ShARe Guidelines for the Annotation of Modifiers for Disorders in Clinical Notes

Latest draft – January 24, 2012
Original draft – January 2011
Primary Author – Noémie Elhadad
Participants – Guergana Savova, Wendy Chapman, Glenn Zaramba, David Harris, Amy Vogel

ShARe/CLEF eHealth 2014 shared task additions – Danielle Mowery and Sumithra Velupillai
Latest additions draft – Dec 13, 2013
Original additions draft – Dec 2013

A little bit of relaxation before we start…

Table of Contents

Overview

The goal of this annotation task is to go through clinical notes, find mentions of disorders and additional information related to the disorders, and annotate them.

Annotating **a disorder** consists of several steps:

(1) **identifying a span of text** in the note that corresponds to the mention of the disorder and **mapping it to a CUI** (Concept Unique Identifier) from the provided terminology (SNOMED CT). The choice for the span of text and the CUI corresponding the disorder are tightly coupled, as the decision for one affects the decision for the other.

(2) within the sentence where the disorder is mentioned, identifying additional information related to the disorder, referred to as **modifiers**. Depending on the modifier, the goal is to **identify a span of text for the modifier and normalize it** by either mapping it to either a CUI or a pre-defined set of choices.

The remainder of this document goes through guidelines for annotating mentions of disorders.

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

For this task, participants will be provided an empty template for each disease/disorder mention; each template consists of the mention's Unified Medical Language System concept unique identifiers (CUI), mention boundaries, and unfilled attribute: value slots (modifiers described above). Participants are asked to develop attribute classifiers that predict the value for each attribute: value slot for the provided disease/disorder mention. There are two attribute: value slot types: normalization and cue. For each attribute in the guidelines, these gray boxes will denote assumptions, defaults, and examples expected for leveraging these annotations for the ShARe/CLEF eHealth 2014 Task 2.

Normalization Note: Final system evaluation will be case insensitive: NULL=null=Null

Cue Note: Contiguous (Disjoint) spans are denoted as follows: START-END (START-END,START-END)

1. Identifying a Mention of a Disorder (span and CUI)

1.1. A disorder mention is a concept in SNOMED-CT part of the Disorder Semantic Group

We define a **disorder mention** if as any **span of text** that can be **mapped to a concept** in the **SNOMED-CT** terminology, which belongs to the **Disorder semantic group**. A concept is in the Disorder semantic group if it belongs to one of the following UMLS semantic types:

- Congenital Abnormality
- Acquired Abnormality
- Injury or Poisoning
- Pathologic Function
- Disease or Syndrome
- Mental or Behavioral Dysfunction
- Cell or Molecular Dysfunction
- Experimental Model of Disease
- Anatomical Abnormality
- Neoplastic Process
- Signs and Symptoms

Please be aware that your notion of what a disorder is may be quite different than what is allowed by this set of semantic types. Be careful not to let your prejudice about what you consider a disorder get in the way of annotating disorder mentions.

Note that this definition of the Disorder semantic group differs slightly from the original definition from the UMLS, as we do not consider the Findings semantic type.

In our current annotation project, we rely on the **SNOMED-CT** terminology, part of the **UMLS version 2011AA**.

Example 1.1.a

This is a gentleman who presents with colon cancer.

The span "colon cancer" is a disorder mention. It corresponds to CUI C0007102 (Preferred term: Malignant tumor of colon) in SNOMED-CT, which is under the semantic type "Neoplastic Process."

Example 1.1.b

No chest pain.

The span “chest pain” is a disorder mention. It corresponds to CUI C0008031 (Preferred term: Chest Pain) in SNOMED-CT, which is under the semantic type “Sign or Symptom.”

Example 1.1.c

The patient's glucose is 107.

There is no disorder mention in this sentence.

Example 1.1.d

The patient was continued on nebulizers while she was admitted.

There is no disorder mention in this sentence.

1.2 A disorder mention is a reasonable synonym of the lexical form of a SNOMED-CT Disorder Group CUI

A disorder mention must be annotated, even if the text span is not one of the lexical forms in the terminology, but is a **reasonable synonym**.

Example 1.2.a

Chief Complaint: left-sided facial droop

The span “facial droop” is a disorder mention. The lexical form is not part of SNOMED-CT, but can be mapped to CUI C0427055 (Preferred term: Facial Paresis) in SNOMED-CT.

Example 1.2.b

His face was weak.

The disjoint span “face … weak” is a disorder mention. The lexical form is not part of SNOMED-CT, but it can be mapped to CUI C0427055 (Preferred term: Facial Paresis).

1.3 A disorder mention is explicit

A concept should not be annotated unless it is specifically mentioned by name. That is **descriptions of disorders should not be annotated**. Similarly, there should **not be any inference** made to decide that a disorder is mentioned.

Example 1.3.a

The patient's chem profile reveals a 138 sodium, 4.0 potassium, 106 chloride, 19 bicarb, 43 and 2.2 are the BUN and creatinine, indicating probable blood in the GI tract.

There is no disorder mention in this sentence, even though it describes the disorder GI bleed.

Example 1.3.b

EF 10-15%.

There is no disorder mention in this sentence, even though an ejection fraction of 10-15% strongly indicates the disorder congestive heart failure.

Example 1.3.d

She had a Crohn flare with symptoms of bowel obstruction that typically resolves with rehydration. Her current symptoms of reminiscent of this.

The first sentence contains two disorder mentions ("Crohn" and "bowel obstruction"). There is no disorder mention in the second sentence: "this," while referring to a disorder, is not a disorder mention.

1.4 A disorder mention is the most specific disorder conveyed in the text

Only the **most specific named disorder** may be annotated.

Example 1.4.a

The patient has a small bowel obstruction.

The sentence contains only one disorder mention, "small bowel obstruction." It corresponds to CUI C0235329 (preferred term: Small bowel obstruction). Only one mention is annotated, even though there are sub-spans of text corresponding to other disorders: the span "obstruction" can be mapped to CUI C0028778 (preferred term: Obstruction), and the span "bowel obstruction" can be mapped to CUI C0021843 (preferred term: Intestinal Obstruction). Because "small bowel obstruction" is more specific than "bowel obstruction" and "obstruction," it is the disorder chosen for this annotation.

Example 1.4.b

The patient was found to have left lower extremity DVT.

The sentence contains only one disorder mention, "lower extremity DVT." It corresponds to CUI C0340708 (preferred term: Deep vein thrombosis of lower limb). The span "DVT" can be mapped to CUI C0149871 (preferred term: Deep Vein Thrombosis) is present, but is not annotated because it is part of a more specific disorder in the sentence.

Example 1.4.c

The patient has severe pre-eclampsia.

The sentence contains only one disorder mention, "severe pre-eclempsia." It corresponds to CUI C0341950 (preferred term: Severe pre-eclampsia). The sub-span "pre-eclampsia" can be mapped to CUI C0032914 (preferred term: Pre-eclampsia) but is not annotated as it is more general.

Example 1.4.d

The patient has chronic gingivitis.

The sentence contains only one disorder mention, “chronic gingivitis.” It corresponds to CUI C0008684 (preferred term: chronic gingivitis). The sub-span “gingivitis” can be mapped to CUI C0017574 (preferred term: gingivitis) but is not annotated as it is more general.

1.5 A disorder mention does not contain any negation in its span

When looking for a concept in the terminology, do not map to concepts that have negations in them.

Reason: The negation will be annotated as a modifier of the disorder.

Example 1.5.a

No pain.

The span “pain” is a disorder mention (C0030193, preferred term: Pain), even though there is a SNOMED-CT CUI for the span “no pain” (C0234225, preferred term: Absence of pain).

1.6 A disorder mention does not contain any mention of temporality in its span

When looking for a concept in the terminology, do not map to concepts that have information about past or history.

Example 1.6.a

Hx of stroke.

The span “stroke” is a disorder mention (C0038454, preferred term: Cerebrovascular accident), even though there is a SNOMED-CT Disorder CUI for the span “Hx of stroke” (C0559159, preferred term: “H/O: CVA”).

1.7 A disorder mention is not linked to any syntactic construct

Most of the time, a disorder mention is either a whole or a subset of a noun phrase, including any prepositional phrase attached to a noun phrase. But in some cases, the disorder mention is not conveyed through a noun phrase only.

Example 1.7.a

The patient had a tumor of the skin removed.

The span “tumor of the skin” is a disorder mention. It corresponds to CUI C0037286 (preferred term: Skin Neoplasms). The determiner at the edge of the span “a” is not part of the annotation, but the preposition “of” is part of the disorder mention.

Example 1.7.b

A tumor was found in the left ovary.

The disjoint span “tumor … ovary” is a disorder mention. It corresponds to C0919267 (preferred term: ovarian neoplasm).

1.8 The span for a disorder can be disjoint

When necessary, the span for a disorder can be disjoint. In such cases, it is important not to annotate the whole sentence as disorder span, but rather annotate only the keywords that describe the most specific concept.

Example 1.8.a

A tumor was found in the left ovary.

The disjoint span “tumor … ovary” is a disorder mention. It corresponds to C0919267 (preferred term: ovarian neoplasm).

1.9 A disorder mention is annotated even if it does not pertain to a patient

Any mention of a disorder should be annotated in a clinical note. The fact that a particular disorder mention does not pertain to the patient is taken care of by one of the modifiers for the disorder (see section on modifiers for more details).

Reason: The goal is to train and test NLP systems that identify any disorder mention. This is different from a retrieval system, which would do some inference about whether the presence of a disorder mention signifies that the patient has the particular disorder.

Example 1.9.a

The patient’s son has schizophrenia.

The span “schizophrenia” is a disorder mention (C0036341, preferred term: Schizophrenia). The fact that the patient’s son has the disorder is coded in one of the disorder’s modifiers.

Example 1.9.b

He should return to the ED immediately if any rash occurs.

The span “rash” is a disorder mention (C0015230, preferred term: Exanthema), even though the patient may never have any, does not have it currently, and may never have any. The fact that this is a conditional disorder is coded in one of the disorder’s modifiers.

Example 1.9.c

The patient was referred to the lupus clinic.

The span “lupus” is a disorder mention (C0409974, preferred term: Lupus Erythematosus), even though the patient may or may not have lupus. The fact that this is part of an institution is coded in one of the disorder’s modifier.

1.10 All the disorder mentions in a note must be annotated

Every disorder mention must be annotated, no matter (i) what section of the note it appears in and (ii) whether a previous mention of the same disorder was already mentioned.

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

Participants will be provided an empty template for each unique disease/disorder mention; each template consists of the mention's Unified Medical Language System concept unique identifiers (CUI) and mention boundaries.

Note: A unique disease/disorder span may contribute to one or more templates when more than one Body Location is expressed. For instance, "facial and abdominal rash" generates two templates– facial rash and abdominal rash.

2. Selecting the Appropriate CUI for a Disorder Mention

Once a span for a disorder mention is identified, there still might be several CUIs available for that particular span. This section guides you through selecting the appropriate CUI for a disorder mention.

2.1 There should always be one CUI per mentioned disorder

A CUI is required for every disorder mention. If it is impossible to find a CUI corresponding to the disorder, **one should leave the slot empty as a last resort solution**.

Example 2.1.a

Atrial ectopy

There is a disorder in the UMLS, but not in SNOMED-CT (similarly for ectopy). This should be annotated as a disorder, but the CUI is left empty.

Example 2.1.b

Thick bronchial secretions

While "bronchial secretions" is a body substance, "thick bronchial secretions" can be interpreted as a disorder. There is a no disorder in SNOMED-CT for "thick bronchial secretions". This should be annotated as a disorder, but the CUI is left empty.

Example 2.1.c

The patient was admitted with low blood pressure.

The span "low blood pressure" is a Finding in SNOMED-CT, and as such does not belong to our definition of the Disorder semantic group. In this case, however, because it does indeed describe a disorder, it should be annotated. The CUI is left empty.

2.2 There should be only one annotation per mentioned disorder

In many circumstances there will be a mentioned disorder, which is difficult to map exactly to a single CUI. It may be tempting to annotate the same disorder with multiple CUIs. However, for this annotation task, you may only annotate a mentioned disorder with a single annotation span and one corresponding CUI. The challenge is to choose the best CUI for the mentioned disorder along with the text span that best corresponds to that CUI.

2.3 Use the context in which the disorder is mentioned to select the most appropriate CUI

Whenever possible, use the context of the note to identify the most appropriate CUI for a given disorder.

Example 2.3.a

The patient complained of watering of the eye.

There are two candidate CUIs: C0152227 (Excessive tearing) and C2233621 (Watering or discharge from eye). Both are from the same semantic class “Sign or Symptom”. The definitions can help identify the most likely CUI for this particular note.

Example 2.3.b

The patient complained of numbness.

There are two candidate CUIs: C0028643 (numbness) and C0235018 (localized numbness). Both are from the same semantic class “Sign or Symptom”. In this case, the preferred term can help decide the most likely CUI for this particular sentence (i.e., the first one).

3. Annotating the Modifiers of a Disorder Mention

The modifiers of a disorder mention are spans of text which complement the disorder mention, as defined in the previous sections. Taken together, the span about a disorder mention and all of its modifiers provide full information about the disorder, even if this union does not belong to SNOMED-CT.

Some modifiers are annotated through selecting spans of text, and others are annotated as indicators.

3.1 Only annotate the modifiers of interest

The following modifiers are in our annotation schema:

- Body Location
- Temporal Expression
- Negation Indicator
- Uncertainty Indicator
- Course
- Severity
- Subject
- Conditional
- Generic
- DocTime

The next sections provide guidelines for the annotation of each modifier, and how to annotate when several modifiers are present for a given disorder.

3.2 Annotate the modifiers conveyed in the same sentence as the disorder mention only

Do not annotate a modifier if it is conveyed in a different sentence from the disorder mention. (This guideline stems from the scope of the project. Cross-sentence relations might exist, but are ignored for now.)

Example 3.2.a

The patient came in with swelling of the eyelid. It progressed to the whole eye.

In the first sentence, the disorder “swelling of the eyelid” is annotated with the corresponding CUI C0560024. The span “of the eyelid” is also annotated as a “Body Location” modifier. However, the span “to the whole eye” is not annotated because it is mentioned in a different sentence than the disorder.

3.3 When annotating a modifier, include prepositions

Prepositions should be part of the span of a modifier.

Example 3.3.a

The patient came in with swelling of the eyelid. It progressed to the whole eye.

The modifier (of the type Body Location) is “of the eyelid,” not “eyelid.”

3.4 The unit of annotation for a modifier is a word

If a modifier is part of an acronym, do not annotate it.

Example 3.4.a

AMI.

The modifier (of the type “Course”) is “acute” (A in the acronym), but is not annotated because it is part of a word.

4. Body Location

4.1 Definition

This modifier creates a relation between the disorder mention and a well-defined Anatomical Site (i.e., body location, body part, or laterality). There are three parts to annotate for this modifier: the **span of text** describing the anatomical site, its **associated SNOMED-CT CUI**, and whether the relation between the anatomical site and the disorder is **negated** or not.

From the CEM definition: The place on the body where the observation is present.

A span of text is considered an anatomical site if there is a concept in SNOMED-CT which corresponds to any of the following semantic types (this follows the Bodenreider and McCray Anatomy group):

- Anatomical structure
- Body location or region
- Body part, organ or organ component
- Body space or junction
- Body substance
- Body system
- Cell
- Cell component
- Embryonic structure
- Fully formed anatomical structure
- Tissue

Example 4.1.a

The patient has a facial rash.

The span "facial" is the Body Location for the disorder "rash". Its CUI is C0015450 (preferred term: Face). There is no negation between the anatomical site "facial" and the disorder "rash".
(Note: "facial rash" is not a SNOMED-CT disorder.)

4.2 Default value

By default, a disorder mention has no associated Body Location.

4.3 The Body Location is the most specific description

Like for disorder mentions, we are interested in the most specific description.

Example 4.3.a

Chief complaint: left knee infection.

The span “left knee” is the Body Location for the disorder “infection”. Its CUI is C0230432 (preferred term: Structure of left knee). This is more specific than the span “knee,” and therefore is chosen as the modifier for this disorder mention.

4.4 The Body Location can contain the relation to the disorder mention

When a preposition specifies a spatial relation between the disorder and its Body Location, you must include it in the annotation of the Body Location. Examples of prepositions are “in”, “behind”, “above”, “outside”, etc.

Example 4.4.a

A tumor was removed from behind his left ear.

The span “behind his left ear” is the Body Location for the disorder mention “tumor”. It is associated to the CUI is C0229299 (preferred term: Left ear structure).

4.5 The Body Location can overlap with or be part of the disorder mention

The span of the Body Location can be outside, overlapping with, or inside the span of the mentioned disorder.

Example 4.5.a

He was admitted with right leg swelling.

The span “right leg” is the Body Location for the disorder “leg swelling”. It is more specific than “leg” on its own.

4.6 There can be several Body Location modifiers associated with one disorder mention

A disorder mention can have several Body Locations associated with it.

Example 4.6.a

The patient had a skin tumor removed from behind his left ear.

The disorder “skin tumor” has two associated Body Locations: “skin” with CUI C1123023 (preferred term: Skin), and “behind his left ear” with CUI C0229299 (preferred term: Left ear structure).

Example 4.6.b

Left-sided chest pain radiating to jaw and left shoulder

The disorder “left-sided chest pain” has three associated Body Location: “left-sided chest” “to jaw” and “left shoulder.”

4.7 The span for a Body Location can be disjoint in the text

A Body Location might be disjoint in the text of a clinical note.

Example 4.7.a

There were cysts in the left and right ovaries.

In this example there are two associated Body Locations to the disorder “cysts”: “left [...] ovaries” (C0227874, preferred term: Structure of left ovary) and “right ovaries” (C0227873, preferred term: Structure of right ovary). The first Body Location is disjoint.

4.8 A Body Location can be negated

Because the Body Location is a relation between an Anatomical Site and a Disorder, it is possible to annotated when the relation is a negation.

Example 4.8.a

A tumor was found outside the pancreas.

In this example the disorder “tumor” has an associated Body Location, with the following: a Negation Indicator (“outside”), and an anotmical site (“pancreas”). Note that while “outside” is a Negation Indicator, it is associated to the Body Relation modifier, not to the Disorder itself.

One way to decide where to attach the Negation Indicator is to ask yourself:

- is the Disorder present? In this case, the tumor is present

- is the Body Location present? In this case, the pancreas is present

- is the Body Location negated? In this case, it is (there is a tumor, there is a pancreas, but the tumor is not in the pancreas)

Example 4.8.b

The pancreas is tumor free.

In this example the disorder “tumor” has two modifiers: a Body Location (“pancreas”) and a Negation Indicator (“free”). In this case, the Body Location is not negated, but the disorder is.

Example 4.8.c

No tumor was found in the eye.

In this example the disorder “tumor” has two modifiers: a Body Location (“eye”) and a Negation Indicator (“No”). In this case, the Body Location is not negated, but the disorder is.

4.8 Normalization

Once a span is identified for an Anatomical Site for a Body Location, its normalization is the particular associated CUI. Like with disorder mentions, there might be several CUIs corresponding to a particular span of text for anatomic location. You must chose the more specific CUI corresponding to the Anatomical Site.

In general, anatomical sites and body parts in particular have several synonym CUIs. For instance, the span "right ear" can be mapped to the CUI C0229298 (Right Ear Structure) and the CUI C1289033 (Entire Right Ear). When such ambiguity occurs, chose the CUI that corresponds to the structure (C0229298 in this case).

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Body Location normalization slot value is assigned one of three potential value types: 1) *NULL (no Body Location annotated), 2) a particular CUI (annotated Body Location in SNOMED-CT), or 3) CUI-less (annotated Body Location not in SNOMED-CT).

Cue Note: A Body Location cue slot value is assigned one of two potential value types: 1) *null (no Body Location annotated) or 2) span offset of lexical cue (start and end spans).

General Note: When a negation cue modifies a Body Location class, the cue and normalization default values may remain for a collocated disorder mention. For instance, "pain outside the pancreas" results in a "pain" disorder template with the default values for Negation cue (null) and normalization (no) and Body Location cue (null) and normalization (null) semantically conveying that pain exists without a particular Body Location.

5. Temporal Expression

5.1 Definition

This modifier refers to any temporal expression mentioned about a disorder (the start date of a disorder, the end date, or the duration). A temporal expression is defined as a TIMEX in the TIMEML formalism.

5.2 Default value

By default, a disorder mention has no associated temporal expression.

5.3 Normalization

A span of text can be mapped to one of four types of time expressions: DATE, TIME, DURATION, and SET. The time expressions follow the TIMEX definition from the TimeML standard.

Example 5.3.a

The patient came in with a rash Friday evening.

There is one Temporal Expression associated with the disorder "rash": "Friday evening" which is a TIMEX DATE.

Example 5.3.b

The rash was present for three days.

There is one Temporal expression associated with the disorder "rash": "for three days" which is a TIMEX DURATION.

5.4 There can be several temporal expressions associated with one disorder mention

Example 5.4.a

By Friday evening, which is approximately three days ago, the patient had a rash.

There are two different temporal expressions associated with the disorder "rash": "By Friday evening" which is a TIMEX DATE, and "three days ago" which is a TIMEX DATE.

5.5 Do not annotate temporal relations between concepts

We only focus on temporal expressions associated with a disorder, as defined by the TIMEX convention. (This guideline stems from the scope of the project. There are temporal relations among disorders evidently, but we ignore them for this project).

Example 5.5.a

MI secondary to ischemia.

There are two disorder mentions in this sentence, but no temporal expression.

Example 5.5.b

The patient had an infection after the CABG.

The relation “after the CABG” is not a TIMEX, and thus is not annotated.

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Temporal Expression normalization slot value is assigned one of five potential value types: 1) *none (no Temporal Expression annotated), 2) DATE, 3) TIME, 4) DURATION, or 5) SET.

Cue Note: A Temporal Expression cue slot value is assigned one of two potential value types: 1) *null (no Temporal Expression annotated) or 2) span offset of lexical cue (start and end spans).

6. Negation Indicator

6.1 Definition

This modifier refers to whether the **presence of a disorder was negated**.

From the CEM definition: Used to indicate that a procedure or assertion did not occur or does not exist. It an indicator, so valid value are Yes and No. Yes means the observation is not present.

Example 6.1.a

The patient has not noticed any numbness.

The disorder "numbness" has a Negated modifier associated with the span "not noticed any".

Example 6.1.b

The patient had a tumor removed from behind his left ear.

See example 14.d for a detailed discussion of this type of example. The disorder "tumor" has no Negation Indicator present. Rather, the disorder has a Course modifier and a DocTime modifier.

6.2 Default value

By default, a modifier has no Negation Indicator. That is, the modifier is left un-annotated.

The default interpretation is that a disorder is not negated. That is, the Negation Indicator is annotated only when there is evidence of negation of the disorder. Explicit evidence for presence of a disorder is left un-annotated, because it is the default interpretation.

Example 6.2.a

The patient presents with SOB.

There is no Negated modifier associated with the disorder "SOB".

6.3 Normalization

Since this modifier is a marker for negation, there is no normalization needed. The absence of Negation marker signifies no negation, and the presence of a span signifies that the disorder is negated.

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Negation Indicator normalization slot value is assigned one of two potential value types: 1) *no (no Negation Indicator annotated with respect to the disorder), or 2) yes (annotated Negation Indicator with respect to the disorder).

Cue Note: A Negation Indicator cue slot value is assigned one of two potential value types: 1) *null (no Negation Indicator annotated) or 2) span offset of lexical cue (start and end spans).

7. Uncertainty Indicator

7.1 Definition

This modifier refers to the uncertainty associated with the mention of a disorder. It only refers to explicit mentions of uncertainty, and does not involve any pragmatics-level reasoning.

From the CEM definition: An introduction of a measure of doubt into a statement.

Example 7.1.a

The patient presents for the evaluation of MI.

The disorder "MI" has an uncertainty indication, as evidenced by the span "evaluation of".

Example 7.1.b

There certainly has been no pain.

The disorder "pain" has no Uncertainty Indicator associated with it.

Example 7.1.c

Patient reported history of pain.

The disorder "pain" has no Uncertainty Indicator associated with it.

Example 7.1.d

There certainly has been no pain.

The disorder "pain" has no Uncertainty Indicator associated with it, although "certainly" indicates certainty it does not indicate uncertainty.

Example 7.1.e

There has been possible pain.

The disorder "pain" has "possible" as Uncertainty Indicator associated with it

7.2 Default Value

By default, a disorder mention has no uncertainty indicated associated with it.

7.3 Normalization

Because this modifier is an indicator of uncertainty, there is no normalization to it. When the modifier is not annotated, it signifies there is no uncertainty related to the

disorder; when there is a span annotated, it signifies that there is an explicit mention of uncertainty related to the disorder.

7.4 Uncertainty is not determined according to the section in which a disorder is mentioned

If a disorder is mentioned in the Chief Complaint or the HPI section of a note, but there is no explicit mention of certainty in the sentence where the disorder is mentioned, the Uncertainty Indicator should be left unmarked.

Reason: The inference about whether "history of present illness" is information reported by patient, and whether it is to be trusted is outside the scope of annotating a disorder and its uncertainty.

Example 7.4.a

CC: left facial droop

The disorder "facial disorder" has no Uncertainty Indicator associated with it. "CC" is not a span related to uncertainty.

Example 7.4.b

Patient denies numbness.

The disorder "numbness" has no Uncertainty Indicator associated with it. "denies" is not a span related to uncertainty (it is annotated as a Negation Indicator) because no inference is made about trusting the patient or not.

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Uncertainty Indicator normalization slot value is assigned one of two potential value types: 1) *no (no Uncertainty Indicator annotated), or 2) yes (annotated Uncertainty Indicator).

Cue Note: A Uncertainty Indicator cue slot value is assigned one of two potential value types: 1) *null (no Uncertainty Indicator annotated) or 2) span offset of lexical cue (start and end spans).

8. Course

8.1 Definition

This modifier refers to the development or alteration of a disorder mention. From CEM definition: An indication of progress or decline of a condition.

Example 8.1.a

The cough worsened over the next two days.

The disorder "cough" has a Course modifier associated with the span "worsened" and mapped to "worsened".

8.2 Default value

By default, a disorder mention has no mention of course. This is interpreted as an unmarked Course.

8.3 Normalization

A span of text for the Course modifier can be mapped to the following: unmarked, changed, increased, decreased, improved, worsened, and resolved. These normalizations follow the CEM definition.

Example 8.3.a

Screen mammogram showed increasing calcifications in right breast.

The disorder "calcification" has a Course modifier associated with the span "increasing" and normalized to "increased". (Note that "calcification" is a CUI-less disorder, since it belongs to the Findings semantic type).

8.4 A Course Modifier can be negated

This modifier refers to the development or alteration of a disorder mention.

Example 8.4.a

Benign Prostatic Hypertrophy is currently asymptomatic.

The disorder "BPH" has a Course modifier associated with the span "currently asymptomatic" and normalized to "changed" and a Negation Indicator set to "yes" with the same span ("currently asymptomatic").

Example 8.4.b

He had a skin tumor with no known recurrence.

The disorder “skin tumor” has a Course modifier associated with the span “no known recurrence” and normalized to “changed” and a Negation Indicator set to “yes” with the same span (“no known recurrence”).

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Course Class normalization slot value is assigned one of seven potential value types: 1) *unmarked (no Course Class annotated), 2) changed, 3) increased, 4) decreased, 5) improved, 6) worsened, or 7) resolved.

Cue Note: A Course Class cue slot value is assigned one of two potential value types: 1) *null (no Course Class annotated) or 2) span offset of lexical cue (start and end spans).

General Note: When a Negation cue modifies a Course class, the cue and normalization default values may remain for a collocated disorder mention. For instance, “He had a skin tumor with no known recurrence” results in a “skin tumor” disorder template with the default values for Negation cue (null) and normalization (no), Course cue (null) and normalization (unmarked), and updated DocTime (Before) semantically conveying that skin disorder exists without Course before the DocTime.

9. Severity

9.1 Definition

The Severity modifier refers to the degree of severity the clinical condition is evaluated to be.

From CEM definition: The relative intensity of a process or the relative intensity or amount of a quality or attribute.

Example 9.1.a

He noted a slight bleeding.

The disorder "bleeding" has a Severity modifier associated with the span "slight" and mapped to "Slight".

9.2 Default value

By default, a disorder has no Severity modifier associated with it. It is interpreted as "unmarked".

9.3 Normalization

A span of text can be mapped to four types of Severity: unmarked, slight, moderate, and severe.

Example 9.3.a

Patient presents with mild sleep apnea.

The disorder "sleep apnea" has a Severity modifier associated with the span "mild" and mapped to "Moderate".

9.4 The severity can overlap with or be part of the disorder mention

The span of the severity can be outside, overlapping with, or inside the span of the mention disorder.

Example 9.4.a

The patient presented with severe pre-eclempsia

The disorder "severe pre-eclempsia" has the modifier Severity associated with the span "severe" and mapped to "Severe".

9.5 The severity should be explicit

If inference is required to determine the severity of a disorder, then no Severity modifier should be annotated.

Example 9.5.a

The patient presented with tiny gallstones.

The disorder "gallstones" has no Severity modifier. The word "tiny" is a qualitative modifier, and mapping it to a severity would require medical inference.

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Severity Class normalization slot value is assigned one of four potential value types: 1) *unmarked (no Severity Class annotated), 2) slight, 3) moderate, or 4) severe.

Cue Note: A Severity Class cue slot value is assigned one of two potential value types: 1) *null (no Severity Class annotated) or 2) span offset of lexical cue (start and end spans).

10. Subject

10.1 Definition

This modifier refers to the entity experiencing the disorder.

10.2 Default value

By default, the Subject modifier is the patient about which the note is written.

Example 10.2.a

The patient has congestive heart failure.

The disorder "congestive heart failure" has no associated Subject modifier, even if there is an explicit mention of the patient being the Subject.

10.3 Normalization

There are six possible values for Subject: Patient, Family_Member, Donor_Family_Member, Donor_Other, Null, and Other.

Example 10.3.a

The patient's son has schizophrenia.

The disorder "schizophrenia" has a Subject modifier associated with "son" and value "Family_Member".

Example 10.3.b

The patient was sent to the Lupus clinic.

The disorder "lupus" has a Subject modifier associated with "clinic" and value "Null".

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Subject Class normalization slot value is assigned one of six potential value types: 1) *patient (no Subject Class annotated), 2) family_member, 3) donor_family_member, 4) donor_other, 5) null, or 6) other.

Cue Note: A Subject Class cue slot value is assigned one of two potential value types: 1) *null (no Subject Class annotated) or 2) span offset of lexical cue (start and end spans).

11. Conditional

11.1 Definition

The Conditional modifier refers to disorders, which could exist under certain circumstances.

From the CEM definition: conditional use of a disorder, eg "if pain is reported, then…"

Example 11.1.a

The patient should come back to the ED if any rash occurs.

The disorder "rash" has a Conditional modifier associated with "if" with the value "True".

Example 11.1.b

The patient should return to the ED immediately for any numbness.

The disorder "numbness" has a Conditional modifier associated with "should return" with value "True".

11.2 Default value

By default a disorder has no Conditional modifier associated with it. Its interpretation is that it is not conditional.

11.3 Normalization

A Conditional modifier can be normalized to "True" or "False".

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Conditional Class normalization slot value is assigned one of two potential value types: 1) *false (no Conditional Class annotated), or 2) true.

Cue Note: A Conditional Class cue slot value is assigned one of two potential value types: 1) *null (no Conditional Class annotated) or 2) span offset of lexical cue (start and end spans).

12. Generic

12.1 Definition

The Generic modifier refers to disorder mentions, which are generic, i.e., not related to the instance of a disorder.

Example 12.1.a

The patient was referred to the Lupus clinic.

The disorder "lupus" has a Generic modifier associated with "clinic" with the value "True."

Example 12.1.b

We discussed possible increased risk of recurrence locally without radiation.

The disorder "recurrence" (mapped to the disorder "recurrent tumor") has a Generic modifier associated with "possible increased risk" with the value "True."

12.2 Default value

By default a disorder has no Generic modifier associated with it. Its interpretation is that it is not generic.

12.3 Normalization

A Generic modifier can be normalized to "True" or "False".

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A Generic Class normalization slot value is assigned one of two potential value types: 1) *false (no Generic Class annotated), or 2) true.

Cue Note: A Generic Class cue slot value is assigned one of two potential value types: 1) *null (no Generic Class annotated) or 2) span offset of lexical cue (start and end spans).

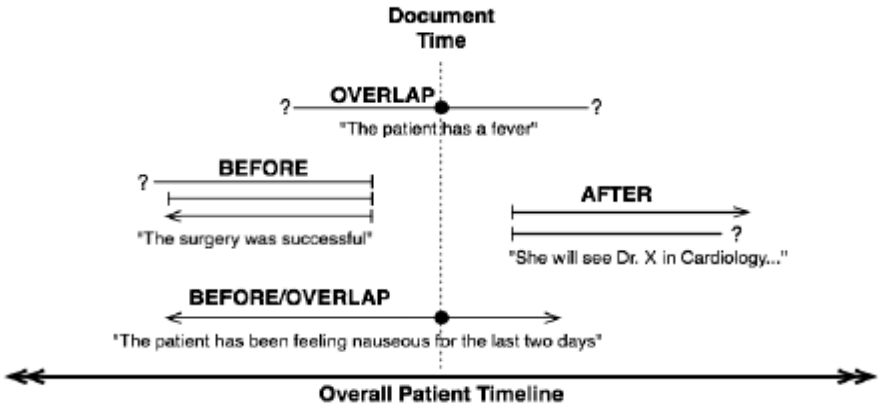
13. DocTime

Please, do not distribute these as we are working on a temporal relations manuscript describing the guidelines. Please, treat this as sensitive material for now!

13.1 Definition

Note: this set of guidelines is fully credited to the temporal relations guidelines developed by Palmer, Savova, Styler, et al. in consultation with James Pustejovsky.

The DocTime modifier refers to temporal relation between a disorder mention and the time when the document was authored. This is different from a temporal expression, as no specific time is specified, but instead an explicit temporal relation is mentioned, which enables the reader to assess whether a disorder occurred in the past (before note time), overlaps with the notes, will occur after the note, or occurred in the past and continue to occur at the time of the note.



The above chart is a representation of how a given disorder occurs relative to the moment that the note was written. Note: The example sentence above for the values BEFORE and AFTER are not about disorder mentions; but the same holds for disorder mentions.

Full temporal annotation is a complex task, which is not in the scope of this project. This modifier (along with the Time Expression modifier) does not cover all the possible temporal relations that can be conveyed in a note.

Example 13.1.a

Past MI.

The disorder "MI" has a DocTime modifier associated with the value Before.

13.2 Default Value

There is no default value associated with DocTime.

13.3 Normalization

There are four possible values for this modifier: Before, After, Overlap, Before-Overlap and Unknown.

13.3 DocTime refers to explicit mentions in the text

Example 13.3.a

MRI on December 12 revealed a stroke.

The disorder "stroke" has a DocTime modifier associated with the value Before.

Example 13.3.b

MRI revealed a stroke.

The disorder "stroke" has a DocTime modifier associated with the value Before, as indicated by the past simple tense in the verb "revealed".

Example 13.3.c

Patient came in and still presents with left-sided weakness.

The disorder "weakness" has a DocTime modifier associated with the value Before-Overlap, because of the phrase "still presents"

Example 13.3.d

Patient has had a tumor for the past two months.

The disorder "tumor" has a DocTime modifier associated with the value Before-Overlap.

Example 13.3.e

Patient came in with a tumor.

The disorder "tumor" has a DocTime modifier associated with the value Before.

13.4 DocTime is determined without any inference

Example 13.4.a

Echo on December 12 revealed congestive heart failure.

The disorder "congestive heart failure" has a DocTime modifier associated with the value Before. That is, the annotation does not make any inference about the disease

course for CHF (and the fact that very likely, the patient has CHF at the time the note is authored).

ShARe/CLEF eHealth 2014 Task 2: Information extraction from Clinical Text: Disease/Disorder Template Filling

*Task default value

Normalization Note: A DocTime Class normalization slot value is assigned one of five potential value types: 1) *unknown (no DocTime Class annotated), 2) before, 3) after, 4) overlap, or 5) before-overlap.

Cue Note: A DocTime Class has no cue slot value.

14. Some Tricky Examples

This section provides examples of complete annotation, where all the rules are put together.

Example 14.a

The patient has an extensive **thyroid** history.

The sentence contains one disorder with the following characteristics:

- Annotated span of the disorder: “thyroid”
- CUI for the disorder: C0040128 (Preferred Term: “Thyroid diseases”). Note that we do not select the CUI C0455487 (Preferred Term: “H/O: thyroid disorder”). See rule 1.6.
- Annotated span for Body Location: “thyroid” and normalized to CUI C0040132 (Preferred Term: “Thyroid Gland”).
- There is no Temporal Expression modifier associated with this disorder. See rule 5.6.
- There is no Negation Indicator modifier associated with this disorder.
- There is no Uncertainty Indicator modifier associated with this disorder.
- There is no Course modifier associated with this disorder.
- Annotated span for Severity: “extensive” normalized to “severe.”
- There is no Subject modifier associated with this disorder (since it is the default value for the patient).
- There is no Conditional modifier associated with this disorder.
- There is no Generic modifier associated with this disorder.
- The DocTime modifier is “Before”

Example 14.b

Intermittent **chest pain**.

"intermittent" is left un-annotated (no Course or Severity modifiers).

Example 14.c

90% RCA **lesion**.

"90%" is left un-annotated, because we do not do any inference for determining modifiers (e.g., Severity). RCA is annotated as a Body Location. "lesion" is a CUI-less disorder.

Example 14.d

The patient had a **tumor** removed.

The patient used to have a tumor. There was an event, and now the patient does not have a tumor. This is represented through the following modifiers:

DocTime = Before (bc of had)

NegationIndicator = Absent

Course= changed ("removed")

Example 14.e

No past **tumors**.

DocTime = Before-Overlap

Negation Indicator = Present

Example 14.f

Patient describes **chest pain** as burning.

"burning" is left unannotated, as it is a qualitative modifier, which is not a modifier type of interest in this annotation project.

Example 14.g

Resting regional wall motion abnormalities include basal and mid inferior and inferoseptal akinesis with apical inferior hypokinesis.

That is just a terribly complicated sentence…

15. Annotation Workflow

This section is for the traditional annotation.

When annotating a note, the following workflow seems to bring the most reliable annotation:

(1) identify all disorder spans and CUIs in the note
(2) for each annotated disorder, annotate its modifiers