regional skin mets, so these specimens could represent new primary or stage III disease (AC2EJBKWJO (possible new primary), TIZXXXVCV9), but unable to definitively determine with available information
    - These 2 patient IDs should have Discrepancy = 1
    - Patient ID AC2EJBKWJO should have Possible New Primary = 1