DRAFT

Drug Product/Dietary Supplement Classification

Contents

Introduction

The purpose of this task is to create a gold standard corpus of tweets that are classified as positive (containing the mention of a drug product or dietary supplement) or negative (no mention of drug product or dietary supplement). The following guidelines will outline for annotators rules to consider when distinguishing between positive and negative tweets. The classification guidelines include examples for each rule to provide clarity for the annotator. In the examples, the drug product/dietary supplement name will be in **bold** text.

Corpus Scope and Definitions

The corpus will contain tweets classified as either (1) contains a mention of a drug product or dietary supplement, herein labeled a 'drug product/dietary supplement tweet' (DPDST) or (2) does not (a 'non-drug product/dietary supplement tweet' (non-DPDST)). The rest of this section identifies and defines the entities that are included when classifying DPDST tweets and entities that are not included in the corpus.

Corpus Entities Defined

Our definition of a drug product follows that stated by the FDA, and includes the differentiation made between a drug and drug product According to the FDA [cite], a **Drug** is defined as:

- A substance recognized by an official pharmacopoeia or formulary.

- A substance intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease.

- A substance (other than food) intended to affect the structure or any function of the body.

- A substance intended for use as a component of a medicine but not a device or a component, part or accessory of a device.

- Biological products are included within this definition and are generally covered by the same laws and regulations, but differences exist regarding their manufacturing processes (chemical process versus biological process.)

A **Drug Product**, is distinguished from a drug as *the finished dosage form that contains a drug substance, generally, but not necessarily in association with other active or inactive ingredients*. For a tweet to be classified as positive, the entity mentioned should be a drug product; that is the finished dosage form.
The drug product name contained in the tweet may be the nonproprietary (generic) name, the trade (brand) name or the drug class. Table 1 lists entity types with their subtypes and examples.

We will also rely on the FDA definition for dietary supplements:

The law defines dietary supplements in part as products taken by mouth that contain a "dietary ingredient." Dietary ingredients include vitamins, minerals, amino acids, and herbs or botanicals, as well as other substances that can be used to supplement the diet.

Dietary supplements come in many forms, including tablets, capsules, powders, energy bars, and liquids. These products are available in stores throughout the United States, as well as on the Internet. They are labeled as dietary supplements and include among others

- *vitamin and mineral products*
- *"botanical" or herbal products—These come in many forms and may include plant materials, algae, macroscopic fungi, or a combination of these materials.*
- *amino acid products—Amino acids are known as the building blocks of proteins and play a role in metabolism.*
- *enzyme supplements—Enzymes are complex proteins that speed up biochemical reactions.*

Entity Type	Subtype	Example
Drug Product	Nonproprietary (generic) name	sertraline
	Trade (brand) name	Zoloft
	Drug Class – Therapeutic	antidepressant
	Drug Class – Mechanism of action	ACE inhibitor
	Drug Class – Mode of action	diuretic
Dietary Supplements	Vitamins	ascorbic acid (vitamin C)
	Minerals	calcium
	Amino acid products	leucine
	Enzyme supplements	amylase
	Botanicals or herbs	St. John's wort

Table 1: Examples of entity types included in corpus

Entities Excluded from Corpus

The following entities are outside the scope of this corpus and should not be considered by annotators when classifying tweets:

1. Cosmetic/Drug Products - Some cosmetic products are also considered drug products due to their having two uses, one of which is therapeutic (e.g. antidandruff shampoo, toothpaste with fluoride, deodorant with antiperspirant). As it may be difficult to discern when a user is referring to a pure cosmetic or a duel cosmetic/drug product, we have excluded these products

2. Illicit Drug Mentions – including those being used for therapeutic purposes (e.g. medical marijuana) unless it is mentioned in an approved drug product form

3. IUPAC and IUPAC-like chemical names

4. Vitamins and Minerals directly from food sources

Drug and Supplement Resources

Determining drug product and dietary supplement names can be a difficult task. Therefore, the following resources may be used by annotators to assist them in resolving questions about whether a name is a possible reference of interest:

- **DrugBank** – a searchable database of drug products by their generic and trade names, and includes dietary supplements. It contains 10,999 entries and includes products in the US, EU, and Canada https://www.drugbank.ca/

- **Drugs@FDA** – a database of FDA approved brand name and generic drug products, OTC human drug products and biological therapeutic products https://www.accessdata.fda.gov/scripts/cder/daf/

- **MedlinePlus** – a database curated by the National Institutes of Health of prescription and OTC drug products, as well as information on supplements and herbs https://medlineplus.gov/druginformation.html

- **Dietary Supplement Label Database** – database containing label derived information from dietary supplement products marketed in the United States https://www.dsld.nlm.nih.gov/dsld/index.jsp

DPDST Case Examples

Drug Products

Positive Examples

Generic, Brand or OTC Names

A tweet should be classified as a DPDST if it contains the name of a drug product, whether the mentions is the generic, trade or OTC name, and this mention is referring to the drug product. The following are examples of DPDSTs:

1) *@KimJ721 **Benadryl cream** and a waterproof Nexcare bandage are how I conquer itchy bug bites.*
2) ***Humira** welt is better 2 days sooner after 2x 25mg **Benadryl** and **topical diclofenac gel**. \ud83e\udd17*

The mention does not have to be a first-hand experience or other relation to the drug product to be classified as DPDST, so long as the mention is referencing an actual drug product:

3) *Sunglasses and **Advil***
4) *AbbVie Ordered by Jury to Pay $15 Million to **Depakote** Victim - Bloomberg https://t.co/SP2I6sBR1d*
5) ***Tamiflu** may have odd side effects, particularly in children, experts say*
6) *FDA approves Genus' **Goprelto** (**cocaine hydrochloride** nasal solution) for local anesthesia... their goes the market for $LCI*

The reference in (3) is from a song lyric, however the song is referencing the drug product, Advil. (4) is referencing a news article that is discussing the drug product Depakote, and (5) is an informational tweet that is speaking of a drug product. The mention in (6) is also an informational tweet and although the active ingredient mentioned is commonly thought of as an illicit drug, in this case it is in an approved drug product and therefore would be classified as positive.

Non US Names

The same drug product may be marketed under another name in other countries. As long as it is referencing a drug product, it should be classified as DPDST.

7) *Really? I'd far prefer to know someone who needed **Escitalopram** (a.k.a. **Cipralex**) was actually taking it. Why should that bar the person from otherwise normal life? Going off the meds would likely be far more risky (suicide above all).*
8) prescribed **escitalopram**(**lexapro**) and hydroxyzine(atarax) for now. if they don't help i'll go to a specialist. doctor gave me a list of therapists that i can start talkin to. neat

Drug Class

Tweets that mention a drug product in terms of its drug class should be classified as DPDST.

9) *It's no exaggeration to describe anti-microbial resistance as an existential threat to modern medicine, so thank you GPs for cutting **antibiotics** prescriptions by 13% since 2012*

10) *I mean, I was even advised against my **anti-convulsant** meds (I'm epileptic), just to be safe, even though obviously this put me at risk of seizures. And yet an experimental hormone regime is OK? I find this story completely bewildering*

11) *Anecdotal,I know,but my husband had 2 heart attacks; despite having low cholesterol levels,was prescribed very high **statin** dosage, stopped taking them of his own volition having looked at the evidence.Still around 8 years later,at the age of 85*

Negative Examples

Homonyms

Tweets should be classified as non-DPDST if it is obvious the drug product name in the tweet is not referring to a drug product. For example, if the tweet is referring to a person, place or thing that shares a name with a drug product:

12) *Just finished helping **Lyza** color her xmas cards (: only made 4 tho lol 2 more to go*

13) *10 dollars for a **codeine** double lp. Thanks urban outfitters Santa Monica . #vinyl #comeups http://t.co/ksrUuW9wMg*

Not referring to a drug product

Tweets that mention a drug product name, but the context of the tweet indicates that the person is not referencing a drug product should be classified as negative.

14) *Tool – **Enema***

Example (14), the band Tool has an album, Aenima so this reference is a misspelling and not referring to a drug.

Endogenous Substances

Endogenous substances are substances developed or synthesized naturally within the body. Mentions referring to endogenous substances should be classified as non-DPDST

15) *In need of some **oxytocin**. Someone let me borrow your dog.*

16) High Fructose Corn Syrup is a poison disguised as food, contributing to diabetes, weight gain, heart attacks, stroke & **insulin** resistance.

In example (15), the user's mention of borrowing a dog indicates that they are looking for a natural release of oxytocin and not a drug product.

Figurative Language

17) *#Hospitalmovies **Enema** Of The State.*

18) *She's **morphine**, queen of my vaccine. My love, my love, love, loveeee. â¤ï¸*

19) *Rain. Nature's **Ambien**.*

Example (17) is a play on words and examples (18) and (19) are uses of figurative language and are not necessarily referring to a drug product.

Dietary Supplements

The considerations taken into account for the classification of tweets containing dietary supplement mentions are similar to those of drug products listed above.

Positive Examples

Product names

20)   Took **melatonin** *so I might actually sleep tonight but also the Super Blue Blood Moon is keeping my werewolf side awake so we'll see what happens. It's so bright outside it's casting shadows.*

21)   *Taking a* **calcium** *supplement each day during one's 20s, 30s, and 40s can keep bones strong and healthy.*

22)   all these people talking about how they never go outside lmao i my blood test results say i have to start taking a **vitamin D** pill or something because i dont receive enough sunlight

23)   I'm too young to take **Centrum Silver** but I'm too old to take Flintstone vitamines.

Class Names

24)   *It's totally acceptable to take your* **vitamins** *and then eat a Kit Kat right?*

25)   *If your goal is gains in muscle size and strength, then ISO100 is your perfect workout partner. Loaded with muscle building* **amino acids***, ISO100 can support even the most serious resistance-training programs. Shop now* https://goo.gl/Eg3od3

Negative Examples

Homonyms

When what the user is discussing has a name in common with a dietary supplement, the tweet should be classified as non-DPDST

26)   *yo* **gaba** *gaba for life*

27)   *I need to build a scraper that can extract Emails from websites using* **Selenium** *Python or Python. (Budget: $250 - $â€¦* https://t.co/2GOaqRWMnZ

Endogenous Substances

Endogenous substances are substances developed or synthesized naturally within the body. Mentions referring to endogenous substances should be classified as non-DPDST

28)   *Atherosclerosis is an accumulation of plaque in your artery. Plaque is consisted of fat, cholesterol,* **calcium** *and other substances in your blood.*

29)   *Most people are deficient in* **Magnesium** *- needed for over 300 processes in the body. Heart palpitations, stress, anxiety, twitches, fatigue & more are just a few of the deficiency symptoms. Add in leafy greens,*

avocados, dark chocolate, pumpkin seeds & many more #magnesium #health

30)   The sun gives you **vitamin D** so being bald is like having a solar panel on the roof.

Special Cases

Non-English Language Tweets

Tweets that are primarily in non-English language should be classified as non-DPDST:

31)   *RT @azrarahman: **Antara** sbb tak baca cepat2 time duduk antara dua sujud â˜ºï¸ â˜ºï¸ â˜ºï¸ â˜ºï¸ â˜ºï¸* *https://t.co/joUbMi1jv2*

32)   *@itsHlulani Chubu yi ba u khoma hala mahlweni u cina-cina vanhu va ku wo **cina** kasi wo ti tshuva \ud83d\ude02\ud83d\ude02* *https://t.co/jwMcaU4REi*

Limited Context

In some tweets, there may be the mention of a drug product or supplement name, however, the tweet does not contain enough contextual clues to determine if the user is referring to a drug product or something else. For example, the following tweets all contain the name of a drug product:

33)   *I like **lyrica***
34)   *I need a **mili** now*
35)   *1 minutes til **Bismuth***
36)   *I need some **stress relief** …*

In (i), (ii), and (iii), however, the user could be referring to a person or band. In (iv), the term "stress relief" is a common phrase and in this case it is unclear to what the user is referring. These instances may appear more frequently with products that share a name with people, musical bands, places, etc.; or when the product name is similar to a general phrase.

For some other products their names may be less frequently associated with anything but a drug product name, thereby eliminating some, but not all, of the doubt that the user is discussing a drug product

37)   ***Metformin** is freaking ridiculous*
38)   *I love you, **albuterol***

In cases with limited context, the annotator will need to use available resources and domain knowledge when determining the classification of the tweet. If the annotator cannot be sure due to lack of context, the tweet should be classified as non-DPDST

Abbreviations

Tweets should be considered DPDST if the relevant mention is in an abbreviated form:

39)   *Caramel Machiatto with an **NSAID** to jumpstart my day*

40)   *Starting off the morning strong, Dr. Rachel Saylor will be presenting her work on serotonin dynamics after* **SSRI** *administration: A voltammetric and microscopic study at 8:30 am in 205C* #Pittcon #weloveserotonin
41)   **MAOI** *helped me. Withdrawal now and I want to die*

Annotators should keep in mind that abbreviations are not unique and be mindful of the fact that it may not be referring to a drug product

42)   #PennState **SSRI** *News Feb 26: Opioid research preproposals due March 1, plus more events, news, and funding opportunities.* http://ow.ly/sAwe30iFtA7

The above tweet (42) would be classified as non-DPDST because SSRI is an abbreviation for Social Science Research Institute.

Street Names/Slang
Some drug products are mentioned in tweets by a slang or informal (street) name. If these mentions are discussing a drug product the tweets should be classified as DPDST:

43)   *I told everyone stop taking* **xanz** *years ago...now 2018 it's a trend.. y'all slow*
44)   *Kids are taking* **z-bars** *and alcohol recreationally, add fentanyl & they die* urbandictionary.com/define.php?ter ...
45)   *Sally looks like she's off her nut on* **ket** #ThisMorning
46)   *Damn 30MG* **addys** *arent shit anymore.... anything stronger out there?*

As with abbreviations discussed above, these are not unique to only drug names and annotators should be mindful of other possible entities being reference.

47)   *My 5 yo gives 1000% more effort with only the prospect* **Z-bars** *and juice boxes. Man, fuck these Nats.*

In example (47), Z-bars is a reference to a type of granola bar and the slang term for Xanax