

Lancaster
University



Corpus Framework Analysis: Integrating computational linguistics, corpus linguistics, and clinical psychology to analyse online posts on personal recovery in bipolar disorder

CADSconf @ Bertinoro, Italy 26/8/22

Glorianna Jagfeld, Chloe Humphreys, Paul Rayson, Fiona Lobban, Steven Jones

Spectrum Centre for Mental Health Research · Lancaster University

g.jagfeld@lancaster.ac.uk  [@glorisonne](https://twitter.com/glorisonne)





Lancaster
University



Introduction & motivation

Personal recovery in bipolar disorder



- Bipolar disorder (BD): severe mental health (MH) issue, characterised by extreme + changing mood states



- Clinical recovery: remove symptoms, restore functioning
- Personal recovery (PR): living a satisfying, hopeful, contributing life even with limitations caused by symptoms¹

1. Anthony WA. Recovery from mental illness: the guiding vision of the mental health system in the 1990s. Vol. 16, Psychosocial Rehabilitation Journal. 1993.16(4):11–23.

Why study PR in BD via online forum posts?



- Peer online support forums: source of information, support, social contacts
- PR research so far: interviews, focus groups: retrospective, researcher influenced
- Online posts: insights into 'an experience as it is lived rather than as it is enacted in the researcher constructed environment'¹
- Opportunity to include voices of people not in contact with services/researchers
- Reddit (reddit.com): large international online discussion forum, subforums (subreddits) for different topics, including MH/BD, all posts in public domain

¹ Seale, C., Charteris-Black, J., MacFarlane, A., & McPherson, A. (2010). Interviews and internet forums: A comparison of two sources of qualitative data. *Qualitative Health Research*, 20(5), p. 600

Bigger picture - PhD thesis „Talking about personal recovery in bipolar disorder“



Personal recovery in bipolar disorder: Systematic review and "best fit" framework synthesis of qualitative evidence – a POETIC adaptation of CHIME (Jagfeld et al., 2021)

Understanding who uses Reddit: Profiling individuals with a self-reported bipolar disorder diagnosis (Jagfeld et al., 2021)

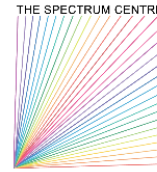
An exploratory analysis of posting patterns in peer online support forums and their associations with emotions and mood in bipolar disorder (under review)

Talking about personal recovery in bipolar disorder: A corpus framework analysis of peer online support forum posts (in progress)

Research question

What can online support forum posts reveal about the processes and experience of personal recovery in bipolar disorder in relation to the POETIC framework?





Lancaster
University



Methods: Corpus construction & analysis

POETIC: Lived experience of personal recovery in bipolar disorder



P	Purpose and meaning	Meaning of mental illness experiences	Paid or voluntary work	Quality of life	Meaningful life and social roles
O	Optimism and hope	Belief in possibility of recovery	Positive thinking and valuing success	Hope-inspiring relationships	Having dreams and aspirations
E	Empowerment	Self-management & personal responsibility	Controversial role of medication	Control over life	
T	Tensions	Balancing acceptance with ambitions	Openness enables support, but also stigmatisation	Ambivalence around (hypo-) mania	
I	Identity	Rebuilding positive sense of self	Overcoming stigma	Dimensions of identity	
C	Connectedness	Support from others	Relationships	Peer support and support groups	Being part of the community

Ethical issues



- Ethical approval by FHM Research Ethics Committee in May 2019
- Remove usernames, paraphrase quotes to protect users' anonymity
- Feedback from people with lived experience throughout project
- Reproducibility: dataset + corpora will be made available to other researchers upon request and signing a data usage agreement

Dataset¹: Reddit users with a self-reported bipolar disorder diagnosis

- User account identification via self-reported BD diagnosis statements



¹ Jagfeld, G., Lobban, F., Rayson, P., & Jones, S. H. (2021). Understanding who uses Reddit: Profiling individuals with a self-reported bipolar disorder diagnosis. Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access at NAACL 2021.

Selecting Posts about PR in BD with an information retrieval approach



Rank	Subreddit	users
1	AskReddit	65%
2	bipolar	42%
3	funny	39%
4	pics	39%
5	todayilearned	33%



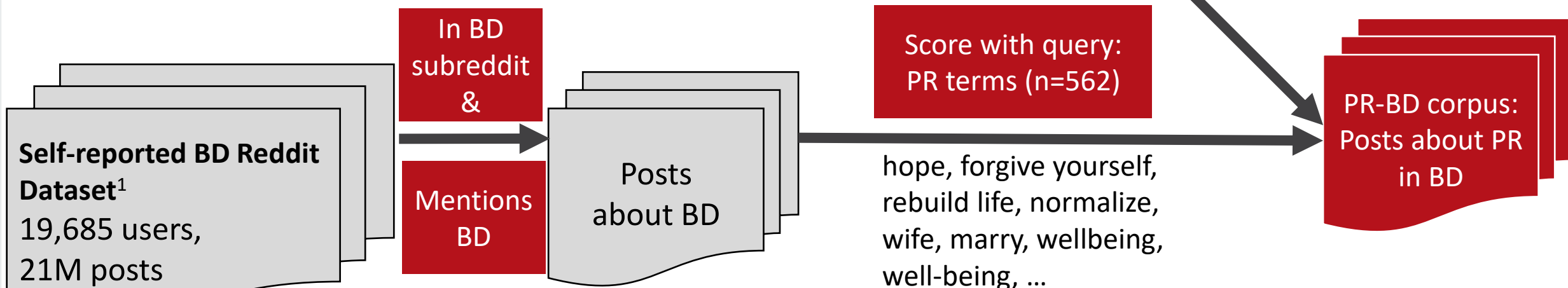
r/bipolar · Posted by u/██████████ 3 months ago

My favorite recovery song

Facing the sun by fritz kalkbrenner.
I love this song, good times and bad times and I think the lyrics tell a lot about myself.

4 Comments Share Save Hide Report

88% Upvoted



¹ Jagfeld, G., Lobban, F., Rayson, P., & Jones, S. H. (2021). Understanding who uses Reddit: Profiling individuals with a self-reported bipolar disorder diagnosis. Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access at NAACL 2021.

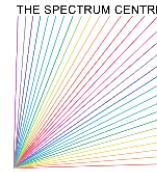
Corpus Framework Analysis (CFA): Integrating computational linguistics, corpus linguistics & health research



- 1) Data familiarisation: read randomly sampled posts in threads
- 2) Generate key lemmas via #LancsBox 6.0¹: significantly overused² in PR-BD corpus compared to reference corpus of non-PR relevant posts, used by at least 5% (n=100) of users
- 3) POETIC framework analysis of 30 random concordances per key lemma
 - a) Expected items: key lemmas that *are* in PR terms list
 - b) New items: key lemmas that *are not* in PR terms list
 - c) Collocation analysis for selected key lemmas (#LancsBox)
- 4) Consider absences³: PR terms underused/missing in the PR-BD corpus

¹ Brezina, V., Platt, W., & McEnery, T. (2021). #LancsBox v. 6.0. <http://corpora.lancs.ac.uk/lancsbox> ² Log likelihood, $p < 0.0001$

³ Duguid, A. and Partington, A. (2018) 'Absence: you don't know what you're missing. Or do you?', in Taylor, C. and Marchi, A. (eds) *Corpus Approaches to Discourse: A Critical Review*. Taylor & Francis Group, pp. 38–59



Results

PR-BD corpus & reference corpus, key lemmas

PR-BD corpus: Posts in BD
subreddits about BD and PR

Posts: 4,462
Words: 1,337,080
Users: 1,982

Reference corpus: Posts in BD
subreddits about BD but not PR

Posts: 25,197
Words: 4,700,834
Users: 6,075

130 key lemmas
used by at least
5% of users

PR score > 0.025,
59% precision in selecting PR relevant posts
(based on 38% PR-relevant posts out of
167 manually coded posts)

PR score < 0.013,
75% precision in selecting non-PR relevant posts

POETIC framework analysis results

P	Purpose and meaning	Meaning of mental illness experiences	Paid/voluntary work or formal education	Quality of life	Meaningful life & social roles	Spirituality	Meaningful life & social goals	Rebuilding of life
O	Optimism and hope	Belief in possibility of recovery		Positive thinking & valuing success		Hope-inspiring relationships		Having dreams and aspirations
E	Empowerment	Self-management & personal responsibility		Controversial role of medication		Control over life		
T	Tensions	Balancing acceptance with ambitions		Openness enables support, but also stigmatisation		Ambivalence around (hypo-) mania		
I	Identity	Rebuilding positive sense of self		Over-coming stigma		Dimensions of identity		
C	Connect-edness	Support from others		Relationships		Peer support and support groups		Being part of the community
X	Not POETIC	Other		Symptoms		Genetics & family history		Diagnosis
Frequency		159	350	1000	1300	2000	4000	6500

Meaningful life or social roles (5,715): Reproductive decision making



- kid 57%, parent_n 53%, daughter 53%, baby* 53%, child 50%, son 47%, bear_v* 43%, raise_v* 37%, **chance** 30%, environment* 30%, **decision** 27%, mom 27%, grow* 23%, **risk_n** 20%, society 20%, father_n 20%, challenge 17%, develop 13%, perspective 13%, possibility 13%, capable 13%, life 13%, **choice** 13%, **responsible** 13% *new key lemma (not PR term)
- Genetic risk + manageability of parenting a child for parents with BD¹
- *"I wouldn't wish bipolar disorder on my worst enemy, let alone could I bear to pass it on to anyone. I've got other reasons, too. I'm terrified of the thought of having to come off my medication for at least nine months, particularly with the hormonal changes and stress of parenting. I see my suicide risk going up and I'm scared I'd be so selfish and end my child's life as well. [...] I've considered it all and having kids just doesn't seem possible for me."*

¹ Sahota, P. K. C., & Sankar, P. L. (2019). Bipolar Disorder, Genetic Risk, and Reproductive Decision-Making: A Qualitative Study of Social Media Discussion Boards. *Qualitative Health Research*.

Spirituality (1,153)



- god 70%, belief 37%, life 10%, teach* 10% *new key lemma (not PR term)
- Very personal, also controversial experiences
- (Hypo-)manic vs. religious experiences
- “I meditate and visualize angel wings enfolding me and being flooded with radiating loving light. I realise this sounds quite bipolar. **Talking about this part of my inner world to a psychiatrist would require a lot of trust for me.** I have always had this active imagination, as if there was another presence just out of reach. I think of it as a private activity. If it does not negatively affect my everyday life, it should be **nobody’s business.**”

Summary & outlook

Summary, implications & limitations



Summary

- Corpus Framework Analysis of online support forum data: innovative approach to study PR in BD
- POETIC framework captures PR experiences shared online well

Implications

- Practice: more awareness of relevant PR issues in BD
- Research
 - POETIC framework validation
 - Apply Corpus Framework Analysis to other research questions

Limitations

- PR terms list enabled but biased corpus construction
- Topic-like POETIC categories easier to capture with single terms than theme-like categories

Thank you – I'm looking forward to the discussion!

Glorianna Jagfeld · Spectrum Centre for Mental Health Research · Lancaster University
g.jagfeld@lancaster.ac.uk  [@glorisonne](https://twitter.com/glorisonne)



With what meanings do Reddit users with a self-reported BD diagnosis use *recover* terms?



Preparatory study: code meaning of 8 most frequent wordforms of recovery in the dataset (= *recover* terms): recovery, recover, recovering, recovered, recovers, recoveries, recoverable, unrecoverable

Meaning	% total (n=377)	% BD subreddits (n=50)	% MH subreddits (n=50)	% non-MH subreddits (n=277)
MH recovery	43.0	80.0	84.0	28.9
- Clinical recovery	66.0	75.0	73.8	57.5
- Personal recovery	17.3	12.5	11.9	14.3
- Cannot determine	16.7	12.5	14.3	20.0
Physical health	22.8	12.0	2.0	28.5
Life event	14.1	6.0	10.0	16.3
Other	20.1	2.0	4.0	26.3

MH issue	% BD subs (n=50)	% MH subs (n=50)
Addiction	15.0	38.1
BD	67.5	0.0
Eating disorders	0.0	28.6
Major depressive disorder	0.0	11.9
Borderline personality dis.	0.0	9.5
Anxiety disorder, PTSD, psychotic disorder, self-harm, multiple, unknown	17.5	11.9

➡ Recovery is not a good search term!

➡ Only BD subreddit posts mentioning BD

Creating the PR terms list



- Manual selection of overused keywords from:
 - POETIC review articles (n=12, 75,553 words) vs. excluded articles (n=80, 409,877 words) that fulfilled all inclusion criteria except PR focus¹
 - Quotes from included POETIC articles from each POETIC domain vs. quotes coded in all other domains²
 - MH recovery posts manually coded in preparator study as as PR (n= 26, 6,368 words) vs. clinical recovery (n=89, 13,790 words)¹
- + Terms informed by clinical knowledge, spelling variants
- PR terms (n=562): hope, forgive yourself, rebuild life, normalize, wife, marry, wellbeing, well-being, ... (automatically lemmatised via spacy³)

¹ Log likelihood, $p < 0.001$

² Log likelihood, $p < 0.01$

³ Honnibal, M., Montani, I., van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. <https://doi.org/10.5281/zenodo.1212303>

Determine PR score cutoff: Manual coding of PR relevance



- Does the post indicate that the user can be regarded as being in personal recovery, i.e. do they share experiences relevant to their personal recovery in the post?
- Moderate agreement (Cohen's kappa (k) = 0.51, 77% observed agreement) for GJ & CH coding 100 posts
- Good agreement (k = 0.71, 87% observed agreement) for 62 posts where both had $\geq 3/5$ confidence
- Calculate precision & recall for various PR score cut-offs of manually coded posts



Absences: What is missing online?



- Underused or missing in the PR-BD corpus: 13% (55/423 unique PR terms (lemmatised, no pronoun/spelling variants))
- Underused: 8 PR terms
 - Log ratios between 0 and -0.7, 5 PR terms significantly underused w $p < 0.0001$
 - Topics – expect that these are underrepresented in a corpus about PR
 - Mania: mania, manic, high mood
 - Medical MH professionals: doctor, pdoc, psychiatrist
 - Symptom: sleep
 - Race: race_n (identity) overused (LR 1.3) but race_v (running) underused (LR – 0.5)
- Missing in PR-BD corpus (zero frequency): 47 PR terms
 - All terms also do not appear in the reference corpus
 - Topics
 - Characteristics of the recovery journey (12: recovery journey, growth process etc.)

Overview of the data: analyse 50 random posts in their threads



- 34 (68%) PR-relevant: ask question/share own experience
 - Romantic relationship advice, bad professional/family support, reproductive decisions, spirituality
- 16 (32%) not PR-relevant:
 - 9 (18%) give advice/general comment without personal experience
 - 6 (12%) vent about/discuss/seek advice for symptoms
 - 1 (2%) genetics: family members with MH issues

Tf-idf weighted cosine similarity



- Given a vocabulary of size N with i representing one term of the vocabulary, a post p is represented by the vector x of length N according to its bag of words with one hot encoding: $x_i = \begin{cases} 1 & \text{if vocabulary term } i \text{ appears in } p \\ 0 & \text{else} \end{cases}$
- Analogously, the PR terms are represented by the vector y , also of length N : $y_i = \begin{cases} 1 & \text{if vocabulary term } i \text{ appears in the PR terms list} \\ 0 & \text{else} \end{cases}$
- The PR score of the post p is calculated as the cosine similarity between the tf-idf weighted vector representation \bar{x} of the vector representation x and the tf-idf weighted vector representation \bar{y} of the PR terms list vector representation y : $\text{PR score}(x) = \text{cosine}(\bar{x}, \bar{y}) = \frac{\sum_{i=1}^N \bar{x}_i * \bar{y}_i}{\sqrt{\sum_{i=1}^N \bar{x}_i^2} * \sqrt{\sum_{i=1}^N \bar{y}_i^2}}$
- The tf-idf weighted vector representation \bar{z} of a vector z and a particular post p is calculated by multiplying the vector entry for each vocabulary term i with its associated weight $w_{i,p}$: $\bar{z} = w_{i,p} * z_i$
- The weights consist of the multiplication of the term frequency $tf_{i,p}$ of vocabulary term i and the inverse document frequency idf_i of term i : $w_{i,p} = tf_{i,p} * idf_i$, where the frequency of term i that appears $\text{count}(i, p)$ times in a post p is defined as follows:
$$tf_{i,p} = \begin{cases} 1 + \log_{10} \text{count}(i, p) & \text{if } \text{count}(i, p) > 0 \\ 0 & \text{else} \end{cases}$$
- The inverse document frequency of term i that appears in df_i posts out of a collection of M posts (traditionally called documents) is $idf_i = \log_{10} \frac{N}{df_i}$

Outline



- 1) Introduction & motivation
- 2) Methods
 - i. Corpus construction
 - ii. Corpus framework analysis (CFA)
- 3) Preliminary results
- 4) Outlook

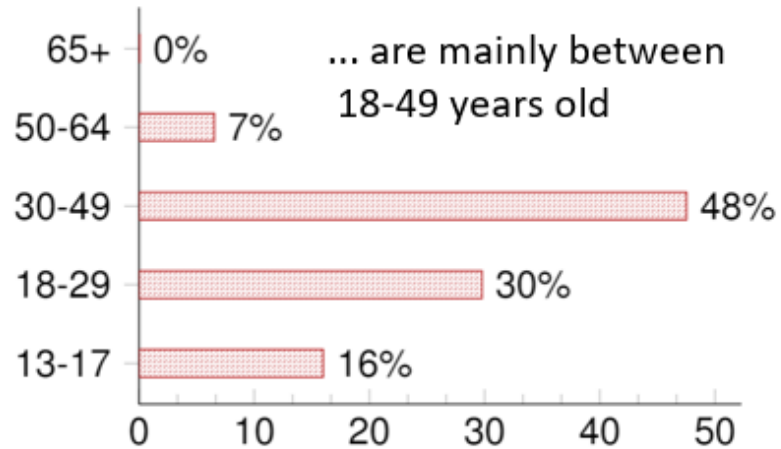
Frequency of top 10 *recover* terms in the dataset



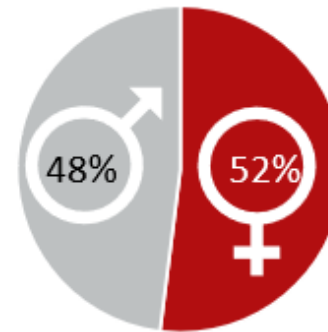
Rank	Term	Frequency	Rank
1	recovery	32425	17
2	recover	15459	18
3	recovering	10399	19
4	recovered	7178	20
5	recovers	822	21
6	recoveries	263	22
7	recoverable	194	23
8	unrecoverable	124	24
9	recovery](suburladdress	40	25
10	recovery-	34	26

Subreddit	Posts in subreddit (n)	Posts in subreddit (%)
stopdrinking	5969	9.80
AskReddit	5233	8.59
Bipolar	2689	4.41
BipolarReddit	1294	2.12
BPD	1031	1.69
relationships	965	1.58
OpiatesRecovery	805	1.32
raisedbynarcissists	780	1.28
depression	656	1.08
REDDITORSINRECOVERY	603	0.99

Reddit users who disclose a BD diagnosis



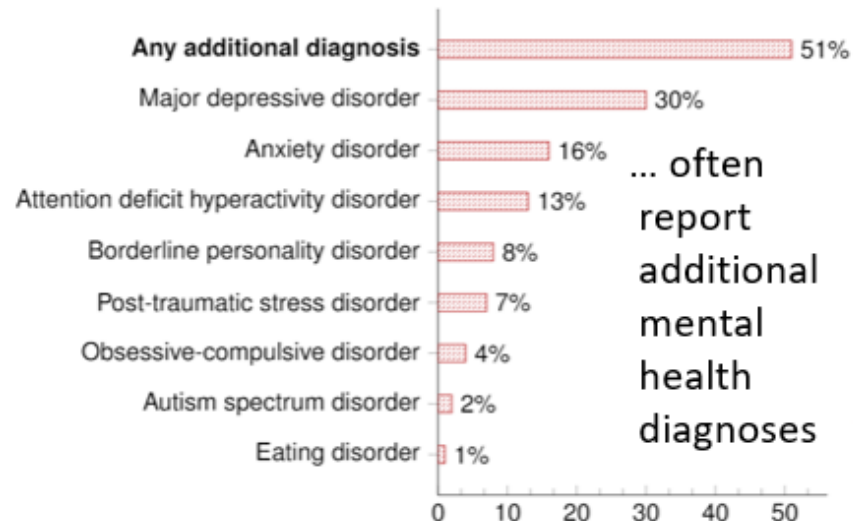
... have a balanced gender distribution*



■ Feminine

■ Masculine

* non-binary gender not considered due to lack of natural language processing methods



... live predominantly in the USA



What's new: 15 not PR-term key lemmas



Often general terms that blend in well with PR terms, few surprises:

- Peer support: support_n (1238), **attend** (123), **network** (114)
- Relationships: family (1988), relationship (1777), parent_n (1275), mom (845), healthy (811), father (435), son (387), member (307), daughter (297), sexual (184), religious (180), **married** (96), **fulfill** (79), vulnerable (73), **sibling** (72), **divorce** (70), rebuild (64), **boundary** (63), value_v (51)
- Meaningful life and social roles: family (1988), parent_n (1275), child (1243), kid (1146), mom (845), decision (566), father (435), choice (433), son (387), daughter (297), responsibility (284), responsible (156), **married** (96), **fulfill** (79), **adopt** (76), parent_v (63)

Connectedness: navigate family & friend support & conflicts via boundaries




- Relationships (4359): relationship (1777), healthy (811), father (435), son (387), daughter (297), sexual (184), religious (180), married* (96), fulfill* (79), vulnerable (73), divorce* (70), rebuild (64), boundary* (63), value_v (51) (*not PR terms/new key lemmas)
- Support from others (3447): support_n (1238), social (599), support_v (464), supportive (421), resource (283), educate (204), active (141), network* (114), spouse (110), religion (110), acceptance (99), loving* (96), uncle (95), christian (76), vulnerable (73), boundary* (63), atheist (61)
- Both: family (1988), parent_n (1275), mom (845), member (307), sibling* (72)
- *Having bipolar doesn't somehow make every emotion I have now illegitimate. If someone tells me to "take my meds" in an emotional situation, my response is usually not good. People in my life know not to use that as a weapon or to ask me if I've taken them if I seem emotional for whatever reason. It's not an open topic for discussion and I've set that **boundary** w everyone in my life.*
- *My **boundaries** change from moment to moment, so I always have to tell my husband that cuddles are fine for now.*

Spirituality (444): hardly discussed in POETIC




- religious (180), belief (177), faith (176), spiritual (144), religion (110), practice* (88), meditate* (80), church* (78), christian (76), atheist (61)
- Relationship of (hypo-)mania and spirituality
*It kind of makes me sad that the de facto assumption for thoughts / feelings relating to the **spiritual**, while bipolar, are automatically assumed to be psychosis.*
- Religious/spiritual beliefs hindering treatment
*I realized that I had a prejudice towards mental illness and taking medication, though I had previously thought I was so understanding of those with it [...]. I thought if I followed all my **church** rules, I wouldn't get it or be unhappy in any way.*
- Spirituality aiding recovery through faith & self-management
*I personally like **meditating** [...] or praying if you're **religious** in the "western" sense. But even **atheists** can meditate. Quite honestly, becoming a Buddhist has helped my BiPolar II more than any medication or therapy ever did. It does not stop the urges to be mad or irritable, and it does not prevent mania or depression, but its been good at alleviating anxiety and reducing future-future-thinking delusions that cause the most problems in my social life. I never thought **religion**/philosophy would help me so much, but it did.*




The Bipolar Meeting Point

r/bipolar



Joined



Posts



Create Post

Hot

New

Top

459

PINNED BY MODERATORS

Posted by u/AutoModerator 1 year ago

MED MONDAYS

2241 Comments

Award

Share

Save

r/bipolar · Posted by u/ 3 months ago

My favorite recovery song

Facing the sun by fritz kalkbrenner.
I love this song, good times and bad times
and I think the lyrics tell a lot about myself.*

4 Comments


Share

Save

Hide

Report

88% Upvoted



11 hr. ago

Mixed Episodes

I love this :)

2

Reply

Give Award

Share

Report

Save

Follow

About Community

A safe haven for diagnosed individuals to discuss bipolar related issues. We are a community here not just a help page. Be a part of something that cares about who you are.

159k

Members

118

Online

Created Oct 26, 2008

Create Post

USER FLAIR PREVIEW

COMMUNITY OPTIONS

r/bipolar Rules

1. NO PERSONAL INFO

2. No asking for a diagnosis.

3. We do not allow users to post or comment as medical professionals.

Reddit as
data source

<https://reddit.com/r/bipolar>
*Made-up post to protect
users' anonymity

33

Quality of life (2,634): Meeting basic needs



- hobby 70%, improve 40%, **money** 23%, parent_n 17%, marry 17%, enjoy 13%, social 10%, exercise_v 10%, activity 10%, husband 10%, manage 10%, **job** 10%, wife 10%, degree 10%, adult* 10%, quality* 10%
* new key lemma (not PR term)
- Meeting basic needs barely came up in POETIC review: no one reported financial struggles - no US studies, mainly UK + EU
- “I have a great job in a pharmacy I like, but some days (most, actually) I just can't get out of bed. My boss had to stop my health insurance about 9 months ago because I was working less than 20 hours per week. I don't know why I can't get out of bed. I don't know why I can't get out of bed. My alarm rings and I'm just like 'not today, don't feel like it'. But without health insurance I can't see a doctor. Well, I could pay out of pocket, but then I can't because I work so few hours. A vicious cycle.”

Online posts offer candid and in the moment experiences



Disclosing secret behaviours (Purpose & Meaning: Spirituality)

*I meditate and visualize angel wings enfolding me and being flooded with radiating loving light. I realise this sounds quite bipolar. **Talking about this part of my inner world to a psychiatrist would require a lot of trust for me.** I have always had this active imagination, as if there was another presence just out of reach. **I think of it as a private activity. If it does not negatively affect my everyday life, it should be nobody's business.****

Posting during a manic episode (Not POETIC: Symptoms)

*Update: **Yesterday I posted here about the realization that I've entered a manic episode.** I drove 600 miles to see an online friend after breaking up with my boyfriend. I told no one where I was going. Aside from my family and this guy, I've destroyed all my relationship in the past days. I hate it - when I get this way, I become so impulsive and insensitive, and I end up feeling super guilty afterwards. I would love to feel normal one day! Positive note: I am safe.**

* Post quotes paraphrased to protect users' anonymity