

Editorial

Are rating scales for bipolar disorders fit for purpose?

Jan Scott and Greg Murray



The DSM-5 definition of bipolar disorder elevates increased activity or energy as a cardinal symptom (alongside mood changes) for mania and hypomania ('hypo/mania'). The ICD-10 likewise requires increases in activity and energy (alongside mood) for hypo/mania, as well as decreases for bipolar depression. Using bipolar disorder as an example, we propose that, when diagnostic criteria are revised, instruments used to measure clinical course and treatment response may need revisiting. Here, we highlight that the 'gold-standard' symptom rating scales for hypo/mania and depression were developed in an era when abnormalities of mood were viewed as the cardinal symptom of bipolar disorder. We contend that archetypal measures fail to give proportionate weighting to activity or energy, undermining their utility in monitoring bipolar disorder and treatment response in clinical and research practice.

Declarations of interest

J.S. and G.M. are members of mMARCH, (Motor Activity Research Consortium for Health), which is led by Dr Kathleen Merikangas, National Institute for Mental Health. J.S. reports being a visiting professor at Diderot University, the Norwegian University of Science and Technology, Swinburne University of Technology and The University of Sydney; receiving grant funding from the UK Medical Research Council and from the UK Research for Patient Benefit programme; and receiving a personal fee from Janssen-Cilag for a non-promotional talk on sleep problems.

Keywords

Bipolar affective disorders; diagnostic criteria; activity and energy; rating scales.

Copyright and usage

© The Royal College of Psychiatrists 2018.

Jan Scott (pictured) is a professor of Psychological Medicine at Newcastle University, UK and a visiting professor at the Institute of Psychiatry, Psychology and Neuroscience, London, UK and also holds posts at the Department of Mental Health, NTNU, Trondheim, Norway and the Brain and Mind Centre, The University of Sydney, Australia. Her interests include bipolar and mood disorders, clinical staging models, phenomenology and development of new psychological interventions. **Greg Murray** is a professor of Clinical Psychology at the Swinburne University of Technology, Australia. He studies the description, explanation and management of mood disorders.

As with other psychiatric conditions, the definition of bipolar disorder appears in two (related but different) types of artefact – diagnostic criteria aimed at optimising binary decisions about the presence of disorder in a given case, and symptom rating scales designed to quantify the disorder's severity at a point in time. We contend here that changes to diagnostic criteria have led to a substantive misalignment with major bipolar disorder rating scales. Recognition of activation as a central component of bipolar disorder, and consequent changes to diagnostic criteria, mean that long-established scales may no longer reflect current understanding and diagnosis of the disorder.

From manic depressive insanity to bipolar disorders (and back again?)

Early descriptions of manic depressive insanity (MDI) identified that the core, shared dimensions of acute illness episodes reflected disturbances of mood, cognition and motor activity.¹ Although Kraepelin¹ did not give primacy to any specific dimension, others suggested that change in psychomotor activity rather than mood was the most striking and consistent feature of MDI.² Over the next half century, however, descriptive attention shifted towards abnormalities of mood. Specifically, MDI was characterised in terms of two mood states, representing opposite poles of an implied pathological axis, and by the 1980s, the term 'bipolar disorder' replaced MDI. DSM-III and later revisions (DSM-III-R,

DSM-IV, DSM-IV-TR) identified mood change as the defining feature (A criterion) of the depressive and hypo/manic mood episodes making up the bipolar disorder diagnosis, with psychomotor disturbances among the seven B criteria (i.e. not essential for the diagnosis). Similar approaches were used in ICD-8 and ICD-9 to diagnose bipolar disorder.

Views of the optimal criteria for diagnosing bipolar disorder and its subtypes continued to evolve and by the turn of the century, the ICD-10 described bipolar disorder as being characterised by repeated episodes in which the patient's mood and activity levels are significantly disturbed.³ The ICD-10 description of bipolar disorder notes that individuals will show increased mood, energy and activity in hypo/mania and decreased mood, energy and activity in depression.⁴ Most recently, the DSM-5 has revised the criteria for bipolar disorder,⁵ stating that the A criterion for mania and hypomania should include 'a distinct period of abnormally and persistently elevated, expansive, or irritable mood *and* abnormally and persistently increased activity or energy' (italics added). Also, DSM-5 allows a specified 'with mixed features' (a descriptor that can be added when a person experiences both symptoms of depression and mania within the same episode). Diverging from ICD-10, the DSM-5 does not elevate activity and energy to the A criterion for bipolar depression (i.e. major depressive episode in the context of bipolar disorder).

Activation as a core dimension of bipolar disorders

The changes in diagnostic criteria for bipolar disorder have received qualified support. For example, in our recent systematic review,⁶ we conceptualised activity and energy, respectively, as the observable and subjective elements of activation in bipolar disorder and found that (a) activation was the most common primary dimension in mania and (b) there were differences in the factor structure of bipolar as compared with unipolar depression (for example psychomotor activity, suicidality and sleep) despite minimal differences in overall levels of symptom severity. Further, we demonstrated evidence of disturbed activation in bipolar disorder studies that

employed real-time monitoring (for example experiential sampling methods and ecological momentary assessment) or objective measurements (for example actigraphy and human behaviour pattern monitoring). Although debate continues around current diagnostic descriptions, we concluded that the elevation of activation to an obligatory symptom of hypo/mania was evidence based.

Given that the concept of bipolar disorder has evolved over the last half-century, it is unclear how well the most widely used clinical scales capture contemporary diagnostic criteria, especially regarding activation (activity and/or energy). Many mania and depression rating scales were developed some 30–40 years ago and demonstrate considerable heterogeneity in the range of assessed symptoms and in underlying assumptions about the nature of bipolar disorder episodes. For example, the influential Young Mania Rating Scale (YMRS)⁷ was introduced around the time the bipolar disorder construct replaced MDI in classification systems, and its items preclude rating an episode as dysphoric mania (i.e. depressed mood with increased activity and other symptoms of mania), one of the most common presentations (such as Goodwin & Jamison⁸) or the formal assessment of the DSM-5 mixed features specifier.

Assessment of activation in symptom rating scales

Given the above, we undertook a preliminary scoping review to identify the archetypal measures of the symptom states of bipolar disorder and examined those scales with sound psychometrics. As a first step in assessing scales' concordance with contemporary diagnostic criteria, we calculated the weighting of activation items, quantified as the percentage of the maximum possible score (POMP). As a point of comparison, POMP scores were also calculated for mood-related items.

As expected, the most commonly used scales aim to measure one of the two major syndromes in a bipolar conceptualisation. The most widely used mania rating scale was the clinician-rated YMRS,⁷ on which a single (increased) activation item has little impact on total score (activation POMP 7%; mood POMP 20%). Activation is weighted higher on the Bech Rafaelson Mania Scale (POMP 18%)⁹ and the self-report Altman Self-Rating Scale (POMP 20%).¹⁰

The most widely used depression rating scale was the 17-item clinician-rated Hamilton Rating Scale for Depression (HRSD).¹¹ The 17-item HRSD score (activation POMP 8%; mood: depression POMP 8% and anxiety POMP 8%) weighs activation less than does the 30-item Inventory of Depressive Symptoms (IDS clinician or self-rated versions: POMP 14%).¹² Also, the Montgomery–Åsberg Rating Scale (MADRS)¹³ gives less priority to activation than to mood (activation POMP 10%, mood POMP 20%). Overall, the depression scale with the maximal weight for an activation dimension is the Quick IDS (QIDS or IDS-16: POMP 22%).¹⁴ Activation is weighted even lower on common self-rating compared with observer-rated depression scales: depending on the version, the Beck Depression Inventory, for example, generated activation POMP of about 4–9%.¹⁵

Our review also identified scales not premised on the assumption that this complex condition is best characterised as two fundamental contrasting mood states. Newer scales like the self-rated MaTHyS (Multidimensional Assessment of Thymic States),¹⁶ for example, quantify points in a theoretically derived multidimensional bipolar disorder state space (emotional reactivity, motivation, sensory-perception, psychomotor activity, cognition speed). The earlier self-reported Internal State Scale has the same logic, but measures four dimensions (activation, well-being, perceived conflict,

depression index).¹⁷ Such multidimensional scales capture more forms of dysregulation and uncouple mood and activation, but they are not yet widely used.

Discussion

Our preliminary review raises concerns about the content validity of the most widely used symptom rating scales, the YMRS, MADRS and the 17-item HRSD. The POMP estimates suggest that none of these scales weight activation in a manner proportionate to its cardinal symptom status in DSM-5 (hypo/mania) and ICD-10 (hypo/mania and depression). Our strategy here of calculating the weighting given to activation is more rhetorical than quantitative, and more rigorous investigation of this issue is now warranted: our group is currently undertaking a systematic review of the psychometric and clinimetric properties, and external correlates of the major scales.

Coincidentally, researchers have recently questioned whether some of the most established rating scales are fit for purpose on other grounds. For example, YMRS scores show non-significant correlations with objective measures of mania.⁶ Also, an item-response theory analysis of the utility of the YMRS and MADRS concluded that both these scales are poorly constructed and inefficient (for example they contained several items that provided little or no psychometric information and only measured a narrow band of severity of symptoms) or that answers were, in part, determined by factors other than symptom severity.¹⁸ Likewise, the utility of the HRSD has been increasingly questioned.¹⁹ Questions about the proportionality of activation weighting (the focus here) may not be separate from other validity issues: for example the HRSD-6 (targeting six symptom dimensions) has the highest activation POMP score among versions of the HRSD (POMP 17%) and is superior to longer HRSD versions in terms of psychometric performance, item-response theory analysis and clinimetric profile.²⁰

Conclusions and recommendations

Recent changes to the diagnostic criteria for bipolar disorder, underpinned by growing research into its dimensional nature, motivate critical attention to the rating scales with which diagnoses are supposed to be paired. This initial scoping review highlights that the so-called gold-standard tools for the assessment of manic and depressive episodes may be limited in the assessment of the evolving bipolar disorder construct as defined by DSM-5 and ICD-10.

There are obvious benefits in using well-benchmarked scales, so in the short-term researchers and clinicians should choose from existing scales, mindful of their match with the diagnostic syndrome they are hoping to ameliorate (for example preferencing mania scales that give more equal weighting to mood and activation and/or uncouple elation from increased activity, depression from retardation). In the longer term, we await rating scales built on improved understanding of the dimensional structure of bipolar disorder, and pragmatic guidance about how such complex measures will be used to monitor and manage the condition in real time.

Jan Scott, MB BS, MD, FRCPsych, Professor, Psychological Medicine, Newcastle University, UK and Visiting Professor, Institute of Psychiatry, Psychology and Neuroscience, UK and Department of Mental Health, NTNU, Norway and Brain and Mind Centre, The University of Sydney, Australia; **Greg Murray**, PhD, FAPS, Professor, Clinical Psychology, Centre for Mental Health, Swinburne University of Technology, Australia

Correspondence: Jan Scott, Academic Psychiatry, Institute of Neuroscience, Newcastle University, Newcastle NE6 4BE, UK. Email: jan.scott@newcastle.ac.uk

First received 4 Apr 2018, final revision 14 May 2018, accepted 12 Aug 2018

References

- 1 Kraepelin E. *Manic-Depressive Insanity and Paranoia*. ES Livingstone, 1921.
- 2 Meyer A. *Collected Papers* (ed E Winters). John Hopkins Press, 1951.
- 3 Kaltenboeck A, Winkler D, Kasper S. Bipolar and related disorders in DSM-5 and ICD-10. *CNS Spectr* 2016; **21**: 318–23.
- 4 World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*. WHO, 1992.
- 5 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, DSM-5 (5th edn)*. American Psychiatric Association, 2013.
- 6 Scott J, Murray G, Henry C, Morken G, Scott E, Angst J, et al. Activation in bipolar disorders: a systematic review. *JAMA Psychiatry* 2017; **74**: 189–96.
- 7 Young RC, Biggs J, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978; **133**: 429–35.
- 8 Goodwin F, Jamison K. *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression (2nd edn)*: 1262. Oxford University Press, 2007.
- 9 Bech P, Bolwig TG, Kramp P, Rafaelsen OJ. The Bech-Rafaelsen Mania Scale and the Hamilton Depression Scale – evaluation of homogeneity and inter-observer reliability. *Acta Psychiatr Scand* 1979; **59**: 420–30.
- 10 Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. *Biol Psychiatry* 1997; **42**: 948–55.
- 11 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.
- 12 Rush AJ, Carmody T, Reimitt PE. The Inventory of Depressive Symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *Int J Meth Psychiatr Res* 2000; **9**: 45–59.
- 13 Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134**: 382–9.
- 14 Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, et al. An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry* 2006; **59**: 493–501.
- 15 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; **4**: 561–71.
- 16 Henry C, M'Bailara K, Mathieu F, Poinot R, Falissard B. Construction and validation of a dimensional scale exploring mood disorders: MAThys (Multidimensional Assessment of Thymic States). *BMC Psychiatry* 2008; **8**: 82.
- 17 Bauer MS, Crits-Christoph P, Ball WA, Dewees E, McAllister T, Alahi P, et al. Independent assessment of manic and depressive symptoms by self-rating. Scale characteristics and implications for the study of mania. *Arch Gen Psychiatry* 1991; **48**: 807–12.
- 18 Prisciandaro JJ, Tolliver BK. An item response theory evaluation of the Young Mania Rating Scale and the Montgomery-Asberg Depression Rating Scale in the systematic treatment enhancement program for bipolar disorder (STEP-BD). *J Affect Disord* 2016; **205**: 73–80.
- 19 Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004; **161**: 2163–77.
- 20 Timmerby N, Andersen JH, Sondergaard S, Ostergaard SD, Bech P. A systematic review of the clinimetric properties of the 6-item version of the Hamilton Depression Rating Scale. *Psychother Psychosom* 2017; **86**: 141–9.

psychiatry
in historySir Thomas More, L^D Chancellor – ‘His discourse was extraordinary facetious’

Greg Wilkinson

Sir – or Saint – Thomas More (1478–1535) was a man of conscience, a lawyer, statesman, and Renaissance humanist: a man for all seasons.^a He wrote *Utopia*, published in 1516, about the political system of an imaginary, ideal island nation, ruled by reason – in contrast to unruly contemporary European politics. More became close to Henry VIII and was Lord High Chancellor of England from October 1529 to 16 May 1532. However, he opposed Henry's separation from the Catholic Church (1532–1534); would not acknowledge Henry as Supreme Head of the Church of England and the annulment of his marriage to Catherine of Aragon; and, after refusing to take the Oath of Supremacy, he was arrested in 1534, convicted of treason and beheaded.

His country-house was at Chelsey ... Where the gate is now, adorned with two noble pyramids, there stood anciently a gate-house, w^{ch} was flatt on the top, leaded, from whence is a most pleasant prospect of the Thames and the fields beyond: on this place the L^d Chancellor More was wont to recreate himself, and contemplate. It happened one time, that a Tom of Bedlam^b came up to him, and had a mind to have thrown him from the battlements, saying, ‘Leap, Tom, leap.’ The Chancellor was in his gowne, and besides ancient, and not able to struggle with such a strong fellowe. My L^d had a little dog with him, sayd he, ‘Let us first throwe the dog downe, and see what sport that will be;’ so the dog was throwne over. ‘This is very fine sport,’ sayd my L^d, ‘fetch him up, and try once more;’ while the madman was goinge downe, my L^d fastened the dore, and called for help, but ever after kept the door shut.

His discourse was extraordinary facetious. One night riding, upon the suddaine he crossed himself, *majori cruce*,^c crying out, ‘Jesu Maria, doe you not see that prodigious dragon in the skye?’ they all lookt-up, and one did not see it, nor the other did not see it, at length one had spyed it, and at last all had spyed; whereas there was no such phantom, only he imposed on their phantasies.¹

Reference

- 1 John Aubrey. *Lives of Eminent Men, Vol. II. Part I (Appendix No. IV)*: 462–3. Printed for Longman, Hurst, Rees, Orme, and Brown; and Munday and Slatter, 1813.

The British Journal of Psychiatry (2018)
213, 629. doi: 10.1192/bjp.2018.156

^a ‘More is a man of an angel's wit and singular learning. He is a man of many excellent virtues; I know not his fellow. For where is the man (in whom is so many goodly virtues) of that gentleness, lowliness, and affability, and as time requires, a man of marvelous mirth and pastimes and sometime of steadfast gravity — a man for all seasons.’ Robert Whittington, *Vulgaria*, 1520. Whittington's likely debt to More's friend Erasmus is elaborated by Dr. Clarence H. Miller on ‘a man for all seasons’. Thomas More Studies Conference, 5 November 2005 (thomasmorestudies.org).

^b ‘Tom o' Bedlam’ was used in early modern Britain to describe beggars and vagrants who had or feigned mental illness. They claimed, or were assumed, to have been former inmates at the Bethlem (Bedlam), founded in 1247 and still extant.

^c greater cross (i.e. a full ‘sign of the cross’).