

# Emory-RLS Collaboration

## Project Overview

This project is a collaborative project between [Emory NLP](#) and [Real Life Science](#), aiming to build an NLP pipeline that recognizes all entities (e.g., [PATIENT\\_CONDITION](#) and [PATIENT\\_TREATMENT](#)) in disease specific social media data. For instance, in the following example, [IBS](#) is the condition that the patient/poster has, and [Digesterol](#) is the medicine that the patient is currently taking. Our goal is to capture all disease-related information in the data to help the patient, the caregiver, and the health care practitioner.

Hi there I am looking to try a new [IBS][PATIENT\\_CONDITION](#) medication called [Digesterol][PATIENT\\_TREATMENT](#)! would appreciate hearing any feedback good or bad re it.

Our dataset is gathered from Reddit using a keyword-based searching method. We now have 50 double annotated posts and a basic annotation guideline. In order to train the NLP model, we plan first to annotate ~2,000 posts.

## Annotation Scheme

We currently have a coarse annotation scheme and will keep updating it during the annotation process. The guideline contains 6 major entity types, including [Patient Condition](#), [Patient Treatment](#), [Caregiver Condition](#), [Caregiver Treatment](#), [Unspecified Condition](#) and [Unspecified Treatment](#).

### Patient Condition

This label refers to the condition that the writer of the passage has or is dealing with. It could be a specific disease or any related symptoms, side effects, or impairment caused by a disease or medication.

I've been dxd with [Narcolepsy][PATIENT\\_CONDITION](#). Excessive [daytime sleepiness][PATIENT\\_CONDITION](#) is its main symptom. This is also dx with a sleep study.

When labeling conditions, general words such as 'symptoms', 'problems', or 'disease' should be excluded unless they are parts of the disease names. For example, in the following sentences (1) and (2), 'problems' and 'symptoms' should be excluded. However, in the sentence (3), 'common' should be labeled since it is a part of the disease name.

(1) I have [dry eye][PATIENT\\_CONDITION](#) problems.

(2) I don't know if it's the [MS][PATIENT\\_CONDITION](#) symptoms.

(3) I was diagnosed with [cvid][PATIENT\\_CONDITION](#) [common variable immunodeficiency][PATIENT\\_CONDITION](#).

If the post contains a disease description, we should label the complete description. As shown in the following example, the whole phrase should be labeled instead of 'pinhead bump' only.

I have now noticed [a yellowish whitish pinhead bump on the inside of my lower lip]PATIENT\_CONDITION.

## Patient Treatment

A **Patient Treatment** is any medical treatment (e.g., medicine, surgery, or even counseling) that the writer of the passage is using or receiving. Note that the condition that the **Patient Treatment** is treating does not have to exist in the text.

I was given a prescription called "[magic mouthwash]PATIENT\_TREATMENT" by my Oncologist. It has [benadryl]PATIENT\_TREATMENT and something else in it.

**Treatment** also include any non-traditional or ill-advised treatments like alcohol or hard drugs.

I medicate with [whiskey]PATIENT\_TREATMENT.

If the post only contains a general reference without indicating specific medication or treatment, we should not label it.

I think I took the right medication. I think I took the right medication ([methotrexate]PATIENT\_TREATMENT).

## Caregiver Condition

Different from **Patient Condition**, **Caregiver Condition** is the condition affecting someone the writer of the passage cares for (e.g., family members or friends).

My brother Austin was diagnosed with [ADHD]CAREGIVER\_CONDITION at 5 and later also DX with [ODD]CAREGIVER\_CONDITION tendencies.

## Caregiver Treatment

Likewise, **Caregiver Condition** is not the condition that the caregiver (assumed to be the writer of the passage) is undergoing, but rather a treatment that someone they care for is taking.

Since getting on the [gluten-free diet]CAREGIVER\_TREATMENT, my child's picky eating habits have improved.

## Unspecified Condition

**Unspecified Condition** appears in the context where the subject of the condition is unclear or general. Usually, the person that suffers from the condition does not fit into the **Caregiver Condition** label (e.g., other users in the previous post threads).

Lots of people with [dementia]UNSPECIFIED\_CONDITION are originally diagnosed with [MCI]UNSPECIFIED\_CONDITION.

## Unspecified Treatment

Similar to **Unspecified Condition**, **Unspecified Treatment** are labeled when the people who receive the treatment are unclear or general.

This is an over the counter medication called [Mucus Relief DM]UNSPECIFIED\_TREATMENT. Its an [expectrant cough suppressant]UNSPECIFIED\_TREATMENT.

Note that **Unspecified Condition/Treatment** could exist in the same text with **Patient/Caregiver** types.

the free clinic treat me for [gonnerhea]PATIENT\_CONDITION and [chlaymedia]PATIENT\_CONDITION. and why is it dangerous to medicate with [doxycycline]UNSPECIFIED\_TREATMENT?