

# **Entity Annotation Guidelines for English COVID-19 Related Tweets<sup>1</sup>**

**YLab  
Zhejiang University  
Version 1.0**

---

<sup>1</sup>The standard refers to Zhang H, Zong Y, Chang B, et al. Medical Entity Annotation Standard for Medical Text Processing [C]//CCL. 2020: 561-571.

# Contents

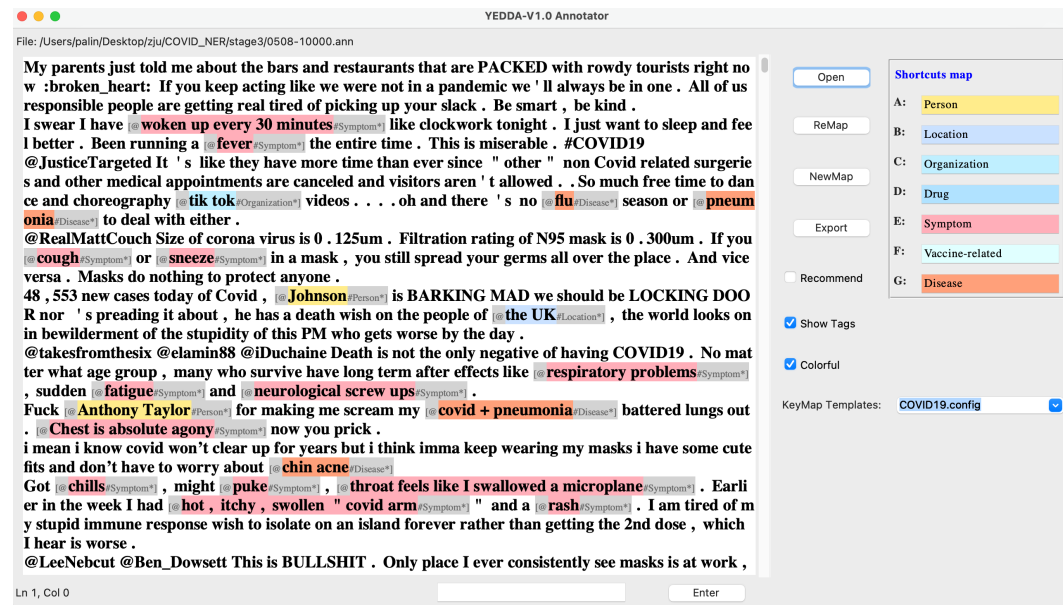
<b>INTRODUCTION .....</b>	<b>3</b>
<b>ANNOTATION TOOL AND PROCESS .....</b>	<b>4</b>
<b>ENTITY TYPE DESCRIPTION.....</b>	<b>5</b>
<b>ANNOTATION GUIDELINES.....</b>	<b>6</b>
GENERAL ANNOTATION SPECIFICATION .....	6
DETAILED ANNOTATION SPECIFICATION .....	7
1. <i>Person</i> .....	7
2. <i>Location</i> .....	7
3. <i>Organization</i> .....	8
4. <i>Symptom</i> .....	9
5. <i>Drug</i> .....	12
6. <i>Vaccine</i> .....	13
7. <i>Disease</i> .....	16
8. <i>@ (Mention)</i> .....	18
<b>ANNOTATION QUALITY EVALUATION .....</b>	<b>20</b>

# Introduction

The COVID-19 pandemic continues to bring up various topics discussed or debated on social media. In order to explore the impact of pandemics on people's lives, it is crucial to understand the public's concerns and attitudes towards pandemic-related entities (e.g., drugs, vaccines) on social media. However, models trained on existing named entity recognition (NER) have limited ability to understand COVID-19-related social media texts because these datasets are not designed or annotated from a medical perspective. Therefore, this annotation guideline was designed to help researchers build a NER dataset for medical research, which could be utilized to develop better medical social media understanding tools and facilitate computational social science research, especially on epidemiological topics.

# Annotation Tool and Process

All the annotation work could be done using the YEDDA annotation platform. The readers could check more details about YEDDA on its GitHub page: <https://github.com/jiesutd/YEDDA>. The screenshot of using YEDDA to annotate entities for tweets is shown as follows:



The annotation process includes 3 phases:

1. In the pre-annotation phase, all annotators are requested to conduct 3 rounds of annotation (with training). F1 value is used as the metric of inter-annotator agreement. Annotators with F1 greater than 80% are selected to enter the formal annotation process. The annotation guidelines are also iteratively updated throughout this process.
2. In the formal annotation phase, annotators label the tweets in pairs (3 pairs in total) to ensure each tweet is annotated twice. When an inconsistency occurs, another annotator steps in to determine the final annotation of the tweet.
3. After the formal annotation phase, the project team conducts a quality control check on the labeled results to ensure that the annotated tweets meet the annotation guidelines' requirements. The final inter-annotator agreement is 85.0% in the F1 value.

## Entity Type Description

We define 7 entity types based on public health research needs, including 3 general entity types and 4 medical entity types for annotation. The description of each entity type is shown as follows:

Entity Type	Coverage	Example
Person	Real people, fictional characters, characters from literature religious characters	Virginia Wade, Harry Potter
Organization	companies, schools, hospitals, etc.	Google
Location	Country, province, city, etc.	France, China
Symptom	Clinical manifestations, generally referring to the patient's discomfort and abnormal manifestations through examination	Fever, headache
Drug	Substances used to prevent, treat and diagnose diseases, including clinical drugs, antibiotics, etc.	Dexamethasone
Vaccine	Include description of vaccine type and vaccine brand.	Moderna
Disease	Cause of the unhealthy state of the patient or the diagnosis made by the doctor on the patient, which can be treated, including disease or syndrome, injury or poisoning, etc.	Cardiopathy, hypertension

# Annotation Guidelines

## General Annotation Specification

1. The annotations should be concrete and specific, rather than abstract or general. For example: "woman, girl, place, location, father, small town, hospital, school, college" should not be annotated.
2. If the English abbreviation or general name of the entity appears, it needs to be annotated.
3. Nested entities are not allowed, in other words, only the longest boundary of one entity is marked, and other entities within should not be annotated.
4. Extra spaces should not be included, spaces should not appear at the beginning or end of the entity range.
5. Words should be complete. Intercepting part of the annotating word are not allowed, a word either belongs to one entity or does not belong to any.
6. When the recommendation mode of YEDDA is on, manual confirmation is required for the recommended entities, otherwise the output mark file will automatically record all the recommended results, which will affect the accuracy of the annotation.
7. Only annotate definite article "the" when it appears before "Location" and "Organization". "the" in other entity types will not be annotated.  
For example: the U.S., the U.K., the WHO, the NIH, the Westchester Country Center should be annotated.
8. Words that describe the severity of symptoms should not be annotated, only the symptoms themselves, but should be annotated when there are body parts.

## Detailed Annotation Specification

### 1. Person

The Person entity type includes the names of real people, fictional characters, characters in literature, and religious characters. If no specific name appears, do not annotate it.

● Some representative examples:

Example	Explanation
The Queen has said...	"The Queen" is not a specific name
Mr <b>[@Weir#Person*]</b> also told	Honorifics are not annotated
allergist <b>[@Troy Baker#Person*]</b> , MD, tells what to expect and explains that serious allergic reactions are extremely rare	Only annotate the personal name, not the identity
<b>[@GOP#Organization*]</b> , MAGA leaders , Qleaders said	No personal name appears, so there is no need to mark
<b>[@Katten#Organization*]</b> lawyers offer <b>[@UK#Location*]</b> employers as the best approaches towards COVID19 vaccination in the workplace.	If no specific name appears, it is not necessary to annotate it. For example, the lawyer of "Katten" does not refer to a specific name, so it is annotated as an organization. Similarly, there is no specific personal name in "UK employer", so it is only necessary to annotate "UK".
How a secret military experiment left <b>Black Georgians</b> wary of COVID-19 vaccines	Black Georgians indicates race, and no specific names appear, so they do not need to be annotated
now <b>god</b> knows who he is spread it to	God should not be annotated
This mother <b>[@Miriam#Person*]</b> is a schismatic who ignores the teaching of the Church,	Only the personal name, rather than role

### 2. Location

The Location entity type includes continent, ocean, country, province, city, county, region, street, township, town, village, airport, military base, military zone, railroad, highway, bridge, strait, bay, harbor, river, lake, park, grassland, coal mine, pasture, farm, concert hall, theater, church, temple, library, museum, art gallery, exhibition center, park, zoo, botanical garden, railroad station, square, building, building, stadium (gymnasium), swimming pool, racing track, shopping mall, supermarket, bookstore and other urban public facilities, as well as certain urban buildings and fictitious premises.

- Some representative examples:

Example	Explanation
joining [ <b>@Chicago city#Location*</b> ]... many [ <b>@african countries#Location*</b> ]...	location+city/province/country, the city/province/country should be annotated together
I know [ <b>@BC#Location*</b> ] COVID updates/reporter questions are minimal,	British Columbia (province name, located in western Canada)
[ <b>@the U.S.#Location*</b> ]	Definite article should be annotated

### 3. Organization

An organization is an entity, such as a **company**, an **institution**, or an **association**, comprising one or more people and having a particular purpose. It usually includes stock (securities) exchange, national or international legislative or executive department, business group (companies, enterprises, factories), television stations, radio station, newspaper and magazine, publishing house, political party or party, school, research institution, hospital, clinic, post office, band, sports team, league, parliaments or congress, army, cafe, bar, restaurant, hotel, and fictitious institution

- Some representative examples:

Example	Explanation
a signature of its [ <b>@Wuhan lab#Organization*</b> ] creators?	Wuhan lab should be annotated together as Organization
We call on the Mayor, [ <b>@TFL#Organization*</b> ] and [ <b>@Unite the Union#Organization*</b> ] to lobby [ <b>@the UK government#Organization*</b> ] with the	The country name + government should be marked as organization as a whole
[ <b>@TNXP#Organization*</b> ] connection to [ <b>@Alzheimer#Disease*</b> ]'s is pretty tenuous.	<b>TNXP is an abbreviation for a pharmaceutical company</b>
Pains me to hear of [ <b>@NHS#Organization*</b> ] and care staff who refuse the Covid vaccine.	National Health Service

- Other hard examples:

Example	Explanation
a volunteer administering at [ <b>@Health Centre#Organization*</b> ] shares	Generally, organizations should be specific, such as xx Hospital, xx Army, and should not be annotated when only the word hospital appears. However, if the first letter of the word is capitalized, such as Health Centre, it is considered to refer to



	an specific institution and is now annotated as an organization.
Calling <b>811</b> or the using [@AHS#Organization*]	811 or 9/11 is not considered as an organization
[@Scotland#Location*] covid update <b>RECAP</b>	RECAP is not an organization
[@DWP#Organization*] and other depts can get access to that record.	Department for Work and Pensions
[@CDC#Organization*] guidelines are not followed in [@Providence#Location*], [@RI#Location*] schools no mandatory testing no vaccines. The risks are is real. Sad!	Centers for Disease Control
partner at on a <b>vaccination rollout guide</b> to help pharmacies navigate the various state approaches to efforts. Read more	Vaccination guidelines are not annotated
With my co-morbidities and my work in <b>healthcare</b> ,	Unnecessary to annotate
<b>Covid Epi Weekly</b> : Best of Times, Worst of Times The third US surge is fading fast but variants, some ominous, are spreading fast. Vaccination is picking up steam but we are failing to address equity and pandemic <b>fatigue</b> is high. We must hang on until most of us are vaccinated. 1/	Unnecessary to annotate
<b>The Holocough</b> - Health care worker dies after second shot of "COVID-19" vaccine via	Unnecessary to annotate
<b>Old Wine in New Bottles</b> : Low-Tech Approaches to a Covid-19 Vaccine	Unnecessary to annotate

#### 4. Symptom

Symptoms are discomfort or pain experienced by patients. They are usually discomfort that patients subjectively experience, such as abdominal pain and dizziness, or pathological changes that patients notices on their own, such as hematuria and hematochezia, or mobility dysfunction.

Common symptoms of COVID-19 include but are not limited to the following table:

COVID symptoms			
fever	muscle aches	red eyes	shortness of breath
dry cough	muscle pain	diarrhea	abdominal pain
tiredness	body aches	skin rash	pneumonia
loss of taste	congestion	fever	dyspnea
loss of smell	runny nose	chills	hypoxia
aches	nausea	cough	respiratory failure
pains	vomiting	shortness of breath	shock

headache	diarrhea	difficulty breathing	multiorgan dysfunction
sore throat	myalgia	breathing difficulties	fatigue
nasal congestion	sputum		

**Note: The above table is just for reference. All words that the annotators consider to be symptoms should be annotated.**

Here are some rules to guide the annotation process for Symptom entity type.

**Rule 1:** When parts of body appear before and after symptoms, they should be annotated together. Adjectives and adverbs that describe the severity is not necessary to be annotated:

- ✧ I was feeling a lot of [**@body ache#Symptom\***] and [**@soreness all over#Symptom\***]
- ✧ Cue splitting [**@headache#Symptom\***], [**@nausea#Symptom\***]and the [**@shivers#Symptom\***]
- ✧ The second COVID vaccine is giving me horrible [**@body chills#Symptom\***]
- ✧ ...my [**@arm was painig#Symptom\***]...
- ✧ I feel normal just a little [**@sore on my arm#Symptom\***] but I am good now
- ✧ My [**@arm was sore#Symptom\***] for 2 days,
- ✧ Some mild [**@fever#Symptom\***]
- ✧ low grade [**@fever#Symptom\***]
- ✧ [**@Pain swelling on the arm#Symptom\***] where you got the shot 2. [**@Fever#Symptom\***] or [**@chills#Symptom\***] 3. [**@Tiredness#Symptom\***] 4. [**@Headache#Symptom\***] Remember to get your 2nd shot.
- ✧ [**@Fever#Symptom\***] has broke but have terrible [**@head and muscle ache#Symptom\***].
- ✧ My father in law has been in agonizing [**@pain#Symptom\***] for the last 24 hours
- ✧ The 2nd dose of the COVID vaccine must have woken all the dormant peach takka and Smirnoff in my immune system from 2015 because I [**@woke up feeling drunk#Symptom\***] asf. I was prepared for a [**@fever#Symptom\***] and [**@headache#Symptom\***]. Absolutely no one told me to be prepared for senior yr of highschool. Wtf is this
- ✧ Am suffering a lot of [**@pain in my left eyes#Symptom\***] and [**@headaches#Symptom\***] since yesterday as a result of taking the Corona virus vaccine.
- ✧ True: Earth spiked a bit of a fever in 2020, partly because of cleaner air
- ✧ My 2nd dose of the covid vaccine is treating me well! Only having [**@arm soreness#Symptom\***] and feeling [**@fatigued#Symptom\***]/pretty [**@sleepy#Symptom\***].

**Rule 2:** When the symptom is plural, past tense, or progressive tense, the complete word should be marked.

- ✧ ...cover [**@coughs#Symptom\***] and [**@sneezes#Symptom\***]...
- ✧ some just had [**@sore arms#Symptom\***] some just

- ✧ [**@body aches**#Symptom\*]
- ✧ [**@headaches**#Symptom\*]
- ✧ [**@fevers**#Symptom\*]
- ✧ [**@coughing**#Symptom\*]

**Rule 3:** The symptoms should be annotated separately when connecting words appear between them, such as “with”, “and”, etc.

- ✧ I feel [**@achey**#Symptom\*] with [**@chills**#Symptom\*] but at least I will never have to worry about going on a ventilator.

Note the differences in the following example:

- ✧ [**@Fever**#Symptom\*] has broke but have terrible [**@head and muscle ache**#Symptom\*].

In this case, “and” connects not the symptom but the body part, and the symptom is “ache”, so it needs to be fully annotated.

**Rule 4:** Implied expressions of symptoms should also be annotated.

- ✧ I asked about the vaccine but she said even though I have a [**@lung condition**#Symptom\*] now
- ✧ I feel [**@freezing from the inside out**#Symptom\*] have a [**@100.5F temp**#Symptom\*]
- ✧ 1st day of shot, felt [**@woozy**#Symptom\*] with a slight [**@headache**#Symptom\*]. 2nd day, mild [**@flu like symptoms**#Symptom\*],
- ✧ [**@Knocked me off my feet**#Symptom\*].
- ✧ I got [**@headaches**#Symptom\*] after dose two but I also had [**@not slept well**#Symptom\*] that night for other reasons.
- ✧ i could [**@barely sleep all night**#Symptom\*]
- ✧ Had [**@fluey symptoms**#Symptom\*] from jab but ...
- ✧ [**@flu-like symptoms**#Symptom\*]
- ✧ since then been v [**@sneezy**#Symptom\*] and a bit [**@sniffly**#Symptom\*].
- ✧ my [**@head is killing**#Symptom\*] me i hope this does not happen with the second dose of the covid vaccine
- ✧ Day 1 after COVID vaccine and so far just a little more [**@tired than usual**#Symptom\*], some mild body aches and a sore arm from the injection. Nothing compared to how awful I felt with COVID.
- ✧ ~21 hours after Dose 2 and I have a [**@100°F fever**#Symptom\*], [**@cutaneous hyperesthesia**#Symptom\*], [**@restless fatigue**#Symptom\*].

**Rule 5:** The “feel like” in the symptom should also be annotated, otherwise it may cause ambiguity.

- ✧ I woke up this morning [**@feeling like I have been hit by a truck**#Symptom\*]
- ✧ Even though I started the day out [**@feeling like I was hit by a Mack truck**#Symptom\*].

● Other hard examples:

Example	Explanation
*cough*	Internet term, not used to express symptoms, so it should not be annotated

With the <b>headache</b> of distribution on top of procurement, how will the world reach the herd immunity levels needed to defeat the virus?	Here "headache" does not indicate a symptom according to our definition, so it is unnecessary to annotate
got the shot: <b>[@Pain#Symptom*]</b> <b>[@Swelling#Symptom*]</b> Throughout the rest of your body: <b>[@Fever#Symptom*]</b> <b>[@Chills#Symptom*]</b>	Be careful not to include "Throughout the rest of your body"
if you add that to lockdown <b>fatigue</b> ,	Here "fatigue" is not used to indicate symptoms, so it should not be annotated
Firstly, there is natural <b>fatigue</b> with Covid measures.	Here "fatigue" is not used to indicate symptoms, so it should not be annotated
Really did make me appreciate the face masks helping me avoid getting <b>sick</b> this last yea	Here "sick" refers to general sickness, so it should not be annotated
QA! Can a COVID-19 vaccine make you <b>sick</b> with COVID-19? No, none of the vaccines in <b>[@the U.S. #Location*]</b> contain the love virus that causes COVID-19. The teaches your immune system how to recognize and fight the virus.	Here "sick" refers to general sickness, so it should not be annotated
the second vaccine will probably make you <b>sick</b> .	Here "sick" refers to general sickness, so it should not be annotated
<b>Much higher antibody titres</b> but elevated rates of adverse effects too.	Not a symptom
<b>[@Fever#Symptom*]</b> post vaccine? Do not <b>panic</b> , treat it!	Panic is not a symptom here
not only pandemic <b>fatigue</b> , but the inevitable selection of vaccine-escape variants if we keep cases high. and there lies	"fatigue" is not a symptom here
unless you apply NPI, cases increase. but people are <b>tired</b> .	"tired" is not a symptom here
WATCH: Dr. <b>[@Fauci#Person*]</b> discusses vaccine safety, COVID <b>fatigue</b> in appearance on News 12	"fatigue" is not a symptom here

## 5. Drug

Drugs are defined as substances which is used to prevent, treat and diagnose diseases in general, include clinical diagnostic reagents.

When annotating drugs, pay attention to following rules:

1. The properties of drugs should not be annotated as "drugs".

- The general name of most drugs should be annotated, such as nutrients, antibacterial drugs, emergency drugs, etc., although these are general names, a certain corresponding therapy range comes along. When words like "commonly used drugs", "drugs" appears alone, the range of them is too broad, which should not be annotated.

Some representative examples:

- ✧ so hopefully that means it was not the [@placebo#Drug\*].
- ✧ i used this same concoction others were mere [@vitamin c#Drug\*], [@zinc tabs#Drug\*] [azithromayicin#Drug\*] which is used to treat [cough#Symptom\*]

## 6. Vaccine

Including vaccine name, vaccine type and vaccine brand, etc.

Common types of Covid vaccines include but are not limited to the following list:

- ✧ mRNA
- ✧ protein subunit
- ✧ subunit
- ✧ vector
- ✧ inactivated virus
- ✧ peptide
- ✧ synthetic
- ✧ conjugate
- ✧ plasmid

Common brands of Covid vaccines and their manufacturer include but are not limited to the following table<sup>2</sup>

Vaccine brand	Another name	Vaccine type	Manufacture
Oxford– AstraZeneca	Vaxzevria	viral vector vaccine	British University of Oxford
	Covishield		British-Swedish company AstraZeneca
			Coalition for Epidemic Preparedness Innovations (CEPI)
Pfizer– BioNTech	Comirnaty	mRNA vaccine	German company BioNTech
			American company Pfizer

<sup>2</sup> This list is organized from [https://en.wikipedia.org/wiki/List\\_of\\_COVID-19\\_vaccine\\_authorizations](https://en.wikipedia.org/wiki/List_of_COVID-19_vaccine_authorizations)

COVID-19 vaccine			
Moderna	COVID-19 Vaccine Moderna	mRNA vaccine	Moderna
			National Institute of Allergy and Infectious Diseases （NIAID）
			Biomedical Advanced Research and Development Authority （BARDA）
			Coalition for Epidemic Preparedness Innovations （CEPI）
Janssen	Johnson & Johnson COVID-19 Vaccine	viral vector vaccine	Johnson & Johnson
	COVID-19 Vaccine Janssen		Janssen Pharmaceutica
			Beth Israel Deaconess Medical Center （BIDMC）
Sinopharm-BBIBP	BBIBP-CorV	inactivated virus vaccine	China National Pharmaceutical Group Corporation （Sinopharm）
	Hayat-Vax		CNPGC
Sputnik V	Sputnik V COVID-19 vaccine	viral vector vaccine	Russian Gamaleya Research Institute of Epidemiology and Microbiology
CoronaVac		inactivated virus vaccine	Sinovac Biotech
Covaxin		inactivated virus vaccine	Bharat Biotech （BBIL）
			Indian Council of Medical Research （ICMR）
Sputnik Light		viral vector vaccine	Russian Gamaleya Research Institute of Epidemiology and Microbiology
Convidecia	Convidicea	viral vector vaccine	CanSino Biologics （CanSinoBIO）
			Beijing Institute of Biotechnology of the Academy of Military Medical Sciences （AMMS）
Sinopharm-WIBP	WIBP-CorV	inactivated virus vaccine	China National Pharmaceutical Group （Sinopharm）
			Wuhan Institute of Biological Products

EpiVacCorona		peptide vaccine	State Research Center of Virology and Biotechnology VECTOR
Zifivax		subunit vaccine	Anhui Zhifei Longcom Biopharmaceutical
Abdala		subunit vaccine	Center for Genetic Engineering and Biotechnology (CIGB)
Soberana 02		conjugate vaccine	Instituto Finlay de Vacunas
CoviVac		inactivated virus vaccine	Chumakov Centre at the Russian Academy of Sciences
QazCovid-in	QazVac	inactivated virus vaccine	Research Institute for Biological Safety Problems
Minhai		inactivated virus vaccine	Minhai Biotechnology Co.
			Shenzhen Kangtai Biological Products Co. Ltd.
COVIran Barakat		inactivated virus vaccine	Shifa Pharmed Industrial Co
Chinese Academy of Medical Sciences	Chinese Academy of Medical Sciences COVID-19 vaccine	inactivated virus vaccine	Chinese Academy of Medical Sciences
Medigen	MVC-COV1901	protein subunit vaccine	Medigen Vaccine Biologics and Dynavax Technologies
ZyCoV-D		DNA plasmid based COVID-19 vaccine	Cadila Healthcare
			Biotechnology Industry Research Assistance Council

**Note:** The manufacturer of the vaccine should be annotated as *Organization* if appears alone, and side effects from the vaccine should be annotated as *Symptoms*.

**Rule 1:** When referring to a vaccine by its manufacturer, it should not be annotated as an organization, but as Vaccine.

- ✧ Alhamdulillah 4 of my family members had the [@Pfizer vaccine#Vaccine\*]
- ✧ Had the [@Pfizer one#Vaccine\*]
- ✧ Urging people to not get the [@JJ vaccine#Vaccine\*] is crazy.

**Rule 2:** 'Vaccine name' + 'COVID-19 Vaccines' should be annotated completely

- ✧ [@Moderna COVID-19 Vaccine#Vaccine\*]
- ✧ post 2nd dose of [@Moderna Covid-19 vaccine#Vaccine\*]
- ✧ the first [@COVAX vaccine#Vaccine\*] shipment reaches [@Ghana#Location\*],

- ✧ [@Pfizer covid-19 vaccine#Vaccine\*]

**Rule 3:** 'disease' + 'vaccine' should also be annotated as Vaccine (including expressions such as 'COVID+vaccine')

- ✧ Taken [@Polio, Measles, Tetanus, Diphtheria vaccines#Vaccine\*] as a baby.
- ✧ Received my 1st [@Covid shot#Vaccine\*] yesterday.
- ✧ I got the second dose of the [@COVID19 vaccine#Vaccine\*] yesterday.
- ✧ I have had [@BCG vaccine#Vaccine\*],[@MMR vaccine#Vaccine\*] ,[@Yellow fever vaccine#Vaccine\*],the [@flu vaccine#Vaccine\*] ,and yesterday I had the [@Pneumonia vaccine#Vaccine\*],

### Some hard examples:

- ✧ 'Location + Vaccine', the location should be annotated as Vaccine, then the location is not annotated separately.
- ✧ Such as: in charge of distributing [@West Virginia's vaccine#Vaccine\*] says it ma
- ✧ Thanks and Was given [@Oxford AstraZeneca vaccine#Vaccine\*].
- ✧ do not use the [@live virus#Vaccine\*] that causes COVID-19.
- ✧ were noted in about 50% of participants in the [@mRNA-1273#Vaccine\*] group after the second dose..
- ✧ QA! Can a COVID-19 vaccine make you sick with COVID-19? No, none of the vaccines in [@the U.S.#Location\*] contain the [@love virus#Vaccine\*] that causes COVID-19. The teaches your immune system how to recognize and fight the virus.

## 7. Disease

The cause of patient's abnormal conditions or the doctor's diagnosis on the patient, including disease, syndrome, injury or poisoning, etc. The details of each type of disease are shown in following table:

Type	Description	Example
Disease or Syndrome	Refer to the name of disease or syndrome.	Hypertension, pneumonia, heart disease, sepsis, deformity, etc.
Injury or Poisoned	After the patient is injured or poisoned, it causes damage to the body, which leads to an unhealthy state.	Alcoholism, Sitootxism
Organ or Cell damage	If the abnormal or damaged organs and cells may endanger body, although they are part of the body, they have become a	Intracranial hemorrhage (ICH), Birth trauma



	pathogenic factor and harm health.	
--	------------------------------------	--

**Note:**

- Generally the names of some diseases can be very long, with prefix like "XX sex", "XX shape", "XX type", and modifications of body parts (one or more). To ensure the completeness and specificity of the disease, it should be annotated with these prefixes.  
Example: Acute viral infection, primary HIV infection, acute icteric hepatitis (AIH), acute icteric hepatitis b, rotavirus infection, seasonal infantile diarrhea, liver, kidney and thyroid diseases.
  - Most of the general terms should be annotated, such as nutritional diseases, metabolic diseases, suppurative and non-suppurative syndromes, etc. Although these are general terms, they corresponds to specific disease categories, so they should be annotated. Special cases: When "common diseases", "frequently-occurring diseases", "diseases" etc. appear individually, such general terms is too broad and should not be annotated.  
Examples: [@Hypertension#Disease] is a common and frequently-occurring disease that seriously harms human health.
  - When there are several types of disease, "disease + type" or "type + disease" is annotated as a whole, but the type is not annotated when it appears alone.  
Examples: Type 2 diabetes
  - Pay attention to some examples of using virus that refers to diseases. Only when it is determined that the virus does not refer to a specific disease should it not be annotated. Otherwise, it can be treated as the disease and annotated as Disease.
- ✧ saying he had earned it over a lifetime of leadership on [@HIV#Disease\*] research and [@AIDS#Disease\*] relief
  - ✧ Yep. [@Zika#Disease\*]? Nothing burger; I
  - ✧ [@Ebola#Disease\*]? Boomslang snake and other reasons for [@hemorrhagic fevers#Disease\*]

**Some hard examples:**

Example	Explanation
a regression w/o confounders like <b>obesity</b> , [@Azithromycin#Drug*],        [@steroids#Drug*], [@comorbidities#Disease*],	Obesity does not necessarily mean disease
Felt like I had [@flu#Disease*]	Only need to annotate Flu as disease
UNCLASSIFIED said mosquitoes not infected [@w/yellow fever#Disease*]	If there is no space between the disease and 'w', it will be annotated together, otherwise, the

	annotate range will not include w/
[@TB#Disease*] cert entering is for applicants staying 6 months	Tuberculosis

### Disease vs. Symptom:

The mean difference between a disease and its clinical symptoms is that: a disease is confirmed by diagnosis, and a disease is essentially a physical damage; while symptoms are essentially a phenomenon that appears after physical damage, such as the patient's discomfort, physical abnormal changes, but these are often superficial phenomena seen by the patient or doctor. As a doctor, it is necessary to confirm the patient's disease through further diagnosis, which means that there are essential differences between the disease and symptoms.

For "infection" related entities, there are a few things to pay attention to:

1. If there is a clear cause of disease (such as viruses, bacteria or body parts) together with "infection" to form a word, it will be annotated as "disease" as a whole. Such as [@HAV infection#disease], [@coccal infection#disease], [@upper respiratory tract infection#disease]
2. When the word "infection" appears alone, if the context clearly indicates that it is a reference to a certain disease, it is annotated as a disease, otherwise it is annotated as "symptom"
3. If the modifier before the word "infection" indicates the degree or frequency, it is annotated as "symptom" as a whole

## 8. @ (Mention)

In English tweets, @mentions are sometimes included as part of the tweet body, affecting the meaning of tweet. Therefore, the @ mentions participating in the sentence expression can be further annotated, so as to expand the sample size of the two types of entities, Person and Organization. Specifically, annotate @mentions in the following conditions:

1. If the @mention appears at the beginning of the tweet, it is necessary to determine whether the first letter of the first word after the @mention is a lowercase. If it is lowercase, the @mention is considered to be part of the sentence, so it needs to be annotated. Otherwise, there is no need to annotate.

**Example 1:** @Pathik\_Trader in some case rtPCR negative but covid positive if fever not reduce in 3 days then please test ct scan

**Explanation:** 'in' first letter is lowercase, so need to annotate corresponding

@Pathik\_Trader

**Example 2:** @TheDavidL81 Here's the part where I explain Trump's strategy.

**Explanation:** 'Here' first letter is uppercase, so no need to annotate @TheDavidL81

2. If the @mention appears in the body, it needs to be annotated.

**Example 3:** The hypocrisy of @JoeBiden and @KamalaHarris is tragic.

**Explanation :** @JoeBiden @KamalaHarris appears in the body, so it needs to be annotated

3. If the @mention appears at the end of the tweet, it is necessary to check the word before @mention. If there is a hashtag (#) or punctuation, it is not necessary to annotate. Otherwise, it is considered as part of the tweet and needs to be annotated.

**Example 4:** Breathing difficulties with covid. #COVID @NHSuk

**Explanation:** @NHSuk before it is a hashtag # , NO need to annotate

**Example 5:** it's #COVID19 coverage has been disgraceful. @BBCTalkback

**Explanation:** @BBCTalkback before it is a punctuation, NO need to annotate

**Example 6:** understand about COVID19. Follow @Cleavon\_MD

**Explanation:** @Cleavon\_MD NEED to annotate

## Annotation quality evaluation

Named entity recognition is generally modeled as a sequence labeling problem, therefore micro-averaging (Micro-F1) is often used as evaluation indicators, the calculation formula is as follows:

$$P = \frac{\text{The number of correctly predicted entities}}{\text{The total number of predicted entities}}$$
$$R = \frac{\text{The number of correctly predicted entities}}{\text{The total number of entities}}$$
$$\text{Micro} - F1 = \frac{2 \times P \times R}{P + R}$$