REVIEW

The International College of Neuropsychopharmacology (CINP) Treatment Guidelines for Bipolar Disorder in Adults (CINP-BD-2017), Part 1: Background and Methods of the Development of Guidelines

Konstantinos N. Fountoulakis, MD; Allan Young, MD; Lakshmi Yatham, MD; Heinz Grunze, MD; Eduard Vieta, MD; Pierre Blier, MD; Hans Jurgen Moeller, MD; Siegfried Kasper, MD

3rd Department of Psychiatry, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece (Dr Fountoulakis); Centre for Affective Disorders, Institute of Psychiatry, Psychology and Neuroscience, King’s College, London, UK (Dr Young); Department of Psychiatry, University of British Columbia, Mood Disorders Centre of Excellence, Djavad Mowafaghian Centre for Brain Health, Vancouver, Canada (Dr Yatham); Paracelsus Medical University, Salzburg, Austria (Dr Grunze); Hospital Clinic, Institute of Neuroscience, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain (Dr Vieta); The Royal Institute of Mental Health Research, Department of Psychiatry, University of Ottawa, Ottawa, Canada (Dr Blier); Psychiatric Department, Ludwig Maximilians University, Munich, Germany (Dr Moeller); Department of Psychiatry and Psychotherapy, Medical University Vienna, MUV, AKH, Vienna, Austria (Dr Kasper).

Correspondence: Konstantinos N. Fountoulakis, MD, 6, Odysseos str (1st Parodos Ampelomon str.), 55535 Pylaia Thessaloniki, Greece (kfount@med.auth.gr).

Abstract

Background: This paper includes a short description of the important clinical aspects of Bipolar Disorder with emphasis on issues that are important for the therapeutic considerations, including mixed and psychotic features, predominant polarity, and rapid cycling as well as comorbidity.

Methods: The workgroup performed a review and critical analysis of the literature concerning grading methods and methods for the development of guidelines.

Results: The workgroup arrived at a consensus to base the development of the guideline on randomized controlled trials and related meta-analyses alone in order to follow a strict evidence-based approach. A critical analysis of the existing methods for the grading of treatment options was followed by the development of a new grading method to arrive at efficacy and recommendation levels after the analysis of 32 distinct scenarios of available data for a given treatment option.

Conclusion: The current paper reports details on the design, method, and process for the development of CINP guidelines for the treatment of Bipolar Disorder. The rationale and the method with which all data and opinions are combined in order to produce an evidence-based operationalized but also user-friendly guideline and a specific algorithm are described in detail in this paper.

Keywords: Bipolar Disorder; anticonvulsants; antidepressants; antipsychotics; evidence-based guidelines; lithium; mania; bipolar depression; mood stabilizers; treatment
Introduction

General Background, Disclosure, and Aim

Treatment guidelines are becoming an ever more important part of medical reality, especially since the translation of research findings to everyday clinical practice is becoming increasingly difficult with the accumulation of complex and often conflicting research findings that are thereafter also included in metaanalysis. Guidelines aim to assist clinicians but also policymakers to arrive at decisions concerning the treatment and care of patients. They set the standard of care and training for health professionals and they also identify priority areas for further research, since they are based primarily on the available evidence, but also in areas where evidence is not available, on expert opinion (Fountoulakis, 2015i).

In the field of Bipolar Disorder (BD), accumulated knowledge is often complex, confusing, and in many instances contrasts with the beliefs and practices that appear to have been set in stone in psychiatric culture and training for the last few decades.

To fulfil this need for expert translation of research findings into clinical practice for the benefit of patients, the International College of Neuropsychopharmacology (CINP) launched an effort to critically appraise the literature and provide guidance to clinicians in the form of a precise treatment algorithm. It is hoped that this algorithm for the treatment of BD will help the clinician to follow the state-of-the-art evidence, thus enabling their clinical practice to be based on an informed decision-making process. This guideline has been commissioned by the CINP, and the workgroup consisted of experts with extensive research and clinical experience in the field of BDs. There was no funding from any source for the development of the guidelines and the activities of the workgroup.

All the members of the workgroup were psychiatrists who are in active clinical practice and were selected according to their expertise and with the aim to cover a multitude of some different cultures. All of them were involved in research and other academic activities, and therefore it is is possible that through such activities some contributors have received income related to medicines discussed in this guideline. All conflicts of interest are mentioned at the end of this paper, which is the introductory paper to the CINP BD guidelines. It should also be noted that some drugs recommended in the guideline may not be available in all countries, and labeling and dosing might vary.

The aim of the current endeavor was to develop a guideline and precise algorithm for treatment of BD in adults for use in primary and secondary care. Children, adolescents, and the elderly are not the focus of this guidance. The guideline and algorithm have been developed after a complete review of the literature and with the use of stringent criteria. Both the guideline and the precise algorithm try to balance research vs clinical wisdom but give primacy to the available evidence.

To comply with the journal’s word limit for manuscripts and for easy readability, the CINP guidelines have been organized and presented as a series of 4 distinct papers. This paper is the first of this series and will cover the general background of the guideline and algorithm, that is, the historical perspective and general clinical and treatment issues followed for the development of the guideline and the algorithm. The second paper summarizes, classifies, and grades the treatment data on BD while the third paper includes the guideline and the treatment algorithm themselves. The fourth and final paper addresses the unmet needs and areas that should be the focus of attention and specific research in the future.

Historical Perspective

Depression and bipolarity were mentioned in Eber’s papyrus in ancient Egypt around 3000 BC (Okasha and Okasha, 2000) and in the Hippocratic texts. Plato (424–348 BC) and Aristotle (384–322 BC) further elaborated on the concept and Aristotle was the first to describe accurately the affections of desire, anger, fear, courage, envy, joy, hatred, and pity. Later, Galen (131–201 AD), Themison of Laodicea (1st century BC) and Areteaus of Cappadocia (2nd century AD) as well as Arab scholars and especially Avicenna (980–1037) further elaborated on the concept of mood disorders (Fountoulakis, 2015b).

Jean-Philippe Esquirol (1772–1840) was the first to clearly point out that melancholia was a disorder of the mood with “partial insanity” (monomania) and used the word “lypemania.” Finally, Jean-Pierre Faret (1794–1870) and Jules Gabriel Francois Baillarger (1809–1890) established the connection between depression and mania and gave it the name of “folie circulaire” or “folie à double forme,” but it was Emil Kraepelin (1856–1926) who established manic-depressive illness as a distinct nosological entity and separated it from schizophrenia on the basis of heredity, longitudinal follow-up, and a supposed favorable outcome (Kraepelin, 1921). His pupil Wilhelm Weygandt (1870–1939) published the first textbook on mixed clinical states (Weygandt, 1899). Following a similar line of thinking, and in spite of some major objections to the Kraepelian approach, Karl Jaspers (1883–1969) described aspects of mixed depressive states that he named “querulant mania,” “nagging depression,” or “wailing melancholia” (Jaspers, 1913), while Eugene Bleuler coined the term “affective illness” and by this he broadened the concept of manic-depression.

In 1957 Karl Leonhard (1902–1988) proposed that the term “bipolar disorder” should replace manic-depression, and he also made a distinction between monopolar (unipolar depression) and bipolar illness (Leonhard, 1957a, 1957b; Leonhard, 1979).

In 1870, Silas Weir Mitchell (1829–1914) was the first to recommend lithium as an anticonvulsant, hypnotic, and as medication for “general nervousness” (Mitchell, 1870, 1877). In 1871, William Alexander Hammond (1828–1900) was probably the first to prescribe a modern and effective psychotropic agent, and this was lithium (Mitchell and Hadzi-Pavlovic, 2000). Carl Lange (1834–1900) and Frederik Lange (1842–1907) had used lithium in the treatment of depression since 1886 (Lenox and Watson, 1994). However, in spite of encouraging results, by the turn of the 20th century, the “brain gout” theory of mood disorders disappeared as a medical entity and the use of lithium in psychiatry was abandoned.

In 1949 John Cade (1912–1980) reported positive results from the treatment of 10 acutely manic patients (Cade, 1949, 2000); however, 2 years later he reported the first death caused by lithium toxicity in a patient whose bipolar illness otherwise responded extremely well to treatment. Later, Mogens Schou (1918–2005) undertook a randomized controlled trial of lithium in mania (Schou et al., 1954; Bech, 2006), and eventually the efficacy of lithium during the maintenance phase was established (Gershon and Yuwiler, 1960; Baastrup, 1964; Baastrup and Schou, 1967; Angst et al., 1969, 1970; Baastrup et al., 1970; Schou et al., 1970; Johnstone et al., 1988; Schioldann, 1999; Mitchell and Hadzi-Pavlovic, 2000; Bech, 2006; Schioldann, 2006; Schioldann, 2011).

Valproate was introduced in 1966 as an anticonvulsant (Lambert et al., 1966) and later carbamazepine (Okuma et al., 1979) followed. Neuroleptics were introduced by Jean Delay...
In terms of individual symptoms, fatigue and psychomotor retardation dominate the clinical picture in 75% of patients during acute bipolar depression. Irritability is present in almost 75% of patients (Winokur et al., 1969), delusions are present in 12 to 66% (Winokur et al., 1969; Carlson and Strober, 1978; Rosenthal et al., 1980; Black and Nasrallah, 1989), and hallucinations in 8 to 50% (Winokur et al., 1969; Carlson and Strober, 1978; Rosenthal et al., 1980; Black and Nasrallah, 1989; Baethge et al., 2005). Psychotic features seem to constitute a stable trait that tends to repeat itself across episodes (Helms and Smith, 1983; Nelson et al., 1984; Aronson et al., 1988a, 1988b). Depending of the study sample composition, changes in appetite for food are seen in almost all patients (Winokur et al., 1969), with one-fourth manifesting overeating and one-fourth losing significant weight (Casper et al., 1985). Almost all bipolar depressed patients experience some kind of sleep problem (Winokur et al., 1969; Casper et al., 1985). A subgroup of bipolar depressed patients (up to 25%) often exhibit excessive sleep and have difficulty getting up in the morning (Winokur et al., 1969). Decreased sexual desire is seen in more than 75% of patients (Winokur et al., 1969; Casper et al., 1985) and concerns both sexes. Approximately two-thirds of bipolar depressed patients present with multiple physical pains and complaints (e.g., headache, epigastric pain, precordial distress, etc.) in the absence of any physical illness, especially in primary care (Winokur et al., 1969).

Euphoria is observed in 30 to 97% of acutely manic patients (Clayton and Pitts, 1965; Winokur et al., 1969; Beigel and Murphy, 1971; Carlson and Goodwin, 1973; Taylor and Abrams, 1973; Winokur and Tsuang, 1975; Abrams and Taylor, 1976; Leff et al., 1976; Loudon et al., 1977; Taylor and Abrams, 1977; Cassidy et al., 1998a), while unrestrained and expansive mood is seen in 44 to 66% (Taylor and Abrams, 1973, 1977; Louden et al., 1977). Patients are dissatisfied and intolerant and the vast majority manifest mood lability and instability (42 to 95%) (Winokur et al., 1969; Carlson and Goodwin, 1973; Abrams and Taylor, 1976; Louden et al., 1977; Taylor and Abrams, 1977; Cassidy et al., 1998a), while some patients manifest a return to the premorbid level of functioning between the episodes and to favorable outcome compared with schizophrenia (Kraepelin, 1921), today we know that this is not always the case (Tohen et al., 1990; Grande et al., 2016). Not only BD is a primary care (Winokur et al., 1969).

Clinical Description

While the basic conception of BD suggested that it is characterized by the alternation of manic and depressive episodes with a return to the premorbid level of functioning between the episodes and to favorable outcome compared with schizophrenia (Kraepelin, 1921), today we know that this is not always the case (Tohen et al., 1990; Grande et al., 2016). Not only BD is a primary care (Winokur et al., 1969).

The fact that often the correct diagnosis is made only after 8 to 10 years have passed because the first episode is psychotic-like or depressive and the correct diagnosis can be made only after a manic or a mixed episode emerges (Angst, 2007) is especially problematic. It has been estimated that more than one-half of hospitalized patients originally manifesting a depressive episode will turn out to be bipolar in the next 20 years (Angst et al., 2005a). It is of utmost importance for both clinicians and researchers to create a biographical chart with the patient’s course over time and including any important event in the developmental history of the patient and emphasizing the main events and hallmarks of his/her life and his/her full psychiatric and medical history. Such a chart clarifies both the diagnosis and the course of the disease and also the response to therapeutic interventions, since any delay in the proper diagnosis also delays proper treatment (Altamura et al., 2010; Drancourt et al., 2013).

Table 1. List of the Multiple Clinical Aspects of Manic-Depressive Illness

<table>
<thead>
<tr>
<th>1. Manic episodes</th>
<th>2. Depressive episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Subthreshold depressive symptoms</td>
<td>6. ‘Mixed’ states and ‘roughening’</td>
</tr>
<tr>
<td>7. Mood lability/cyclothymia/‘personality-like’ behavior</td>
<td>8. Predominant polarity</td>
</tr>
<tr>
<td>15. Self-destructive behavior and suicidality</td>
<td></td>
</tr>
</tbody>
</table>
Cassidy et al., 1998b; Serretti and Olgiati, 2005) and psychotic features in 33 to 96% of patients (Winokur et al., 1969; Carlson and Strober, 1978; Rosenthal et al., 1980; Black and Nasrallah, 1989).

Overall, psychotic features are so common that acute mania should be considered primarily a psychotic state (Koukopoulos, 2006). Delusions are present in 24 to 96% of manic patients, and it is interesting that persecutory ideas are equally frequent with delusions of grandiose (Bowman and Raymond, 1932; Rennie, 1942; Astrup et al., 1959; Clayton and Pitts, 1965; Winokur et al., 1969; Beigel and Murphy, 1971; Carlson and Goodwin, 1973; Taylor and Abrams, 1973, 1977; Murphy and Beigel, 1974; Abrams and Taylor, 1976; Jeff et al., 1976; Loudon et al., 1977; Carlson and Strober, 1978; Rosenthal et al., 1980; Winokur, 1984; Black and Nasrallah, 1989; Serretti et al., 2002; Keck et al., 2003; Goodwin and Jamison, 2007). Hallucinations are less frequent and present in 13 to 66% of cases; they can either be congruent or noncongruent, with auditory, visual, and olfactory ones being almost in 13 to 66% of cases; they can either be congruent or noncongruent, with auditory, visual, and olfactory ones being almost equally frequent (Lange, 1922; Bowman and Raymond, 1932; Astrup et al., 1959; Winokur et al., 1969; Taylor and Abrams, 1973, 1977; Abrams and Taylor, 1976; Carlson and Strober, 1978; Rosenthal et al., 1980; Winokur, 1984; Black and Nasrallah, 1989; Serretti et al., 2002; Keck et al., 2003; Goodwin and Jamison, 2007). Psychotic symptoms in BD are predictive of a more detrimental course, including a higher rate of rehospitalizations (Caetano et al., 2006; Ozylidirm, et al., 2010).

Almost one-third of acutely manic patients are “confused” and 46 to 75% are violent (Carlson and Goodwin, 1973; Taylor and Abrams, 1973, 1977; Abrams and Taylor, 1976; Cassidy et al., 1998b). The term confused refers to manic disorganization and not to organic drop in the level of consciousness. As many as 14 to 56% of patients manifest severe regression, catatonia, posturing, and negativism, often making differential diagnosis the classic rapid-cycling includes cycles with duration of weeks to months. Ultra-rapid cycling is reported when mood cycling somewhere between 15% and 50% of BD patients are reported to manifest some type of seasonal variation of symptomatology (Hunt et al., 1992; Faedda et al., 1993; Goikolea et al., 2007; Shand et al., 2011). Two opposing seasonal variations have been described: fall-winter depression with or without spring-summer mania or hypomania; and spring-summer depression with or without fall-winter mania or hypomania (Faedda et al., 1993). Most studies support the first subtype (Walter, 1977; Parker and Walter, 1982; Mulder et al., 1990; Peck, 1990; Parton and Lonnqvist, 1996; Clarke et al., 1999; Lee et al., 2007; Murray et al., 2011).

The concept of rapid cycling appeared for the first time in the 70s in a landmark paper by Dunner and Fieve (1974). In general the classic rapid-cycling includes cycles with duration of weeks to months. Ultra-rapid cycling is reported when mood cycling has frequency of weeks to days, and ultradian cycling when there is significant mood variation within a 24-hours period (Kramlinger and Post, 1996). Other terms include ultra-ultra rapid and ultradian rapid and refer to weekly or daily cycling, which is not uncommon in BD patients (Kramlinger and Post, 1996). Most studies suggest a 5 to 33.3% up-to-1-year prevalence
In terms of neurocognitive function, the literature suggests that the neurocognitive deficit in BD patients concerns almost all domains and phases of the illness with only a few exceptions. Its magnitude is at the severe range during the acute episodes and at the medium range during euthymia, while the origin of the deficit remains unclear. In terms of neurocognitive function, BD patients do quantitatively better than patients with schizophrenia, but the qualitative pattern of the deficit is similar in the 2 disorders. There are no clear differences between BD subtypes. The deficit is present early in the course of the disorder. At least in some patients it might emerge before the onset of the first mood episode, and in the majority of patients it progresses probably in relationship with the manifestation of psychotic symptoms. The verbal memory and executive function deficits probably constitute endophenotypes, while the role of medication as a causative factor is limited (Tsitsipas and Fountoulakis, 2015; Cullen et al., 2016).

Finally, in contrast to the original conceptualization of BD by Emil Kraepelin a century ago, unfortunately it seems that only a minority of BD patients achieve complete functional recovery (Goldberg et al., 1995a, 1995b; Keck et al., 1998; Strakowski et al., 1998; Daban et al., 2006; Martinez-Aran et al., 2007; Mur et al., 2007).

Classification

ICD and DSM include BD as a diagnostic entity but with significant differences between them (Fountoulakis, 2015b). It is important to note that almost all the research literature follows the DSM classification, while almost all countries worldwide have the obligation to use the ICD in their official documents, including hospital records, etc. The ICD-10-CM helps to bridge these 2 different classification systems for administration purposes. In ICD-10 (WHO, 1992, 1994), BD is included in the chapter on mood (affective) disorders (F30-F39). While in previous editions of the DSM, both unipolar and bipolar disorders were grouped under the chapter on mood disorders, on the contrary in DSM-5 (American Psychiatric Association, 2013), BDs were separated from unipolar depression. The “bipolar” chapter includes BD and cyclothymic disorder, while the “depression” chapter includes disruptive mood dysregulation disorder, major depressive disorder, persistent depressive disorder (dysthymia), and premenstrual dysphoric disorder. Both chapters include “unspecified,” “other,” and “due to” categories.

Another important difference between the 2 classification systems is that ICD requires the presence of at least 2 episodes of pathological disturbance of mood while DSM does not. DSM recognizes the presence of 2 subtypes of BD, that is, of BD-I (BD with manic episodes) and BD-II (BD with hypomanic but not manic episodes). BD-II is not part of the ICD-10 diagnostic list, which accepts hypomania as a diagnostic entity (F30.0), but it is considered simply a low-severity mania.

In ICD-10 a mixed affective episode (F38.0) is defined as an affective episode of at least 2 weeks duration that is characterized by either a mixture or a rapid alternation (usually within a few hours) of hypomanic, manic, and depressive symptoms. In DSM-5 a radical change was the abolishment of the concept of mixed episodes. In previous versions of the DSM, mixed episodes were defined as the coexistence of full-blown manic and depressive episodes simultaneously. Although such a coexistence is rather rare, almost one-third of patients recruited in pharmaceutical trials of acute mania were diagnosed as mixed. Thus there exists ample data, although neither properly analyzed nor published. Instead of the diagnosis of mixed episodes, DSM-5 introduced the mixed features specifier concept. According to this, a mood episode (either manic or depressed) has mixed features if at least 3 criteria of the opposite pole (from a specific list) coexist. It is important to note that according to DSM-5, mixed features can also be attributed to a unipolar major depressive episode without changing the diagnosis to BD.

Another important change in the DSM-5 is the introduction of the anxious distress specifier, which demands the presence of at least 2 criteria from a list of 5 (tension, restlessness, concentration difficulties, worry, fear of losing control).

The ICD-10 classification accepts the presence of “somatic syndrome,” which seems analogous but it is not identical to “melancholic features” of DSM-5 (Fountoulakis et al., 1999). The typical features, rapid cycling, and anxious distress are described in DSM-5 but not in ICD-10. Also, ICD-10 does not differentiate psychotic symptoms into mood congruent vs mood incongruent. The other specifiers, catatonia, peripartum onset, and seasonal pattern, are not included in the ICD-10 either. It is also important to note that ICD-10 recognizes catatonia only in the frame of schizophrenia, while DSM-5 uses this specifier also in affective disorders.

There is an issue concerning the diagnosis of cases with subthreshold manic symptoms or long-lasting hyperthymia. While the traditional bipolar vs unipolar distinction is widely used and adopted by classification systems, it is doubtful whether it can capture the essence of the huge heterogeneity observed in mood disorders and their dynamic nature with frequent switches and changes in the clinical profile. The greatest disadvantage of both classification systems is that they perform better (and focus) when interepisode remission is present; instead, the everyday real-life patient is more likely to suffer from a chronic disorder with residual and mixed symptoms. The term spectrum was first used in psychiatry in 1968 for the schizophrenia spectrum (Kety et al., 1968).

The proposed mood spectrum models unify categorical classification, which is essential, with a dimensional view, which is true to nature; both are needed and both are empirically testable. Today the term bipolar spectrum is mainly used in 2 complementary senses: (1) a spectrum of severity, which embraces psychotic and nonpsychotic major and minor BDs (including bipolar dysthymia, recurrent brief and minor depressive, cyclothymic disorders, hypomania and, at its broadest, even borderline disorders and cyclothymic temperament; (2) a proportional mood spectrum, which considers the 2 components, mania and depression, on the level of major and minor mood disorders. This proportional model is an extension of Kleist’s concept of BD as a combination of the 2 monopolar disorders of depression and mania (Kleist, 1937). Thus these 2 approaches to spectrum reflect 2 distinct continua: from normal to pathological and from unipolar to bipolar.

An important part of the bipolar spectrum is cyclothymic disorder, which is considered to be an attenuated form of BD. Their behavior is characterized by the alternation of extremes (Akiskal et al., 1977). Depending on the threshold of traits used in determining the presence of hyperthymia, cyclothymic patients may constitute 10 to 20% of those with major depressive disorder. Also, cyclothymia is often a prodromal of BD (Akiskal et al., 1979). Another important part of the bipolar spectrum are those patients who experience an antidepressant-induced switch.
Thus, many patients with so-called unipolar depression are actually pseudounipolar.

Some authors suggest that a significant part of the literature consists mostly of expert opinion overemphasizing various links between bipolar and unipolar mood disorders and personality disorders (Paris et al., 2007; Patten and Paris, 2008). Recently, the first solid international epidemiological data in support of the bipolar spectrum have been published (Merikangas et al., 2007, 2011; Angst et al., 2010). According to these authors there is a direct association between increasingly restrictive definitions of BD and indicators of clinical severity, including symptom severity, role impairment, comorbidity, suicidality, and treatment. For example, the proportion of mood episodes rated as clinically severe increased from 42.5% for subthreshold BD to 68.8% for BD-II to 74.5% for BD-I. However, since clinical diagnosis and severity share confounding factors and definitions overlap, it is also important to note that these studies also showed that the proportion of cases reporting severe role impairment ranged from 46.3% for subthreshold BD to 57.1% for BD-I (Merikangas et al., 2011).

On the basis of both epidemiological data and clinical wisdom, a limited number of models reflecting the structure of the bipolar spectrum have been proposed. The first effort was a dimensional concept (from normal to pathological) proposed by Kretschmer in 1921 for schizophrenia (schizothymic-schizoid-schizophrenic) and for affective disorders (cyclothymic temperament-cycloid ‘psychopathy’-manic-depressive disorder). Bleuler suggested a similar concept in 1922. In 1977 Akiskal proposed a cyclothymic-bipolar spectrum (Akiskal et al., 1977). A simple model system was introduced in 1978 by Jules Angst (1978, Angst et al., 1978), who used the the following codes: M for severe mania, D for severe depression (unipolar depression), m for less severe mania (hypomania), and d for less severe depression. In 1981 Gerald Klerman suggested a manic spectrum (Klerman, 1981, 1987) and in the late 1990s Akiskal proposed 6 subtypes, some of which are further subdivided according to their unique clinical features. A summary of his proposed subtype schema is as follows (Akiskal and Pinto, 1999; Akiskal and Benazzi, 2005; Ng et al., 2007; Fountoulakis, 2008).

Epidemiology

In the last few decades there has been an increasing interest in psychiatric epidemiology. For BD, a point that plays a major role in the estimation of the prevalence rates is the definition of hypomania and of mixed, irritable, or dysphoric forms of manic episodes. This is further complicated by the presence of inaccurate recall and the low sensitivity of the interview instruments concerning subthreshold symptomatology and nonclassical clinical pictures (Kessler et al., 1997a).

A number of important studies exist and provide important but inconclusive information. The Amish study (Egeland and Hostetter, 1983; Egeland et al., 1983; Hostetter et al., 1983) reported similar prevalence rates between unipolar depression and bipolar illness and also similar rates between genders. It is impressive that 79% of patients with BD-I were previously diagnosed as suffering from schizophrenia. The Epidemiological Catchment Area study (ECA) (Eaton et al., 1981; Regier et al., 1984, 1988, 1993; Bourdon et al., 1992) reported a lifetime prevalence of 0.8% for BD-I (0.3–1.2%) and an annual prevalence of 0.6% (0.2–1%) with similar prevalence for males and females. The annual incidence was 0.4% (0.1–0.6%) of cases, which corresponds to approximately 3.2 (0.8–4.8) per 100,000 residents. The median age at onset was 18 years. A reanalysis of the ECA data with the addition of subthreshold bipolarity produced a total lifetime prevalence of 6.4% with 0.5% being a lifetime prevalence of BD-II (Judd and Akiskal, 2003). The National Comorbidity Survey (NCS) (Kessler et al., 1993, 1994a, 1994b, 1995, 1996, 1997b; Blazer et al., 1994, Wittchen et al., 1994; Warner et al., 1995; Kendler et al., 1996; Magee et al., 1996) reported a lifetime prevalence of 1.7% for BD-I and an annual prevalence of 1.3% with similar prevalence for males and females. The median age at onset was 21 years. The NCS-R (Kessler et al., 2004, 2005a, 2005b, 2012a, 2012b; Kessler and Merikangas, 2004; Merikangas et al., 2007; Angst et al., 2010; Nierenberg et al., 2010) reported a lifetime prevalence of 1.0% for BD-I and an annual prevalence of 0.6% with again similar prevalence for males and females. The median age at onset was 19 years. For BD-II the lifetime prevalence was 1.1% and the annual prevalence was 0.8% with similar prevalence for males and females. The median age at onset was 20 years. There was a small difference between males and females in the BD-II rates, with female rates being slightly higher. The Cross National Collaboration Group study included data from 7 countries (US, Canada, Puerto Rico, Germany, Taiwan, South Korea and New Zealand) (Weissman et al., 1996) and reported variable rates for different countries, but overall the rates seemed moderately consistent cross-nationally. The Zurich study (Angst et al., 1984, 2005b; Wicki and Angst, 1991) reported an annual prevalence of BD-I of 0.7% and a lifetime prevalence for the bipolar spectrum of 5.5%. The Nottingham study (Brewin et al., 1997) reported the 2-year incidence rate reported for BD was 0.005%, which corresponds to an annual incidence of 2.5/100,000. The Netherlands study (Bijl et al., 2002; Regeer et al., 2002; ten Have et al., 2002) suggested a lifetime prevalence of BD equal to 2.0%. The annual incidence was equal to 2.7/100,000. There was no significant difference between males and females. The Australian National Survey reported the year prevalence of euphoric BD (combined BD-I and BD-II) was 0.5% (Mitchell et al., 2004). The Butajira study from Ethiopia reported a lifetime prevalence of BD-I disorder of 0.5%, with the rate being 0.6% for males and 0.3% for females. The mean age of cases was 29.5 years, with no significant sex difference. The mean age of first recognition of illness was 22 years. There was no significant sex difference in the age at onset of manic or depressive phases (Negash et al., 2005). A more recent cross-sectional, face-to-face, household survey in 11 countries in the Americas, Europe, and Asia reported that the lifetime prevalence of BD was 0.6% for BD-I and 0.4% for BD-II, while the year prevalence was 0.4% and 0.3%, respectively (Merikangas et al., 2011). A few studies report on the epidemiology of bipolar spectrum and suggest that in the adult population the lifetime prevalence of the bipolar spectrum is between 3 and 8.3% (Weissman and Myers, 1978; Angst et al., 1984, 2005b; Oliver and Simmons, 1985; Wicki and Angst, 1991; Heun and Maier, 1993; Angst, 1998; Szadoczky et al., 1998; Hirschfeld et al., 2003a, 2003b; Judd and Akiskal, 2003; Moreno and Andrade, 2005; Faravelli et al., 2006; Kessler et al., 2006).

Overall and according to the WHO, BD affected an estimated 29.5 million persons worldwide in 2004 (WHO, 2008). The available data suggest that the life prevalence of BD-I is around 1%, with probably a similar rate concerning BD-II. The full bipolar spectrum probably has lifetime prevalence around 5%. There are no striking differences between genders. However these figures should be considered as only indicative, since important discrepancies exist among studies and countries, as mentioned above. The rather small difference between annual and lifetime rates suggests that BD is both an episodic but also a chronic mental disorder with high recurrence rates.
The various studies from around the world suggest that the age at onset is late adolescence or early adulthood, around the age of 18 to 20 years, but also they suggest that approximately one-fourth of BD patients have the onset before the age of 13 (Perlis et al., 2004; Post et al., 2008; Stringaris et al., 2010; Merikangas et al., 2012), and among other things this suggests caution in the use of stimulants for the treatment of children with ADHD and worse overall outcome (Agnew-Blais and Danese, 2016).

Staging

After the introduction of operationalized diagnostic criteria for all contemporary classification systems, the need to define and rate seriousness, progression, changes in physiology, and damage made and the extent and the specific characteristics of the disease emerged. Staging is the term that defines this procedure (Fountoulakis, 2015k). The field in medicine where staging is most successful and enjoys great importance is that of clinical oncology. Since 1993 there were many attempts to arrive at a staging model for psychiatry (Fava and Kellner, 1993; Yung and McGorry, 1996, 2007; McGorry et al., 2006, 2007, 2010; McGorry, 2007, 2010b; Vieta et al., 2011; Cosci and Fava, 2013). The concept of staging if and when applied has a number of implications. Almost by definition it suggests that early stages are easier to treat, while later stages are rather refractory to treatment. Thus these later stages might need the application of treatment options with more adverse events, higher risk, and less overall benefit (Post et al., 2010) or some kind of palliative care should be considered.

The earliest research contribution to the effort of staging BD was the description of the stages of mania in the early 1970s when Carlson and Goodwin not only described discrete stages in the development and course of acute mania, but also they described a rollback phenomenon that is the clinical condition improves by manifesting the same stages but at a reverse order (Carlson and Goodwin, 1973). Up to n, 5 major staging models have been proposed for BD (Berk et al., 2007a, 2007b; Kapczinski et al., 2009; Post, 2010; Post et al., 2012; Cosci and Fava, 2013; Frank et al., 2014). Although there is some support for the proposed staging models, the research base is thin, the heterogeneity of the data is significant, and the studies include small sample sizes. A number of vicious cycles could be in place. Most of the data are cross-sectional (Kapczinski et al., 2014), and the need for a transdiagnostic and longitudinal research approach is prominent (Lin et al., 2013).

The data so far support the presence of an asymptomatic at-risk phase and a nonspecific prodromal phase. This prodromal phase seems to be common for a number of mental disorders, and prediction is extremely difficult on the basis of current knowledge. The literature is also supportive of the presence of an early stage of the full-blown illness, during which the episodes are well defined and there are no or very few inter-episode residual symptoms, good response to treatment, and little disability. It also supports the presence of a late stage that is associated with a more chronic and refractory disease, probably with depressive predominant polarity, psychotic features, and significant disability. It is disappointing that there is little research on the treatment effect at late stages (Berk et al., 2012), with only a few exceptions (Torrent et al., 2013). The use of biomarkers might, in the near future, facilitate the validation of staging systems and their therapeutic utility (Vieta, 2015).

Therapeutic Issues

The treatment of BD is complex (Fountoulakis, 2008) and for several decades the treatment of BD was theoretically based on the concept of mood stabilizers. This term was originally used during the 1950s to refer to a combination of amphetamine and a barbiturate to treat patients with neurotic instability but not patients with BD. The term mood stabilizer was proposed by Mogens Schou for lithium (Schou, 1963), but eventually the stabilizer concept prevailed, probably because the focus of research with lithium was on the long-term prophylaxis.

During the last decade, however, there was a plethora of data, mainly because of the introduction of atypical antipsychotics as possible treatment options. However, this gave the chance also for older substances to be tested under rigorously defined research conditions. These studies revealed that the treatment could be more complex than previously believed and several issues exist. The clinician should be aware of many specific indications, contraindications, details, and traps (Fountoulakis et al., 2005, 2007a, 2007b, 2008, 2015m; Gonda et al., 2009).

The concept of mood stabilizers is disputed, since the data do not support an equally wide efficacy for different compounds like lithium, valproate, or carbamazepine to warrant such a label. On the contrary, there are negative data concerning specific areas, while our knowledge is quite limited concerning other areas. One particular problem, which only recently has been acknowledged, is that probably some facets of the disorder are refractory to treatment. Another important problem is that not only is the evidence limited concerning the treatment of specific facets and issues of BD (Fountoulakis et al., 2010; Fountoulakis et al., 2012, 2013), but also continued scientific training and reading is inadequate. Thus, research findings are not making it to everyday clinical practice. Focused educational intervention might be necessary to change this attitude. Part of this problem is reflected in the common practice among clinicians to use medication on the basis of a class effect. This means that they consider that a whole class of medications possesses a specific action. This class effect is often considered in combination with a syndromal approach, which means that irrespective of the nosological entity, a specific kind of symptoms respond to a specific class of medication.

For example, according to this combined approach all antipsychotics are equally effective against psychotic symptoms irrespective of disorder diagnosis, and the same holds for all antidepressants against depressive symptoms. This is the most commonly used approach in everyday clinical practice and has a huge impact on public mental health. Its significant advantage is that it provides the clinician with fast and simple rules to determine treatment. On the other hand, its greatest problem is that this approach has been proven false, especially in the case of BD where it is specifically combined with a very broad mood stabilizers concept (Fountoulakis et al., 2011). The extent to which this truly influences the everyday clinical practice worldwide is unknown but is probably significant. The extent to which this concept influences the outcome of BD is similarly unknown, although theoretically a more evidence-based approach should improve the overall outcome of BD patients.

It is important to note that with the introduction of the second-generation antipsychotics, antipsychotics became a cornerstone for the treatment of BP also according to treatment guidelines. On the contrary, a number of studies showed that the usefulness of antidepressants that were traditionally seen in Europe as a meaningful treatment option for bipolar depression is questionable (Pacchiarotti et al., 2013b). Additionally, the maintenance/long-term treatment became more complex, since it has proven that agents previously considered to be mood stabilizers were essentially more effective for one pole than the other (Popovic et al., 2012). This is definitely a fast-moving field,
and it is certainly difficult for a clinician to follow new findings and incorporate them into his or her everyday clinical practice.

On the other hand, the data on the usefulness of psychosocial interventions are limited, and their value against specific symptoms and problems remains unknown (Fountoulakis et al., 2009; Reinares et al., 2014).

One very special issue is agitation and its treatment. There is a significant number of published papers on the pharmacological (Citrome, 2004; Battaglia, 2005; Nordstrom and Allen, 2007; Nordstrom et al., 2012) but also on the nonpharmacological treatment of agitation (Marder, 2006; Amann et al., 2013), while recently a consensus paper on how to treat agitation in BD patients has been published (Garriga et al., 2016).

**Special Issues**

Since BD is characterized by phases that respond to a completely different way to treatment and comorbidity (Fountoulakis, 2015f; Vrublevska and Fountoulakis, 2015), it is relatively easy to define acute either manic/hypomanic or depressive episodes. However the terms continuation and maintenance are often interchangeably used in the terminology of RCTs and thus create significant confusion (Frank et al., 1991; Gaemi et al., 2004). Continuation treatment lasts up to 12 months, and the duration depends on an estimate of when the episode would have remitted spontaneously. On the other hand, maintenance treatment starts after remission and thus after continuation and covers several years. Although a strict definition demands at least 2 months of sustained recovery for the patient to be considered in remission (American Psychiatric Association, 2000), the reality is that only a minority of patients in RCTs achieve complete remission. This makes the use of terms (relapse vs recurrence and continuation vs maintenance) problematic. In the nomenclature of RCTs, the terms relapse and maintenance are preferred. The FDA policy is to accept data, based on patients in remission for <2 months, thus adding to the continuation vs maintenance confusion of definitions (Calabrese et al., 2006).

The term relapse is also problematic in BD. A narrow definition suggests that relapses are of the same polarity with the index episode, and they tend to occur within the first months of improvement. However with a polymorphic disease like BD, it might be inappropriate not to include in relapses the early emergence of an episode of the opposite pole. It is important to note that licensing authorities accept the latter approach.

The acute episode after which BD patients are enrolled into maintenance trials is called the index episode. To date, most maintenance trials follow an enriched design, that is, only patients who have remitted under the investigation agent during the acute phase are enrolled into the double blind maintenance phase. This design has interesting consequences, since it biases the sample both towards a specific predominant polarity and also towards a favorable response to the specific agent (Cipriani et al., 2014). These 2 comments constitute important limitations in the generalizability of the results and make very difficult the translation of research findings into the everyday clinical practice in the case of patients who, rather than continued on the same medication, are switched to another one during the maintenance phase (Grande et al., 2014).

**Economic Considerations**

It is very difficult to calculate the true economic cost of a polymorphic disorder like BD. The cost includes direct spending due to hospitalizations and medication, cost of supporting infrastructure of the various National Health Systems, somatic comorbidity, indirect and out-of-pocket costs, as well as the absenteeism from work and premature death (Fountoulakis, 2015g).

For the UK the total cost has been estimated to be £2.055 billion in 1999/2000 prices (Das Gupta and Guest, 2002). It is interesting that 86% of this cost was the result of productivity loss and unemployment, while only 10% was cost related with NHS services. Medication costs in primary care were approximately £9.5 million, corresponding to 0.4% of total cost and 4.3% of NHS cost. A more recent study showed that the NHS cost has been doubled, with medication costs rising disproportionally and reaching £25.2 million, that is 7.4% of NHS cost (Young et al., 2011). In the US, the cost of medication was rather very low during the 1990s and reached 2% of the total cost after 2000, but the exact figure is unknown (Wyatt and Henter, 1995; Begley et al., 2001; McCrone et al., 2008; Dilsaver, 2011). In Germany the total annual cost was calculated to be 5.8 million euros, with 98% being due to productivity loss (Runge and Grunze, 2004). Similar estimations come from other areas of the world, with the calculations basing on different prevalence rates and health systems and societal structures (Hakkaarten Roijen et al., 2004; Fisher et al., 2007; Ekman et al., 2013).

It is clear that the cost of medication treatment constitutes a very small percentage of the total cost of BD (Hidalgo-Mazzei et al., 2015). Medication treatment is, however, the intervention with the greatest impact on the course of the illness and the intervention that makes possible other actions to exist by resolving acute episodes in a reliable way. Furthermore, it decreases the long-term impairment and improves insight and collaboration by the side of the patient. Nevertheless, it also seems clear that medication cost is disproportionally rising, at least in some places of the world and for periods of time, and this constitutes an additional factor of concern. One should be very careful, because a small reduction in medication costs as a consequence of giving priority to cheaper agents and disregarding clinical data could easily result in a significant and disproportional increase in the total cost of the disease.

The CINP workgroup decided not to take medication cost or availability of medication into consideration. It chose to rely exclusively on clinical data, leaving the cost and availability issues to local and national groups who would like to implement the CINP guidelines in a specific country or region and would be obliged to take into consideration also the local socio-political and economic environment.

**Methodology**

The workgroup decided after consensus to follow the following methodology for the development of the treatment algorithm with the steps listed below:

a) Defining the sources of data and choosing which to use  
b) Development of a grading method  
c) Search of the literature  
d) Grading of the data  
e) Defining the clinical parameters to take into consideration  
f) Development of a precise treatment algorithm  
g) Development of the clinical guideline

**Defining the Sources of Data and Choosing Whom to Use**

**Randomized Controlled Trials (RCTs)**

This type of study constitutes the main source of evidence. Without them it is impossible to say whether an agent or
method possesses efficacy or not, since it is impossible to control for confounding variables any other way. Randomization of patients to parallel treatment arms, including placebo, allows one to attribute confidently observed differences in efficacy between these arms to the effects of the treatments (McAlister et al., 1999a, 1999b; Pocock and Elbourne, 2000).

However, with BD there is an important problem. Because of ethical, practical, and most often economic limitations, it is not always possible to apply the RCT method across all facets and issues of BD. For many of them, data are available only on the basis of post hoc analyses or secondary outcomes. Another major limitation is that this kind of study is very expensive, and thus most of them are industry sponsored with the objective to control for confounding variables any other way. Randomization of methods possesses efficacy or not, since it is impossible to obtain the label for the specific product. Although such trials follow the regulatory agencies’ design, they have limitations on generalizability. Also it is well known that only a small minority of highly selected patients is eligible to enter these studies, and thus the generalizability of results is problematic. The study duration is often relatively short and this is true also for maintenance trials, in part because the existence of a placebo arm carries a high attrition rate.

An important pitfall concerns the actual results of the RCT, which often are different from those published. It is not unusual that when a trial is negative on the basis of its primary outcome, a publication is done on the basis of positive secondary outcomes. This is essentially misleading, but fortunately it is a phenomenon that has been less frequent during the recent years.

Meta-Analysis
Meta-analysis is a technique that combines data (not simply pooled) from several trials and returns a specific quantitative answer to a specific question that usually is which treatment is superior in comparison with others or placebo. Sometimes but not always it also provides an absolute estimate of the treatment effect size.

There are a number of significant limitations for the meta-analytic methods (Huf et al., 2011a, 2011b). There is a need for the studies included in the meta-analysis to be similar in design and with sufficient information being available. Meta-analytic studies often violate this rule and include a diverse group of trials in the analysis (e.g., studies of monotherapy and combination treatment, fixed and flexible dosage studies, etc.) with unknown consequences (Fountoulakis et al., 2014).

Common problems of meta-analyses include small sample sizes, inadequate power, study heterogeneity, lack of extractable data, lack of interchangeable measurement instruments and definitions of outcomes, and other differences in the design of studies whose data are utilized. Negative trials are often not published and this poses an important limitation to the meta-analytic approach. Today the trials sponsored by official foundations can be traced in trial repositories. However, their detailed results are unlikely to be retrieved and even if they are retrieved, they have not undergone the essential peer review process (which adds credibility) like those published, and their quality could be questionable.

The question whether it is appropriate to use data from the largest possible number of disparate studies vs the need for including data only from essentially identical studies is a matter of debate and has also been discussed specifically concerning acute mania trials where these different approaches gave conflicting results (Yildiz et al., 2010; Cipriani et al., 2011). Practically, all meta-analytical studies utilize compromises to deal with the above problems and limitations. These compromises might have profound effects on the validity and generalizability of their results (Noble, 2006; Mismetti et al., 2007; Huf et al., 2011b).

Some authors consider meta-analysis to be on the top of the evidence-based pyramid of data sources. This approach suggests that its results are superior to the results of the RCTs, and subsequently it is meant that a positive meta-analysis is superior to a number of negative RCTs even in the case of the absence of any positive RCTs. However, the authors of the current paper consider that in most cases meta-analysis has a lower evidence level than RCTs and therefore graded it below them, primarily because of a significant number of limitations and drawbacks that often make the results of meta-analysis equivocal.

Open Trials
Open trials do not utilize the double blind design and they are not placebo controlled. Therefore they are easier to conduct, their number and size are greater, and the quality of patients enrolled is closer to that seen in the real world. Their great limitation is that their open nature induces significant bias, and thus they are by no means considered to be even close to being the gold standard or a reliable source of evidence data. Their role should be considered complementary. It is not unusual that treatment modalities with many positive open trials fail in RCTs, with topiramate in BD being a striking example (Suppes, 2002).

Review and Opinion Papers
Review and opinion papers mainly constitute educational tools, which attempt to translate the research findings into ready-to-use tools for the everyday clinical practice. They are extremely useful for the average clinician; however, they usually echo the opinion of the author, and thus they might contain significant bias. Their overall reliability and validity is questionable and only a few add significantly to our understanding by critically analyzing the existing data. Their ever-increasing number in the literature might constitute a problem, since they often obscure research findings by reproducing widely established biases and misconceptions. This is an important problem especially in the field of BD treatment.

Sources to Include
The authors decided by consensus to include only RCTs and meta-analyses in the development of the current treatment algorithm, since they have the highest validity for judgment. The authors reserved the privilege to judge and use the second and third source on an individual basis and according to their research and clinical experience for the latter steps of the algorithm where a Delphi method to arrive at decisions was utilized.

Development of a Grading Method
The authors decided to develop a grading method for the evaluation of available data concerning the treatment of BD. Such methods have existed since the early 1980s (Fletcher and Spitzer, 1980), but the de novo development of such a method was judged to be absolutely necessary, because the existing grading methods were not sufficiently appropriate for use in this particular set of data. In the frame of this process, the most widely accepted grading methods were studied, and their advantages and disadvantages were identified and taken into consideration in relationship to the specific needs of the current study. All grading methods include a method to assess the quality of data and a method to arrive at recommendations on the basis of the extent to which we can be confident that the desirable effects of an intervention outweigh the undesirable effects. The values
and preferences factor as well, but the cost was not taken into consideration by the workgroup.

Starting in 1992, 5 steps were developed to summarize the process of individual-level decision making and they were published in 2005 (Dawes et al., 2005). They include:

a. The formulation of a precise and answerable question and avoiding uncertainty and vague statements (Richardson et al., 1995; Schlosser et al., 2007).

b. The performance of a systematic search and retrieval of the evidence available (Rosenberg et al., 1998).

c. The critical review and classification of the retrieved evidence with the recognition of the presence of systematic errors, various types of bias, confounders, reliability and validity issues, etc. The clinical significance and the generalizability of the results should also be taken into account (Parkes et al., 2001; Horsley et al., 2011).

d. Application of results in practice.

e. Evaluation of performance (Jamtvedt et al., 2003, 2006a, 2006b; Ivers et al., 2012).

It is important to assess the quality of the evidence that comes from the sources described above. The quality assessment is based on the strength of their freedom from the various biases that beset medical research. In this frame, triple-blind, placebo-controlled trials with allocation concealment and complete follow-up involving a homogeneous patient population and medical condition should be considered to constitute the highest grade, while case reports should be considered to constitute the lowest grade. Expert opinion should not be considered to be a source of evidence, although it could be a valuable tool for the development of guidelines (Tonelli, 1999).

Until recently there were a number of grading systems for assessing the quality of evidence that were developed by different organizations. One of them is the U.S. Preventive Services Task Force (U.S. Preventive Services Task Force, 1989; Sherman et al., 2011) and another system is the Oxford (UK) Center for Evidence Based Medicine Levels of Evidence, which also is useful for the grading of diagnostic tests, prognostic markers, or harm (Oxford (UK) Center for Evidence Based Medicine Levels of Evidence Working Group) and constituted the basis for the use of the BCLC staging system for diagnosing and monitoring hepatocellular carcinoma in Canada (Paul et al., 2012). Another method to grade data is the Patient Outcomes Research Team (PORT) method (Lehman and Steinwachs, 1998), which has been used by the World Federation of Societies of Biological Psychiatry for the development of the WFSBP guidelines (Grunze et al., 2002, 2003, 2004). In 1992 the Agency for Health Care Policy and Research and the National Institute of Mental Health established a PORT for Schizophrenia at the University of Maryland School of Medicine and the Johns Hopkins University School of Public Health. The PORT investigators adopted the criteria on levels of evidence used for development of the Agency for Health Care Policy and Research Depression Guidelines.

The most detailed and precise modern method seems to be the GRADE method (short for Grading of Recommendations Assessment, Development and Evaluation) for the development of guidelines (Guyatt et al., 2008b; Jaeschke et al., 2008), which clearly separates quality of evidence from level of recommendation and suggests it is necessary to include a clear question that should include all 4 components of clinical management (patients, an intervention, a comparison, and the outcomes of interest) (Oxman and Guyatt, 1988) and to grade the outcomes into those who are critical for the decision making and those who are not (Schunemann et al., 2006). In this frame, the assessment of the quality of evidence is important, since it reflects the confidence whether the effect is adequate to support recommendations. The determinants of quality that are study limitations, inconsistency of results, indirectness of evidence, imprecision, and reporting bias (Guyatt et al., 2011a, 2011b, 2011c, 2011d, 2011e, 2013). There is some option to upgrade the quality when the effect size is very high (Guyatt et al., 2011f). The GRADE method provides guidance to grade the data from a variety of sources (Guyatt et al., 2008a), but it is not sensitive for datasets that focus solely on RCTs like the dataset of the current workgroup. According to the GRADE grading system, all the data included in the current effort to develop guidelines are of high quality. From the limitations recognized by the GRADE (lack of allocation concealment, lack of blinding, large losses to follow-up, failure to adhere to an intention to treat analysis, and stopping early for benefit or failure to report outcomes), only large losses to follow-up and stopping early for benefit or failure to report outcomes could be applicable to the current study. A comparison of all the grading methods is shown in Table 3.

### Table 3. Comparative Presentation of Different Grading Methods

<table>
<thead>
<tr>
<th>USPSTF</th>
<th>OCEBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level I: Evidence obtained from at least one properly designed randomized controlled trial. Level II-1: Evidence obtained from well-designed controlled trials without randomization. Level II-2: Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group. Level II-3: Evidence obtained from multiple time series designs with or without the intervention. Dramatic results in uncontrolled trials might also be regarded as this type of evidence. Level III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.</td>
<td>Systematic review of randomized trials or n-of-1 trials Randomized trial or observational study with dramatic effect Nonrandomized controlled cohort/ follow-up study Case-series, case-control studies, or historically controlled studies</td>
</tr>
<tr>
<td></td>
<td>High quality Medium quality Low quality Very low quality</td>
</tr>
<tr>
<td>PORT</td>
<td></td>
</tr>
<tr>
<td>Level A: Good research-based evidence, with some expert opinion, to support the recommendation</td>
<td>Level B: Fair research-based evidence, with substantial expert opinion, to support the recommendation</td>
</tr>
</tbody>
</table>
The recommendation methods constitute a step forward and are determined by the balance of risk vs benefit of the intervention and the level of evidence on which this information is based. A comparison of the recommendation methods of the U.S. Preventive Services Task Force uses (Sherman et al., 2011) that utilizes a 5-levels system and the GRADE system that has only 2 categories concerning recommendations and characterizes them as strong (conditional) and weak (discretionary) (Guyatt et al., 2008b, 2008c) and also considers cost (Brunetti et al., 2013) is shown in Table 4.

As defined previously, only RCTs were taken into consideration, a fact that puts all the data at the highest grading according to all systems. However, the workgroup was concerned about a number of issues, including inconsistency of results between RCTs, conflicting results between RCTs and meta-analyses, issues explored only on the basis of secondary outcomes, etc. After recognizing all these sources of problematic quality, 32 individual scenarios were identified and are listed in Table 5. Afterwards they were ranked after consensus and grouped into levels. Two solutions were proposed. The ranking the 4- and 5-levels solution, and the final grading system are shown in Table 6. The description of the grading and the recommendation systems are shown in Table 7.

At this point it is important to note that the absence of evidence is not identical with the presence of negative data.

### Table 4. Comparative Presentation of Recommendation Methods

<table>
<thead>
<tr>
<th>USPSTF</th>
<th>GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level A: Good scientific evidence suggests that the benefits of the clinical service substantially outweigh the potential risks.</td>
<td>Strong</td>
</tr>
<tr>
<td>Level B: At least fair scientific evidence suggests that the benefits of the clinical service outweighs the potential risks.</td>
<td>Strong</td>
</tr>
<tr>
<td>Level C: At least fair scientific evidence suggests that there are benefits provided by the clinical service, but the balance between benefits and risks are too close for making general recommendations.</td>
<td>Weak</td>
</tr>
<tr>
<td>Level D: At least fair scientific evidence suggests that the risks of the clinical service outweighs potential benefits.</td>
<td>Weak</td>
</tr>
<tr>
<td>Level I: Scientific evidence is lacking, of poor quality, or conflicting, such that the risk versus benefit balance cannot be assessed.</td>
<td>Weak</td>
</tr>
</tbody>
</table>

### Table 5. The 32 Different Scenarios That Were Identified, Listed, and Graded

#### Primary Outcome Scenarios
1. At least 1 positive 2-active arm RCTs vs placebo exist, plus positive 1 active arm RCTs. No negative RCTs
2. At least 2 positive RCTs vs placebo exist. No negative RCTs
3. One positive RCT vs placebo exists. No negative RCTs
4. Some positive plus some negative RCTs vs placebo. Positive all meta-analyses
5. Some positive plus some negative RCTs vs placebo. Mixed results from meta-analyses
6. Some positive plus some negative RCTs vs placebo. Negative all meta-analyses
7. More positive but some negative RCTs vs placebo. Positive all meta-analyses
8. More positive but some negative RCTs vs placebo. Mixed results from meta-analyses
9. More positive but some negative RCTs vs placebo. Negative all meta-analyses
10. More negative but some positive RCTs vs placebo. Positive all meta-analyses
11. More negative but some positive RCTs vs placebo. Mixed results from meta-analyses
12. More negative but some positive RCTs vs placebo. Negative all meta-analyses
13. Only 1 negative trial exists vs placebo
14. Only negative trials exist vs placebo. Meta analyses all negative
15. Only negative trials exist vs placebo. Meta analyses all positive
16. Only negative trials exist vs placebo. Meta analyses mixed

#### Posthoc scenarios
17. Only 1 positive from posthoc analyses vs placebo
18. At least 2 positive from posthoc analyses vs placebo
19. Only 1 negative from posthoc analyses vs placebo
20. At least 2 negative from posthoc analyses vs placebo. Positive all meta-analyses
21. At least 2 negative from posthoc analyses vs placebo. Negative all meta-analyses
22. At least 2 negative from posthoc analyses vs placebo. Mixed meta-analyses
23. More negative than positive from posthoc analyses vs placebo. Positive all meta-analyses
24. More negative than positive from posthoc analyses vs placebo. Negative all meta-analyses
25. More negative than positive from posthoc analyses vs placebo. Mixed meta-analyses
26. More positive than negative from posthoc analyses vs placebo. Positive all meta-analyses
27. More positive than negative from posthoc analyses vs placebo. Negative all meta-analyses
28. More positive than negative from posthoc analyses vs placebo. Mixed meta-analyses

#### Other scenarios
29. Only 1 failed trial, no other data
30. At least 2 failed trials, no other data
31. Only prematurely terminated trials
32. Although trials exist, the data are not available in a way to arrive at reliable conclusions

Table 6. The Ranking, the 4- and 5-levels Solution, and the Final Grading System for the 32 Different Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Rank</th>
<th>5-Grade</th>
<th>4-Grade</th>
<th>Grade system</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least 2 positive 2-active arm RCTs vs placebo exist, plus positive 1 active arm RCTs. No negative RCTs</td>
<td>1</td>
<td>A</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>At least 2 positive RCTs vs placebo exist. No negative RCTs</td>
<td>1</td>
<td>A</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>One positive RCT vs placebo exists. No negative RCTs</td>
<td>2</td>
<td>A</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>More positive but some negative RCTs vs placebo. Positive all meta-analyses</td>
<td>2</td>
<td>A</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>Some positive plus some negative RCTs vs placebo. Positive all meta-analyses</td>
<td>3</td>
<td>B</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>More negative but some positive RCTs vs placebo. Positive all meta-analyses</td>
<td>4</td>
<td>B</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>Only negative trials exist vs placebo. Meta analyses all positive</td>
<td>4</td>
<td>B</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>At least 2 positive from posthoc analyses vs placebo</td>
<td>5</td>
<td>B</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>Only 1 positive from posthoc analyses vs placebo.</td>
<td>5</td>
<td>B</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>Some positive plus some negative RCTs vs placebo. Mixed results from meta-analyses</td>
<td>6</td>
<td>C</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>More positive but some negative RCTs vs placebo. Mixed results from meta-analyses</td>
<td>6</td>
<td>C</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>More positive than negative from posthoc analyses vs placebo. Positive all meta-analyses</td>
<td>7</td>
<td>D</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>More negative than negative from posthoc analyses vs placebo. Positive all meta-analyses</td>
<td>7</td>
<td>D</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>At least 2 negative from posthoc analyses vs placebo. Positive all meta-analyses</td>
<td>8</td>
<td>D</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>More positive than negative from posthoc analyses vs placebo. mixed meta-analyses</td>
<td>9</td>
<td>E</td>
<td>D</td>
<td>4</td>
</tr>
<tr>
<td>Only negative trials exist vs placebo. Meta analyses mixed</td>
<td>9</td>
<td>E</td>
<td>D</td>
<td>4</td>
</tr>
<tr>
<td>At least 2 negative from posthoc analyses vs placebo. Mixed meta-analyses</td>
<td>10</td>
<td>E</td>
<td>D</td>
<td>4</td>
</tr>
<tr>
<td>More negative than positive from posthoc analyses vs placebo. Mixed meta-analyses</td>
<td>10</td>
<td>E</td>
<td>D</td>
<td>4</td>
</tr>
<tr>
<td>Some positive plus some negative RCTs vs placebo. Negative all meta-analyses</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>More positive but some negative RCTs vs placebo. Negative all meta-analyses</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>More negative but some positive RCTs vs placebo. Negative all meta-analyses</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>Only 1 negative trial exists vs placebo</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>Only negative trials exist vs placebo. Meta analyses all negative</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>Only 1 negative from posthoc analyses vs placebo.</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>At least 2 negative from posthoc analyses vs placebo. Negative all meta-analyses</td>
<td>5</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>More negative than positive from posthoc analyses vs placebo. Negative all meta-analyses</td>
<td>5</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>More positive than negative from posthoc analyses vs placebo. Negative all meta-analyses</td>
<td>5</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>Only prematurely terminated trials</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>Although trials exist, the data are not available in a way to arrive at reliable conclusions</td>
<td>neg</td>
<td>neg</td>
<td>neg</td>
<td>5</td>
</tr>
<tr>
<td>Only 1 failed trial, no other data</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>At least 2 failed trials, no other data</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td></td>
</tr>
</tbody>
</table>

All treatment agents were graded also in terms of safety and tolerability. All combination options were graded at best with 2, since they put the patient at a higher risk for manifesting adverse events.

Search of the Literature

The workgroup decided that the PRISMA method (Hopewell et al., 2008; Liberati et al., 2009; Moher et al., 2009a, 2009b) should be followed in the search of the literature, which will include 3 kinds of papers:

i. RCTs (placebo controlled as well as clinical trials with an active comparator with the compounds used as monotherapy or add-on therapy).
ii. Posthoc analyses of RCTs
iii. Meta-analyses and review papers
iv. Treatment guidelines papers

The search strategies will include:

1. To locate RCTs, the combination of the words ‘bipolar,’ ‘manic,’ ‘mania,’ ‘manic depression,’ and ‘manic depressive’ and ‘randomized’ will be used.
2. Webpages containing lists of clinical trials will be scanned. These sites include http://clinicaltrials.gov and http://www.clinicalstudyresults.org as well as the official sites of all the pharmaceutical companies with products used for the treatment of BD.
3. Relevant review articles will be scanned and their reference lists will be utilized.
4. The MEDLINE will be searched with the combination of keywords ‘guidelines’ or ‘algorithms’ with ‘mania,’ ‘manic,’ ‘bipolar,’ ‘manic-depressive,’ or ‘manic depression.’
5. The treatment guidelines will also be scanned and their reference lists will be utilized.
6. Only papers in English language will be included.

Additionally, an unstructured search of the literature will be performed concerning the adverse events and other safety issues of treatment options

The workgroup considered the fact that it is difficult to locate unpublished studies, especially old ones, and even more difficult to retrieve their results. Thus it was decided that the focus should be put mainly on published studies which are definitely peer reviewed, are of higher quality, and provide more details than meeting abstracts or report sheets. However, whenever an unpublished trial should be located, it is mentioned in the specific part of the manuscript. The authors decided not to seek for additional information concerning unpublished trials from manufacturers, because this might increase the retrieval bias.

Grading of the Data

The grading of the data will follow their retrieval and will be done according to the method developed and described in the
Defining the Clinical Parameters to Take into Consideration

In the real-world setting, the therapist encounters patients with specific clinical features that often determine the choice of treatment on the basis of clinical experience and wisdom rather than evidence. These features include the so-called core manic and core depressive features, psychotic features, anxiety, the co-occurrence of manic and depressive symptoms in a variety of combinations that often do not correspond to concepts accepted by modern classification systems, agitation, and rapid cycling. It is interesting to address the complete constellation of symptoms instead of a specific group. The problem is that the data often focus on the second rather than the first option. It is important also to consider the predominant polarity and subtype of BD (BD-I vs BD-II), the personal history of the patient, and more specifically previous response or refractoriness to treatment and adverse events (including switch).

The data will be scanned concerning the treatment of all the above conditions and modifiers and relevant conclusions will be made concerning whether they can be used as clinical cues for the selection of appropriate treatment.

Development of a Precise Algorithm

The development of a precise algorithm for experimental reasons will be the first task. This algorithm will be based exclusively on the evidence and will be the next step after the data and the interventions are graded in terms of recommendation. This algorithm will be based on the data in a narrow and strict sense and might provide with very precise but limited treatment options for the everyday clinical practice. There will be no trade between the evidence-based approach and clinical utility; the first will be absolutely dominant. This algorithm will reflect the exact state of the art concerning hard data but will lack any clinical wisdom, and it is expected that its application in everyday clinical practice will be problematic. Therefore it should be considered as experimental, and clinicians who will wish to apply it in their clinical practice should do so by taking into consideration these advantages and disadvantages. The algorithm will be included in the second paper concerning the CINP guidelines for BD, and it will be accompanied by a detailed table with the grading recommendation of all available interventions during all the phases of BD and in relevance with the presence of specific clinical features.

At a later time point a software application will be developed by the CINP to assist with the use of the algorithm.

Development of the Clinical Guideline

The development of the guideline will follow after the data and the interventions have been graded and the precise algorithm has been developed. The guideline will be included in the third paper concerning the CINP guidelines for BD. The workgroup decided after consensus on the following rules for the development of the guidelines:

1. Overall the guideline should be based on existing research hard evidence, but also it should make sense for the everyday clinical practice and should be user friendly. Although their nature will be based on the evidence-based approach, this should not go too far concerning the interpretation of the research findings and the potential clinical implications.

Table 7. Summary of the Method for the Grading of the Data and Recommendation as Decided by the Workgroup on the Basis of Both Efficacy and Safety Tolerability

<table>
<thead>
<tr>
<th>Grading on Basis of Efficacy</th>
<th>Level 1</th>
<th>Good research-based evidence, supported by at least 2 placebo controlled studies of sufficient magnitude and good quality. In case of the presence of negative RCT’s, positive RCT’s should outnumber negative ones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 2</td>
<td>Fair research-based evidence, from one randomised, double-blind placebo controlled trial. Also in case one or more trials exist, however, they fail to fulfil all the criteria above (e.g., very small sample size or no placebo control) as well as in case of positive meta-analysis alone.</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>Some evidence from comparative studies without placebo arm or from posthoc analyses.</td>
</tr>
<tr>
<td></td>
<td>Level 4</td>
<td>Inconclusive data or poor quality of RCT’s</td>
</tr>
<tr>
<td></td>
<td>Level 5</td>
<td>Negative data</td>
</tr>
<tr>
<td>Grading on the basis of safety and tolerability</td>
<td>Level 1</td>
<td>Very good tolerability, few side effects which are not enduring, they do not cause significant distress and are not life-threatening and they do not compromise the overall somatic health of the patient</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>Moderate tolerability, many side effects which could be enduring, and cause significant distress but they are not life-threatening although they could compromise the overall somatic health of the patient. Agents with very good overall tolerability but with rare life-threatening adverse events, could be classified here only if the lethality risk can be essentially considered to be negligible with the application of procedures and protocols (e.g., laboratory testing, titration schedules, etc.)</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>Poor tolerability, many side effects which are enduring, cause significant distress, compromise the overall somatic health of the patient or are life-threatening. Agents with moderate overall tolerability and rare life-threatening adverse events should be classified here even in cases the lethality risk can be essentially considered to be negligible with the application of procedures and protocols (e.g., laboratory testing, titration schedules, etc.)</td>
</tr>
</tbody>
</table>

Recommendations for treatment (combination of efficacy and safety/tolerability)

<table>
<thead>
<tr>
<th>Level</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Level 1 or 2 for efficacy and 1 or 2 for safety/tolerability</td>
</tr>
<tr>
<td>Level 2</td>
<td>Level 1 or 2 for efficacy and 2 for safety/tolerability</td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 3 for efficacy and 1 or 2 for safety/tolerability</td>
</tr>
<tr>
<td>Level 4</td>
<td>Level 4 for efficacy or 3 for safety/tolerability</td>
</tr>
<tr>
<td>Level 5</td>
<td>Level 5 for efficacy (not recommended)</td>
</tr>
</tbody>
</table>
ii. Agents and treatment modalities with proven efficacy across all 3 phases of the illness (acute mania, acute bipolar depression, and maintenance phase concerning the prevention of both manic and depressive episodes) should be given priority.

iii. No economic and availability issues will be taken into consideration. National bodies that might wish to utilize the CINP guidelines could add such analyses tailored to the specific country or region.

Discussion

The current paper sets the frame for the development of the CINP treatment guidelines for BD. It contains all the background information, including important clinical features, staging methods, and important treatment issues and details. It also elaborates on the methodology to be used and describes the development of a grading system that will be suitable for use with the kind of data under consideration.

The overall aim of the workgroup was to push guidelines one step further by evaluating the available data in depth and also by identifying clinical issues that need specific interventions that could be supported by the data. A significant contribution is expected to be the precise experimental algorithm that will constitute an option for further study.

Acknowledgment

The authors thank Professor Guy Goodwin for his valuable input in the authoring of this manuscript.

Statement of Interest

K.N.F. has received grants and served as consultant, advisor, or CME speaker for the following entities: AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Ferrer, Gedeon Richter, Janssen, Lundbeck, Otsuka, Pfizer, the Pfizer Foundation, Sanofi-Aventis, Servier, Shire, and others.

E.V. has received grants and served as consultant, advisor, or CME speaker for the following entities: Allergan, AstraZeneca, Bristol-Myers Squibb, Dainippon Sumitomo Pharma, Ferrer, Forest Research Institute, Gedeon Richter, Glaxo-Smith-Kline, Janssen, Lilly, Lundbeck, Otsuka, Pfizer, Roche, Sanofi-Aventis, Servier, Shire, Sunovion, Takeda, the Brain and Behaviour Foundation, the Spanish Ministry of Science and Innovation (CIBERSAM), the Seventh European Framework Programme (ENBREC), and the Stanley Medical Research Institute. A.H.Y. is employed by King’s College London; is Honorary Consultant SLaM (NHS UK); has paid lectures by and participated in advisory boards for all major pharmaceutical companies with drugs used in affective and related disorders; and has no share holdings in pharmaceutical companies. He was lead Investigator for Embolden Study (AZ), BCI Neuroplasticity study, and Aripiprazole Mania Study; investigator initiated studies from AZ, Eli Lilly, Lundbeck, and Wyeth; and has received grant funding (past and present) from: NIHR-BRC (UK); NIMH (USA); CIHR (Canada); NARSAD (USA); Stanley Medical Research Institute (USA); MRC (UK); Wellcome Trust (UK); Royal College of Physicians (Edinburgh); BMA (UK); UBC-VGH Foundation (Canada); WEDC (Canada); CCS Depression Research Fund (Canada); MSFHR (Canada); and NIHR (UK).

H.G. within the last 3 years received grant/research support from: NIHR UK, MRC UK, NTW, and NHS Foundation Trust; receipt of honoraria or consultation fees from: Gedeon-Richter, Lundbeck, and Hofmann-LaRoche; and participated in a company-sponsored speaker’s bureau at BMS, Ferrer, Janssen-Cilag, Otsuka, Lundbeck, and Pfizer.

L.Y. has been on speaker/advisory boards for, or has received research grants from Allermes, Allergan, AstraZeneca, Bristol Myers Squibb, CANMAT, CIHR, Eli Lilly, Forest, GlaxoSmithKline, Intas, Janssen, the Michael Smith Foundation for Health Research, Pfizer, Servier, Sumitomo Dainippon, Sunovion, and the Stanley Foundation.

S.K. within the last 3 years received grants/research support, consulting fees, and honoraria from Angelini, AOP Orphan Pharmaceuticals AG, AstraZeneca, Eli Lilly, Janssen, KKRA-Pharma, Lundbeck, Neuraxpharm, Pfizer, Pierre Fabre, Schwabe, and Servier.

H.J.M. received honoraria for lectures or advisory activities or received grants by the following pharmaceutical companies: Lundbeck, Servier, Schwabe, and Buyer. He was president or in the executive board of the following organizations: CINP, ECNP, WFSBP, EPA, and chairman of the WPA-section on Pharmacopsychiatry.

P.B. has received research grants, honoraria for participation in advisory boards, and/or gave presentations from Allergan, AstraZeneca, Bristol Myers Squibb, Canadian Institute for Health Research, Eli Lilly, Lundbeck, Janssen, Ontario Brain Institute, Meda-Valeant, Merck, Otsuka, Pierre Fabre Medicaments, Pfizer, Shire, Sunovion, and Takeda.

References


Leonard K (1957b) [Cycloid psychoses, often erroneously considered as schizophrenia]. Psychiatr Neurol Med Psychol (Leipzig) 9:359–365.


and study population characteristics. Arch Gen Psychiatry 41:934–941.


