

# Guidelines for Annotating Biomedical Entities and Relations in Tweets

Amelie Wühl

August 2021

## 1 Introduction

The goal of this annotation process is to create a corpus of Twitter posts with annotations of biomedical entities, and the relations between them. Entities include all types of medical conditions, treatments, environmental factors, and diagnostic instruments. Relations include e.g., prevents, treats, causes, or prescribes.

We annotate such entities on token level and label the relation that they share. We use INCEpTION<sup>1</sup> as our annotation environment.

## 2 Annotation

### 2.1 Biomedical Entities

We always start annotating a document by identifying biomedical entities that are mentioned in the texts. Note that we focus on human biomedical issues. We therefore consider documents that discuss health related issues in animals or pets as off-topic. We want to label the following groups of entities:

1. Medical Conditions
2. Treatments
3. Environmental Factors
4. Pathogens
5. Biochemical Entities
6. Diagnostic Instruments
7. Subjective Quality of Life Assessments

---

<sup>1</sup>[https://inception-project.github.io/releases/0.19.3/docs/user-guide.html#sect\\_annotation](https://inception-project.github.io/releases/0.19.3/docs/user-guide.html#sect_annotation)

Each group contains a respective tag or subset of tags which we want to identify and assign to the entities in the text. Refer to Table 2 for the full list of entity tags and examples of each group. The following sections contain brief descriptions of each entity group.

### 2.1.1 Entity Groups

**Medical Conditions** Medical conditions include all mentions of diseases, symptoms, side effects, and medical events or incidents or descriptions thereof.

**Treatments** The treatments category covers mentions of any kind of treatment. This includes drug names, both generic and brand names, as well as all types of therapy or even prevention methods. Anything that is intended as a treatment is annotated as such regardless of how we judge truth value of the method/treatment.

**Environmental Factors** Environmental factors include entities that somehow influence, cause or contribute to a medical condition. As is the case with all entities, we also annotate descriptions thereof. We differentiate between socio-economic (age, gender, ethnicity, social background etc.), geographic/climatic (issues that relate to geography, climate, weather etc.), dietary, habitual (smoking, exercise, stress etc.) or pollution-related (air, water pollution, Smog, electro-magnetic, UV, nuclear etc. radiation) factors.

**Pathogens** Pathogens are organisms that cause diseases. This includes mentions of bacteria, viruses, fungi or parasites.

**Biochemical Entities** Biochemical entities include mentions of biochemical substances that are can not be categorized as drug treatments. This can be mentions of active ingredients, proteins, hormones etc. This category also includes biochemical processes such as biological, pathogen-related or chemical mechanisms.

**Diagnostics** The diagnostics category includes mentions of tests or other diagnostic instruments that are used to diagnose or test for a medical condition.

**Quality of Life Assessments** This category includes descriptions of subjective assessments of one’s quality of life (QoL). This also includes descriptions of medical conditions that are somewhat fuzzy or vague. Descriptions like “Thanks depression for *making me feel like I can’t function* today” should be assigned the QoL entity label. If you can not decide if an entity should be annotated as a medical condition or a QoL assessment check if the condition is listed on ICD-10 <https://icd.who.int/browse10/2019/en>. If yes, annotate it as a medical condition and make sure to add the condition’s ICD-10 ID in the text field. We can use this information later on in the project.

Group	Entity Tags	Examples
medical conditions	medC	<b>autism</b> is not connected to the MMR vaccine <b>breathlessness</b> can be a sign of long-covid hay fever meds make me <b>drowsy</b> the flu causes many <b>deaths</b>
treatments	treat_drug treat_therapy	his eczema responds well to <b>Eucerin cream</b> a daily <b>yoga</b> practice helps with the back pain
environm. factors	env_socio-economic env_geo-climate env_dietary env_habitual env_pollution	there is a link between obesity and <b>education</b> covid is not as common in <b>hotter climates</b> <b>drinking</b> can be the cause of liver failure <b>working out</b> eases anxiety <b>smog</b> is a major cause of lung diseases
pathogens	pathogen	covid-19 is caused by the <b>corona virus</b>
biochem. entities	biochem_substance biochem_process	<b>Nicotine</b> contributes to lung cancer They researching the treatment of acute myeloid leukemia by looking into <b>epigenetic mechanisms</b>
diagnostics	diag	they noticed it when the doctor <b>palpated my abdomen</b> and then it was confirmed by a <b>blood test</b> <b>chest x-ray</b> shows temporal lung changes
quality of life	qol	the new meds made me <b>feel like I'm dying</b> recalled vividly at the time <b>my brain felt like a raw pork shoulder inside a spinning fishbowl</b> if not for acetaminophen I <b>wouldn't have been able to function enough to go to work</b>
other	other	The vaccine will hopefully give us <b>immunity</b>

Table 2: Entity groups and entity tagset for biomedical NE annotation. In each example the entity with the respective tag is in bold face. (Note, however, that there might be more than one NE in the example phrase.)

### 2.1.2 Workflow

1. Log onto INCEpTION with your username and password.
2. Click on the respective project.
3. Click on “Annotation” in the menu on the left.
4. Open the document you want to annotate.
5. In the Layer drop down menu at the top of the panel on the right-hand side, select BiomedEntities.
6. Read the document carefully. Before you start annotating anything, take a moment and ask what the main message/claim/statement is in the tweet is trying to convey.
7. Start to take note of any biomedical entity.
8. Mark the whole span of the entity that you found.
9. Once you have highlighted the token(s) that make up the entity, select the respective entity tag from the value drop-down menu on the right-hand side.
10. That’s it!

### 2.1.3 Notes

- Authorial intention: Make sure to annotate an entity with regards to how it is used by the author. Consider the mentions of **UV radiation** in the phases below. In the first example the term is intended as an environmental factor, while in the second example it is used and recounted as a treatment.
  - **High UV radiation** often is the cause of skin cancer. [env\_geo-climate]
  - They said **UV radiation** will help with my low vitamin D levels. [treat\_therapy]
- Descriptions of biomedical entities:
  - We annotate medical conditions, treatments and all other entities not only if they are nouns or noun phrases, but we also annotate descriptions of entities. Those can be formulated as longer phrases. It can be helpful to try and summarize longer descriptions to test if they refer to a single entity.
    - \* After taking the pain medication I usually **can’t sleep**. [medC: sleeplessness.]
    - \* I **lost four friends** to brain cancer in the last two years. [medC: death]
    - \* Luckily, they found that my **tumor was not cancerous** [medC: un-cancerous/benigne tumor]

- \* The reason why pancreatic cancer kills so many people is that it is normally **not detected early enough to be treated**. [medC: late-stage diagnosis]
  - In some cases it can be enough to annotate an adverb or adjective:
    - \* I was a bit **bloated** at first, but **not nauseous**. [medC, medC]
- Groups of treatments: Mentions which describe a group of treatment methods and which can be used for to treat multiple diseases are annotated separately from the condition they are being applied to.
  - **Antibody treatments** for **Covid** are promising. [treat\_therapy, medC]
  - In most cases, working with **chemotherapy** against **breast cancer** is successful. [treat\_therapy, medC]

Similarly, we annotate the medical events like *death* separately from the medical condition that caused it if there is a clear causal understanding implied:

- **death** from **Covid** [medC *cause\_of* medC]
  - 94 % of **covid deaths** in the US had **underlying conditions**. [medC1 *not\_cause\_of* medC2, medC3 *cause\_of* medC2]
- Including entity details: Be sure to make the annotations as explicit as possible. This means that we include all tokens that help specify or identify e.g. a certain medication (including the brand name), the type or location of a condition, dosages, amounts or levels of entities:
  - His eczema responds well to **Eucerin baby eczema cream**. [treat\_drug]
  - Wrote about **Effexor XR withdrawal** today, [...]. [medC]
  - My child is **vaccine injured** from the MMR shot. [medC]
  - Stomach cancer is caused by the **growth of cancerous cells in the lining of the stomach**. [biochem\_process]
  - Because **smallpox** is like the **flu** and the **vaccine** also left a **scar on your arm**. [medC, medC, treat\_therapy, medC]
  - They attribute a diffuse collection of symptoms to **low levels of exposure to electromagnetic fields at home**. [env\_pollution]
  - **A cup of Amel Susan Cocoa** can prevent heart disease. [treat\_drug]
  - My prostate cancer is being treated with **radioactive seeds (brachytherapy)**. [treat\_therapy]
- Unspecific or general mentions and entity descriptions: We annotate mentions of medical conditions, treatments etc. even if they appear unspecific or general. We annotate the whole token sequence the way the author intended them. If someone mentions the “symptoms of X” we annotate the whole token sequence, instead of just the mention of the disease name “X”:

- We use this as a chance to raise awareness of the **symptoms of ovarian cancer**. [medC]
- Sometimes Covid tests are false. The resulting **improper treatment** causes many people to die. [treat\_therapy]
- Have you had any issues getting your **diabetes care supplies** during Covid? [treat\_therapy]
- No one has had **serious side effects** from the Covid vaccine. [medC]
- Her dad got diagnosed with brain cancer last year. His brother is in **treatment now for the same**. [treat\_therapy]
- Off-topic documents:
  - Entities. We annotate biomedical entities (all classes) in all documents, even if they appear off-topic or if statements are not worth fact-check. The same goes for entities and relations (see rules for special consideration of relations in bullet below) that are expressed in quotes, or if the document/tweet consists of the headline or title of an article, or might be an ad or simply click bait. If entities are ambiguous (cancer: medC vs. star sign) and refer to their non-medical meaning, we do not annotate them. The same goes for medical terms used e.g. as insults (her cancerous behavior).
    - \* They announced that he **passed away** last week. He **died** of **pancreatic cancer**. [medC, medC, medC]
    - \* Otherwise they will say “well **mental health** can cause **physical symptoms**”. [medC, medC]
    - \* Certainly interesting, especially because it’s just a **Baclofen** analogue with some affinity to the **GHB receptor**. [biochem.substance, biochem.substance]
  - Relations. We annotate relations in off-topic documents except when the document is both off-topic and the resulting claim is not fact-check-worthy. If both these conditions are met, we only annotate the entities and move on. In short: off-topic + no fact-check-worthiness = no relation
    - \* There is a concept album about a **cancer** patient coming to terms with his **death**. [medC, medC]
    - \* Putin wants extradites him to Iran, and Trump **dies** in an Iranian prison in 2022 of untreated **COVID19** and **heart disease**. [medC, medC, medC]
- Biomedical mentions in hashtags: Mentions in hashtags should be considered for annotation, but only if it is a standalone entity and not part of a longer hashtag.
  - They put in place new regulations to address the consequences of **#covid #USAfightscovid** [medC]

If the entity is ambiguous (e.g., medC vs. pathogen or medC vs. env) and it has been referred to before in the document, give it the same entity tag as you did in the earlier mention.

- **Psychological distress symptoms** show elevated rates in people with **higher stress jobs** #stress. [medC, env\_habitual, env\_habitual]

Tweets which spell out their message only using hashtags, and said message contains biomedical entities and/or relations, are annotated as well. Make sure not to include the actual hashtag.

- #New #**Air** #**Filtration** #**Models** #For #Schools #Prevent #**COVID19**
- Fuzzy terminology: The terminology around Covid-19 and corona virus is used very interchangeably. Strictly speaking, specific mentions of the “corona virus” should be categorized as pathogens. However, we annotate such mentions the way the author has intended them. If they misuse the terminology and refer to the situation caused by the corona virus (the disease, pandemic, lockdown etc.) consider annotating as medical condition or environmental factor instead. Standalone mentions of corona virus in hashtags we annotate as pathogen. For phrases like “covid infections” only annotate the whole phrase when the author actually refers to the infection event. If they use this phrase to actually refer to the disease, only annotate “covid” [medC].
  - They are developing new therapies for treating **Coronavirus**. [medC]
  - **Corona virus** are the cause of Covid. [pathogen]
- Combined entity mentions: Generally, we annotate entities individually. Try and split up mentions of combined entities whenever possible unless it is clear that the author has specifically intended a particular combination to be interpreted almost as one entity. Multiple mentions for which splitting might mean losing some information (see last example), we add the information in the comment box (the whole phrase as it should be annotated) along with the entity tag. Acronyms and abbreviations for entities are included within the entity span.
  - He was given **injections** and **medicines** to help with the pain. [Annotate separately: treat\_drug for each one]
  - I’ve had a hard time with meds, so my psychiatrist practitioner put be on **buspar/buspirone** and it honestly has been a lifesaver with my **anxiety**. [treat\_drug, treat\_drug, medC]
  - There’s a case of **alpha lipoic acid supplementation** causing **Hirata Disease (insulin autoimmunity)**. [env\_dietary, medC, medC]
  - Reported symptoms include **suicide** and **depression, nausea, fatigue and loss of libido**. [tricky to decide for “suicide and depression” because the extra “and” appears to imply a connection between the two. When in doubt, try and annotate separately: medC for each one]

- A **serological or antibody test** is not typically used to diagnose an active coronavirus infection. [Annotate as one entity: diag]
  - Whether **bacterial, viral or fungal infection**, these will help. [infection—medC, infection—medC, medC]
  - **Corona virus disease (Covid)** often comes with a severe **cough**. [medC, medC]
- Annotating treatments: If you are unsure if a mentioned treatment is considered a drug or a therapy, chose `treat_therapy`. We consider this tag to be slightly more general than `treat_drug`. If the author is being unspecific, go with the more general option. Vaccines can be both preventative as well as acute treatments, so we annotate them as *treat\_therapy*.
    - She’s getting **treatment for her allergies** right now. [treat\_therapy]
    - The **vaccine** will help with **Covid**. [treat\_therapy, medC]
- Biomedical mentions as part of other entities: Biomedical mentions which are part of other entities (e.g., organizations, company names etc.) are not to be annotated since they do not refer to the biomedical concept.
    - This was a study from Cancer Research UK.
    - I got treated in the Michigan Academy of #sleep medicine.
    - Would you want to go to a Breast Cancer March?
- Incomprehensible documents: Documents which are not understandable, confusing or incomplete can be difficult to annotate because of the missing context or our lack of medical knowledge etc. In such cases we annotate the entities and relations as much as possible without spending an extensive amount of time to research potential meaning. A quick Google search is fine, but if this does not clear things up, move on to the next document. The same goes for documents that are incomplete or end in the middle of a sentence.
- General biomedical mentions: Domain-related biomedical terms like “pandemic”, “epidemic” or “quarantine” are generally not to be annotated unless they are specifically described e.g. as a contributing factor to a medical condition (see first two examples). The same goes for standalone mentions of terms like “mental health”. Only annotate it (e.g. as medC) if it is actually understood as an issue of health problem. It helps substituting with “physical health” or just “health”. If you would not annotate those, general mentions of “mental health” should not be annotated either.
    - During the **pandemic** his **mental health problems** continuously got worse. [env\_socio-economic][medC]

- As we deal with the **health crisis** caused directly by **COVID19**, we must not lose sight of its knock on effects on other health targets like **cervical cancer screenings**. [medC, medC, treat\_therapy]
- Your mental health is just as important as your physical health.
- Medical events belong to medC: Entities that broadly refer to medical events or incidents such as accidents, hospitalization, pregnancy etc. belong to the medC category:
  - Peoples’ **hospitalization** make their families **immobilized with fear**. [medC, qol]

Note that non-actions or an omission to do something can also be a medical event and therefor a condition:

- **Not going to the hospital** is the reason so many actually **die**. [medC, medC]
- My **depression wasn’t diagnosed** until later that year. [medC, medC]
- Biochemical processes: Biomedical states or processes that are not listed in ICD-10 as medical conditions should be annotated as biochemical process (first example). However, if the author understands the entity as a condition or it is included in ICD-10, we annotate it as a medical condition (second example). If you are not able to decide how it was intended or if it is generally unclear if something is a regular biochemical process or rather a condition, choose the stronger label (medC).
  - Every month I get **depressed** because of my **period**. [medC, biochem\_process]
  - My **sporadic period** decided to show up yesterday after all. [medC]
- Mental health conditions vs. moods: The language around mental health conditions like depression is often (incorrectly) used to describe moods or emotional states. In cases for which we can not decide if they refer to the actual medical condition, be bold and assign the label anyway.
  - Watching this commercial with all this wholesome family activities is making me really **depressed**. [medC]
- Other entities: Entities which do not fit into one of the other entity classes or cases which are highly unclear, can be assigned to the **other** category. This includes mentions of family histories with a certain condition or hereditary factors. All relations in the relations tag set can potentially be used to connect to such an entity.
  - This **vaccine** will make us **immune**. [treat\_therapy, other]
  - **Nature** increases **immunity**. [env\_geo\_climate *pos.influence\_on* other]

- The preliminary data suggests that people with **blood type 0** are protected against the severity of **Covid19** infections. [other *pos\_influence\_on* medC]
- We have a family friend who had a **mild heart attack** - his doctor said it was caused by his recent **vaccination** ... never mind that he had 98 % **blockage of valve** and **heart disease runs in his family**. [medC, tread\_therapy, medC, other]
  - \* vaccination *not\_cause\_of* mild heart attack
  - \* blockage of valve *cause\_of* mild heart attack
  - \* heart disease runs in his family *cause\_of* mild heart attack
- Quality of Live assessments: The qol entity class captures descriptions that “assess the impact of disease and its management, including interventions, on the well-being of the patient”<sup>2</sup>. Potential description can be quite broad, so we only annotate them if there is a clear connection or reference to the condition or treatment that is impacting and causing the assessment (first example). If the connection is unclear (second example) we do not annotate potential candidates as qol.
  - I suffer from **depression** and **routines are really hard for me**. [medC, qol]
  - Was so miserable that I cried every single day and got so **depressed** that I had to be put on meds. [medC]
- Determiners: We generally do not include determiners like *a* or *the* within the span of the entity.
- Environmental and risk factors: This category includes external factors on medical conditions, treatments etc. The different kinds of factors can be tricky to differentiate. Be sure to consider the authorial intent when tagging the instance.
  - **Silicosis** is an **interstitial lung disease** caused by **breathing in tiny bits of silica**. [medC, medC, env\_pollution]

Unless clearly intended as a treatment, we categorize mentions of supplements as env\_dietary:

- There’s a case of **alpha lipoic acid supplementation** causing **Hirata Disease (insulin autoimmunity)**. [env\_dietary, medC, medC]

## 2.2 Relations

Once we have identified all mentions of biomedical entities in the document, it is time to look into how they relate to each other. Table 3 contains all types of relations that we are interested in. A relation always connects two named entities. One entity can

---

<sup>2</sup><https://www.ncbi.nlm.nih.gov/books/NBK235120/>

be related to multiple other entities. It can have both incoming and outgoing relational connections.

### 2.2.1 Relation Types

There are several types of relations which can connect biomedical entities. Table 3 lists the different relation types as well as the corresponding relation tags that are available for the annotation. Along with each relation tag, it also contains examples to illustrate how these relations might look in practice. The labels are divided into groups according to which types of entities are being connected by a relation.

Relation Types	Relation Tags	Examples
<b>treat – medC</b>		
treats	treats	I just had a cortisone injection to sort out my allergies If your primary problem is a burning sensation, antacids are your best bet
	does_not_treat	Even the vaccine by Pfizer is debated for treating Covid effectively
	worsens	Magnesium can worsen your diarrhea
prevents	prevents	My county’s health dept. is also reporting very high vaccine effectiveness at preventing hospitalization. I waited too long to take my Zoloft and now I have been feeling irritated all day.
	does_not_prevent	Those antibiotics they gave me did not prevent any infection.
causes	cause_of	You can’t undo surgery, you’ll make scar tissue
contraindicates	is_contraindicated_drug	antacids are your best bet – unless you have a heart or kidney condition
prescribes	prescribed_for	Ketamine nasal spray could be prescribed for depression within the year.
influences	pos_influence_on	People who smoke marijuana are less likely to be obese. It’s no cure, but taking more vitamins has improved my overall mental state.
	neg_influence_on	In the UK Drs prescribe HRT for many GenitoUrinary conditions despite the higher risk of Endometrial Cancer.
<b>medC – treat</b>		
side effect	side_effect_of	Paxil withdrawal update: still on 1/2 pill every other day, so sleepy, dizzy and a little sad.
<b>env/pathogen/biochem – medC</b>		

Relation Types	Relation Tags	Examples
causes	cause_of not_cause_of	I have stress acne but some virus can cause the cancer Bacteria have nothing to do with Covid inflammation
influences	pos_influence_on neg_influence_on	The mountain air helps with my anxiety A high #BMI increases your risk for heart disease.
prevents	prevents does_not_prevent	working out regularly prevents stress All those chemicals don't guarantee you won't get Covid!
<b>medC – medC/biochem</b>		
symptom	has_symptom	Panic attacks got me fainting
co-occurrence	cause_of	Venous thrombosis can be part of COVID infections.
similarity	not_cause_of is_similar_to	You can't die directly from covid. The common cold and Covid have pretty much the same symptoms.
<b>treat – treat</b>		
interactions	neg_interaction	"Cortisol added to insulin is a bad combination and the most problematic hormonal combination for belly fat." ...
	pos_interaction	bless the holy combination of Prozac and Wellbutrin, folks.
similarity	is_similar_to	Ibuprofen and Naproxen essentially do the same thing
<b>diag – medC/pathogen</b>		
diagnoses	may_diagnose	so that's when I had the MRI and then told that it was menieres which I knew anyway. My fluctuating blood sugar levels became evident from the blood tests they administered
	may_not_diagnose	A serological test is not typically used to diagnose an active coronavirus infection hr-hpv ... the sexual transmitted virus isn't tested for in sexual health screenings
<b>pathogen – biochem</b>		
causes	cause_of	Coronavirus could be destroying pancreatic beta cells.
	not_cause_of	This particular hormone isn't the reason for lower oxygen levels in the blood.
<b>medC/treat/env/diag – qol</b>		

Relation Types	Relation Tags	Examples
influences	pos_influence_on	Thanks to Trikafta it's getting easier to walk up the stairs without taking a nap afterwards.
	neg_influence_on	I'm feeling super weak today. Might be from chemo therapy
causes	cause_of	I suffer from depression and routines are really hard for me.
	not_cause_of	For once the reason I can't work out is not my chronic back pain, but just lack of motivation
<hr/>		
<b>general</b>		
type of	is_type_of	Leukemia is a type of blood cancer FDA approves remdesivir, rheumatoid arthritis drug to treat Covid patients.
other	other	The flu is much more dangerous than Covid.

Table 3: Relation classes and tags between the different types of entities.

### 2.2.2 Workflow

To annotate the relation between two medical entities, e.g., in an instance like

- Totally anecdotal but I started taking **vitamin d3** and when I got **covid my symptoms** where much less severe,

we suggest the following workflow:

1. Note which groups of entities (see Table 2) the annotated entities belong to.
  - Say you have identified the mention of a drug and a disease, the respective entity groups which are involved in this pair are *treatment* and *medical condition*.
2. Refer to Table 3 and find the relation category that corresponds to your entity pair.
  - For our example pair of drug – disease, we consequently identify the first row of Table 3 (treat – medC) as the relevant one.
3. Check out the major relations listed on the left-hand side of Table 3, and decide which option is the one you are looking for. Once you have settled on a major relation, take a look at the corresponding annotation tags listed in the right column of the table. Pick the tag that most accurately represents the relation that is being expressed in the text.

- In our example, we first chose prevents as a major relation, and consequently *prevents* as the tag.
4. Go back to INCEpTION to annotate the relation you just identified. Note that we annotate directed relations. This means that in order to draw the connection we need to make out which is the outgoing and the incoming entity. Connect the two entities accordingly: Click on the entity from which the relation goes out and drag the arrow that appears towards the second entity.
  5. From the drop-down menu that appears in the panel on the right-hand side, chose the relation tag that you have identified earlier.
  6. You made it!

### 2.2.3 Notes

- Note that it is possible that there is in fact no relation between our entities.
- You might have a hunch which relation tag you want to assign between two entities right away. However, especially at the beginning of annotation, we would like to ask you to still follow the procedure outlined above. This process ensures consistency and let's you get familiar with the relation types that are available.
- Some relations are stated (only) indirectly, or a relation is mentioned later in a document. For such instances we infer the respective relation and annotate accordingly.
  - Question with potential indirect claim: Is **steroid** induced **psychosis** a thing? Inferred relation: *is\_sideeffect*
  - Relation stated in different sentence: I take Paxil for my anxiety. It helps quite a bit. Inferred relation: *treats*
- Relation-entity triples can be interconnected to model complex relations:
  - FDA approves remdesivir, rheumatoid arthritis drug to treat Covid patients.
    - \* remdesivir [treat\_drug] *is\_type\_of* rheumatoid arthritis drug [treat\_drug]
    - \* remdesivir [treat\_drug] *treats* Covid [medC]

When using the *is\_type\_of* relation, we draw it from the more specific to the more general or generic term. In some cases this allows us to have entities inherit relations from the more general entity that they are connected to:

- \* **#AFib** and **#cancer** are both **progressive diseases** with a **range of treatment options**. [medC, medC, medC, treat\_therapy]
  - AFib *is\_type\_of* progressive disease
  - cancer *is\_type\_of* progressive disease
  - range of treatment options *treats* progressive disease

- Say we have documents with multiple mentions of the same entity (entity1) that carry different types of additional information as well as another entity (entity2) that related to (both mentions of) entity1. In such cases, we draw a *is\_type\_of* relation between the different versions of entity1 and connect entity2 to the more specific version of entity1:
  - \* **Pancreatic cancer** is a leading cause of **death** largely because there are no **detection tools** to diagnose the **disease in its early stages**.  
[medC, medC, diag, medC]
    - disease in its early stages *is\_type\_of* Pancreatic cancer
    - detection tools *may\_not\_diagnose* disease in early stages
    - Pancreatic cancer *cause\_of* death
- Similarity relations: When using the *is\_similar\_to* relation we always draw the relation from the first entity to the second entity.
- Ambiguity: If you're struggling to decide on the relation tag because it might be ambiguous, it can help to think about the content with regards to the fact-checking task. Ask yourself which entity-relation combination/claim should to be fact-checked in the document you're looking at. Cases in which two options remain, chose the more general relation:
  - **Physical activity** is effective in the prevention and treatment of a variety of **chronic diseases**, like **heart disease**, **hypertension**, **obesity**.  
[env\_habitual, medC, medC, medC, medC] env\_habitual *pos\_influence\_on* medC
- Relations within and across sentence borders: In general, relations can cross sentence borders. If statements get more complex, sometimes the same entity is referenced multiple times. This means we have to decide which of the mentions we chose to draw the relation to or from. To model a claim in such cases, we generally try and stay within sentence borders, and draw the relation to the entity instance which is closest (first example). However, if connecting two separate entities to the same third entity instead of connecting to two separate entities helps to model the main claim of a document, definitively do so (second example).
  - **CoronamedC1** is a dangerous disease especially for **old people**env\_socio-economic. People need to understand that if you are **already sick**medC2 with something else **CoronamedC3** will make it worse.
    - \* env\_socio-economic *neg\_influence\_on* medC1
    - \* medC2 *neg\_influence\_on* medC3
  - But **deaths**medC1 from **pneumoniamedC2**, **heart diseasemedC3** and the **flumedC4** are down. So what's going on? [...] **Deaths**medC5 previously attributed to other diseases are now blamed on **Covid**medC6.
    - \* medC2,3,4 *cause\_of* medC5

- \* medC6 *not\_cause\_of* medC5
- medC - medC relations: As a non-medical expert it can be tricky to decide if a medical condition is considered a symptom (*has\_symptom*) or if it might be caused by and co-occur with an other medical condition. If this is not clear from the text, we chose *has\_symptom* over *cause\_of*.
  - **Cytokine surge in Covid causes lung fibrosis.** [biochem\_process, medC, medC]
    - \* Covid *has\_symptom* Cytokine surge
    - \* Cytokine surge *cause\_of* lung fibrosis
- Influences vs. causes: Negative influences are often similar to causal relations. If there is a clear causal relation between say a environmental factor and a medical condition, use *cause\_of*. However, especially when entities are negated *neg\_influence\_on* can help to model those claims. In such cases we only annotate the medC (not the negation) and connect them by drawing a *neg\_influence\_on* relation from the other involved entity:
  - A full **lockdown** since January would have resulted in massive **suicide, depression**, missed **cancer** screenings, missed treatments for **cardiovascular disease** etc. [env\_socio\_economic, medC, medC, medC, medC]
    - \* lockdown *cause\_of* suicide
    - \* lockdown *cause\_of* depression
    - \* lockdown *neg\_influence\_on* cancer
    - \* lockdown *neg\_influence\_on* cardiovascular disease
  - **Cytokine surge in Covid** in long run can cause deleterious effects on **health**. [biochem\_process, medC, medC]
    - \* Covid *has\_symptom* Cytokine surge
    - \* Cytokine surge *neg\_influence\_on* health
- treat - medC relations: The *treats* relations implies that a treatment works or is successful. For this to apply, there has to be an existing condition that is being treated.
  - **Losacar** is used for the treatment of **essential #hypertension**. [treat\_drug *treats* medC]
  - **buspirone** [...] has been a lifesaver for my **anxiety**. [treat\_drug *treats* medC]

In contrast to that, we use *prevents* when someone talks about preventing, averting or warding off a condition ahead of time. We chose *prevent* over *pos\_influence\_on* when the claim is presented strongly and this preventive treatment is described as successful, or if the term *prevents* is assertively used:

- My county’s health dept. is also reporting very high **vaccine** effectiveness at preventing **hospitalization** [...] [treat\_therapy *prevents* medC]

Indicators for *pos\_influence\_on* (as well as *neg\_influence\_on*) are statements that are non-committal, evasive or relative (hedging):

- Treatment X reduces risk of/lowers the likelihood of/increases the chances of etc.
- **Losacar** is used for the treatment of essential #hypertension & reduction in the risk of #**stroke**. [treat\_drug *pos\_influence\_on* medC]
- In general, we can not and do not want to fact-check death counts. While this is strongly related to a disease (especially to the discussion around Covid), it is not directly a biomedical claim that is within the scope of our task. In most cases, this means that we consider such documents as off-topic and only annotate the entities. However, we want to and can capture claims that state that deaths might not be caused by Covid, but by another disease instead (Covid *not\_cause\_of* deaths, seasonal flu *cause\_of* deaths).
- Underlying or contributing condition: There are cases in which the not just one, but a certain combination of entities (e.g., an underlying condition and a trigger) are the cause of another condition, event (or another biomedical entity). For such instances we draw the cause relation from the trigger to the resulting entity, and connect the trigger with a suitable relation to the underlying or contributing condition:
  - People with **Brugada Syndrome** are advised to avoid **hot baths** because they can trigger **cardiac arrest**!
    - \* hot baths [env\_habitual] *cause\_of* cardiac arrest [medC]
    - \* hot baths [env\_habitual] *neg\_influence\_on* Brugada Syndrome [medC]
  - She has a **peanut allergy** and so the **peanuts** in the dish caused her to go into **anaphylactic shock**.
    - \* peanuts [env\_dietary] *cause\_of* anaphylactic shock [medC]
    - \* peanuts [env\_dietary] *neg\_influence\_on* peanut allergy [medC]