

BIPOLAR AND UNIPOLAR DEPRESSION

Jonathan Rogers^{2,4} & Mark Agius^{1,3,5}

¹Department of Psychiatry University of Cambridge, Cambridge, UK

²School of Clinical Medicine University of Cambridge, Cambridge, UK

³South Essex Partnership University Foundation Trust, UK

⁴Caius and Gonville College University of Cambridge, Cambridge, UK

⁵Clare College University of Cambridge, Cambridge, UK

SUMMARY

Since Kraepelin grouped affective disorders under the title of 'manic-depressive insanity', there has been controversy over whether the bipolar and unipolar entities within this are distinct affective disorders or whether they are merely two ends of an affective continuum. In order to bring some clarity and goal-posts to this argument, we define the criteria that must be fulfilled by diseases in order to be considered as part of a spectrum. We analyse bipolar disorder and major depressive disorder with respect to these criteria and find the model fits in many respects but fails to account for either the poor correlation in severity of manic and depressive symptoms or for the apparent discontinuity in the distribution of familial mania. A one-dimensional spectrum is thus too simple and a two-dimensional approach is required; this also fits much better with our current understanding of the genetic picture.

Key words: bipolar disorder – unipolar depression - bipolar spectrum - mania

* * * * *

INTRODUCTION

Bipolar and unipolar depression have much in common: the presence of hypothymic symptoms, the overlapping familial pools, common treatment strategies and the possibility for individuals to move from the latter to the former. Thus it was that Kraepelin grouped them both under the title of 'manic-depressive insanity' (MDI). However, as Kraepelin later admitted (Kraepelin 1913 cited in Angst et al. 2000), MDI was far from an homogenous entity and subsequent research has demonstrated that there are considerable differences between what are now termed bipolar disorder (BP) and major depressive disorder (MDD); these include the presence of mania in bipolar alone, differences in the symptomatology of depressive episodes, the earlier age of onset of bipolar and nuances in the inheritance patterns. Thus it was that, noting the development of full hypomanic, manic and depressive episodes in a cohort of cyclothymic patients, Akiskal postulated the existence of a 'cyclothymic-bipolar spectrum' (Akiskal 1977). This has since been expanded in several ways (Akiskal 2005, Angst 2007, Kelsoe 2003), but the general format in which it is presented and conceived in research literature is as a continuum from bipolar I (BP-I) to MDD, going via bipolar II (BP-II) and cyclothymia on the way. Additional intermediates have been proposed, such as bipolar III (BP-III), which is defined by hypomanic features that only appear after treatment with an antidepressant, and Angst's depressive state superimposed on a hyperthymic personality (Akiskal 2005). This schema has been proposed as a model for systematically understanding the similarities and differences between BP and MDD. However, since its inception, this model has faced considerable criticism.

Some disparagement is to be expected of any attempt to shun distinct classifications in favour of a continuum on account of the difficulties this causes clinicians in both diagnosis and treatment (Goldberg 2000). However, criticisms have also abounded regarding the scientific validity of a spectral model of the affective disorders on the grounds that BP and MDD are too different to be considered together, that the spectrum model lacks consistency and that there are gaps in the so-called continuum that cannot be crossed.

In the light of this debate, in the article we intend to examine the question of whether BP and MDD can be conceived as part of a spectrum and whether there is a better way of understanding them. As such, our first task is to define the criteria by which one should judge whether two or more diseases form a continuum. We shall then proceed to evaluate the evidence with respect to these criteria, establishing to what extent the putative BP-MDD spectrum meets these requirements, before considering the evidence as a whole to determine if a spectrum is indeed the best model in which to frame the information we have.

CRITERIA FOR A DIAGNOSTIC SPECTRUM

The concept of a spectrum is commonly used in medicine to describe a gradation in severity, as in the concept of acute coronary syndrome, which models the continuum between angina and STEMI via NSTEMI. Of course there are distinct cut-off points and these may have clinical significance, as in this example, but with others, such as hypertension, the distinction between mild and moderate manifestations of the disease can

appear somewhat arbitrary. There are also cases where a spectrum doesn't describe severity as much as it models the transition from one form of the disease to another. For instance, the concept of COPD has been born out of the two conditions of emphysema and chronic bronchitis in the recognition that they consist of a set of overlapping and graded symptoms and pathological processes. In order to establish whether BP and MDD may be considered as part of a spectrum, we must first consider what is distinct about the properties of a disease spectrum, which we have endeavoured to do below. The criteria are summarised in Table 1.

Table 1. Criteria for a Diagnostic Spectrum

1. The presence of intermediates without any significant gaps
2. The ability of individuals to move between entities
3. Either consistency or graded unidirectional across the following properties:
 - a. Epidemiology
 - b. Pathology
 - c. Symptomatology
 - d. Treatment

First, to place two diseases on a spectrum, it must be clear that they are not entirely distinct from each other; it would be difficult, for example, to place gallstones disease and pulmonary fibrosis on a spectrum because there are no obvious intermediates that could be placed between them. A spectrum implies that there is continuity between two extremes, so it is clear that there must be evidence of at least some intermediate points. It would be impossible to satisfy the requirement for demonstrating full continuity, as this would require an infinite number of contiguous points, so we shall stipulate that some intermediates must be demonstrated and that there must not be any significant gaps.

Secondly, a spectrum consists of variables manifested to varying degrees. Since the conditions in the human body are in a state of constant flux due to internal and external pressures, it would be expected that some of these variables would change from time to time. This would sometimes have the result of moving an individual from one entity on the putative spectrum to another. Thus we must require that there be evidence of patients moving between different points on the continuum. However, there is an important addition to make to this criterion due to the fact that in any two unrelated diseases there may be comorbidity (or that by chance one may cease to have one condition around the same time as acquires another). Hence, to use an earlier example, a patient may be unfortunately endowed simultaneously with both gallstones disease and pulmonary fibrosis, despite the lack of obvious relationship between them. Alternatively, this patient may cease to be in the throes of gallstones just as pulmonary fibrosis takes root, yet this does not necessarily imply that the gallstones has metamorphosed into pulmonary fibrosis.

This is clearly a question of statistical significance, for while there may be a few cases of the termination of one disease coinciding with the initiation of another, it is necessary to demonstrate that this is above the population rate. It is this test that would clarify that the phenomenon of visuospatial neglect giving way to extinction is the transfer to some residual form of the disease, while gallstones giving way to pulmonary fibrosis is merely bad luck.

Thirdly, in addition to the first criterion - that intermediates must be present - it must also be demonstrated that they are systematically related in some way. It is not enough that blood pressures of between the thresholds for mild and severe hypertension exist, but they must be shown to be related in a coherent way to their neighbours. This may manifest itself in two ways: either the variables must be constant (it is probable that all entities in a disease spectrum will have some commonalities, which indeed give them their place on the spectrum), or the variables must change progressively as conditions traverse the spectrum. The ways in which this phenomenon must be seen to operate need be limited to no more than the basic properties of a disease: its epidemiology, pathology, symptomatology and treatment.

BIPOLAR AND UNIPOLAR DEPRESSION AS A SPECTRUM

Presence of Intermediates

As we have already demonstrated, there have been several proposed models for a unipolar-bipolar spectrum. Only 15% of the population (the so-called 'super-normal') experiences no affective symptoms over their lifetime (Angst 2007), leaving a majority with sub-threshold symptoms. This provides a connection between our spectrum and the baseline, but there are also numerous gradations that have been illustrated. Angst et al. claim that current DSM criteria only catch the 'tip of the iceberg' when it comes to diagnosing bipolar, postulating the existence of minor bipolar disorder (MinBP) on the basis that many MDD sufferers also experience subthreshold hypomanic symptoms; by their calculations, 25% to 50% of all patients with a current diagnosis of MDD should actually be reclassified with some form of BP (Angst et al. 2002). On the backdrop of research like this, it has seemed straightforward to place recurrent depression, depression with a background hyperthymic temperament, hypomania on depression in the context of antidepressants or substance abuse, and depression with a background cyclothymic temperament on the scale between MDD and BP-II. Additionally, conventional bipolar disorder can be subdivided to fit a spectral model, giving rise to entities such as schizobipolar disorder and depression with protracted hypomania (Akiskal 2005).

However, a spectrum that incorporates all these elements has to involve a certain amount of rather

inconsistent vagueness, for while parts of it are coherent, as a whole it lacks an overall structure. Why, for example, should a fairly mild cyclothymic temperament find itself posited between MDD and BP-I, two much more serious diseases? It is hard to see any rule that holds throughout the spectrum. However, in spite of these objections, which shall be discussed later, the bipolar spectrum does still nonetheless have some considerable evidence behind it.

Movement along the Spectrum

It is in the area of transferral between conditions that the bipolar spectrum can become very helpful for clinicians and patients as a way of understanding the disease. The evidence is, moreover, robust that patients can move from one condition to another over the course of their illness. It has been demonstrated that, when followed for 4.5 years, patients initially diagnosed with Bipolar Not Otherwise Specified and Cyclothymia tended to progress to BP-II, while patients who began with a diagnosis of BP-II had a tendency to move to BP-I (Alloy et al. 2012). Moreover, there is conversion rate of approximately 1% per year of patients with depression to hypomania/mania (Angst & Sellaro 1995 cited in Angst et al. 2000). Since population prevalence of BP is generally estimated to be around the 1% mark, one would expect this incidence of BP among MDD sufferers to be considerably above levels in the community. What is interesting about the data is that they show a drift from MDD towards BP-I rather than movement in the opposite direction. This does not detract from the fact that these conditions are obviously linked, but it does raise intriguing questions about how this connection may operate.

Continuity of Properties

Here we endeavour to ascertain if a putative bipolar-unipolar spectrum yields either constant, or progressively changing variables across a number of scientific areas.

Epidemiology

Firstly, in terms of prevalence, a trend may be seen from BP-I to MDD, as the presence in the population increases dramatically (Kelsoe 2003); this is also a gradual change with BP-II being much more common than BP-I, so this fits well with a spectral model. Equally encouraging for the proponents of a continuum are the results of familial studies for depression, since it has been well demonstrated that relatives of probands with both unipolar and bipolar depression have higher rates of MDD than in the general population (Winokur, et al. 1995). However, there are two significant problems with this and other similar studies. The first is that, as hypomania is hard to diagnose, there may well be BP patients misdiagnosed in the MDD group. The second is based on the phenomenon of conversion from MDD to BP, as discussed above: Here it is possible that

some of the patients in the MDD group will at a future point convert to BP, so for the purposes of this type of research, they are more fittingly seen as pre-bipolar patients than as pure MDD. These problems could be eliminated only with difficulty, by conducting frequent diagnostic screening on patients to check for conversion over a number of years, retrospectively removing any patients who convert from the original MDD group.

It also seems that a spectrum including the normal range is supported by the fact that even relatives of probands with sub-threshold bipolar are at greater risk of having both MDD and BP (Angst et al. 2002). This last study also makes the case for a broader concept of bipolar in the light of this familial overlap.

There is, however, some significant epidemiological evidence that is claimed militates against the concept of a bipolar spectrum. On the other hand, before mentioning this, it is important to note that our concept of a spectrum is such that merely demonstrating differences between MDD and BP does not rule out such a model, as some seem to suggest; instead, it is necessary to demonstrate discontinuity in a variable that may be measured in the populations. For instance, the assertion that the earlier age of onset of BP compared to MDD necessarily rules out a spectrum (Winokur et al. 1993) is misleading, for it has also been demonstrated that BP with an earlier age of onset is a more severe form of the disease than BP commencing later in life (Goldstein et al. 2006, Leverich et al. 2007), which is exactly the result that would be expected in a spectrum. One epidemiological result that puts the idea of a spectrum into question though is the finding that rates of familial mania in MDD are no higher than in the general population (Winokur 1991 cited in Winokur et al. 1995); a spectrum would predict familial mania to be lower in MDD than BP but still higher than the prevalence in the community. To make the results all the more surprising, the authors also point out that familial mania is no higher in BP-I than BP-II.

Pathology

In terms of the pathogenesis of bipolar and unipolar depression, a theory based on temperaments in sub-threshold individuals is gaining credence. Akiskal defines Bipolar IV as being a depression superimposed on a hyperthymic temperament, while Bipolar II½ is depression superimposed on cyclothymic temperament (Akiskal 2005). These are, however, more than theoretical speculations with evidence now emerging to support the claim that extremes of personality traits may be related to affective disorders. For instance, it has been demonstrated that monozygotic twins of BP probands are more likely to be 'moody' people than the general population, as well as being predisposed to the development of schizophreniform psychoses (Akiskal 2005). Moreover, conversion from MDD to BP-II is strongly predicted by character traits such as high neuroticism, orality, emotional reliance and lack of self-confidence (Akiskal et al. 1995). On a more biological

level, it is possible that there are differences in cerebral metabolism between BP and MDD (Schwarz et al. 1987 cited in Winokur et al. 1995), but whether this necessarily represents discontinuity it is not possible to say.

Symptomatology

The main argument against symptomatic continuity between MDD and BP is the complete absence of mania or hypomania in MDD (Winokur et al. 1993). However, what is actually observed is that only a small proportion of the population as a whole manifests no affective symptoms (Angst 2007) and the rest is distributed over a scale from sub-threshold hypomania (Angst et al. 2002) to psychotic mania. This argument often implicates debate over the definition of bipolar disorder, since the presence of sub-threshold hypomanic symptoms does seem to have familial significance (Angst et al. 2002). Moreover, sub-threshold hypomania also predicts conversion from MDD to BP (Fiedorowicz et al. 2011), as do transient manic symptoms (Nadkarni et al. 2010). The perceived importance of these symptoms has led some to consider up to half of all current MDD patients eligible for re-diagnosis as some form of BP (Angst et al. 2002). For our purposes, however, these semantics are irrelevant, but what does matter is that it is evident that mania and hypomania are not discrete entities, for they are merely the end of a continuous chain of symptoms. The problem with this approach is that it separates mania from depression, rendering it difficult to include them in the same spectrum. Similarly, it is hard to conceive the concept of a mixed state on a linear axis from unipolarity to bipolarity.

Treatment

It is true that in the main pharmacological therapy for unipolar and bipolar depression varies significantly; this, in part at least, is why the correct diagnosis is so important and why there is so much controversy surround definitions. However, an interesting point to add to this area is that antidepressants are often used (with caution) in the depressive phases of BP, while lithium may be of use in some cases of MDD. In fact, Akiskal labels Bipolar V as recurrent depression that fails to respond to antidepressants but does respond to lithium. It is possible that this entity, which he describes as 'phenotypically unipolar' but 'genotypically bipolar', does indeed represent a therapeutic intermediate between MDD and BP (Akiskal 2005).

DISCUSSION

To summarise our findings so far, epidemiology points to depression being continuous but not mania; pathology illustrates a potential personality continuum; symptomatology shows a convincing continuum that struggles to remain within one dimension, and treatment demonstrates on a small scale the possible existence of an intermediate. The main problems for the spectral model then are as follows:

- Familial mania is not continuously distributed across the spectrum.
- Depression and mania are not always correlated.

An obvious way to resolve these problems, which has been adopted by many (Cuellar et al. 2005), is to consider the affective disorders as two spectra - one for depression, the other for mania. The advantages of this are manifest, for it allows one to consider mania alone as a disease and it permits diseases of different severities to exist alongside the gradation from depression to mania. Moreover, it provides a more satisfactory understanding of mixed states, wherein both manic and depressive components are high, and it additionally allows the representation of sub-threshold states. Examples of how it might represent different diseases are shown in Figure 1.

An alternative two-spectrum model has also been proposed (Angst 2007) in which the two axes are severity and depression-mania, as illustrated in Figure 2. In actual fact, these two models are conceptually the same, since the severity of the second one may merely be calculated from the sum of the depressive and manic components.

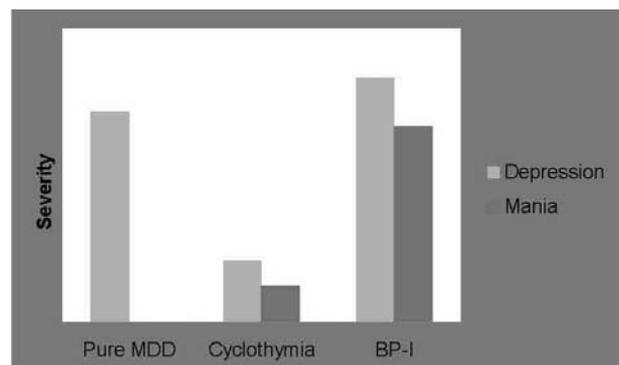


Figure 1. A Two-Spectrum Model of Affective Disorders

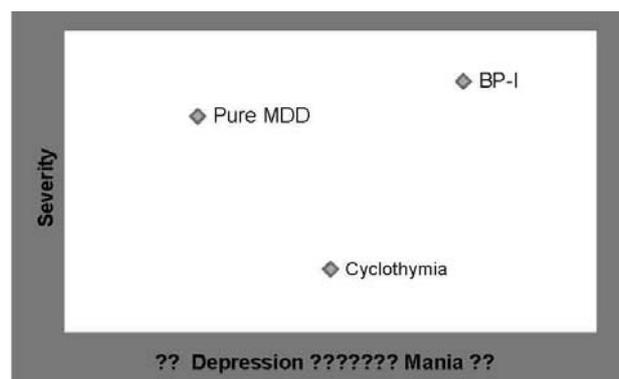


Figure 2. A Two-Spectrum Model of Affective Disorders

However, while this two-spectrum model is clearly more satisfactory it still has shortcomings. Primarily, it rather neglects the fact that mania and depression are so frequently comorbid by separating them into two separate spectra.

At this point it is useful to refer to a very helpful review by Kelsoe (2003), which examines the genetic

aspect of this question. His thesis is two-fold: firstly that the genetics of bipolar seems to consist of both some major loci as well as a great deal of polygenicity; secondly, there are some genes that are specific to bipolar, some specific to unipolar and some that predispose towards both. These genes interact both with each other through epistatic processes and with the environment. With respect to the models we have been examining these findings are significant in several ways. In the first place, the polygenetic element affirms the idea of a spectrum, since the concept is of many genes each exerting a small effect, which may be approximated to a continuous spectrum. Equally striking is the concept of the differential relevances of genes with respect to the two conditions, since the idea of unipotent genes predisposing for either depression or mania is consistent with a two-spectrum model, but pluripotent genes that predispose to BP and MDD more resemble the initial one-spectrum model (though this would clearly fail on closer analysis, as the accumulation of these pluripotent genes does not necessarily push the person towards BP and away from MDD). Thus, in order to incorporate these findings into our model, we must conceive of three distinct genetic effects that genes may have on a person's position on this spectrum:

- They may raise the depressive component;
- They may raise the manic component; or
- They may raise both components (note that this is identical to raising the severity component, if we are to use the terminology of Figure 2).

Additionally, this would only describe a person's genetic tendency towards morbidity and environmental factors would then interact with this.

Unfortunately, however, due to the epistatic phenomena at work, even this would be a gross oversimplification, since it is quite probable that the action of a gene A is augmented in the presence of a gene B and diminished in the presence of a gene C. Some account would also need to be taken of the presence of major loci, conferring strong familial heredity.

CONCLUSION

A two-dimensional model of an affective spectrum is a good fit for the current epidemiological, pathological, symptomalogical and therapeutic evidence. This highlights the necessity for clinicians to cease considering patients as immovable entities, fastened to an absolutely defined diagnosis. Instead, diagnostic descriptions should be flexible enough to be shaped by patients not vice versa. To use a nautical analogy, patients must be viewed not only as docking at particular ports but also as travelling in the complex space in between, driven by their own intrinsic characteristics and the state of their surroundings. It must be continually acknowledged that any strictly defined diagnostic entity is only an approximation to the

patient's illness, which will often fall between any such criteria. Clinicians must thus not be deterred from treating patients on the basis that they may not entirely fit the conventional categories.

Acknowledgements: None.

Conflict of interest: None to declare.

REFERENCES

1. Akiskal H *The Bipolar Spectrum: History, Description, Boundaries, and Validity (Book Section)/ Handbook of Bipolar Disorder: Diagnosis and Therapeutic Considerations/ book auth. Kasper S Hirschfeld R. - Boca Raton: Taylor & Francis Group, 2005.*
2. Akiskal HS et al. Switching from 'unipolar' to bipolar I/An 11-year prospective study of clinical and temperamental predictors in 559 patients. *Arch Gen Psychiatry*. 1995; 52: 114-23.
3. Akiskal HS Cyclothymic disorder: validating criteria for inclusion in the bipolar affective group. *American Journal of Psychiatry* 1977; 134:1227-1233.
4. Alloy LB et al. Progression along the Bipolar Spectrum: A Longitudinal Study of Predictors of Conversion from Bipolar Spectrum Conditions to Bipolar I and II Disorders. *J Abnorm Psychol* 2012; 121:16–27.
5. Angst J et al. Major Depressive Disorder with Sub-threshold Bipolarity in the National Comorbidity Survey Replication. *Am J Psychiatry* 2010; 167:1194–1201.
6. Angst J and Gamma A A new bipolar spectrum concept: a brief review. *Bipolar Disord* 2002 4 Suppl 1.; 11-4.
7. Angst J and Preisig M Course of a clinical cohort of unipolar, bipolar and schizoaffective patients. Results of a prospective study from 1959 to 1985. *Schweiz Arch Neurol Psychiatr* 1995; 146:5-16.
8. Angst J and Sellaro R Historical perspectives and natural history of bipolar disorder. *Biol Psychiatry* 2000; 48:445-57.
9. Angst J The bipolar spectrum. *Br J Psychiatry* 2007; 190:189-91.
10. Angst J, Gamma A and Lewinsohn P The evolving epidemiology of bipolar disorder *World Psychiatry* 2002; 1:146-148.
11. Cuellar AK, Johnson SL and Winters R Distinctions between bipolar and unipolar depression *Clin Psychol Rev* 2005; 25:307-339.
12. Fiedorowicz JG et al. Subthreshold Hypomanic Symptoms in Progression From Unipolar Major Depression to Bipolar Disorder *Am J Psychiatry* 2011; 168:40–48.
13. Goldberg D Plato versus Aristotle: categorical and dimensional models for common mental disorders *Comprehensive Psychiatry* 2000; 41(2 suppl):8-13.
14. Goldstein BI and Levitt AJ Further evidence for a developmental subtype of bipolar disorder defined by age at onset: results from the national epidemiologic survey on alcohol and related conditions. *Am J Psychiatry* 2006; 163:1633-6.
15. Kelsoe J Arguments for the genetic basis of the bipolar spectrum. *J Affect Disord* 2003; 73:183-97.

16. Leverich GS et al. The poor prognosis of childhood-onset bipolar disorder. *J Pediatr* 2007; 150:485-90.
17. Middeldorp CM et al. The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data *Transl Psychiatry* 2011; 1:e50.
18. Nadkarni RB and Fristad MA Clinical course of children with a depressive spectrum disorder and transient manic symptoms *Bipolar Disord* 2010; 12:494–503.
19. Winokur G et al. A family study of manic-depressive (bipolar I) disease. Is it a distinct illness separable from primary unipolar depression? *Arch Gen Psychiatry* 1995; 52:367-73.
20. Winokur G et al. A prospective follow-up of patients with bipolar and primary unipolar affective disorder. *Arch Gen Psychiatry* 1993; 50:457-65.

Correspondence:

Mark Agius, MD
SEPT at Weller Wing, Bedford Hospital
Bedford, Bedfordshire, MK42 9DJ, UK
E-mail: ma393@cam.ac.uk