

Manual evaluation of inferred information about Reddit users

Thank you for your help in manually checking automatically inferred or extracted information about Reddit users with a self-reported diagnosis of bipolar disorder (BD). Please provide your annotations in the respective sheets of [Annotation excel sheet].

1. BD diagnosis (100 users) – sheet “BD diagnosis”

Please label for each user in the sheet “BD diagnosis” (see Table 1) whether they report to have received a BD diagnosis in the column ‘include?’, considering the “Yes” and “No” criteria below and the examples in Table 2. Provide comments in the ‘comment’ column if necessary.

Table 1 Sheet 'BD diagnosis' to label whether users report a BD diagnosis

Column	id	# self-reported diagnosis posts	include?	comment	Self-reported diagnosis post 1	Self-reported diagnosis post 2 ... 16
Explanation	User Id in database	Number of identified self-reported BD diagnosis posts for this user	Select yes/no – see below for criteria	Provide reason for ‘no’ label, add comment if deemed necessary	Id, time posted at and text of one identified self-reported BD diagnosis post	Further identified self-reported BD diagnosis posts (random order, maximum number is 16)
Example	1	1			1(2016-12-16 15:31:43): i can relate with your rage, I am diagnosed with ...	

Criteria for self-reported BD diagnosis

Yes:

- At least one identified self-reported diagnosis post indicates that the user was diagnosed with BD by a health professional (e.g. psychiatrist, psychologist, doctor) at some point in their life.
- To indicate an official diagnosis a user needs not necessarily mention what professional made the diagnosis (e.g. by a doctor/psychiatrist), but could also indicate, for example, the time when the diagnosis was made (“I was diagnosed with BD seven years ago/with 30”) or other context in the posts could indicate this is not a guess/self-diagnosis (see examples below).
- Any type of BD diagnosis is fine: manic depression, BD-1, BD-2, BD not otherwise specified (NOS), rapid cycling bipolar disorder, cyclothymia
- As long as the user reports that some professional diagnosed them with BD at some point in their life (including childhood), label as ‘yes’ even if the author states that they do not agree with the diagnosis (anymore), or that other professionals provided a different/no diagnosis. The reason for this is that it is impossible to judge from a users’ posts if past or present diagnoses were/are warranted according to diagnostic manuals.

No:

- The text indicates that someone else than the author was diagnosed

- The diagnosis is something different than BD (e.g. BPD mostly refers to a different diagnosis of Borderline Personality Disorder, except when used with 1/2 as BPD-1, BPD 2, or BPD I, BPD II)
- The diagnosis is not official (self-diagnosed, guess, a friend said so, ...)

Table 2 Example Yes/No labels for self-reported BD diagnosis illustrated with tweets (paraphrased)

Tweet	Label	Reason/Explanation
I was diagnosed at 16 w severe depression; I'm not ashamed to say I have wished for that finality many times. Since then bi-polar; PTSD.	Yes	The depression diagnosis seems to have been made by professional due to the age provided. In this context "since then bi-polar" indicates that this diagnosis was also made by a professional.
I've been diagnosed with bipolar for almost a year now.	Yes	Time indicates that diagnosis was made by professional and is not self-diagnosis.
thinking about that time an actual licensed pyschiatrist told me that my bipolar 2 diagnosis wasn't valid because 'bipolar 2 isn't real' - i'm still so mad about it	Yes	User reporting a bipolar disorder type 2 diagnosis to psychiatrist (even though that psychiatrist thinks BD does not exist) indicates that it was made by a professional and is not merely self-diagnosis.
This makes me so angry. I was diagnosed with bipolar and put on mood stabilizers for YEARS when really it was just my birth control. 🙄	Yes	Although BD diagnosis was not valid according to the tweet author, it seems to have been made by professional at some point and was treated with medication.
My new doctor is better than my old one, but still not great. The old doctor diagnosed me as bipolar and treated for that for 5 years. For me to find that's incorrect	Yes	BD diagnosis was made by doctor, although new doctor disagrees.
i was diagnosed with schizoaffective bipolar disorder 🤪	No	Refers to schizoaffective disorder, bipolar type (a different diagnosis from bipolar disorder that is actually an exclusion criterium for a bipolar disorder diagnosis)
I feel you sis. I'm not bipolar but I was diagnosed with borderline personality disorder and its SO exhausting.	No	Other diagnosis but not BD
I am diagnosed with aspergers syndrome and ODD. I am undiagnosed with but probably have anxiety, paranoia, bi polar, depression and unkown others.	No	Guesses BD diagnosis, not confirmed by clinician

2. Gender and age (116 users) – sheet “gender+age”

Age and gender information was extracted from post titles of users that matched a specific regular expression, intended to capture cases like this: 'Me [17f] just broke up with partner [17m]'. This indicates that the user is female and 17 years old at the time of posting. All posts of a user that matched the regular expression are taken together to check for the consistency of the extracted information and summarised in the 'gender+age' excel sheet (see Table 3).

Please annotate in the four bold faced final columns (see Table 4) whether the extracted information is correct. Some users in the sheet only have either a gender or date of birth value assigned. Only use the information presented in the sheet, do not check Reddit user profiles for this.

Table 3 Sheet 'gender+age': columns with extracted gender and age information

Column	id	posts dates	post titles	age + gender self-reports in parentheses	age + gender	set of extracted dates of birth	extracted gender	extracted date of birth
Explanation	User id in database	Dates of posts whose titles matched regular expression	Post titles from which age + gender was extracted	Age + gender information in parentheses that matched regular expression		Age (years, months, days) subtracted from post date	m or f if set if age + gender only contains one gender value or one value and None; otherwise ?	Earliest date of birth (day/month/year) if range of set of extracted dates of birth <= 1 year; otherwise ?
Example 1	427	{'2017-11-25 07:06:19'}	{'[28][M4F] Hoping to find love'}	{'[28]'}	{'28 None'}	{datetime.date(1989, 11, 25)}	?	25/11/1989
Example 2	2531	{'2017-04-04 15:50:13'}	{'[Help][Aldrich][PC][ng SL 50]'}	{'[Help][Aldrich][PC][ng SL 50]'}	{'50 None'}	{datetime.date(1967, 4, 4)}	?	04/04/1967

Table 4 Sheet 'gender+age': columns for labelling

Column	gender correct?	date of birth correct?	corrected gender	corrected date of birth	Comment
Explanation	yes / no ? only if extracted gender is ?	yes / no ? only if extracted date of birth is ?	f or m if "gender correct?" is "no" or "?" and correct gender can be inferred from post titles, ? otherwise if "gender correct?" is "no" or person identifies as non-binary gender/transgender	Age subtracted from post date if "date of birth correct?" is "no" or "?" and correct age can be inferred from one or more post titles, ? otherwise if "date of birth correct?" is "no"	Any comment on this user
Ex.1	?	yes	m		[M4F] indicates that user is male looking for female partner
Ex. 2	?	no		?	

3. Country of residence (100 users) – sheet “country”

Please try to determine the country a Reddit user seems to live in based on their posts and the subreddits they posted in. See Table 5 for an explanation of the annotation sheet. As a help, the file [Excel sheet] lists all posts of users that contain a location mention (see Table 6). Note that this information is automatically extracted, so it contains some noise.

If a username does not appear in the file, this means that no locations could be extracted from their posts. In this case, please look up the user on reddit ([https://www.reddit.com/user/\[username\]](https://www.reddit.com/user/[username])). First look at their thread starts ([https://www.reddit.com/user/\[username\]/posts/](https://www.reddit.com/user/[username]/posts/)) in their order of recency, to see if any of their texts indicates a location (e.g. 'as an American') or if they were posted in a location-specific subreddit (e.g. r/BerlinJobs). If you cannot determine a country from the posts, also look at a users' comments, searching for clues in a similar way ([https://www.reddit.com/user/\[username\]/comments/](https://www.reddit.com/user/[username]/comments/)).

It may not be possible to infer a country for all users. If a user has > 50 posts or comments, you do not need to check all of them for clues.

Further conventions for country of residence annotation:

- If a user indicates moving, enter the most recent country, even if this does not seem to be the country where they have been living longest so far.
- If the user indicates to stay somewhere temporarily (on vacation, studying abroad), enter the normal country of residency if you can determine it, otherwise "?".
- Just enter a users' country, even if they provide more specific city/state information.

Table 5 Sheet 'country' for country of residence labelling

Column	Username	Country	Reason
Explanation + Example	Reddit username	Please enter the latest country of residence of this user as a two letter ISO Alpha-2 code (see https://www.iban.com/country-codes). ? if you cannot determine the country	Evidence that user is from suggested country: Short excerpt + link to the post (e.g. "I'm from Germany" reddit.com/r/...)

Table 6 user-country-posts.xlsx sheet to support country of residence labelling.

Column	Username	Locations	Post date	Post url	Post
Explanation	Reddit username	List of automatically extracted locations in the post	Post date. Posts for each user are listed in descending order of recency	Prefix url with reddit.com/ to look at the thread with this post on reddit. The term after r/ indicates the subreddit	Full post text
Example	[username]	['mascara'] Example for noise, Mascara is a also a city in Algeria	22/08/2018 05:53:41	[link]	Shopping sucks, I was looking for clear mascara today. [...]