# ONCOLOGY PROGRESS NOTES ANNOTATION GUIDELINES

General BRAT instructions: [brat features](https://brat.nlplab.org/features.html)

You will need to log into brat before adding any annotations. You will find the option to log in on the top right of the screen (please hover over that portion to see a login option). Please use the credentials we have shared individually to log in.

## Cohort selection

We will prioritize the notes for annotation using the following order of priority:

1. Identify the cohort of all patients suffering from a given cancer by matching their diagnoses codes and their descriptions.
2. Filter the patients to retain only those who have a history of cancer care on the site. This can be identified by limiting the encounter department specialties to those mentioning either cancer centers or medical oncology departments.
3. Prioritize patients that contain a structured data entry for cancer staging information, if available. This is done to both prioritize patients who are potentially receiving care at UCSF, as well as to potentially train a stage-detection model later.
4. Select a uniform distribution of patients from different stages of the disease (identified using the structured data or cancer registry if available).
   1. Additionally include patients who progress from a lower stage to higher within the care history. This would incorporate more temporal and multi-document information for complex care timelines.
   2. If further filtering of patients needs to be performed (to limit the size of the corpus we annotate), we prioritize patients containing longer notes over shorter ones. This is done under the assumption that longer notes would provide more information than shorter notes. We recognize that this may not always be the case, as longer notes can be an artifact of copy-forwarding as well. However, we do not attempt to circumvent that yet.
5. Always annotate the first progress note within the medical oncology department. This note is assumed to contain the entire history of the present illness of the patient.
6. Additionally, for the same patients, annotate the notes closest to the time/encounter at which staging was done (if different from the first HPI note and if present). This note is presumed to contain the most information required to assign a stage.
7. [For future] For the same patients, annotate the notes when the treatment (regimens) were switched, also as identified from structured data.

## Annotation Guidelines

All the documents will be pre-highlighted with highlights ‘Annotate’ or ‘Skip’ to indicate whether a section should be annotated or can be skipped. The sections of interest may have been mentioned under different headers, which could have been missed by our rule-based pipeline for section detection. The best guesses should be made about the content of a paragraph or section based on their individual headings in the text.

Sections to annotate

We will annotate the following sections in clinical notes:

* Sections describing the patient’s history
  + History of present illness
  + Past Medical History
  + Past Surgical History
  + Subjective
  + Interval history
  + Oncology history
* Assessment and plan (Can also be mentioned as impression and plan)
* Review of systems:  If this is mentioned as semi-structured data, ignore everything that is mentioned as N/A or not on file. Annotate the information that is known.
* Social history narrative: If this is mentioned as semi-structured data, ignore everything that is mentioned as N/A or not on file. Annotate the information that is known.
* Family history: If this is mentioned as semi-structured data, ignore everything that is mentioned as N/A or not on file. Annotate the information that is known.

Sections to skip

* The section ‘Physical Examination” that describes the results of a physical exam (e,g, PHYSICAL EXAM, OBJECTIVE ASSESSMENT, etc.).
* The section containing semi-structured information about medications (E.g. CURRENT (OUTPATIENT) MEDICATIONS).
* The section containing semi-structured content about lab tests (e.g. Lab data, LABORATORY RESULTS, etc.).
* The section containing a review of allergies (e.g. ALLERGIES, Allergen Reactions, etc.).
* Radiology and Pathology sections (e.g. sections IMAGING, RADIOGRAPHIC AND PATHOLOGY RESULTS, SURGICAL PATHOLOGY REPORT, etc.)
  + Any radiology reports pasted within the note directly should be skipped as well.
* Copy forwarded information that has already been annotated.

The entire sections that are skipped should be annotated with the **SectionSkip** entity. The corresponding attribute *‘SectionSkipType’* should be used to indicate the type of section that was skipped.

Annotation schema details

The entity and relation annotations should be completed in the sequential order of documents. All the documents will be pre-highlighted with section headers ‘Annotate’ or ‘Skip’ entities of 3 types: problems, treatments, and tests, as recognized by publicly-available models. These are present as a guide rather than as the entities that we need to annotate ourselves.

The entities comprise everything of interest in the categories mentioned below in this schema. For all the entities, the minimum span that conveys the requisite information should be annotated. This would mean that adjectives that express additional information about an entity or an event should be retained, for example, “serious”, however, articles such as “a”, “an”, “the” or other extraneous information should be excluded. To the extent feasible, a single span should be used for a single entity. Abbreviations should be annotated in an equivalent manner to their expanded forms.

All entities should specify details of negation and modality markers (*NegationModalityVal*), as well as temporality (as a *Temporal* relation). The attribute *NegationModalityVal* takes a few different values: negated, affirmed, uncertain\_in\_present, uncertain\_in\_past, planned\_in\_future, and hypothetical\_in\_future. ‘Affirmed’ would be used by default whenever no other value has been set, so please take special care for indicating other values when relevant.

* **Datetime:** Includes date and time expressions, as well as frequency such as “December 2019”, “12th December”, “2pm” as well as indirect references to time like “last year” or “2 months ago”. A single span should be added for the entire DateTime expression. The minimum span that conveys the information should be used. For example, “last year” should be annotated instead of “in the last year”. The attribute *‘DatetimeVal’* is used to additionally mark if the mentioned DateTime is of types duration, frequency (for example *weekly*), age, or others (default). The *Temporal* relation should be used to relate any entities with the DateTime relations.
* **PatientCharacteristics** This generic category of PatientCharacteristics should be used for all mentions of relevant information other than Symptoms, ClinicalCondition, social determinants for health factors, and any other information explicitly listed as entities ahead. Some examples would be information about menopause, number of children, etc. that may be relevant for the cancer in question, or that may be relevant as an inclusion or exclusion criteria for a treatment or a clinical trial. *SiteOf* relation can be added for all symptoms and clinical conditions to add information about their site.
  + **Symptom:** This includes all mentions of symptoms and complaints that a patient presents with, for example, *fatigue, nausea, breathing difficulties*. Minimal entity spans that convey the information should be annotated, for example, *nausea* should be annotated instead of *the patient presents with nausea*. If the severity of the symptom is also mentioned, if relevant, should be included with the symptom entity (for example *minor breathing difficulties*).
    - The attribute *ContinuityVal* should be set to indicate whether it is a new symptom, a stable symptom, or whether the symptom has improved, worsened, or stopped.
    - The attribute *ChronicVal* should be set to indicate whether it is chronic or non-chronic.
    - Furthermore, the attribute *IsPresentOnFirstCancerDiagnosis* should be selected to indicate that the symptom was present when the patient was diagnosed with a new tumor for the first time. For this attribute, the value ‘unclear’ would be assumed by default if either yes or no is not indicated.
    - The attribute *IsCausedByDiagnosedCancer* should be set to indicate that the symptom was caused due to a cancer diagnosis. For this attribute, the value ‘unclear’ would be assumed by default if either yes or no is not indicated.
    - The attribute *NegationModalityVal* should be used to indicate any negation and modality markers jointly with all the entities.

The relation *TreatmentAdministeredForProblem* should later be added between symptom entities and treatment names if the treatment was administered for a given symptom.

The relation *TreatmentDiscontinueBecauseOf* should later be added to relate treatment entities with symptoms if the symptoms were the reason for discontinuing a treatment.

* **ClinicalCondition:** This includes the diagnosis attached to the symptoms that the patient has presented with. This is usually stated by the provider and not the patient. Conditions like breast cancer and colon cancer are also included in this category. The attribute *ExperiencerVal* should be set to indicate whether the experiencer is the patient, family of the patient, or others. We can use ‘patient’ as experiencer by default, so please take special care to indicate this attribute if the experiencer is anyone other than the patient. Furthermore, the attribute *ChronicVal* should be set to indicate whether the clinical condition is a chronic condition or a non-chronic condition. The attribute *ContinuityVal* should additionally be set to indicate whether it is a new condition, continuing condition, whether the condition has progressed, or it has stopped. As always, *NegationModalityVal* should be used to highlight negation/modality.

The relation *TreatmentAdministeredForProblem* should later be added between clinical condition entities and treatment names if the treatment was administered for a given clinical condition.

The relation *TreatmentDiscontinueBecauseOf* should also be added to relate treatment entities with a clinical condition if it was a reason for discontinuing a treatment.

If a clinical condition causes a symptom or another condition, the relation *ConditionOrTreatmentCausesProblem* should be added.

* **Allergy:** Any existence of allergy within the sections that are annotated should be marked as the Allergy entity. The same relations as Symptom and ClinicalCondition are also present for allergy. The attribute *ContinuityVal* is used to determine whether it is new, stable, improving, worsening, or stopped.
* **SDoH:** This includes different social determinants for health factors. The broad-level category of SDoH should be used when a more specific category (Alcohol, Drugs, Tobacco, ConsumptionQuantity, PhysicalActivity, Employment, LivingCondition, Insurance, SexualOrientation, and MaritalStatus) is not present. The attributes *NegationModalityVal* and *ExperiencerVal* should be used as needed. The attributes *IsStoppedOrContinuing* (default: continuing)should be used to indicate whether these values are in the past (for example, the patient has stopped drinking), or continuing in the present (the patient drinks).
  + **Alcohol:** Mentions of alcohol use.
  + **Drugs:** Mentions of drug use.
  + **Tobacco:** Mentions of tobacco consumption in any form. Smokeless tobacco would be included in this category in addition to smoking. Vaping, if it contains nicotine, would be included.
  + **PhysicalActivity:** All phrases describing physical activity should be marked as such, for example, swimming, walking 2 miles per day, etc.
  + **Employment:** Mentions of employment status. This can include mentions of the profession, for example, “doctor”, or mentions of employment status, for example “employed in the private sector”. The entity spans should cover the details of the profession when available, but no extra information.
  + **LivingCondition:** Information about the living condition of the patient, for example, “lives alone”, “lives with family”, “homeless”, etc.
  + **Insurance:** Any information about the insurance, for example, information about the insurance plan, or the presence or absence of insurance.

The relation *TreatmentDiscontinueBecauseOf* should later be added to relate treatment entities with insurance status if it was a reason for discontinuing the treatment.

* **SexualOrientation:** Any mention of the patient’s sexual orientation.
* **MaritalStatus:** Any mentions of the patient’s marital status.
* **SDoH Modifier:** Any important modifiers for SDoH, particularly those not listed explicitly, should be annotated as this entity. For example: type of tobacco use. These entities should further be related with the corresponding SDoH with an *SDoHDesc* relation.
  + **ConsumptionQuantity:** Any mentions of consumption quantity for alcohol, drugs, or tobacco, for example, 3 packs per day or 1 pint of beer a day. Any mentions of consumption frequency should also be encoded within the same entity.
  + A relation between this entity and the corresponding Alcohol/Drugs/Tobacco entity should be added at the subsequent step (*ConsumptionQuantityRel*).
* Location: The attribute *NegationModalityVal* should be used for the following entities as needed.
  + **Site**: This includes the site (organ/body part) of a test, the site of the tumor on the body, or the site of a surgical procedure, for example ‘breast’. If laterality is present, it should be annotated separately.  For example, in “left breast”, only “breast” should be annotated as the site. When the exact site is mentioned, for example, *10 o’clock from the nipple*, the entire minimal span that conveys the information should be annotated. If the exact location of a specimen on an organ is not present, only the broader topography should be annotated, in line with the ICD-O3 topography guidelines [here](https://athena.ohdsi.org/search-terms/terms?vocabulary=Cancer+Modifier&vocabulary=ICDO3&page=1&pageSize=15) (or in the following guide:<https://apps.who.int/iris/bitstream/handle/10665/96612/9789241548496_eng.pdf>; starting on page 44; breast cancer on page 54).

Site is later related to the name of the test/procedure/tumor that it is a site of using the relation *SiteOf*.

If the site is mentioned in an ambiguous manner, such as ‘left lumpectomy’, please annotate ‘left’ as the site and not laterality.

* **Laterality:** Laterality of the site where either a test was conducted, where the tumor was found, or where a surgical procedure was performed. for example ‘left’ in ‘left breast’.

Laterality should later be related to the corresponding site (*LateralityOfSite*).

* **TumorTest:** This includes any tests conducted to either screen or confirm whether a tumor is present. The generic category of TumorTest will be used only when a test specification is not precisely either covered under Radiology, Pathology, GenomicTest, or DiagnosticLabTest. The site/laterality of the test should be annotated as separate entities and should not be a part of this same entity called TumorTest.

The attribute value under *IntentVal* should be used for each TumorTest to indicate whether the test was conducted for screening, staging, treatment assessment, or other reasons. Similarly, the attribute value under *VenueVal* should be used to indicate whether the test happened at UCSF, at other locations, or is unclear (default). If the tests would be conducted in the future, the attribute value ‘hypothetical\_in\_future’ should be used for *NegationModalityVal*. If the tests were negated, for example, “did not undergo a biopsy”, “biopsy” should be annotated, and the negation attribute should be selected under *NegationModalityVal*.  The attribute *ExperiencerVal* should be used to indicate whether the experiencer is the patient (default), family, or others.

(These relations are also discussed later at the end of the document: ) The site of the test should be added with the relation *SiteOf*.

If a test was conducted in response to an existing problem, tumor characteristics, or disease progression, the relation *TestOrProcedureConductedForProblem* should be used to relate them.

The details of the test results should be related to this entity using the relation *ResultOfTest* described later.

If a test reveals a malignant tumor, clinical condition, or disease progression, the test name should be related to the relevant entity using the relation *TestOrProcedureReveals*.

* **Pathology:** Any mentions of pathology tests conducted should be annotated. If the type of biopsy is specified, for example, “core needle biopsy”, it should be annotated as a part of the same entity.
  + Specific cases:
    - FISH (Fluorescence in situ hybridization): Please label this as a Pathology test
* **Radiology:** The annotations should follow similar strategies as that of the Pathology annotations. All screening, as well as diagnostic tests, should be annotated. For example, “Mammography”, “Ultrasound”, “US”, ‘MRI’, ‘CT’ should all be annotated under this category.
* **DiagnosticLabTest:** This includes lab tests that are specifically conducted to diagnose a tumor. Some examples include PSA, CEA, CA99, AFP, etc.
* **GenomicTest:** This entity should be added for all types of gene tests. Additionally, the attribute*GenomicTestType* should be added to indicate whether the genomic test is of types germline test, somatic test, gene expression test (for example Oncotype Dx or Mammaprint), others, or the type unclear from text. Remember that for any hypothetical discussions of tests to be conducted in the future, the *NegationModalityVal* attribute value “hypothetical\_in\_future” should additionally be used.
* **TestResult**: This is the generic category for the tumor test result. This category should be reserved for any tests except pathology, radiology, genomic test, or diagnostic lab test. The attribute *ExperiencerVal* should be used to indicate whether the test result is about the patient (default), family, or someone else.

The relation *ResultOfTest* should be used to relate the test name with its corresponding result (this result can either be the generic category of TestResult, or a more specific category of RadPathResult, GenomicTestResult, or LabTestResult).

* **RadPathResult:** This category is for results related to both pathology and radiology tests. Attribute value *RadPathResultVal* should be set in addition to highlighting the corresponding entity to indicate whether the test result refers to the no disease, initial disease diagnosis, stable disease, disease progression, treatment response, mixed response (for example one tumor is improving, the other is worsening), others, or is unclear.
* **GenomicTestResult:** This entity is for all results of genomic tests. The attributes *NegationModalityVal*, *ExperiencerVal,* and *GenomicTestType* should be set to indicate the corresponding relevant values.
* **LabTestResult**: This entity is reserved for annotating the results of DiagnosticLabTest. In addition, the attribute value *LabTestResultVal* should be selected to indicate whether the test is ‘normal’, ‘abnormal’, or ‘unclear’.
* **TumorCharacteristics**: This includes any modifiers not covered in the list mentioned below. The attribute *ExperiencerVal* should be set to indicate whether the patient (default), family members, or others have these tumor characteristics. Modifiers of old tumors should be accompanied with the value “history” of the attribute *HistoryVal* (default value: new).  Similarly, all TumorModifier entities should also indicate an *EpisodeDescription* attribute after deciphering whether it is the histology at the “FirstOccurrence” (default), “progression”, “recurrence” or “others”. Any negation and modality should be expressed with the *NegationModalityVal* attribute (default: affirmed).

The relation *TestOrProcedureReveals* would later be added for any tumor properties that have been identified with the help of a test or a procedure.

The relation *TestOrProcedureConductedForProblem* should later be added for any test or procedure that was conducted for any tumor characteristics.

The relation *SiteOf* should be used to relate site entities to the corresponding tumor properties.

The relation *TreatmentAdministeredForProblem* should later be added to describe any relations between treatments administered for any of the tumor characteristics.

* **Histology**: This includes the histology of the tumor either revealed by a test or the histology of a tumor in a patient’s history. Unless a more specific description is available, the tumor histology should also follow the ICD-O3 histology, which can be referenced [here](https://athena.ohdsi.org/search-terms/terms?vocabulary=Cancer+Modifier&vocabulary=ICDO3&page=1&pageSize=15) (and by using ‘ICD-O3’ ‘Search’ option after selecting the relevant text span in BRAT).
* **Metastasis:** Entity to annotate the mentions of tumor metastasis (to other locations than lymph nodes) in text. If metastasis is negated, the corresponding attribute for *NegationModalityVal* should be selected. Relations should later be added to describe details related to the metastasis event, including the Datetime of metastasis (*Temporal*), site of metastasis (*SiteOf*), and the corresponding tumor histology, if any (*MetastasisDesc*).
* **LymphNodeInvolvement**: Entity to highlight mentions of lymph node involvement in text. The number of lymph nodes involved should be included as a part of this entity if mentioned, same for micrometastasis or any other relevant specific detail. The site of lymph nodes should not be included as a part of this entity, but should separately be annotated as Site, and later related to this entity using the *‘SiteOf'* relation. Note that we will not annotate the lymph nodes examined at this stage, only the lymph nodes involved.
* **Stage:** Numeric mentions of the tumor stage, as well as mentions like “early” stage, should be annotated here, for example, ‘IV’ in “Stage IV’. The prefix “Stage” should not be annotated, and only the number or the stage description should be. If the stage is mentioned as a range, for example ‘Stage I-IIA’, ‘I-IIA’ should be annotated.
* **TNM:** The TNM stages should be annotated here. This includes all TNM stages either as a single entity or as individual entities. For example, ‘T2a’, ‘pT1N0M0’, ‘pN0’, ‘pMX’, ‘pT1N1M0(i+)’, ‘ypT1’, ‘cT1N1M1’ should all be annotated. Note that we include both pathological and clinical stages, along with details of TNM such as prefixes for change of stage after therapy (y), as well as suffixes about the procedure, such as (i+).
* **Grade:** Tumor grade. This includes both numeric values such as ‘1’ (do not include the prefix ‘grade’), as well as textual mentions such as ‘low’ in ‘low grade’, ‘intermediate’ in ‘intermediate grade’, etc.
* **Size:** The size of the tumor, along with the corresponding units of measurement, should be annotated. For example ‘7\*9\*5mm’ should be annotated as a single entity.
  + Size can be for any pathological/surgical/radiological procedure and then related to that procedure through *TestOrProcedureReveals*.
* **LocalInvasion**: Annotate all mentions of whether there is a local invasion. This would include LVI (microscopic), as well as other types of invasion, for example, ‘invades chest walls’ etc.
* **BiomarkerName:** Any names of cancer biomarkers should be annotated. For the sake of generalization, our definition of a biomarker is very loose. For example, we would include ‘ER’, ‘PR’ as well as ‘HER-2’ as biomarker names for breast cancer. PSA for prostate cancer would be annotated as a DiagnosticLabTest and not a biomarker. Please annotate only the final test results if the test was conducted multiple times. If the results are present as history, please highlight it as the corresponding attribute value *‘HistoryVal’*. Please note that any terms indicating the results, such as ‘+’ in ‘ER+’ should be annotated as a separate result and would not be a part of the BiomarkerName entity. Any details of the test should not be included in the annotation span and instead should be labeled as pathology test.

Immunohistochemistry stains should be labeled here under biomarker. The name of the gene or protein that is being tested for should be the selected entity

These results would be related with the corresponding entities using the *BiomarkerRel* temporalrelation.

* **BiomarkerResult:** The results of a biomarker, for example positive, negative, equivocal, ‘+’, ‘-’ etc. should be annotated as the BiomarkerResult. Any percentages, if mentioned, should be included as a part of the result, for example, ‘+ 100%’. The attribute value *BiomarkerResultVal* should additionally be used to indicate a positive, negative, low positive, or unclear result.

A relation *BiomarkerRel* between the BiomarkerName and the BiomarkerResult should subsequently be added to link different biomarker names with biomarker results.

* Procedure:
  + **ProcedureName:** Name of a procedure. This includes all types of diagnostic procedures, screening procedures, as well as treatment-related surgeries, such as ‘lumpectomy’, ‘mastectomy’, ‘colonoscopy’, etc. All procedures get the attribute *IntentVal* to indicate the intent of the procedure: whether it is for screening, staging, diagnosis, treatment assessment, treatment (curative, palliative, or others), diagnosing as well as treatment, or others. *TreatmentTypeVal* entity should be set to indicate whether it is an adjuvant procedure, neoadjuvant, maintenance, local, or other procedures (if the procedure is related to treatments). If the procedure is being discussed for the future, the corresponding *NegationModalityVal* attribute should be set. *IsTumorRemaining* entity should be used to indicate whether any tumor was remaining after the procedure or not if the procedure was to remove the tumor (i.e. treatment procedure). This attribute can be left blank if the procedure was not a procedure done with the intent of treating the patient. The attribute *TreatmentCategory* should be used to indicate whether it is an antineoplastic treatment (default), supportive treatment, or others.

If a procedure was conducted in response to an existing problem, tumor characteristics, or because of disease progression, the relation *TestOrProcedureConductedForProblem* should be added.

 If a procedure reveals a malignant tumor, clinical condition, or disease progression, the ProcedureName name should be related to the relevant entity using the relation *TestOrProcedureReveals*.

If any problem such as a symptom or a clinical condition was caused by a procedure, the relation *ConditionOrTreatmentCausesProblem* should be added.

* **ProcedureModifier:** Any modifiers for the procedure that are not described by either the outcome or the margin status should be annotated as this generic entity.

For all the modifiers (including ProcedureOutcome and MarginStatus), the relation *ProcedureDesc* should be added to describe which procedure are they modifiers for.

* **ProcedureOutcome**: Any outcome or result of a procedure, such as results of a biopsy that is not pathology can be annotated as the ProcedureOutcome entity. This can further be related to the ProcedureName entity.
* **MarginStatus:** Value of the surgical margins, along with corresponding units if any. If there are multiple resections, only the final margin should be annotated. The MarginStatus entity takes the attribute *MarginVal* to indicate whether the margin is positive, more than 2mm (negative), less than 2 mm, or unclear. Only the final margin should be annotated in the case of multiple mentions.
* Treatment:Treatments, broadly categorized under the following categories: MedicationName, MedicationRegimen, RadiationTherapyName, or TreatmentType. For all of these entities, a few attributes should be set. The attribute *TreatmentContinuityVal* should be used to indicate whether a treatment was started (immediately), planned (in the future; not immediately), finished, discontinued early, or is continuing currently. The default value “started” would be used if not indicated otherwise. Similarly, under the attribute *TreatmentIntentVal*, we should add the intent of therapy: whether it was curative, palliative, or others. The default value of “curative” would be used if not indicated otherwise. The attribute *VenueVal* should be used to indicate whether the treatment happened at UCSF, at other locations, or is unclear (default). The default value ‘UCSF’ would be used unless indicated otherwise, so special care should be taken that any treatments outside UCSF should be indicated as such. The attribute *TreatmentTypeVal* should be used to indicate whether the treatment is neoadjuvant, adjuvant, maintenance, local, or others. A relation should be added between all treatment entities and their corresponding modifiers (*TreatmentDesc*). The attribute *NegationModalityVal* should be used as needed. Relation (*TreatmentAdministeredForProblem*) should be added between treatment names and the clinical condition, symptoms, tumor characteristics, or disease progression entities that they have been administered for.
  + **MedicationName**: This includes the name of all cancer-specific medication therapies for all modes of administration, including chemotherapy, hormone therapy as well as immunotherapy (but not radiation therapy). Every individual medication name should be annotated as a separate entity. Medications administered for managing secondary symptoms should be annotated as “supportive medications. There is a separate category for medication modifiers like dosage, so only the name should be included under this entity. An attribute *TreatmentCategory* should be set to indicate whether it is an antineoplastic medication, supportive medication, or others (others can be skipped unless it seems to be particularly important).

If any problem such as a symptom or a clinical condition was caused because of a medication, the relation *ConditionOrTreatmentCausesProblem* should be added between the problem and the medication name.

If a medication was discontinued because of a symptom, clinical condition, disease progression, insurance, or hospice, the relation *TreatmentDiscontinuedBecauseOf* should be added between the reason and the medication name.

* **MedicationRegimen**: If mentioned explicitly in the text, add the name of the regimen for a medication. If the text also mentions the name of medications under this regimen, a relation (*RegimenForName*) should be added between the MedicationRegimen and the MedicationName entities.

If any problem such as a symptom or a clinical condition was caused because of a medication regimen, the relation *ConditionOrTreatmentCausesProblem* should be added between the problem and the medication regimen.

If a medication regimen was discontinued because of a symptom, clinical condition, disease progression, insurance, or hospice, the relation *TreatmentDiscontinuedBecauseOf* should be added between the reason and the medication regimen.

* **RadiationTherapyName:** This includes the name of the radiation therapy.

If any problem such as a symptom or a clinical condition was caused because of radiation therapy, the relation *ConditionOrTreatmentCausesProblem* should be added between the problem and the radiation therapy name.

If radiation therapy was discontinued because of a symptom, clinical condition, disease progression, insurance, or hospice, the relation *TreatmentDiscontinuedBecauseOf* should be added between the reason and the radiation therapy name.

* **TreatmentDosage**: This includes dosage for medications as well as RadiationTherapy. The units for dosage should also be included. If the discussions are related to dose reduction instead of explicit mention of dosage, for example 20%, that should be annotated as the next entity, ‘TreatmentDoseModification’. A relation should later be added between the TreatmentDosage and their corresponding name (*TreatmentDesc*).
* **TreatmentDoseModification:** As mentioned earlier, this includes modifications to the usual dosages, for example, ‘20%’, ‘200%’, or ‘3/4th’. A relation should later be added between the TreatmentDoseModification and their corresponding name (*TreatmentDesc*).
* **TreatmentType:** The type of therapy, for example, adjuvant therapy, chemotherapy, etc. An attribute value to indicate whether they are neoadjuvant, adjuvant, maintenance, local, or another type of therapy is also provided as *‘TreatmentTypeVal’*.
* **MedicationModifier:** Any modifier apart from those mentioned next, as well as apart from dosage, and dose modification, should be annotated as the generic entity MedicationModifier. A relation should later be added between all modifiers and their corresponding name (*TreatmentDesc*). The duration and frequency can be annotated as *Datetime* expressions, with the right type set under the *DatetimeVal* attribute.
  + **Cycles:** Number of cycles that a medication is administered for, for example, ‘5’ in ‘5 cycles’. Please only highlight the number, and none of the terms that talk about the cycle itself. If it is a range, please annotate the entire range.
    - The attribute *CycleType* should be added to provide information about whether the number of cycles annotated is the ongoing cycle number, the number of completed cycles, or the target number of cycles.
* **RadiationTherapyModifier:** Any modifiers for the RadiationTherapy apart from dosage, dose modification, duration, and type (adjuvant, neoadjuvant, maintenance, local, others), should be highlighted as this entity. A relation should later be added between the RadiationTherapyModifier and RadiationTherapyName (*TreatmentDesc*).
* **ClinicalTrial:** Names of clinical trials discussed in the text should be annotated. These names should be modified with the *NegationModalityVal* attribute as needed to indicate either any hypothetical discussions or to indicate whether the patient participated in the trial or not.

Furthermore, the relations *‘InclusionCriteriaFor’* and *‘ExclusionCriteriaFor’* should be used to relate all other entities to the trial if they qualify as inclusion and exclusion criteria respectively. There is an additional option to add temporal relations for ClinicalTrial to indicate the DateTime of the trial.

* DiseaseState:
  + **Remission:** This entity should be selected for any mentions of tumor remission. If remission is negated, the corresponding attribute for *NegationModalityVal* should be selected.
  + **DiseaseProgression:** Text snippets suggesting disease progression should be highlighted as this entity, for example, “progression”. If progression is negated, then this entity should be marked as well, along with its corresponding *NegationModalityVal* attribute value.

This entity is used for adding many different relations, such as whether a test or procedure was conducted due to disease progression (*TestOrProcedureConductedForProblem*), whether a test or procedure revealed disease progression (*TestOrProcedureReveals*), whether a treatment was discontinued due to progression(*TreatmentDiscontinuedBecauseOf*), whether a treatment was administered because of disease progression *(TreatmentAdministeredForProblem),* or to indicate *InclusionCriteria, ExclusionCriteria, and Temporal relations.* These relations should be added in the second phase.

* **Hospice**: Any mentions of hospice should be annotated. The attribute *NegationModalityVal* should be set to indicate whether it is negated.

A relation should be added between entities elucidating the treatment name and the Hospice entity if Hospice is a reason for treatment discontinuation (*TreatmentDiscontinuedBecauseOf*).

* **UnspecifiedEntity:** This category will be used whenever none of the existing entities match what should be annotated, but the information is relevant and should be added to the schema. Please use the comments box to describe this unspecific entity: i) what it is, and ii) why it is relevant.

**Relation annotations**

All relations are symmetric and can be added in either direction.

* **Temporal relation for all entities:** The following relations are allowed between the (Datetime) entity and all other entities in our data. Different types of temporal relations allowed are described next. The detailed description of these relations can be accessed in the [THYME annotation guidelines](https://clear.colorado.edu/compsem/documents/.svn/text-base/THYME%20Guidelines.pdf.svn-base) under ‘TLINK’ descriptions (Section 6.2).
  + Event **BeginsOnOrAt** Datetime (BEGINS)
  + Event **EndsOnOrAt** Datetime (ENDS)
  + Event **HappensAtOnDuring** Datetime (DURING/CONTAINS)
  + Event **HappensBefore** Datetime (BEFORE)
  + Event **HappensAfter** Datetime (AFTER)
  + Event **HappensOverlapping** Datetime (OVERLAP). This refers to the situation when an event happens overlapping a datetime mention. These are more common in relations between two events though.
* **Descriptive relations:**
  + **For SDoH:**
    - **ConsumptionQuantityRel:** Relation between the (ConsumptionQuantity) entity and (Alcohol, Drug or Tobacco) entities.
    - **SDoHDesc**: Relation between any SDoHModifier and the SDoH entities.
  + **Relations related to site and laterality:**
    - **LateralityOfSite:** Relation from the (Laterality) entity to the (Site) entity, which indicates the laterality for the mentioned site.
    - **SiteOf:** Relation from the (Site) entity to any of the test name or result, procedure name or result, tumor modifier, and problem entities (TumorTest, Radiology, Pathology, GenomicTest, DiagnosticLabTest, ProcedureName, TestResult, RadPathResult, GenomicTestResult, LabTestResult, Histology, Metastasis, LymphNodeInvolvement, Stage, TNM, Grade, Size, LocalInvasion, BiomarkerName, TumorCharacteristics, Symptom, Clinical condition).

Note that if multiple sites of tests exist and different sites have different results, this relation should be linked to the test result in addition to the test name.

We have allowed this relation for all tumor modifiers, instead of only histology, so that they can be used if needed. For example, there can be cases where the test conducted isn’t mentioned again when (site of) tumor modifiers are mentioned, or the site for the test isn’t mentioned, although the site for a biomarker is present in a note.

* **ResultOfTest:** Relation from results of tests, which includes either of these entities: (TestResult, RadPathResult, GenomicTestResult, LabTestResult) to the entities describing diagnostic tests conducted for finding tumors (TumorTest, Radiology, Pathology, GenomicTest, DiagnosticLabTest, ProcedureName). Note that relations with procedure name are added as this type only if the procedure provides a RadPathResult, a GenomicTestResult, or a LabTestResult. Any other procedure results would be annotated as the ProcedureOutcome entity, and related with the ProcedureDesc relation instead.
* **ProcedureDesc:** This relates a procedure that was conducted (either surgical or diagnostic) and its modifiers. The relation is added between entity pairs (ProcedureName) and the entities (MarginStatus, ProcedureOutcome, ProcedureModifier).
* **BiomarkerRel:** Between (BiomarkerName) and its result (BiomarkerResult).
* **TumorDesc:** A bidirectional relation between the entity TumorCharacteristics (used to indicate terms like cancer) to all tumor modifiers (Histology|Metastasis|LymphNodeInvolvement|Stage|TNM|Grade|Size|LocalInvasion|BiomarkerName). This relation should be added particularly if a test or procedure was not conducted for this tumor and there exists a dangling relation between tumor modifiers and its parent otherwise.
* **TreatmentDesc:** Between Treatment name entities (Treatment, MedicationName, MedicationRegimen, RadiationTherapyName, TreatmentType) and Treatment modifiers (TreatmentDosage, TreatmentDoseModification, MedicationModifier, Cycles, AdministrationMode, RadiationTherapyModifier).
* **RegimenFor:** Relation between (MedicationRegimen) and (MedicationName or TreatmentType) entities.
* **Advanced relations:**
  + **TestOrProcedureConductedForProblem:** Relation that indicates that a test was conducted because of a symptom or a clinical condition that the patient presented with. This relation is added between the entities (TumorTest, Pathology, Radiology, GenomicTest, DiagnosticLabTest, ProcedureName) and (Symptom, ClinicalCondition,  Allergy, DiseaseProgression,  Histology, Metastasis, LymphNodeInvolvement, Stage, TNM, Grade, Size, LocalInvasion, BiomarkerName, TumorCharacteristics).
  + **TestOrProcedureReveals:** Relation between different test entities (TumorTest, Pathology, Radiology, GenomicTest, DiagnosticLabTest, ProcedureName) and the entities that have revealed a clinical condition, disease progression, or tumor characteristics (Histology, Metastasis, LymphNodeInvolvement, Stage, TNM, Grade, Size, LocalInvasion, BiomarkerName, TumorCharacteristics, DiseaseProgression, ClinicalCondition, Allergy).
  + **TreatmentDiscontinuedBecauseOf:** Relation that indicates the reason for discontinuing treatment, if any.  This relation is added from the entity group (MedicationName, MedicationRegimen, RadiationTherapyName) to the reasons, which encompass most other entities.
  + **ConditionOrTreatmentCausesProblem:** This relation is added to indicate that any treatment caused a new symptom or clinical condition. The relation would be added between entity pairs (MedicationName, MedicationRegimen, RadiationTherapyName, ProcedureName, ClinicalCondition) and (ClinicalCondition, Symptom, Allergy).
  + **TreatmentAdministeredForProblem:** This relation is added to indicate that the treatment was administered for a problem (symptom, clinical condition, allergy, or tumor) that was experienced by a patient. The relation would be added between entity pairs (MedicationName, MedicationRegimen, RadiationTherapyName) and (ClinicalCondition, Symptom,  Allergy, DiseaseProgression, Histology, Metastasis, LymphNodeInvolvement, Stage, TNM, Grade, Size, LocalInvasion, BiomarkerName, TumorCharacteristics). If any of the modifiers is not mentioned along with treatment, then no need to relate it across several notes.
  + **NotUndergoneBecauseOf:** Relation to indicate that a treatment, test, procedure, or trial was not undergone because of any reason such as insurance, distance (annotated as PatientCharacteristics), disease progression, etc. This relation can be added between the entities (MedicationName|MedicationRegimen|RadiationTherapyName|TreatmentType|TumorTest|Radiology|Pathology|GenomicTest|DiagnosticLabTest|ProcedureName|ClinicalTrial) and **MOST** other entities.
  + **Relations about clinical trials:**
    - **InclusionCriteriaFor**: This should be used to relate ALL other entities to either of the entities (ClinicalTrial, MedicationName, MedicationRegimen, RadiationTherapyName, ProcedureName) if they qualify as the inclusion criteria for this trial/treatment/procedure. If multiple entities make up a single inclusion criterion, all of them should be related to the trial name individually. The only exception is when there exists a transitive relation between the entities, for example when medication dosage is linked to medication name, we can only link medication name to the trial.
    - **ExclusionCriteriaFor**: This should be used to relate ALL other entities to either of the entities (ClinicalTrial, MedicationName, MedicationRegimen, RadiationTherapyName, ProcedureName) if they qualify as the inclusion criteria for this trial/treatment/procedure.