Neurosci Bull October 1, 2015, 31(5): 572–588. http://www.neurosci.cn
DOI: 10.1007/s12264-014-1534-0

·Review·

Important clinical features of atypical antipsychotics in acute bipolar depression that inform routine clinical care: a review of pivotal studies with number needed to treat

Keming Gao[1], Chengmei Yuan[2], Renrong Wu[3], Jun Chen[2], Zuowei Wang[4], Yiru Fang[2], Joseph R. Calabrese[1]

[1]*Department of Psychiatry, Case Western Reserve University School of Medicine and Mood Disorders Program, University Hospitals Case Medical Center, Cleveland, Ohio, USA*

[2]*Division of Mood Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China*

[3]*Institute of Mental Health of Second Xiangya Hospital of Central South University, Changsha 410011, China*

[4]*Mood Disorders Program of Hongkou District Mental Health Center of Shanghai, Shanghai 200083, China*

Corresponding author: Keming Gao. E-mail: Keming.gao@uhhospitals.org

© Shanghai Institutes for Biological Sciences, CAS and Springer-Verlag Berlin Heidelberg 2015

English-language literature cited in MEDLINE from January, 1980 to October 30, 2014 was searched by using terms of antipsychotic, generic and brand names of atypical antipsychotics, “bipolar depression/bipolar disorder”, “placebo”, and “trial”. The parameters of response (≥50% improvement on MADRS, Montgomery-Asberg Depression Rating Scale total score), remission (either ≤12 or 8 on MADRS total score at endpoint), discontinuation due to adverse events (DAEs), somnolence, ≥7% weight gain, overall extrapyramidal side-effects (EPSs), and akathisia, were extracted from originally published primary outcome papers. The number needed to treat to benefit (NNT) for response and remission or harm (NNH) for DAEs or other side effects relative to placebo were estimated and presented with the estimate and 95% confidence interval. Olanzapine monotherapy, olanzapine-fluoxetine combination (OFC), quetiapine-IR monotherapy, quetiapine-XR monotherapy, lurasidone monotherapy, and lurasidone adjunctive therapy were superior to placebo with NNTs for responses of 11–12, 4, 7-8, 4, 4–5, and 7, and NNTs for remission of 11–12, 4, 5–11, 7, 6–7, and 6, respectively. There was no significant difference between OFC and lamotrigine, and between aripiprazole or ziprasidone and placebo in response and remission. Olanzapine monotherapy, quetiapine-IR, quetiapine-XR, aripiprazole, and ziprasidone 120–160 mg/day had significantly increased risk for DAEs with NNHs of 24, 8–14, 9, 12, and 10, respectively. For somnolence, quetiapine-XR had the smallest NNH of 4. For ≥7% weight gain, olanzapine monotherapy and OFC had the smallest NNHs with both of 5. For akathisia, aripiprazole had the smallest NNH of 5. These findings suggest that among the FDA-approved agents including OFC, quetiapine-IR and -XR, lurasidone monotherapy and adjunctive therapy to a mood stabilizer, the differences in the NNTs for response and remission are small, but the differences in NNHs for DAEs and common side-effects are large. Therefore, the selection of an FDA-approved atypical antipsychotic for bipolar depression should be based upon safety and tolerability.

Keywords: bipolar depression; atypical antipsychotic; number needed to treat; efficacy; tolerability; weight gain; somnolence; extrapyramidal side-effects; akathisia

Introduction

Patients with bipolar disorders spend more time in the depressive phase than in the manic/hypomanic phase[1, 2]. They also suffer more impairments in work, family, and social life and an increased risk for suicide during the

depressive phase than in the manic/hypomanic phase[1-3]. Atypical antipsychotics have played a major role in the acute treatment of bipolar depression and mania, and the maintenance treatment of bipolar disorders[4-6]. All three of the US FDA-approved pharmacological agents for acute bipolar depression are atypical antipsychotics: quetiapine monotherapy, lurasidone monotherapy, the combination of olanzapine and fluoxetine (OFC), and lurasidone adjunctive therapy with lithium or valproate[5].

In a previous review of typical and atypical antipsychotics in acute bipolar depression[4], we found that olanzapine, OFC, and quetiapine immediate-release (IR) have different effect sizes in bipolar depression. Our reviews have also shown that patients with bipolar depression are more sensitive and less tolerant to atypical antipsychotics than those with mania or schizophrenia[7-9]. More importantly, different antipsychotics have clinically relevant differences in side-effect profiles during the acute treatment of bipolar depression[10]. With the recent approval of lurasidone for acute treatment of bipolar depression, atypical antipsychotics will undoubtedly continue to play a major role in the treatment of bipolar disorders, especially bipolar depression[5]. In addition, since only a limited number of medications have demonstrated superior efficacy relative to placebo in acute bipolar depression[5, 6], the off-label use of other atypical antipsychotics and other pharmacological agents is inevitable. To maximize the benefits and minimize the risks of atypical antipsychotics and other commonly-used pharmacological agents in bipolar depression, it is essential to understand their relative efficacy and safety.

Citrome and colleagues compared the benefits and risks of lurasidone and other FDA-approved agents in the acute treatment of bipolar depression, with number needed to treat to benefit (NNT) for response and remission and the number needed to treat to harm (NNH) for discontinuation due to adverse events (DAEs) and other common side-effects[11]. They found that the NNTs for response and remission of lurasidone monotherapy (5 and 7, respectively) were comparable to those of quetiapine monotherapy (6 for both) and OFC (6 and 5, respectively). The NNHs for the DAEs of lurasidone monotherapy and OFC were 642 and –37, respectively, but that of quetiapine was 10. In contrast, the NNHs for ≥7% weight gain of OFC, quetiapine, and lurasidone monotherapy 20–60 mg/day were 6, 17, and 29, respectively. These data suggest that among the FDA-approved agents, the differences in efficacy relative to placebo are minimal, but the differences in safety and tolerability vary widely.

However, the safety and tolerability of agents not approved by the FDA remain unclear. In the current review, we used published pivotal (Phase III) studies to estimate the efficacy and safety of all atypical antipsychotics relative to placebo in acute bipolar depression, with NNT and NNH to compare the magnitudes of differences.

Selection of Studies and Outcome Measures

Study Selection

The English-language literature published from January, 1980 to October 30, 2014 and cited in MEDLINE was searched using the following terms: “antipsychotic”, “bipolar depression”, “placebo”, and “trial”. A second search was conducted for terms including “atypical antipsychotic”, “clozapine (Clozaril)”, “olanzapine (Zyprexa)”, “risperidone (Risperdal)”, “paliperidone (Invega)”, “quetiapine (Seroquel)”, “ziprasidone (Geodon)”, “aripiprazole (Abilify)”, “Iloperidone (Fanapt)”, “asenapine (Saphris)”, “lurasidone (Latuda)”, “bipolar depression”, “bipolar disorder”, “manic-depressive illness”, “placebo”, and “trial”. A manual search was also conducted through the lists of references in qualified publications.

All pivotal trials of atypical antipsychotics in acute bipolar depression were included, regardless of the FDA-approval status. Studies not registered for FDA approval, but using the change in the MADRS (Montgomery-Asberg Depression Rating Scale) total score from baseline to endpoint as a primary or secondary outcome measure were also included. This was justified on the basis that all pivotal studies in bipolar depression used the change in MADRS total score as the primary outcome and that the MADRS is more accurate in assessing depression severity than the 17-item Hamilton Depression Rating Scale (HAMD-17)[12]. Studies that did not use MADRS were excluded. Because safety and tolerability were not primary outcomes of any clinical trials, the sample sizes for safety measures were not properly estimated. To reduce the potential over- or under-estimation of safety measures, only studies with large samples (n ≥100 in each arm) were included. Open-label studies, case reports, and retrospective studies were excluded due to potential reporting biases or lack

of a comparison group. Studies enrolling children and adolescents were also excluded because they were out of the scope of this review.

Efficacy Measures: Response and Remission Rates and Effect Size

Response and remission rates are commonly used as secondary outcome measures, and have been reported in almost all clinical trials of bipolar depression. Response was defined as ≥50% improvement in MADRS total score from baseline to study endpoint in all trials, and remission was defined as ≤12 or 8 points on the MADRS total score at study endpoint. Cohen's d effect size is commonly used to measure the magnitude of differences between treatments in continuous outcome measures and is calculated by the improvement of active treatment over placebo divided by pooled standard deviation. Effect sizes of 0.2–0.49 are viewed as being small, 0.5–0.79 moderate, and ≧0.8 large[13]. The primary outcome measure, the difference in the change of MADRS total scores from baseline to endpoint between two groups, was reported in different ways across the studies, and re-calculation of the effect sizes of the drugs in some studies was not possible. Therefore, the effect sizes of active treatments relative to placebo in the included studies were obtained from original publications without further analysis.

Safety and Tolerability Measures

To compare the safety and tolerability of different antipsychotics relative to placebo, we chose the risks for the discontinuation due to adverse events (DAEs), reported somnolence, ≥7% weight gain, overall extrapyramidal side-effects (EPSs), and akathisia as safety and tolerability outcome measures. Although some DAEs in a study are not related to the side-effects of the studied drug, previous studies have shown that the majority of patients with bipolar depression discontinued due to somnolence/sedation and weight gain[7]. Understanding the risks for the DAEs, somnolence and ≥7% weight gain will help clinicians to select a better antipsychotic. The risk for acute EPSs differs among atypical antipsychotics in bipolar disorder, and acute EPSs may have long-term implications[8]. Therefore, understanding the risk for overall EPSs and akathisia among atypical antipsychotics not only helps clinicians to properly manage these side-effects, but also potentially prevents long-term consequences. In contrast, some common side-effects such as dry mouth and other gastrointestinal events were not chosen because they have limited impact on treatment outcomes.

Measures for Comparison: Number Needed to Treat

The NNT is defined as the number of patients one would expect to treat with "T" to have one more success (or one less failure) than if the same number were treated with "C". The "T" refers to treatment and the "C" refers to control. Therefore, according to the outcome of success or failure relative to control, the NNT can be estimated for benefit, commonly presented as NNT, or harm, commonly presented as NNH. Mathematically, NNT = 1/absolute risk reduction (benefit) and NNH = 1/absolute risk increase (harm). The NNT and NNH can only be used for categorical outcome measures and are believed to provide more clinically-relevant information than relative risk reduction or odds ratios[14, 15].

Data from studies with a similar design such as randomization scheme, inclusion and exclusion criteria, study duration, and dose schedules of an antipsychotic were combined when possible. However, data from studies with a similar design, but in different populations, were analyzed separately. The NNT or NNH was presented with the estimate and 95% confidence interval.

Comparisons of Efficacy, Safety, and Tolerability between Active Treatment and Placebo or a Comparator

An initial search generated 567 citations, but only 13 studies/publications met the inclusion criteria for analyses. Five atypical antipsychotics were studied in the acute treatment of bipolar depression in at least two large, randomized, placebo-controlled trials either as monotherapy, combination therapy, and/or adjunctive therapy (Table 1). There was only one head-to-head comparison study of OFC *versus* lamotrigine. One quetiapine-IR study included a lithium and one included a paroxetine arm.

Olanzapine and Olanzapine-Fluoxetine Combination

The efficacy and safety of OFC and olanzapine monotherapy in acute bipolar depression were studied in a total of 833 inpatients and outpatients with moderately severe bipolar I depression (MADRS ≥20 points)[16] (Table 1). Patients

Table 1. Number needed to treat for response or remission of atypical antipsychotics relative to placebo in acute bipolar depression

Agents and Trials	Patients	Treatment arms	Duration (weeks)	Total N	Response (≥50% improvement in MADRS)		Remission (MADRS ≤ 8 or 12)	
					N	Mean (95% CI)	N	Mean (95% CI)
Aripiprazole Thase *et al.*, 2008[25]	Bipolar I depression	Aripiprazole 5–30 mg/day	8	337	148	44 (10, ∞, -20)	98[b]	53 (12, ∞, -21)
		Placebo		353	147		96	
Lurasidone[a] Loebel *et al.*, 2014[29]	Bipolar I depression	Lurasidone 20–60 mg/day	6	161	85	4 (3, 8)	68	6 (6, 15)
		Lurasidone 80–120 mg/day		162	83	5 (3,10)	65	7 (4, 22)
		Placebo		162	49		41	
Lurasidone[a] Loebel *et al.*, 2014[30]	Bipolar I depression MS for 4 weeks	MS + lurasidone 20–120 mg/day	6	179	102	7 (4, 24)	90	6 (4, 20)
		MS + placebo		161	68		56	
Olanzapine and OFC Tohen *et al.*, 2003[16]	Bipolar I depression	Olanzapine 5–20 mg/day	8	351	137	12 (6, 63)	115	12 (7, 63)
		OFC 6/25, 6/50, or 12/50 mg/day		82	46	4 (3, 7)	40	4 (3, 8)
		Placebo		355	108		87	
Olanzapine Tohen *et al.*, 2012[17]	Bipolar I depression	Olanzapine 5–20 mg/day	6	343	180	11 (6, 2866)	132[b]	11 (6, 213)
		Placebo		171	74		50	
OFC *vs* LTG Brown *et al.*, 2006[19]	Bipolar I depression	OFC 6/25, 6/50, 12/25 or 12/50 mg/day	7	205	141	16 (7, ∞, -62)	76	16 (7, ∞, -34)
		Lamotrigine 150–200 mg/day		204	122		63	
Quetiapine-IR[a] Calabrese *et al.*, 2005[20] Thase *et al.*, 2006[21]	Bipolar I or II depression	Quetiapine-IR 300 mg/day	8	327	166	7 (5, 15)	171	5(4, 8)
		Quetiapine-IR 600 mg/day		321	165	7 (5, 14)	169	5 (4, 8)
		Placebo		330	121		108	
Quetiapine-IR McElory *et al.*, 2010[24]	Bipolar I or II depression	Quetiapine-IR 300 mg/day	8	229	153	7 (4, 31)	148	11 (5, ∞, -70)
		Quetiapine-IR 600 mg/day		232	156	7 (4, 27)	159	8 (4, 39)
		Paroxetine 20 mg/day		118	65	46 (7, ∞, -10)	67	71 (7, ∞, -7)
		Placebo		121	64		67	

(To be continued)

(Continued)

Quetiapine-IR Yong *et al.*, 2010[23]	Bipolar I or II depression	Quetiapine-IR 300 mg/day	8	255	175	8 (4, 38)	178	7 (4, 22)
		Quetiapine-IR 600 mg/day		263	183	7 (4, 27)	185	7 (4, 19)
		Lithium 600–1800 mg/day		136	85	15 (5, ∞, -20)	85	12 (5, ∞,-29)
		Placebo		129	72		71	
Quetiapine-XR[a] Suppes *et al.*, 2008[22]	Bipolar I or II depression	Quetiapine-XR 300 mg/day	8	133	87	4 (3, 10)	72	7 (4, 35)
		Placebo		137	59		54	
Ziprasidone Lombardo *et al.*, 2012[26]	Bipolar I depression	Ziprasidone 40–80 mg/day	6	158	84	23 (7, 15)	n/a	n/a
		Ziprasidone 120–160 mg/day		166	76	-34 (13, ∞, -7)	n/a	n/a
		Placebo		162	79			
		Ziprasidone 40–160 mg/day	6	180	95	58 (8, ∞, -12)	n/a	n/a
		Placebo		190	97			
Ziprasidone Sachs *et al.*, 2011[29]	Bipolar I depression MS for ≥4 weeks	MS + ziprasidone 40–160 mg/day	6	145	n/a	n/a	n/a	n/a
		MS + placebo		145				

[a]Pivotal studies leading to approval by the FDA of the USA for the acute treatment of bipolar I or II depression; [b] Remission was defined as MADRS ≤8. BPI, bipolar I disorder; BPII, bipolar II disorder; CI, confidence interval; IR, immediate release; LTG, lamotrigine; MADRS, Montgomery-Asberg Depression Rating Scale; MS, lithium, valproate, or lamotrigine in therapeutic doses; N, number; OFC, olanzapine-fluoxetine combination; XR, extended release.

treated with olanzapine monotherapy or OFC for 8 weeks exhibited significant improvements in depressive symptoms compared to those treated with placebo. Significant improvement occurred from week 1 onwards. The OFC arm was significantly better than placebo and olanzapine monotherapy[16]. The NNTs for response and remission relative to placebo (≤12 on MADRS at the end of study) were both 12 for olanzapine, and both 4 for OFC (Table 1). The Cohen's d effect size was small (0.32) for olanzapine monotherapy, but moderate (0.68) for the OFC arm[16].

Similarly, a recent study of olanzapine monotherapy *versus* placebo for bipolar I depression in which most patients were from Asian countries also showed that olanzapine reduces depressive symptoms with an effect size of 0.22[17]. The NNTs for response and remission (≤8 on MADRS total scores) were both 11 (Table 1). However, there was no significant difference in remission rates between olanzapine and placebo when remission was defined as ≤12 on MADRS total scores[17].

In a *post-hoc* analysis of only patients from Japan, olanzapine monotherapy (n = 104) was superior to placebo (n = 52) in reducing MADRS total scores and 6 subscale scores with an overall effect size of 0.42[18]. However, there were no significant differences in the rates of response (54.8% *versus* 40.4%) or remission (43.3% *versus* 34.6%).

Both OFC and olanzapine were relatively well tolerated

compared to placebo. In the first study[16], there was no significant difference in the risk of DAEs between OFC and placebo (Table 2). In contrast, there was a small, but significantly increased risk of DAEs for olanzapine relative to placebo, with an NNH of 24 (Table 2). In the second study[17], there was no significant difference in DAEs between olanzapine monotherapy and placebo (Table 2). Similarly, *post-hoc* analysis in Japanese patients did not find a significant difference in DAEs between them[18].

The NNH for somnolence was 12 for OFC, and 6[16] and 9[17] for olanzapine, relative to placebo (Table 3). However, olanzapine and OFC caused significantly more weight gain than the placebo (2.59 ± 3.24 kg and 2.79 ± 3.23 kg, *versus* –0.47 ± 2.62 kg)[16]. In Japanese patients, olanzapine also caused more weight gain than the placebo (2.12 ± 0.21 *versus* –0.36 ± 0.29 kg). The number of patients with a ≥7% weight gain was significantly higher with olanzapine or OFC than with placebo. The NNHs for ≥7% weight gain of OFC and olanzapine relative to placebo were both 5 (Table 3). The NNH for ≥7% of olanzapine relative to placebo in Japanese patients was also 5 (95% CI 4, 137). The rates of akathisia and overall EPSs were not reported because the incidence of these events did not exceed 10% in any group in the first study[16], and 5% in any group in the second study[17].

OFC versus Lamotrigine

The efficacy and safety of OFC *versus* lamotrigine were investigated in a 7-week head-to-head comparison study of patients with bipolar I depression[19]. There were no significant differences between the two treatments in the rates of response and remission (based on ≤12 MADRS scores) (Table 1) and in the DAEs between OFC and lamotrigine (Table 2). However, the incidences of reported somnolence and ≥7% body weight gain were significantly higher in the OFC group than in the lamotrigine group. The NNHs for ≥7% body weight gain and reported somnolence of OFC relative to lamotrigine were 4 and 10, respectively.

Quetiapine-IR

The safety and efficacy of quetiapine-IR in the acute treatment of bipolar depression were studied in patients with bipolar I or II disorder[20, 21]. All participants were depressed outpatients with at least moderate severity (HAMD-17 ≥20). Both studies demonstrated that quetiapine-IR at 600 mg/day and 300 mg/day produced significantly greater improvement in depressive symptoms than placebo from week 1 onwards. The effect sizes for quetiapine-IR 600 mg/day and 300 mg/day were 0.81 and 0.67 in the first study and 0.54 and 0.61 in the second. The response rates and other secondary outcome measures also showed significant superiority of quetiapine-IR over the placebo. The NNTs for response of quetiapine-IR 300 mg/day and 600 mg/day relative to placebo were 7 for both doses; and NNTs for remission were 5 for both (Table 1).

In a subtype analysis of the first study, the effect sizes for quetiapine-IR 600 mg/day and 300 mg/day in the bipolar I subgroup were 1.09 and 0.91, respectively, but in the bipolar II subgroup were only 0.39 and 0.28[20]. However, in the second study, the effect sizes for quetiapine-IR 600 mg/day and 300 mg/day in bipolar I depression were 0.51 and 0.67, respectively; and in bipolar II depression were 0.64 and 0.56[21]. In these pivotal studies, there were minimal differences in efficacy between fixed dose 600 mg/day and 300 mg/day, but there was a higher rate of DAEs in the quetiapine-IR 600 mg/day group than in the 300 mg/day group[10] (Table 2). The NNHs for the DAEs with quetiapine-IR 600 mg/day and 300 mg/day were 8 and 14, respectively (Table 2). The most common causes for discontinuation were sedation, somnolence, and dizziness[20, 21].

The NNHs for somnolence with quetiapine relative to placebo were 6 for 300 mg/day and 5 for 600 mg/day (Table 3). The mean weight gain was 1.6 kg for quetiapine-IR 600 mg/day, 1.0 kg for quetiapine-IR 300 mg/day, and 0.2 kg for placebo[20]. The NNHs for ≥7% weight gain with quetiapine-IR 600 mg/day and 300 mg/day were 20 and 27, respectively (Table 3). The risk for overall EPSs including akathisia was significantly higher in the quetiapine groups relative to the placebo group with NNHs of 20 for 300 mg/day and 19 for 600 mg/day (Table 3).

Quetiapine-XR

Quetiapine extended-release form (quetiapine-XR) at 300 mg/day also was superior to placebo in reducing depressive symptoms in bipolar I and II depression[22]. The findings from this study were similar to those for quetiapine-IR with an effect size of 0.61 for the entire group, 0.64 for bipolar I, and 0.46 for bipolar II. The NNTs for response and remission for quetiapine-XR relative to placebo were 4 and 7, respectively (Table 2).

Similar to quetiapine-IR, quetiapine-XR 300 mg/day

Table 2. Number needed to treat for discontinuation due to adverse events of atypical antipsychotics relative to placebo in acute bipolar depression

Agents and Trials	Patients	Treatment arms	Duration (week)	Discontinuation Due to Adverse Event (DAEs)		
				Total N	No. of DAEs	Mean (95% CI)
Aripiprazole	Bipolar I depression	Aripiprazole 5–30 mg/day	8	373	50	14 (1, 35)
Thase *et al.*, 2008[25]		Placebo		376	24	
Lurasidone[a]	Bipolar I depression	Lurasidone 20–60 mg/day	6	166	11	-641(19, ∞, -17)
Loebel *et al.*, 2013[29]		Lurasidone 80–120 mg/day		169	10	181(19, ∞, -17)
		Placebo		170	11	
Lurasidone[a]	Bipolar I depression	MS + lurasidone 20–120 mg/day	6	183	11	52 (13, ∞, -28)
Loebel *et al.*, 2013[30]	MS at therapeutic levels for 4 weeks	MS + placebo		165	13	
Olanzapine and OFC	Bipolar I depression	Olanzapine 5–20 mg/day	8	370	34	24 (13, 224)
Tohen *et al.*, 2003[16]		OFC 6/25, 6/50, or 12/50 mg/day		86	2	37 (17, ∞, -30)
		Placebo		377	19	
Olanzapine	Bipolar I depression	Olanzapine 5–20 mg/day	6	343	30	87 (17, ∞, -158)
Tohen *et al.*, 2012[17]		Placebo		171	13	
OFC *vs*. LTG	Bipolar I depression	OFC 6/25, 6/50, 12/25 or 12/50 mg/day	7	205	17	-103 (23, ∞, -16)
Brown *et al.*, 2006		Lamotrigine 150–200 mg/day		205	15	
Quetiapine-IR[a]	Bipolar I or II	Quetiapine-IR 300 mg/day	8	353	43	14 (9, 35)
Calabrese *et al.*,	depression	Quetiapine-IR 600 mg/day		349	63	8 (6, 12)
2005[20]		Placebo		349	18	
Thase *et al.*, 2006[21]						
Quetiapine-IR	Bipolar I or II	Quetiapine-IR 300 mg/day	8	245	21	158 (16, ∞, -16)
McElory *et al.*,	depression	Quetiapine-IR 600 mg/day		247	30	24 (10, ∞, -36)
2010[24]		Paroxetine 20 mg/day		122	15	10 (8, ∞, -30)
		Placebo		126	10	
Quetiapine-IR	Bipolar I or II	Quetiapine-IR 300 mg/day	8	265	26	58 (14, ∞, -21)
Yong *et al.*, 2010[23]	depression	Quetiapine-IR 600 mg/day		268	35	20 (9, ∞, -55)
		Lithium 600–1800 mg/day		133	12	107 (13, ∞, -17)
		Placebo		136	11	
Quetiapine-XR[a]	Bipolar I or II	Quetiapine-XR 300 mg/day	8	139	17	9 (6, 20)
Suppes *et al.*, 2008[22]	depression	Placebo		138	2	
Ziprasidone	Bipolar I depression	Ziprasidone 40–80 mg/day	6	176	63	33 (8, ∞,-15)
Lombardo *et al.*,		Ziprasidone 120–160 mg/day		186	80	10 (5, 455)
2012[26]		Placebo		174	57	
		Ziprasidone 40–160 mg/day		192	73	14 (4, ∞, -42)
		Placebo		200	62	
Ziprasidone	Bipolar I depression	MS + ziprasidone 40–160 mg/day	6	148	25	13 (7, ∞, -621)
Sachs *et al.*, 2011[28]	MS at therapeutic levels for ≥4 weeks	MS + placebo	6	150	14	

[a]Pivotal studies leading to approval by the FDA of the USA for the acute treatment of bipolar I or II depression. Abbreviations: BPI, bipolar I disorder; BPII, bipolar II disorder; CI, confidence interval; DAEs, discontinuation due to adverse events; IR, immediate release; LTG, lamotrigine; MS, lithium, valproate, or lamotrigine in therapeutic doses; N, number; OFC, olanzapine-fluoxetine combination; XR, extended release.

Table 3. Number needed to treat for reported somnolence, ≥7% weight gain, overall EPSs, and akathisia for atypical antipsychotics relative to placebo in acute bipolar depression

Agents and Trials	Treatment arms	Duration (week)	Reported Somnolence		≥7% Weight Gain		Overall EPS		Akathisia	
			Patients Case/Total	Mean (95% CI)	Patients Case/Total	Mean (95% CI)	Patients Case/Total	Mean (95% CI)	Patients Case/Total	Mean (95% CI)
Aripiprazole	ARI 5–30 mg/day	8	27/360	29 (14, ∞, -6789)	17/360	69 (22, ∞, -67)	43/360	19 (11, 113)	88/373	5 (4, 6)
Thase *et al.*, 2008[25]	Placebo		15/367		12/367		25/367		14/376	
Lurasidone[a]	LUR 20–60 mg/day	6	7/164	984 (20, ∞, -22)	7/164	27 (13, 500)	8/164	40 (14, ∞, -155)	13/164	18 (9, 139)
Loebel *et al.*, 2013[29]	LUR 80–120 mg/day		11/167	41 (13, ∞, -38)	1/167	28058 (36, ∞, -36)	15/167	15 (8, 62)	18/167	12 (7, 32)
	Placebo		7/168		1/168		4/168		4/168	
Lurasidone[a]	MS+LUR 20–120 mg/day	6	16/183	22 (10, ∞, -103)	6/183	38 (16, ∞, -159)	28/183	18 (8, -61)	14/183	30 (12, ∞,-52)
Loebel *et al.*, 2013[30]	MS+ placebo		7/163		1/163		16/163		7/163	
Olanzapine and OFC	OLA 5–20 mg/day	8	104/370	6 (5, 10)	65/347	5 (4, 6)	10/360[b]	-107 (57, ∞, -27)	n/a	n/a
	OFC		18/86	12 (5, 326)	16/82	5 (3, 8)	7/86	22 (8, ∞, -289		
Tohen *et al.*, 2003[16]	Placebo		47/377		1/355		14/377			
Olanzapine	OLA 5–20 mg/day	6	59/343	9 (6, 21)	79/341	5 (4, 6)	n/a	n/a	n/a	n/a
Tohen *et al.*, 2012[17]	Placebo		11/171		2/169					
OFC *vs* LTG	OFC 6/25, 6/50, 12/25, 12/50 mg/day	7	38//205	10 (6, 28)	48/205	4 (3, 6)	n/a	n/a	n/a	n/a
Brown *et al.*, 2006[19]	LTG 150–200 mg/day		17/204		0/204					
Quetiapine-IR[a]	QUE-IR 300 mg/day	8	95/350	6 (4, 8)	20/350	27 (15, 122)	33/350[c]	20 (11, 75)	n/a	n/a
Calabrese *et al.*, 2005[20]	QUE-IR 600 mg/day		99/348	5 (4, 8)	27/348	20 (12, 50)	33/348	19 (11, 72)		
Thase *et al.*, 2006[21]	Placebo		33/347		7/347		15/347			

(to be continued)

(Continued)

Table 3. Number needed to treat for reported somnolence, ≥7% weight gain, overall EPSs, and akathisia for atypical antipsychotics relative to placebo in acute bipolar depression

Agents and Trials	Treatment arms	Duration (week)	Patients Case/Total	Mean (95% CI)	Patients Case/Total	Mean (95% CI)	Patients Case/Total	Mean (95% CI)	Patients Case/Total	Mean (95% CI)
Quetiapine-IR[a]	QUE-IR 300 mg/day	8	46/243	9 (6, 30)	16/177	20 (9, ∞, -50)	20/243[c]	24 (11, ∞, -62)	n/a	n/a
McElroy *et al.*,	QUE-IR 600 mg/day		43/244	10 (6, 48)	22/194	13 (7, 163)	24/244	17 (9, ∞, -667)		
2010[24]	Paroxetine 20 mg/day		7/121	-39 (26, ∞, -11)	3/90	-127 (17, ∞,-14)	5/121	1000 (17, ∞, -18)		
	Placebo		10/124		4/97		5/124	85 (19, ∞,-25)		
Quetiapine-IR[a]	QUE-IR 300 mg/day	8	47/260	7 (5, 13)	11/240	75 (19, ∞, -26)	12/240[c]	27 (12, ∞, -55)	n/a	n/a
Young *et al.*,	QUE-IR 600 mg/day		47/267	7 (5, 13)	20/240	20 (10, ∞, -191)	18/240	23 (10, 59)		
2010[23]	Lithium		12/136	20 (9, ∞, -94)	3/127	-112 (26, ∞,-17)	11/136			
	Placebo		5/131		4/123		5/131			
Quetiapine-XR[a]	QUE-XR 300 mg/day	8	40/137	4 (3, 7)	11/137	14 (8, 39)	4/137[c]	45 (15, ∞, -68)	n/a	n/a
Suppes *et al.*,	Placebo		8/140		1/140		1/140			
2008[22]										
Ziprasidone	ZIP 40–80 mg/day	6	39/288	10 (7, 17)	5/239	89 (26, ∞, -92)	2/288	690 (50, -73)	8/288	51 (22, ∞, -791)
Gao *et al.*,	ZIP120–160 mg/day		41/232	7 (5, 11)	5/197	64 (20, ∞, -129)	4/232	85 (26, -166)	8/232	38 (17, 316)
2013[27]	Placebo		13/364		3/308		2/364		3/364	
Ziprasidone	MS+ZIP 40–160 mg/day	6	33/147	6 (4, 10)	n/a	n/a	n/a	n/a	n/a	n/a
Sachs *et al.*,	MS + placebo		7/147							
2011[28]										

[a]Pivotal studies leading to approval by the FDA of the USA for the acute treatment of bipolar I or II depression. [b]There were no data on overall EPSs or akathisia in the original study. In this analysis, the rates of anticholinergic use were used for comparison. [c]In all quetiapine studies, the overall EPSs included akathisia. Abbreviations: ARI, aripiprazole; BPI, bipolar I disorder; BPII, bipolar II disorder; CI, confidence interval; EPS, extrapyramidal symptoms; IR, immediate release; LTG, lamotrigine; LUR, lurasidone; MS, lithium or valproate in therapeutic doses, or lamotrigine 100-200 mg/day; n/a, not available; OFC, olanzapine-fluoxetine combination; OLA, olanzapine; QUE, quetiapine; ZIP, ziprasidone; XR, extended release.

had a significantly higher rate of DAEs with an NNH of 9[10] (Table 2). The NNH for somnolence was 4 (Table 3). Weight gain in the quetiapine-XR group (1.3 kg) was greater than that in the placebo group (–0.2 kg). The NNH for ≥7% weight gain was 14[10] (Table 3). There was no significant difference in overall EPSs between quetiapine-XR and placebo.

Quetiapine-IR and Lithium

Since the efficacy of lithium in the treatment of acute bipolar depression had never been established, a quetiapine-IR study including lithium was conducted in acute bipolar depression[23]. This study was conducted in 110 centers throughout Europe, Canada, and Asia in patients with bipolar I (*n* = 499) or II (*n* = 303) depression who were randomly assigned to quetiapine-IR 600 mg/day (*n* = 268), quetiapine-IR 300 mg/day (*n* = 265), lithium 600–1800 mg/day (*n* = 136), or placebo (*n* = 133) for 8 weeks. The superiority of quetiapine-IR 600 mg/day and 300 mg/day in bipolar depression over placebo was replicated in this study. Lithium had numerically greater improvement in MADRS total score compared with placebo, but without statistical significance[23]. *Post-hoc* analyses according to a lithium concentration of ≥0.8 or <0.8 mEq/L revealed no significant difference from placebo in the improvement of MADRS total score. Quetiapine 600 mg/day, but not 300 mg/day, significantly improved MADRS total score compared with lithium. The NNTs for response were 8 for quetiapine 300 mg/day and 7 for quetiapine 600 mg/day; and the NNTs for remission were 7 for both doses (Table 2).

In bipolar subtype analyses, both doses of quetiapine-IR were associated with significant improvements over placebo in MADRS total score in patients with bipolar I depression, but only numerical improvements over placebo in bipolar II depression. However, patients treated with lithium did not show a significant improvement over placebo in those with bipolar I or II depression[23]. The proportion of DAEs was 13.9% for quetiapine-IR 600 mg/day, 10.4% for 300 mg/day, 8.8% for lithium, and 8.4% for placebo, with no significant differences among the groups (Table 1).

The NNHs for somnolence were 7 for both doses of quetiapine-IR. The rates of somnolence with lithium and placebo were not significantly different (Table 3). The overall EPSs with lithium were also significantly higher than placebo with an NNH of 23, but the rates of overall EPS with quetiapine were not significantly different from placebo (Table 3). There were small but significant differences in weight gain and body mass index increases between placebo and the 3 active treatment arms[23]. However, the proportion of patients with ≥7% weight gain was not significantly different between the active arms and placebo and among the active arms (Table 3).

Quetiapine-IR and Paroxetine

Controversy continues regarding the efficacy and safety of antidepressants in the acute treatment of bipolar depression, especially antidepressant monotherapy. For this reason, a double-blind, placebo-controlled study of quetiapine-IR in acute bipolar depression included paroxetine[24]. This study was conducted in 83 centers in the USA, the European Union, Turkey, Central and South America, South Africa, and Australia in patients with bipolar I (*n* = 478) or II (*n* = 262) depression. Treatments included quetiapine-IR 600 mg/day (*n* = 247), quetiapine-IR 300 mg/day (*n* = 245), paroxetine 20 mg/day (*n* = 122), and placebo (*n* = 126) for 8 weeks.

Similar to the previous studies of quetiapine-IR in bipolar I or II depression, quetiapine-IR 600 mg/day and 300 mg/day were more effective than placebo in reducing depressive symptoms. However, paroxetine did not result in statistically significant improvement in depression compared with placebo at any time point during the study[24]. The NNTs for response were 7 for quetiapine-IR 300 mg/day and 600 mg/day (Table 1). However, the remission rates were only significantly different between quetiapine 600 mg/day and placebo with an NNT of 8 (Table 1). Bipolar subtype analysis did not find superiority of paroxetine over placebo in bipolar I or II depression although both doses of quetiapine-IR were superior to placebo in both types of bipolar depression[24].

The rates of DAEs were not significantly different between the active arm and placebo (Table 2). The NNH for somnolence was 9 for quetiapine-IR 300 mg/day and 10 for quetiapine-IR 600 mg/day (Table 3). Quetiapine-IR 600 mg/day also caused a small but significant weight gain (≥7%) relative to the placebo, with an NNH of 13 (Table 3).

Aripiprazole

Aripiprazole in acute bipolar depression was investigated in two large identical randomized, double-blind, placebo-

controlled studies of patients with bipolar I depression[25]. However, these studies did not show superiority of aripiprazole 5–30 mg/day over placebo in reducing depressive symptoms at the end of 8-week study. Aripiprazole significantly reduced depressive symptoms from baseline to week 1 through week 6 compared with placebo in study 1 and from baseline to week 1 through week 3, and to week 5 in study 2[25]. There were also no significant differences between aripiprazole and placebo in key secondary outcome measures including response and remission rates (Table 1).

The most common side-effect was akathisia, which was significantly higher with aripiprazole than placebo with an NNH of 5[10, 25] (Table 3). The overall EPSs were also significantly higher in the aripiprazole group than in the placebo group with an NNH of 19 (Table 3). The rate of DAEs was also significantly higher in the aripiprazole group than in the placebo group, with an NNH of 14 (Table 2). There was no significant difference in the mean weight gain or the proportion of patients with ≥7% weight gain between the two groups[25] (Table 3).

Ziprasidone Monotherapy

There were two randomized, double-blind, placebo-controlled studies of ziprasidone monotherapy in acute bipolar I depression[26]. Like aripiprazole, ziprasidone was transiently superior to placebo in reducing depressive symptoms. In study 1, ziprasidone 40–80 mg/day, but not 120–160 mg/day, significantly reduced the depressive symptoms compared to placebo only at week 1. In study 2, ziprasidone 40–160 mg/day significantly reduced the depressive symptoms from week 1 to week 3, but not at the end of the study. There were also no significant differences between ziprasidone and placebo in secondary outcome measures including response and remission rates (Table 1).

The investigators then debated whether the negative finding was due to a failed trial or represented true negative trials. Exploratory analyses of these studies found that about half of the patients in each study did not meet the minimal severity threshold required for enrollment[26]. In addition, the investigators also found that 3% in study 1 and 5% in study 2 met the criteria for remission at baseline.

Ziprasidone was well tolerated in these studies. There was no significant difference in the rates of DAEs between ziprasidone and placebo regardless of the higher (120–160 mg/day) or lower doses (40–80 mg/day)[26] (Table 2). The rate of somnolence was significantly higher with either higher or lower doses of ziprasidone than that with placebo. The NNHs were 10 for the lower doses, and 7 for the higher. At the higher doses, there was a slightly higher, but statistically significant, risk for akathisia relative to placebo with an NNH of 38 (Table 3). There was no significant difference in weight gain or overall EPSs between ziprasidone and placebo (Table 3).

Ziprasidone Adjunctive Therapy to Mood Stabilizer

Ziprasidone adjunctive therapy to a mood stabilizer in the acute treatment of bipolar depression was studied in 78 centers in Australia, India, and the USA[28]. Bipolar I patients with moderate depressive severity at screening and randomization (HAMD-17 ≥20) were randomized to receive ziprasidone 40–160 mg/day or placebo. Before randomization, all patients had to be on a stable dose of lamotrigine 100–200 mg/day or blood concentrations of lithium (0.6–1.2 mEq/L) or valproate (50–125 μg/ml) for at least 4 weeks. There were no significant differences between ziprasidone and placebo in any outcome measure at any study point.

Poor quality rating at baseline by subgroups of patients (subject inflation) and raters (rater inflation) were observed[28]. Like the ziprasidone monotherapy, the ziprasidone adjunctive therapy to mood stabilizer was well tolerated. There was no significant difference between ziprasidone and placebo in the risk for DAEs (Table 2). However, there was a significant difference between them in the risk for somnolence with an NNH of 6 (Table 3).

Lurasidone Monotherapy

The efficacy and safety of lurasidone 20–60 mg/day and 80–120 mg/day were studied in patients with bipolar I depression[29]. Both doses resulted in significantly greater reductions of MADRS score compared with placebo. The effect size was 0.51 for both 20–60 and 80–120 mg/day lurasidone. Both active groups differed significantly from placebo from week 2 onwards. Both doses of lurasidone showed significant improvement over placebo on other secondary outcome measures[29]. The NNTs for response were 4 for 20–60 mg/day and 5 for 80–120 mg/day. The NNTs for remission were 6 for 20–60 mg/day and 7 for 80–120 mg/day (Table 1).

The rates of DAEs were similar in the lurasidone

and placebo groups (Table 2). The proportions of patients with ≥7% weight gain were similar in patients treated with lurasidone 80–120 mg/day and those treated with placebo, but patients treated with 20–60 mg/day had a small, but significantly increased, risk for ≥7% weight gain relative to placebo with an NNH of 27 (Table 3). Both doses also had a significant risk for akathisia relative to placebo with an NNH of 18 for 20–60 mg/day and 12 for 80–120 mg/day. Lurasidone 80–120 mg/day also had a significant risk for overall EPSs relative to placebo with an NNH of 15 (Table 3).

Lurasidone Adjunctive Therapy to Mood Stabilizer

The efficacy and safety of lurasidone adjunctive therapy in acute bipolar depression were studied in patients with bipolar I depression who had an inadequate response to a minimum of 28-day treatment with 0.6–1.2 mEq/L lithium or 50–125 μg/mL valproate[30]. At the end of the study, lurasidone 20–120 mg/day adjunctive to either lithium or valproate was superior to placebo in reducing depressive symptoms. The significant differences between lurasidone and placebo appeared from week 3 and onwards. Significant improvements over placebo were also observed for other secondary outcome measures including response and remission rates. The NNTs for response and remission of lurasidone adjunctive to a mood stabilizer relative to placebo were 7 and 6, respectively (Table 1). There were also no significant differences between the two groups in ADEs, ≥7% weight gain, overall EPSs, and akathisia (Tables 2 and 3).

Discussion

In this review, we found that among three US FDA-approved atypical antipsychotics for acute bipolar depression, the differences in the NNTs for response and remission relative to placebo were small (Table 1). However, the differences in NNHs for DAEs, somnolence, ≥7% weight gain, overall EPSs, and akathisia varied widely (Tables 2 and 3). The lurasidone studies were 6 weeks rather than 8 weeks[29, 30], which might introduce uncertainty as to whether the shorter duration affected its efficacy, safety, and/or tolerability. However, the results from the olanzapine studies[17, 18] suggest that the impact of 6-week *versus* 8-week duration on the efficacy and safety of lurasidone might be small (Tables 1, 2, and 3).

Overall, the smaller NNTs for response and remission than the corresponding NNH for DAEs of each FDA-approved medication/dose suggest that initiating any FDA-approved agents is more likely to help than harm. However, the magnitude of the help likely depends on the risk for DAEs and common side-effect(s). The likelihood to be helped or harmed (LHH), the ratio of NNH to NNT, has been used to quantify the benefits and risks for evidence-based medicine[11, 15, 31]. The LHH can provide an estimate of the trade-offs between benefits and risks. As shown in Table 1, the NNTs for response and remission of FDA-approved agents were comparable, but the LHH would be quite different if the risks were taken into consideration. For example, to compare the benefits for response and the risks for DAEs, the LHH for lurasidone 80–120 mg/day monotherapy had the largest ratio (181:5 = 36). In contrast, the LHH for quetiapine-XR 300 mg/day monotherapy had the smallest ratio (9:4 = 2.25), suggesting that the benefits and risks for lurasidone treatment response *versus* DAEs far surpass those of quetiapine-XR 300 mg/day. However, such indirect comparison is based on the assumption that placebo has the same or a similar effect on patients in different studies. For a non-FDA-approved agent, aripiprazole, in bipolar depression, the NNTs for response and remission were larger than the NNH for DAEs; therefore, using aripiprazole to treat acute bipolar depression is more harmful than helpful. Clearly, the risks and benefits should be seriously considered when using off-label agents for bipolar depression.

Although NNT, NNH, and LHH can provide useful information to help clinicians practice evidence-based medicine, one of the limitations of these measures is that they only compare two categorical variables at a time. In this review, most NNTs and NNHs were relative to placebo. We must be cautious when comparing the results from different studies, especially those with different designs and/or different drugs, which might not only affect the response to active drugs and/or placebo, but also potentially affect the risk for adverse events. Among all the studies reviewed here, only that of olanzapine and OFC included patients with psychotic features[16, 17]. Although baseline psychotic features were not associated with the response to olanzapine[32], it remains unclear if psychotic features had a similar effect on the response to placebo and olanzapine.

Similarly, only quetiapine studies included patients with bipolar II depression[20-24]. The first quetiapine study showed a larger difference in effect sizes in bipolar I than in bipolar II depression, 1.09 *versus* 0.39 for quetiapine-IR 600 mg/day and 0.91 *versus* 0.28 for quetiapine-IR 300 mg/day[20]. Although the difference in effect sizes between bipolar I and II depression became smaller in the following studies[21-24], it appeared that the inclusion of patients with bipolar II depression reduced the overall effect size of quetiapine in bipolar depression. In contrast, quetiapine had a significantly higher risk for somnolence compared to placebo (Table 3), which might compromise the integrity of double-blinded design and inflate the rating on MADRS. Exploratory analyses found that patients with or without somnolence treated with quetiapine had similar degrees of improvement in depressive symptoms; MADRS scores of −18.8 *versus* −19.3 points[20, 21]. Patients with somnolence/sedation treated with placebo had degrees of improvement similar to quetiapine, but those without somnolence/sedation had much less improvement in depression; MADRS scores of −18.9 *versus* −11.7 points[20, 21]. These data suggest that an inflated rating due to somnolence/sedation in the quetiapine group cannot be totally ruled out. Other factors such as depression severity, episode frequency, age/gender, family history of bipolar disorder, and medical comorbidity could also affect the efficacy and side-effects differently in those receiving active treatment *versus* those with placebo. Unless all these factors are controlled across studies, the potential impact on the outcome of efficacy and safety will never be known. Therefore, the safety and efficacy of different drugs will never be fairly compared.

There will never be a perfect solution to managing the differences across various studies. Unless all medications for bipolar depression or other psychiatric conditions are included in one study with a randomized, double-blind design, a fair comparison among commonly used psychotropics in bipolar depression or other conditions will never be achieved. Such a study is not impossible, but will be difficult to conduct for reasons of cost and feasibility. Without direct comparison(s) between/among psychotropics, indirect comparison of the benefits and risks relative to placebo among the atypical antipsychotics as in this review can help clinicians to understand the potential differences among these agents and choose the proper treatment. Meanwhile, future pivotal studies in bipolar depression should use a uniform design to reduce variations among studies and to provide a “fair” platform for comparisons.

Since the NNTs were similar (Table 1), proper management of side-effects will be key to maximizing the benefits and reducing the risks for DAEs. Prior studies have shown that somnolence/sedation and weight gain are common reasons for DAEs[15, 19, 20]. A recent *post-hoc* analysis of asenapine and olanzapine in the acute treatment of bipolar mania found that the median time to the onset of somnolence was 1 day for asenapine and 2 days for olanzapine and the median duration of somnolence was 7 days for asenapine and 8.5 days for olanzapine[33]. It appears reasonable that if somnolence/sedation continues after 3–4 weeks even with a lower dose, other medications should be considered.

More importantly, patients in an acute depressive phase were more sensitive to and less tolerant of antipsychotics than patients with bipolar mania or schizophrenia[7-10]. All atypical antipsychotics were superior to placebo in reducing manic symptoms with NNTs from 3 to 7[6]. In contrast to bipolar depression, all atypical antipsychotics in bipolar mania had a similar risk for DAEs[7]. The LHH of being helped by the same antipsychotic in bipolar mania would be larger than that in depression. For example, the NNTs for response to olanzapine, quetiapine, and aripiprazole in bipolar mania were 4–5, 6, and 5 respectively[6], and the NNHs for the DAEs from these 3 antipsychotics in mania were 138, −156, and −97, respectively[7], suggesting that the benefits of an antipsychotic in bipolar mania are far greater than the risks compared to the same medication in bipolar depression.

Starting with a lower dose or a slower titration may reduce the risk for DAEs. As shown in Table 1, there were minimal differences in the efficacy between lower and higher doses of lurasidone and between quetiapine-IR 300 mg/day and 600 mg/day. Clearly, for those who cannot tolerate quetiapine-IR 300 mg/day, a lower dose and/or a slower titration schedule is worth a try. On the other hand, for those who do not respond to quetiapine-IR 300 mg/day and do not have intolerable side-effects, a dose of quetiapine-IR higher than 300 mg/day should be tried since

one study showed that quetiapine-IR 600 mg/day, but not 300 mg/day, was superior to placebo in the remission rate (Table 2). More importantly, it remains unclear how these short-term side-effects, especially weight gain and EPSs, affect the long-term safety outcomes.

In addition to atypical antipsychotics, lithium, anticonvulsants, and antidepressants are commonly used off-label for acute bipolar depression[5, 6]. The results of the head-to-head comparison of OFC and lamotrigine support the use of lamotrigine as a first-line medication for acute bipolar depression[5, 19, 34]. The insignificant difference between lithium and placebo in reducing depressive symptoms (Table 1) confirms previous speculation and a finding from maintenance studies, i.e., lithium is more effective for manic symptoms than for depressive symptoms[35, 36]. Although *post-hoc* analyses using lithium concentrations of ≥0.8 or <0.8 mEq/L revealed no significant difference from placebo in the improvement of MADRS total score[23], only ~64% of patients treated with lithium had levels of 0.6–1.2 mEq/L. It remains unclear whether the subtherapeutic levels of lithium affected its overall efficacy. A previous study showed that lamotrigine adjunctive therapy to lithium was superior to lithium alone in bipolar depression[37], suggesting that these medications can be used together for acute bipolar depression.

The finding that paroxetine was not superior to placebo in reducing depressive symptoms (Table 2) is consistent with previous studies[5, 6]. However, paroxetine was only at 20 mg/day. It remains unclear if a higher dose would affect the efficacy outcome. Moreover, antidepressants are a diverse group. The efficacy of other antidepressant monotherapy in bipolar depression needs to be studied separately before a firm conclusion can be drawn. A risk for TEM (treatment-emergent mania) with paroxetine similar to the placebo should not be considered evidence of safety for using antidepressant monotherapy in bipolar depression. Patients with more frequent cycling courses and recent substance use disorder were excluded from this study[24]. These groups of patients are more likely to have TEM than those without these characteristics[38, 39]. At the present time, using antidepressants in bipolar depression should follow the recommendations of the International Society for Bipolar Disorders task force report on antidepressant use in bipolar disorders[40].

Epidemiological data showed that the prevalence of bipolar II is higher than that of bipolar I[41]. However, only the quetiapine studies included patients with bipolar II depression. Although the first pivotal study found that quetiapine-IR had a larger effect size in bipolar I than in bipolar II[20], the remaining quetiapine-IR studies demonstrated that the efficacy differences of quetiapine in bipolar I and II were much smaller[21-24], suggesting that patients with bipolar I or II depression may have a similar benefit from the same medications. In terms of safety and tolerability, one quetiapine-IR study reported that patients with bipolar II depression were more like to discontinue the study due to adverse events than those with bipolar I depression[20], suggesting the patients with bipolar II depression are less tolerant to the same medication than those with bipolar I depression.

The present study also found a disconnection between the NNT and Cohen's d effect size, two commonly used measures for evidence-based medicine. As mentioned earlier, the differences in the NNTs for response and remission among US FDA-approved agents were small. However, the Cohen's d effect sizes of active treatments *versus* placebo varied more widely. According to the original publications, the effect sizes were 0.68 for OFC[16], 0.54 to 0.81 for quetiapine-IR 300 mg/day[20, 21], 0.61 to 0.67 for quetiapine-IR 600 mg/day[20, 21], 0.61 for quetiapine-XR 300 mg/day[22], 0.51 for both lurasidone 20–60 mg/day and 80–120 mg/day[29], and 0.34 for lurasidone adjunctive therapy to mood stabilizer[30]. Clearly, the smallest effect size of lurasidone adjunctive therapy to a mood stabilizer among these US FDA-approved treatments was not consistent with its NNTs for response and remission (Table 1). In contrast, olanzapine monotherapy was not approved for acute bipolar depression, and its Cohen's d effect size and NNTs for response and remission showed a similar result, i.e., a small Cohen's d effect size and a large NNT for response and remission (Table 1).

In most cases, Cohen's d effect size and NNT supplement each other to reflect the magnitude difference between two groups. Whereas the Cohen's d therapeutic effect size is an important continuous clinical outcome measure, NNT analyses are particularly useful when attempting to quantify categorical measures such as response and remission. Some view NNT analyses as

having more pragmatic clinical relevance since response and remission may be of greater relevance to the clinician and patient[14]. Therefore, quantifying the safety and efficacy with NNH and NNT as in the present study will serve to assist the clinician during the clinical management of patients with acute bipolar depression. In addition to understanding these differences between Cohen's d effect size and NNT, clinicians should be aware of the definition of remission (≤12 on MADRS) in these and future studies. An MADRS total score of ≤12 has been regarded as partial remission[17]. More importantly, even among those defined as in remission based on a rating scale, more than half of patients still have depressive symptoms[42].

Limitations

The findings from this review are limited by not including unpublished data, which is more likely to overestimate the treatment-related benefits and underestimate the treatment-related harm because of publication biases[43]. With the exception of one study that used COSTART (Coding symbols for Thesaurus of Adverse Reaction Terms) to describe side-effects[16], all other studies used the MedDRA (Medical Dictionary for Regulatory Activities) preferred terms to describe side-effects, i.e., to split a particular event or syndrome such as somnolence, drowsiness, sedation, and sleepiness, which all probably refer to the same event, into somnolence and sedation[44]. Because of splitting reporting of somnolence and sedation, the risk for somnolence of a studied drug would be underestimated. Therefore, the NNH for somnolence in the present review is more likely larger than the NNH for somnolence if both somnolence and sedation were reported as one category. We chose the rates of response and remission as the measures of benefits, and these have been arbitrarily defined by different researchers in different studies. These measures did not reflect patients' functioning or well-being. Similarly, using DAEs as a measure of harm was also arbitrary although a majority of patients discontinued studies due to somnolence/sedation, and/or weight gain[7]. Therefore, clinicians should be cautious when interpreting LHH because of the uncertainty of these proxy measures.

Conclusions

Among the FDA-approved medications for acute bipolar depression, the differences in the NNTs for response and remission were small, but the differences in the NNHs for DAEs relative to placebo were much larger. Similarly, the risk for somnolence, ≥7% weight gain, overall EPSs, and akathisia with active treatments relative to placebo also varied widely. These data suggest that tolerability and safety should be prioritized when choosing a FDA-approved medication for acute bipolar depression. The risks and benefits should be seriously considered when prescribing off-label atypical antipsychotics and other pharmacological agents for acute bipolar depression. Since different studies included varying groups of patients, such as patients with or without psychotic features, bipolar I *versus* II depression, and other clinical correlates which could also affect the treatment response and side-effects, we should be cautious when comparing the NNTs and NNHs between different antipsychotics and different studies.

ACKNOWLEDGEMENTS

We thank Mrs. Mary Beth Serrano, a research manager, in the Mood Disorders Program at the University Hospitals Case Medical Center for her editing and proof-reading the manuscript.

Received date: 2014-06-04; Accepted date: 2015-03-04

REFERENCES

[1] Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, *et al.* The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry 2002, 59: 530–537.

[2] Judd LL, Akiskal HS, Schettler PJ, Coryell W, Endicott J, Maser JD, *et al.* A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. Arch Gen Psychiatry 2003, 60: 261–269.

[3] Calabrese JR, Hirschfeld RM, Reed M, Davies MA, Frye MA, Keck PE, *et al*. Impact of depressive symptoms compared with manic symptoms in bipolar disorder: Results of a U.S. community-based sample. J Clin Psychiatry 2004, 65: 1499–1504.

[4] Gao K, Gajwani P, Elhaj O, Calabrese JR. Typical and atypical antipsychotics in bipolar depression. J Clin Psychiatry 2005, 66: 1376–1385.

[5] Gao K, Wu R, Grunze H, Calabrese JR. Pharmacological Treatment of Acute Bipolar Depression. In: Yildiz A, Ruiz P, Nemeroff C (Eds.). Bipolar Book: History, Neurobiology, and Treatment. New York: Oxford University Press (in press).

[6] Gao K, Kemp DE, Wu R, Calabrese JR. Mood stabilizers. In: Tasman AJ, Lieberman J, Kay J, First M, Riba M (Eds.). Psyschiatry. 4th Ed. Chichester, West Sussex, UK: Wiley-Blackwell (in press).

[7] Gao K, Ganocy SJ, Gajwani P, Muzina DJ, Kemp DE, Calabrese JR. A review of sensitivity and tolerability of antipsychotics in patients with bipolar disorder or schizophrenia: focus on somnolence. J Clin Psychiatry 2008, 69: 302–309.

[8] Gao K, Kemp DE, Ganocy SJ, Gajwani P, Xia G, Calabrese JR. Antipsychotic-induced extrapyramidal side effects in bipolar disorder and schizophrenia: a systematic review. J Clin Psychopharmacol 2008, 28: 203–209.

[9] Wang Z, Kemp DE, Chan PK, Fang Y, Ganocy SJ, Calabrese JR, *et al.* Comparisons of the tolerability and sensitivity of quetiapine-XR in the acute treatment of schizophrenia, bipolar mania, bipolar depression, major depressive disorder, and generalized anxiety disorder. Int J Neuropsychopharmacol 2011, 14: 131–142.

[10] Gao K, Kemp DE, Fein E, Wang Z, Fang Y, Ganocy SJ, *et al.* Number needed to treat to harm for discontinuation due to adverse events in the treatment of bipolar depression, major depressive disorder, and generalized anxiety disorder with atypical antipsychotics. J Clin Psychiatry 2011, 72: 1063–1071.

[11] Citrome L, Ketter TA, Cucchiaro J, Loebel A. Clinical assessment of lurasidone benefit and risk in the treatment of bipolar I depression using number needed to treat, number needed to harm, and likelihood to be helped or harmed. J Affect Disord 2014, 155: 20–27.

[12] Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, *et al.* The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. Eur Neuropsychopharmacol 2006, 16: 601–611.

[13] McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. Ann Intern Med 1997, 126: 712–720.

[14] Wu R, Kemp DE, Sajatovic M, Zhao J, Calabrese JR, Gao K. Communication of potential benefits and harm to patients and payers in psychiatry: a review and commentary. Clin Ther 2011, 33: B62–76.

[15] Cohen LS. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum, 1988.

[16] Tohen M, Vieta E, Calabrese J, Ketter TA, Sachs G, Bowden C, *et al.* Efficacy of olanzapine and olanzapine-fluoxetine combination in the treatment of bipolar depression. Arch Gen Psychiatry 2003, 60: 1079–1088.

[17] Tohen M, McDonnell DP, Case M, Kanba S, Ha K, Fang YR, *et al.* Randomised, double-blind, placebo-controlled study of olanzapine in patients with bipolar I depression. Br J Psychiatry 2012, 201: 376–382.

[18] Katagiri H, Tohen M, McDonnell DP, Fujikoshi S, Case M, Kanba S, *et al.* Efficacy and safety of olanzapine for treatment of patients with bipolar depression: Japanese subpopulation analysis of a randomized, double-blind, placebo-controlled study. BMC Psychiatry 2103, 13: 138.

[19] Brown EB, McElroy SL, Keck PE Jr, Deldar A, Adams DH, Tohen M, *et al.* A 7-week, randomized, double-blind trial of olanzapine/fluoxetine combination versus lamotrigine in the treatment of bipolar I depression. J Clin Psychiatry 2006, 67: 1025–1033.

[20] Calabrese JR, Keck PE Jr, Macfadden W, Minkwitz M, Ketter TA, Weisler RH, *et al.* A andomized, double-blind, placebo-controlled trial of quetiapine in the treatment of bipolar I or II depression. Am J Psychiatry 2005, 162: 1351–1360.

[21] Thase ME, Macfadden W, Weisler RH, Chang W, Paulsson B, Khan A, *et al.* Efficacy of quetiapine monotherapy in bipolar I and II depression: a double-blind, placebo-controlled study (the BOLDER II study). J Clin Psychopharmacol 2006, 26: 600–609.

[22] Suppes T, Datto C, Minkwitz M, Nordenhem A, Walker C, Darko D. Effectiveness of the extended release formulation of quetiapine as monotherapy for the treatment of acute bipolar depression. J Affect Disord 2010, 121: 106–115.

[23] Young AH, McElroy SL, Bauer M, Philips N, Chang W, Olausson B, *et al.* A double-blind, placebo-controlled study of quetiapine and lithium monotherapy in adults in the acute phase of bipolar depression (EMBOLDEN I). J Clin Psychiatry 2010, 71: 150–162.

[24] McElroy SL, Weisler RH, Chang W, Olausson B, Paulsson B, Brecher M, *et al.* A double-blind, placebo-controlled study of quetiapine and paroxetine as monotherapy in adults with bipolar depression (EMBOLDEN II). J Clin Psychiatry 2010, 71: 163–174.

[25] Thase ME, Jonas A, Khan A, Bowden CL, Wu X, McQuade RD, *et al.* Aripiprazole monotherapy in nonpsychotic bipolar I depression: results of 2 randomized, placebo-controlled studies. J Clin Psychopharmacol 2008, 28: 13–20.

[26] Lombardo I, Sachs G, Kolluri S, Kremer C, Yang R. Two 6-week, randomized, double-blind, placebo-controlled studies of ziprasidone in outpatients with bipolar I depression: did baseline characteristics impact trial outcome? J Clin Psychopharmacol 2012, 32: 470–478.

[27] Gao K, Pappadopulos E, Karayal ON, Kolluri S, Calabrese JR. Risk for adverse events and discontinuation due to adverse events of ziprasidone monotherapy relative to placebo in the acute treatment of bipolar depression, mania, and schizophrenia. J Clin Psychopharmacol 2013, 33: 425–431.

[28] Sachs GS, Ice KS, Chappell PB, Schwartz JH, Gurtovaya

O, *et al*. Efficacy and safety of adjunctive oral ziprasidone for acute treatment of depression in patients with bipolar I disorder: a randomized, double-blind, placebo-controlled trial. J Clin Psychiatry 2011, 72: 1413–1422.

[29] Loebel A, Cucchiaro J, Silva R, Kroger H, Hsu J, Sarma K, *et al*. Lurasidone monotherapy in the treatment of bipolar i depression: a randomized, double-blind, placebo-controlled study. Am J Psychiatry 2014, 171: 160–168.

[30] Loebel A, Cucchiaro J, Silva R, Kroger H, Sarma K, Xu J, *et al*. Lurasidone as adjunctive therapy with lithium or valproate for the treatment of bipolar i depression: a randomized, double-blind, placebo-controlled study. Am J Psychiatry 2014, 171: 169–177.

[31] Citrome L, Kantrowitz J. Antipsychotics for the treatment of schizophrenia: likelihood to be helped or harmed, understanding proximal and distal benefits and risks. Expert Rev Neurother 2008, 8: 1079–1091.

[32] Tohen M, Katagiri H, Fujikoshi S, Kanba S. Efficacy of olanzapine monotherapy in acute bipolar depression: a pooled analysis of controlled studies. J Affect Disord 2013, 149: 196–201.

[33] Gao K, Mackle M, Cazorla P, Zhao J, Szegedi A. Comparison of somnolence associated with asenapine, olanzapine, risperidone, and haloperidol relative to placebo in patients with schizophrenia or bipolar disorder. Neuropsychiatr Dis Treat 2013, 9: 1145–1157.

[34] Suppes T, Dennehy EB, Hirschfeld RM, Altshuler LL, Bowden CL, Calabrese JR, *et al*. The Texas implementation of medication algorithms: update to the algorithms for treatment of bipolar I disorder. J Clin Psychiatry 2005, 66: 870–886.

[35] Bowden CL, Calabrese JR, Sachs G, Yatham LN, Asghar SA, Hompland M, *et al*. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. Arch Gen Psychiatry 2003, 60: 392–400.

[36] Calabrese JR, Bowden CL, Sachs G, Yatham LN, Behnke K, Mehtonen OP, *et al*. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. J Clin Psychiatry 2003, 64: 1013–1024.

[37] van der Loos ML, Mulder PG, Hartong EG, Blom MB, Vergouwen AC, de Keyzer HJ, *et al*. Efficacy and safety of lamotrigine as add-on treatment to lithium in bipolar depression: a multicenter, double-blind, placebo-controlled trial. J Clin Psychiatry 2009, 70: 223–231.

[38] Wu R, Gao K, Calabrese JR, Grunze H. Treatment induced mood instability: treatment – emergent affective switches and cycle acceleration. In: Yildiz A, Ruiz P, Nemeroff C (Eds.). Bipolar Book: History, Neurobiology, and Treatment. New York, USA: Oxford University Press (in press).

[39] Chen J, Fang Y, Kemp DE, Calabrese JR, Gao K. Switching to hypomania and mania: differential neurochemical, neuropsychological, and pharmacologic triggers and their mechanisms. Curr Psychiatry Rep 2010, 12: 512–521.

[40] Pacchiarotti I, Bond DJ, Baldessarini RJ, Nolen WA, Grunze H, Licht RW, *et al*. The International Society for Bipolar Disorders (ISBD) task force report on antidepressant use in bipolar disorders. Am J Psychiatry 2013, 170: 1249–1262.

[41] Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, *et al*. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry 2007, 64: 543–552.

[42] Zimmerman M, Martinez JA, Attiullah N, Friedman M, Toba C, Boerescu DA, *et al*. Why do some depressed outpatients who are in remission according to the Hamilton Depression Rating Scale not consider themselves to be in remission? J Clin Psychiatry 2012, 73: 790–795.

[43] Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008, 358: 252–260.

[44] Food and Drug Administration Center for Drug Evaluation and Research (CDER). Review Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review. Rockville, Maryland, USA: Food and Drug Administration, 2005.