# Pharmacovigilance from social media: Annotation guidelines

## DIEGO LAB

**Arizona State University**
**Version 1.2**

**Authors:**

Abeed Sarker, Ph.D[1]
Karen O' Connor[1]
Rachel Ginn[1]
Robert Yao[1]
Karen Smith, Ph.D[2]
Azadeh Nikfarjam, MS[1]
Graciela Gonzalez, Ph.D[1]

**Past and current annotators:**

Karen O' Connor[1]
Rachel Ginn[1]
Robert Yao[1]
Mari Firago[1]
Amanda Goodin[1]
Azadeh Nikfarjam, MS[1]
Karen Smith, Ph.D[2]

[1]Department of Biomedical Informatics, Arizona State University, Scottsdale, Arizona
[2]Rueckert-Hartman College for Health Professions, Regis University, Denver, CO

**Website:** diego.asu.edu

**Last updated:** February, 2016

| Version Tracking | | |
|---|---|---|
| **Date** | **Changes** | **Author** |
| **April 18, 2013** | **Initial Guidelines** | **Robert Yao** |
| **June 16, 2014** | **Reformatting and reorganization of original. Modification of rules for clarity. Additional instructions for binary annotations.** | **Karen O'Connor** |
| **Jan 13, 2015** | **Modification of specific details..** | **Abeed Sarker** |
| | | |
| | | |

**Note:** This document is updated sporadically. Current annotation guidelines may change over time.

**Contacts:**

Karen O'Connor: Karen.Oconnor@asu.edu
Abeed Sarker: abeed.sarker@asu.edu

# Annotation Tool

For annotation tasks, this project used the BEAT version 1.2 annotation tool. This Java based tool was designed and is maintained in-house. Improvements were made to the tool during the annotation project to improve performance and increase accuracy of annotations.

# Overview of Annotation Process

1. Select minimal tokens for type: ADR, indication, beneficial, interaction, drug, or other
2. Select type that corresponds to token
>    2a. Indication vs beneficial: prescriber and patient intent (indication) overrides unintended secondary effect (beneficial, e.g. sleep).
3. Select concept that directly corresponds to type

# Annotation Guideline

This guideline describes the types of information that should be annotated for the ADR project. It covers the general annotation guidelines explaining the parts of the text that should be annotated and defines that concept types that should be annotated. Examples presented in the guideline will mark text that should be annotated as **BLUE** text and text that should not be included in the annotation will be marked as **RED** text. The concept types will be indicated in brackets [ ] following the annotations.

# General Annotation Guidelines

## Full Text Annotation

For each concept annotated select the minimum number of tokens that conveys the concept. Examples:

>    Have not **lost any weight [IND]** yet
>    Started **hallucinating [ADR]**... NOT cool !!!
>>    This helps keep day to day **pain [IND]** at a lower level . However if I use the full dose it makes my **brain fill like mush[ADR]**.
>>    who need alcohol when you have **gabapentin[DRG]** and **tramadol[DRG]** that makes you **feel drunk[ADR]** at 12oclock

**Questions to ask**: Do the tokens annotated capture the concept? Would the meaning be changed if any of the tokens were removed?

If the concept is present in the input lexicon verbatim or similar to the annotated text select the concept that is closest syntactically and semantically. Do not include unnecessary modifiers unless the modified phrase has its own concept.

Examples:
>    Couldn't take because of **allergic reactions [ADR]**.
>>    I still take it for **chronic pain [IND]** from fibromyalgia **[IND]** and it helps **[OTH]** tremendously !

This has helped **[OTH]** with the **shortness of breath** **[IND]** when I am active . I have been using it for almost a month **[OTH]** now .

Exclude:

Need to find examples

**Questions to ask**: Is there an equivalent concept in the lexicon to the compound annotation (modifier(s) + basic concept)?

Adjectives and adverbs should be included in the annotation if the adjective (adverb) modifies a basic concept and the combination (adjective (adverb) + basic concept) is different than the basic concept alone, the entire minimum annotation phrase should include the adjective (adverb) + the basic concept.

Example:

Need to find examples

Exclude:

Adverbs and adjectives that do not alter the basic concept

o **Severe** **headaches[ADR]** , **nausea[ADR]** , and **dizzyness** **[ADR]**. o
I 've noticed she is **more** **tired[ADR]**, other than that nothing .

**Questions to ask**: Would removing the adjective/adverb from the annotation change the meaning of the basic concept?

When idioms, metaphors, or similes are used, select the entire phrase that captures the semantics in the least number of tokens

Did not help me at all. Made my **heart feel like it was flip flopping[ADR]**

made my **stomach walls feel like they were sticking together** **[ADR]**, so no.

@reneereed37849 ugh i'm trying to figure out how i got **infectious colitis[IND]**... the **flagyl** **[DRG]** and **levaquin[DRG]** have **knocked me on my butt[ADR]**

**knocked me clean out for 20 hours** **[ADR]**a day ... and made me **put on weight[ADR]**

**Questions to ask**: Are all the tokens annotated needed to capture the meaning of the expression? Would removing any of the tokens change the meaning of the annotation?

so far , get real **dizzy** been a **week** and still cant drive , cant sleep . take 2mg 4 to 5 times a day

When concepts are presented in a list or clustered, annotate each mention individually, mapping the individual annotation to the compound concept.

Example:

Works**[OTH]** well as an antipsychotic **[IND]** , puts on weight **[ADR]** though , when decreasing dosage have **pain in back** **[ADR Concept: back pain]**, **neck[ADR**

**Concept: neck pain]** , **arm[ADR Concept: arm pain]** and **leg[ADR Concept: leg pain]**

severe allergid reaction **[ADR]**, **Tongue** **[ADR Concept: tongue swollen]** and **throat swelled** **[ADR Concept: throat swollen]** up

## Binary Annotation

Binary annotation does not require any token selection. Its purpose is to determine and mark whether the text contains a mention of an ADR (see definition in Concept Types section). If the text does contain an ADR mention, select 'Comment has ADR' checkbox.

# Concept Types

There are five concept types that are being annotated: ADR, Indication, Beneficial, Drug and Other. The definition and examples for each type are explained in the following sections.

## Adverse Drug Reaction (ADR)

An ADR is an effect of the drug that is not desired. This includes a worsening of the indication, drug withdrawal and its associated symptoms, and a loss of effectiveness of the medication. When not clear if something is the indication but shows up after taking the drug, assume it is an ADR.

Examples:

It worked **[OTH]** , but since I started the treatment I 've noticed a lot of **hair loss [ADR]**

Seems to have opposite effect, increased **agitation[ADR]**

does anyone get **headache [ADR]** an hour after tysabri infusion ? I don't know if vyvanse was a good idea #**nosleep [ADR]**

Put me to **sleep[ADR]** and when I wasn't sleeping, I was a **nervous wreck[ADR]**. → "nervous wreck " is ambiguous could be indication or ADR

side affects of this **withdrawal[ADR]** contains **confusion[ADR]** , **deliriousness[ADR]** , **nauesousness[ADR]** and **insomnia[ADR]** #**withdrawal[ADR]** #trazodone **[DRG]**

#omgicantwait #help → when withdrawal is mentioned, annotate 'withdrawal' as an ADR as well as any specific withdrawal symptoms mentioned

#cymbalta **[DRG]**      **withdrawal [ADR]** has reached a peak, **lost vision [ADR]** and almost crashed my car from a **brain zap [ADR]**. thanks a zillion #elililly #bigpharma

Seemed like it helped **[OTH]**      for a few weeks but then **stopped working [ADR]**   → tachyphylaxis (medication worked and then did not work) is an ADR

Exclude:

Mentions that are from texts that are not first person accounts, e.g. repeated from a commercials, literature, ads.

o  so latuda is for people who are bi-polar , but a side effect is increase of  **suicidal thoughts**. what the fuck .

o  fosamax side effects :  **femur fractures esophagus problems hardening of the jaw** 1.888.520.5202 . se habla espanol

- o Cymbalta is legal, an antidepressant.....side effects: **nausea, dry mouth, constipation**, **diarrhea**, **fatigue** , &amp; **THOUGHTS OF SUICIDE**
- o **Osteonecrosis of Jaw** with Denosumab: 3cases in yr1-3 vs 5cases in yr4-7 #EULAR2013 #osteoporosis
- o effient is more effective than plavix ... but causes more **bleeding** . that's why it's not for patients with a prior stroke or tia , or over 75 .

  Mentions that are non-specific or cannot be interpreted to a specific concept
  - o @SkylerTunes oh man levofloxacin? I've had that same exact pill **it hated me** though. Hope it does you justice!
  - o Oh my gosh ... I took Lyrica , the fibromyalgia `` miracle drug '' for a little over a month . I had gotten up to 150 mg per day . THE **SIDE EFFECTS** WERE ABSOLUTELY HORRIBLE .
  - o I'm not sure if Excedrin is the right way for me to go for my TMJ headaches. Vyvanse + Caffeine is pretty **rough**. Ideas?

A detrimental side effect should be annotated as an ADR even if the patient sees it as a positive effect. (NOTE: Under certain circumstances these could be beneficial, see Beneficial Effects)
Examples:

> Vyvanse **[DRG]** make me so **hyper [ADR]** and creative and i think of so many tweets
> @river__tam olanzapine **[DRG]** is amazing srsly, the **dreams are immense[ADR]**, no addiction, its gorgeous, like brave new world soma to me
> its good I do n't have any noticeable side effects and it made me **lose weight [ADR]** , only thing is if your insurance does n't pay for it it costs almost 500 dollars

## Interaction

Interaction type is used when an ADR is caused by an interaction of two or more drugs as indicated by the patient. To annotate:

1) create an annotation instance for each drug responsible for the ADR
2) for every instance, select the type as interaction
3) select the corresponding semantic concept for the ADR
4) for each ADR annotation instance, select a different target drug
5) repeat 2 thru 4 for each drug responsible

Example:

> Created **mood swings [INT] (annotate 2x's)** in combination with Celexa **[DRG]** → set the target drug of second annotation to Celexa

# Indication

An indication is the sign, symptom, syndrome, or disease that is the reason or the purpose for the patient taking the drug or is the desired primary effect of the drug. Additionally, the indication is what the patient, prescriber, etc believes is the main purpose of the drug.
Examples:

> Take at bedtime for **bipolar type 2 [IND]** . Keeps **manic episodes [IND]** from occurring
> seems to have slowed mom 's **dementia [IND]**

Up early for my infusion so I can enjoy the pool/sun the rest of the day. #faith #infusion #tysabri **[DRG]** #**MS[IND]** #fighter

If a general drug class such as a drug agent is listed, then the type should be classified indication versus drug
Example:
@leah_michal i'm on gabapentin for that and nerve pa**i**n **[IND]**. increasing dose tonight . also on an **antipsychotic [IND]**. nothing helps atm .
@goatpox have you ever tried viibryd **[DRG]**. yet ? i have hella issues with **antidepressants [IND]**. but that's what i'm on right now and it's kinda a miracle

# Beneficial Effect

A beneficial effect is the syndrome or disease or sign or symptom that is resolved but is not related to the reason or the purpose for the patient taking the drug or a secondary effect of the drug. If a negation of an ADR is found, select the concept that is the affirmative or opposite semantic match as the beneficial type.

Examples:
At least it helps take the edge off the panic attacks **[IND]** when I get them and it does stop my **face from twitching [BEN]** when I feel the attacks building .
caused **weight loss[BEN]** , **decreased appetite[BEN]** - which for me was good because I was an emotional eater . Gives a sense of calm w/o causing **sleepin[BEN]** →
in this instance 'weight loss' and 'decrease appetite' are beneficial and not ADR because they resolve an issue of the patient
helps anxiety**[IND]** , sleep aid **[IND]** but i do n't feel **groggy[BEN]** in the morning→ negation of an ADR

# Drug
Any mention of a drug name in the text should be annotated.
Example:
Same as her **Exelon [DRG]** . She cant take anything for pain **[IND]** It seems to make her see things **[ADR]** more
Dr put her on both **Aricept [DRG]** & this medication . Just recently taken off **Namend [DRG]** as dr felt it was to strong with **Aricept [DRG]** . Mom became very angry **[ADR]** and combative **[ADR]**

# Other
The other type is used to capture important information about the drug or the patient's experience with the drug that does not fall under the four other types. Some of the information covered by other is the dosage of the drug, the efficacy of the drug, the duration of treatment

**Other drug dosage**

When the strength of a drug is mentioned, the mention should be
annotated Example:

On pretty much lifetime maintence of **32mg** **[OTH: drug dosage]**-LRB- 8 Buphen/2

Nalox qid -RRB-

Exclude:

Non-specific mentions of dosages

o   I take 2 a day and it helps quite a bit

o   works pretty good if I take it all the time....up to 16 pills a day

## Other - Treatment effectiveness/Lack of effect

Anytime a concept token talks about whether the drug's intended effect was felt or no effect was
felt, select tokens that semantically indicate the treatment was effective/not effective and where
possible also include the indications that the drug was effective for.
Example:

It **helped me come down from a manic episode** **[OTH: treatment effectiveness]** →

**Note: '*manic episode*' would also be annotated as 'indication'**

**did not help** **[OTH: Lack of drug effect]** only worsened

## Other – duration of treatment

The length of time a patient reports being on the medication
Example:

now I have been taking this for a **year** **[OTH: Duration of treatment]** best yet .

I am **six weeks** **[OTH: Duration of treatment]** into Alendronate **[DRG]** 70mg **[OTH:
drug dosage]** and in a lot of pain **[IND]** . Have emailed my primary and waiting for a
response .

# Mapping Annotated Text to Concept ID

The following are some general guidelines for selecting Concept IDs for the annotations:

In cases where there is ambiguity due to the context being unclear and more than one
concept is equally valid, select the concept that is the more concrete fit rather than the
abstract concept

o   Example: "Did not take" → select 'non compliance' versus 'lack of drug effect'

When a negated concept is not in the concept option list:

1)   annotate the most basic token which appears in the concept option list without
including the negation

2)   select the "type" based on the full context

3)   then select the concept in the concept option list that most closely matches the
semantic idea

Example: helps anxiety, sleep aid but **i don't feel groggy** in the morning → annotate
groggy and select concept ID for 'groggy'

To generalize excessively specific terms: Select the the concept ID that is the most
general term of the annotated tokens. Examples: mania and manic or weight gain (vs

excessive weight gain), discontinue (vs stop medication), depression, delusional (vs delusional disorder), lack of drug effect

# Appendix A - Dictionaries

http://idioms.thefreedictionary.com
http://www.idiomconnection.com/medical

# Appendix B - Glossary

Akathisia:

Define: It is the ADR dictionary that is the basis of the concept option list

Duration: the amount of time in which a drug is being taken.

Effectiveness: in medicine, the abilitty of an intervention to produce the desired beneficial effect or drug works as indicated or has the desired effect in real world circumstances

Efficacy: under perfect circumstances, people take it everyday at the exact times that they are instructed to take it

Patient noncompliance: when the patient decides to and stops taking the medication.

Related Drug: Any drug mentioned associated in any way with a comment (page on the Daily Strengths website or Twitter)

Research: effectiveness vs efficacy work in real world, produce an effect given how people people take it.

Sleep: assume beneficial unless explicitly stated that it was prescribed for sleep.

Tachyphylaxis: just quits working (worked and then doesn't work)

Target Drug: The drug which "type" refers or is related to.

Titration: changing th drug dose to its optimal dosing regimen (too much, titrate down, too little, titrate up)

Tolerance: need more to get the same effect (work and then needed more to work)

Type: Input lexicon

Weaning: process of getting off is different than weaned

# Appendix C - Idiomatic Hall of Fame

| | |
|---|---|
| Lost it | agitation mental 0302807 |
| comfortable in own skin | Self-confidence |
| no longer human | |
| calmed me down | indication: mania |
| weight gain + any modifier | weight gain |
| can't stay awake = | "Sleepiness vs sluggishness?<br><br>Sleepiness = falling asleep involuntarily<br>Sluggishness = sloth like, still awake but it is a challenge (KS)" |
| bring him down | (def) restore to a normal mood or attitiue |
| stoned | mental stupor |
| Sleep modifier that indicates excessive sleep or any hours mentioned that appears relatively excessive | excessive sleep |
| can't get to sleep, not sleep enough | insomnia |
| knocked me out | excessive sleepiness |
| <mark>manic symptom vs manic episode vs manic</mark> | |
| bring him down (to normal not depressed), in control | indication: agitation<br>effect: normal |
| eating new things | increased appetite |
| not on any longer | discontinued |
| Out of wack, didn't feel like myself | feeling strange |
| PRN, pro renata, as needed | as required |
| obsessed with food | hunger abnormal |
| take it on a regular basis | compliant |
| feeling of not being here | feeling detached |
| hard to function | brain fog |

| | |
|---|---|
| empty shell | depression |
| mask like facies | flat affect |
| quiet the severe symptoms | effectiveness |
| lost days of my life | memory loss |
| switched | change of medication |
| worked well at first then does not work as well | tachyphylaxis (not tolerance b/c tolerance is immune tolerance)<br>lose effectiveness, tolerance (not lack of drug effect) |
| knocked me out | excessive sleepiness |
| go night night | sleep |
| stopped working | tolerance |
| hating my life | unhappy |
| didn't want to do anything | lack of motivation |
| not do anything stupid | impulse control |
| nuts and bolts were missing to my head | insanity |
| wired, activating | excitability |
| finally able to see how beautiful I really am | self esteem |
| adjusting | adaptation |
| zonked out the morning after | groggy and sluggish |
| black out my mind | racing thoughts as indication |
| crawl out of my skin | agitation |

**References to publications related to this data set**

A Sarker, R Ginn, A Nikfarjam, K O'Connor, K Smith, S Jayaraman, T Upadhaya, G Gonzalez. Utilizing social media data for pharmacovigilance: A review. Journal of Biomedical Informatics (JBI). Volume 54. Pages 202-212. DOI: http://dx.doi.org/10.1016/j.jbi.2015.02.004. 2015.

A Nikfarjam, A Sarker, K O'Connor, R Ginn, G Gonzalez. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. Journal of the American Medical Informatics Association (JAMIA). DOI: http://dx.doi.org/10.1093/jamia/ocu041. Pages: 0-11. 2015.

A Sarker, G Gonzalez. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. Journal of Biomedical Informatics (JBI). Volume 53. Pages: 196-207. DOI: 10.1016/j.jbi.2014.11.002. 2014.

K O'Connor, A Nikfarjam, R Ginn, P Pimpalkhute, A Sarker, K Smith, G Gonzalez. Pharmacovigilance on Twitter? Mining Tweets for Adverse Drug Reactions. In Proceedings of the American Medical Informatics Association Annual Symposium (AMIA). 2014.

R Ginn, P Pimpalkhute, A Nikfarjam, A Patki, K O'Connor, A Sarker, G Gonzlez. Mining Twitter for adverse drug reaction mentions: a corpus classification benchmark. In Proceedings of the Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing (BIOTXTM). 2014.

A Patki, A Sarker, P Pimpalkhute, A Nikfarjam, R Ginn, K O'Connor, K Smith, G Gonzalez. Mining Adverse Drug Reaction Signals from Social Media: Going Beyond Extraction. In Proceedings of BiolinkSIG. 2014.