Expert Rater Agreement for Symptoms and Diagnosis of Bipolar Disorder in Youth

Daniel Azzopardi Larios MD, [1] Cathryn A. Galanter MD,[1,2] Jarrod M. Leffler PhD,[3] Frances Shi,[2] Peter S. Jensen MD,[4] Leah Casuto MD,[5] Paul E. Croarkin DO,[3] Kathryn R. Cullen MD,[6] Anne Duffy, MSc MD FRCPC,[7] Bonnie Klimes-Dougan PhD,[8] Robert M. Post MD,[9] Jennifer L. Vande Voort MD,[3] Karen Dineen Wagner MD PhD,[10] Dwight V. Wolf MD[10]

[1]Kings County Hospital Center, Brooklyn, NY; [2]SUNY Downstate College of Medicine, Brooklyn, NY; [3]Mayo Clinic, Rochester, MN; [4]The REACH Institute, New York, NY; [5]Lindner Center of Hope & University of Cincinnati College of Medicine, Cincinnati, OH; [6]University of Minnesota Medical School, Minneapolis, MN; [7]Queen's University, Kingston, ON, Canada; [8]University of Minnesota, Minneapolis, MN; [9]George Washington University, Washington, DC; [10]University of Texas Medical School Branch, Galveston, TX

DISCLOSURES: CA Galanter: Advisory Board: Patient-Centered Mental Health in Pediatric Primary Care (PPP) Mini-Fellowship at the REACH Institute; Honorarium for participation in AACAP Annual Meeting; Royalties: American Psychiatric Association Publishing. PS Jensen: Advisory Board: Shire, The REACH Institute; Grant Support: Shire; Research Support: Division of Health Services, Arkansas, National Institute of Mental Health; Royalties: American Psychiatric Association Publishing, Guilford Press, Oxford University Press, Penguin Random House; Stock or Equity: Catch Services, Inc. JL Vande Voort: Co-PI on an investigator initiated study that has a grant-in-kind for supplies: Assurex Health, Inc. PE Croarkin: Grant Support: Neuronetics, Pfizer Inc., The Brain and Behavior Research Foundation; Research Support: Assurex Health, Inc., NeoSync, Inc., NeuroStar Advanced Therapy, Neuronetics, Pfizer Inc., The Brain and Behavior Research Foundation. KD Wagner: Advisory Board: Anxiety Disorders Association of America; Columnist: UBM Medica LLC; Honorarium for meeting participation: AACAP, APA, American Society of Clinical Psychopharmacology, Klingenstein Third Generation Foundation, Las Vegas Psychiatric Society, Nevada Psychiatric Association, Oxford University Press, Partners Healthcare, Physicians Postgraduate Press, UBM Medica LLC, University of Wisconsin, Weill Cornell Medicine; Scientific Council: Brain & Behavior Research Foundation.

Abstract

Objectives: Clinicians and researchers have varied in their assessment of pediatric bipolar disorder (BD) and their interpretation and implementation of the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for BD in children. Research diagnostic interviews[1] vary in their application of DSM BD criteria[2]. Researchers and authors of DSM-5 made efforts to clarify the criteria for bipolar and bipolar spectrum disorders. To better understand clinician decision making practices, we examined how expert coinvestigators participating in the pediatric bipolar biobank (PBB) varied in their interpretation and application of BD criteria and diagnoses when reviewing detailed clinical narratives. **Methods**: We enrolled 14 children and adolescents in the multisite PBB. As part of the assessment, investigators wrote a clinical narrative using a standardized format. From 2014-2017, study co-investigators reviewed each narrative and marked a visual analog scale with their confidence in whether the narrative reflected the presence of a bipolar mood episode and each BD criterion during the most illustrative (hypo)manic episode. We analyzed raters' mean confidence ratings and inter-rater agreement using interclass correlation (ICC). **Results**: Ratings for confidence in diagnosis and symptoms varied greatly. Standard Deviations (SDs) for some symptoms and diagnoses were fairly consistent (e.g. Impairment, Mania), while others had a much larger spread (e.g. Distractibility, Hypomania). Diagnoses and symptoms with good Interclass Correlation (ICC) ranging from 0.60-0.74 included inflated self-esteem/grandiosity and decreased need for sleep. Those with fair ICCs (0.40-0.59) included increased activity or energy, increased/pressured speech, increased goal-directed activity or psychomotor agitation, excessive involvement in (pleasurable) activities, and mania. Those with poor ICCs (<0.4) included Hypomania and Bipolar Not Elsewhere Classified (NEC)/Not Otherwise Specified (NOS). **Conclusions**: In spite of efforts made to refine BD criteria with DSM-5, our findings indicate that investigators working with pediatric BD may still vary in their diagnostic decision making, thus warranting further research and education in this area. Next steps include investigating which factors contribute to clinician uncertainty that resulted in high variation in confidence ratings and comparing ratings to the Schedule for Affective Disorders and Schizophrenia for School Aged Children (KSADS) diagnoses.

Objectives

- To better understand clinician decision making practices for pediatric BD.
- To examine how expert coinvestigators participating in the pediatric bipolar biobank (PBB) varied in their interpretation and application of BD diagnostic criteria and diagnoses when reviewing detailed clinical narratives.
- We hypothesized that our findings would result in less robust interclass correlation than that in the acceptable agreement range between raters.

Methods

Participants

- 14 children and adolescents enrolled in pediatric bipolar biobank study; 2 excluded from this study due to insufficient data.
- Pediatric Bipolar Biobank Consortium: 12 child and adolescent psychiatrist and psychologist investigators and pediatric BD experts.

Narratives

- Clinical histories (narratives) summarized descriptively and written using a standardized format, including identifying information, source, history of present illness, past psychiatric history, substance abuse/use, developmental and medical history and family history.
- Narrative contained specific prompts to elicit clinical descriptions of the most illustrative (hypo)manic episode: episode (i.e., duration and co-occurrence of symptoms), elevated or expansive mood coinciding with increased energy, irritable mood, persistently increased activity or energy, inflated self-esteem/grandiosity, decreased need for sleep, more talkative than usual, flight of ideas/racing thoughts, distractibility, increased goal directed activity/increased energy, excessive involvement in (pleasurable) activities with potential for painful consequence.

Procedure

- Narratives de-identified, revised for clarity, and distributed to study co-investigators.
- From 2014-2017, co-investigators reviewed each narrative and marked a visual analogue scale (0-10, from highly unlikely to highly likely) with their confidence on whether the narrative reflected the presence of a bipolar mood episode and each BD criterion during the most illustrative (hypo)manic episode (see Figure 1 for the Visual Analog Scale).
- Responses were measured, standardized, and entered.

Data Analysis

- Mean ratings and standard deviations were calculated and presented.
- Inter-rater agreement was analyzed using interclass correlation (ICC)[6].

Results

- We demonstrated the mean and standard deviation (SD) of confidence ratings for all the diagnoses and symptoms of all 12 cases as rated by all 12 investigators (see Table 1).
- Standard deviations (SDs) for some symptoms and diagnoses were fairly consistent (e.g. Impairment, Mania), while others had a much larger spread (e.g. Distractibility, Hypomania).
- Interclass correlation (ICC) for diagnoses ranged from poor to fair, with fair ICC for mania (0.41, CI: 0.19 – 0.76) and poor ICC for hypomania (0.10, CI: -0.01 – 0.47) and BD NOS/NES (0.09, CI: -0.02 – -0.05) (see Table 2).
- ICC for symptoms ranged from good to poor. ICC was good for inflated self-esteem (0.69, CI: 0.43 – 0.93) and decreased need for sleep (0.66, CI: 0.47 – 0.87), but was poor or fair for other criteria (e.g., impairment 0.23, CI: 0.09 – 0.60) (see Table 2).

Table 1. Diagnosis and Symptom Confidence Ratings

		Mania		Hypomania		BP NOS		Episode		Elated		Irritable		Energy		Self-Esteem		Sleep		Talkative		FOI		Distracted		Agitated		Activities		Impaired	
Age	Gender	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
10	F	4.31	2.82	6.55	2.58	6.61	3.11	5.65	1.89	7.06	2.29	7.94	2.53	7.57	1.51	3.60	2.06	8.09	2.01	7.64	1.46	7.18	2.02	6.61	2.69	6.84	1.94	5.07	2.12	8.77	1.68
17	M	6.31	3.28	4.24	3.15	5.37	4.36	5.91	2.80	7.55	2.62	8.43	1.14	6.36	2.83	6.75	2.67	8.47	2.48	6.02	2.28	7.00	2.83	8.00	1.25	7.35	2.70	7.46	2.79	9.07	1.37
16	M	5.89	3.32	6.05	3.63	4.38	3.35	6.65	3.38	5.63	3.33	8.46	1.77	8.343	1.78	6.55	2.22	6.15	3.32	6.13	3.10	7.85	1.49	6.91	3.01	8.09	1.97	8.30	1.55	8.99	1.20
14	F	7.74	2.94	4.87	3.40	2.81	2.69	8.46	1.78	8.84	1.09	5.20	2.49	8.55	1.81	6.42	3.23	7.08	2.71	8.92	1.35	7.64	2.73	7.74	2.40	8.51	1.80	8.60	1.81	8.86	1.19
11	F	7.04	3.51	5.87	3.57	2.87	3.66	7.40	2.85	7.80	2.96	7.57	1.81	6.98	2.84	1.27	2.85	8.26	2.86	8.30	2.14	8.42	2.97	8.46	1.99	7.96	1.95	5.58	2.67	8.74	1.57
13	F	7.45	2.90	4.63	3.20	4.02	3.57	8.42	1.83	8.47	1.41	7.16	2.08	8.22	1.94	7.53	1.95	9.04	1.43	8.84	1.73	7.52	1.61	8.25	1.55	3.04	4.11	8.44	1.90	9.42	0.76
17	F	8.47	3.04	3.73	4.33	1.24	1.75	8.27	2.80	8.76	1.76	6.01	2.49	8.18	2.75	8.54	2.15	8.42	2.01	8.06	2.11	7.74	2.78	7.49	1.50	8.57	1.51	8.58	2.00	9.34	0.80
17	F	7.23	3.63	6.17	3.53	4.67	3.67	9.32	0.81	8.51	1.56	5.74	1.64	8.28	2.35	8.10	2.62	8.96	2.29	5.81	2.32	9.15	0.98	9.13	1.05	7.77	2.93	4.73	2.33	8.56	1.50
14	M	1.92	1.78	3.17	1.89	4.38	2.86	4.10	2.16	4.39	2.20	7.09	1.60	5.00	2.49	4.20	1.48	4.85	2.25	3.63	1.84	4.01	1.49	6.03	1.82	4.73	2.48	3.16	2.11	7.23	2.15
11	F	6.39	3.48	5.85	3.40	4.66	3.62	7.49	2.84	7.57	3.01	7.81	2.10	7.52	2.15	8.32	1.98	1.87	3.10	8.47	2.13	8.38	1.60	8.66	1.24	8.67	1.66	5.81	2.10	9.23	0.78
17	F	2.84	2.44	3.94	2.70	4.29	3.13	4.25	2.43	5.39	2.00	6.34	2.52	4.00	2.37	1.41	2.26	5.56	2.10	3.90	2.09	6.06	2.75	6.50	2.29	2.97	3.26	3.64	2.42	6.63	2.14
16	F	1.78	1.94	4.57	2.63	4.57	2.53	3.93	3.48	6.66	2.21	5.73	2.46	4.61	2.82	3.64	2.18	0.65	1.09	5.52	2.24	4.75	1.60	4.82	0.30	4.57	1.55	4.16	2.02	6.26	1.83

Table 2. Interclass Correlation

Symptom/Diagnosis	ICC	95% CI
Mania	0.41*	0.19 – 0.76
Hypomania	0.10	-0.01 – 0.47
Bipolar NOS/NEC	0.09	-0.02 – 0.50
Episode	0.37	0.18 – 0.69
Elevated or expansive mood	0.36	0.16 – 0.70
Irritable mood	0.29	0.09 – 0.75
Persistently increased activity or energy	0.41*	0.20 – 0.74
Inflated self-esteem	0.69**	0.43 – 0.93
Decreased need for sleep	0.66**	0.47 – 0.87
More talkative than usual/pressured speech	0.53*	0.32 – 0.79
Flight of ideas/racing thoughts	0.38	0.17 – 0.74
Distractibility	0.29	0.09 – 0.70
Increased goal-directed activity or psychomotor agitation	0.47*	0.26 – 0.76
Excessive involvement in pleasurable activities w/ potentially painful consequences	0.48*	0.25 – 0.81
Impairment	0.23	0.09 – 0.60

Less than 0.40 – poor,
* Between 0.40 and 0.59 – fair
** Between 0.60 and 0.74 – good
*** Between 0.75 and 1.00 – excellent

Figure 1. Visual Analog Scale

Case#______ Rater______________________________ Date___________________

Read the vignette, then rate your confidence that the following bipolar symptoms, criteria, or mood episode are present DURING the most illustrative (hypo)manic episode in the narrative, by making a mark on the visual analogue scale. Then, if you were ***not absolutely certain*** (i.e., "highly likely") about the presence of each item (symptom, criterion or mood episode), make an "X" in the boxes to the right, indicating any factors which negatively impacted your confidence in that item. First rate the items for symptoms/criteria then the items for mood episodes.

	Highly Unlikely — Unclear — Highly Likely	Not Mentioned	Not described w/enough Detail	Insufficient Duration	Insufficient Severity	Age, Immaturity, IQ deficits	Not within an episode	OTHER
Distinct period of change of mood & energy for most of the day, nearly every day lasting either ≥4 days or ≥7 days or needing hospitalization: →								
Elevated or expansive Mood: →								
Irritable Mood: →								
Persistently increased activity or energy: →								
Inflated self esteem/grandiosity: →								
Decreased need for sleep: →								
More talkative than usual/pressured speech: →								
Flight of ideas/racing thoughts: →								
Distractibility: →								
Increased goal-directed activity or psychomotor agitation: →								
Excessive involvement in (pleasurable activities w/potential painful consequences: →								
Impairment: →								
Mania: →								
Hypomania: →								
Bipolar NOS/NEC: →								

Other factors ___________________ Comments ______________________________________

Conclusion

- Ratings for confidence in diagnosis and criteria varied greatly. SDs for confidence ratings of symptoms and diagnoses indicated great variation from case to case.
- Most ratings had low ICCs. This was in contrast to our pilot data, where ICCs ranged from high to unacceptable.
- Limitations include a small sample size, limited generalizability to non-BD research clinicians, and need for clarity of ratings using a visual analog scale.[5]
- In spite of efforts made to refine BD criteria with DSM-5, our findings indicate that investigators working with pediatric BD may still vary in their diagnostic decision making, thus warranting further research and education in this area.
- Next steps include investigating which factors contribute to clinician uncertainty that resulted in high variation in confidence ratings and comparing ratings to the Schedule for Affective Disorders and Schizophrenia for School Aged Children (KSADS) diagnoses.

References

1. Galanter CA, Hundt SR, Goyal P, Le J, Fisher PW. Variability among research diagnostic interview instruments in the application of DSM-IV-TR criteria for pediatric bipolar disorder. J Am Acad Child Adolesc Psychiatry 2012;51:605-621
2. Dubicka B, Carlson GA, Vail A, Harrington R. Prepubertal mania: diagnostic differences between US and UK clinicians. Eur Child Adolesc Psychiatry 2008;17:153-161
3. Galanter CA, Patel VL. Medical decision making: a selective review for child psychiatrists and psychologists. J Child Psychol Psychiatry 2005;46:675-689
4. Menard T, Galanter CA, Jensen PS, Youngstrom EA, Duffy A, Leffler J, Post R, Wall C, Seymour LR, Wagner KD, Carlson GA. Strategies for Improved Classification of Pediatric Bipolar Biobank Participants (abstract). American Academy of Child and Adolescent Psychiatry Annual Meeting, October 2013
5. Paul-Dauphin A, Guillemin F, Virion JM, Briancon S. Bias and precision in visual analogue scales: a randomized controlled trial. Am J Epidemiol 1999;150:1117-1127
6. Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. In Kazdin AE (ed): Methodological Issues and Strategies in Clinical Research. Washington, DC: American Psychological Association 1992; 233-256
7. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994;6(4):284–290.

Corresponding Author

Cathryn A. Galanter MD, Cathryn.Galanter@downstate.edu