Evidence-based Practice Center Systematic Review Protocol

Project Title: Treatment for Bipolar Disorder

I. Background and Objectives for the Systematic Review

Bipolar disorder is a serious mental illness. Prevalence studies estimate about 1 to 4% of the population has bipolar disorder, with relatively similar prevalence in men and women, and across cultural and ethnic groups.[1,2] Recurrent episodes of mania and depression can cause serious impairments in functioning and psychosocial morbidity.[3,4] People with Bipolar disorder have an increased risk of suicide; between 25% and 50% of bipolar disorder patients will attempt suicide.[5] Substance abuse is also a common comorbid condition; of all psychiatric clinical disorders, bipolar disorder is the most likely to co-occur with alcohol or drug abuse.[6] The disease burden is heavy, with lifelong treatment requirements.

Bipolar disorder, also known as manic-depressive illness, is a disorder that causes unusual shifts in mood, energy, activity levels, and the ability to carry out day-to-day tasks. According to the DSM 5, Bipolar I disorder is mainly defined by the presence of manic or mixed episodes that last at least seven days, or by manic symptoms that are so severe that the person needs immediate hospital care. Usually, the person also has depressive episodes, typically lasting at least two weeks. The symptoms of mania or depression must be a major change from the person's normal behavior. For a bipolar I, mixed episodes diagnosis, the DSM-5 removes language specifying that individuals meet the full criteria for both mania and a major depressive episode, and instead adds a new specifier, "with mixed features", that can be applied to episodes of mania or hypomania when depressive features are present, and to episodes of depression when features of mania/hypomania are present.[7,8] The associated symptom of psychosis can also shift the episode type from hypomania to mania. Bipolar II disorder is defined by a pattern of depressive episodes shifting back and forth with hypomanic episodes, but no full-blown manic or mixed episodes.

Treatment of bipolar disorder generally begins with the goal of bringing a patient with mania or depression to symptomatic recovery and stable mood. Once stable, the goal progresses to reduction of subthreshold symptoms and relapse prevention. Pharmacologic treatment is challenging because treatments that alleviate depression can cause mania, hypomania, or rapid cycling (four or more episodes in 12 months), and treatments that alleviate mania may cause rebound depressive episodes. Nonpharmacologic psychotherapeutic techniques are applied to enhance medication adherence, reduce episode relapse, ameliorate the psychosocial and relationship damage that can occur with acute episodes, and to improve social and occupational functioning. However, treatment decisions are made more complex by differential responses to treatment by bipolar type (type I or type II), polar episode (depression or mania), and phase of treatment. Further, the characteristics of the disorder itself can create challenges to patient readiness and adherence to treatment plans. Layered over this is the greater complexity of delivering pharmacologic and nonpharmacologic treatments in a coordinated and integrated fashion. Examples of pharmacologic treatments are summarized in Table 1. Recently, comprehensive treatment programs that provide multicomponent treatments

incorporating pharmacological, psychological, and social components in an integrated fashion have been developed to address this complexity and improve how treatment is delivered in clinical practice.

When choosing a treatment option, patients and practitioners need to consider the balance of benefits and harms of treatments from the alternative treatment options. For example, when considering a specific treatment option versus no treatment, they would need to consider the risks from direct harms from treatment against the harms that might come from not treating the patients.[9] Of particular importance to bipolar disorder is the difference between "direct" harms versus the absence of benefits. We define a "direct" harm as harm a patient risks by virtue of receiving the treatment and is part of the trade-off considered in establishing a net benefit. Absence of benefit, on the other hand, can include worsening of the condition because the patient was not receiving a treatment that would provide a benefit. Another way to look at this issue is from the perspective of direct harm increasing the burden of treatment as oppose to the absence of benefit increasing the burden of illness.

A considerable body of systematic reviews, both completed and in progress, on bipolar disorders exists. The systematic review literature itself rests on a large body of experimental literature. However, no systematic review of comprehensive treatment programs has been proposed or conducted. Further, the value of the published systematic reviews is bounded by the discrepancy between the limited experimental protocols found in the empirical literature testing individual treatment components in isolation and the coordinated nature of treatment in natural treatment settings.

Bipolar treatment is an active research area. New treatment approaches in the last decade include new medications and adjunctive psychological approaches such as cognitive behavioral therapy or family-focused therapy, and comprehensive programs. Clinicians and patients face numerous decisions regarding treatment. These decisions take the following general form: 1) What are the right pharmacologic or somatic treatments for this individual?; 2) What can we do to mitigate the side effects of pharmacological treatments that create other medical risks?; 3) Shall we provide anything other than the medications typically prescribed for this individual?; and 4) Shall we add a specific nonpharmacological intervention for this individual in place of the counseling/psychotherapy they are currently receiving?

For this review, we draw several scope boundaries to maintain a focus on treatments for adults with bipolar disorder most likely to be provided in psychiatric or comprehensive care settings.

- Exclude botanicals and nutritional supplements. These are part of a broader class of remedies patients may take on their own for symptom relief. It does remain an important topic. These forms of therapy may interact with prescribed medications. For example, St. John's work may cause switching to mania.[10]
- Exclude caregiver outcomes. There is some evidence that a person with bipolar disorder has a harder time following a treatment plan if the caregiver is under stress.[11]

Table 1. FDA approved medications for bipolar disorder

Generic Name	Trade Name (Pharmaceutical Co)	Manic	Mixed	Mainte-nance	Depression
Mood Stabilizers					
Lithium*◊		X		X	
Chlorpromazine*◊	Sonazine (Sandoz), Promapar (Parke Davis)	X			
Anticonvulsants					
Divalproex sodium*◊ or valproate	Depakote (ABBVIE)	X			
Lamotrigine*◊	Lamictal (GlaxoSmithKline)			X	
Carbamazepine*◊	Carbetrol (Shire), Epitol (TEVA), Equetro (Validus Pharms), Tegretol (Novartis), Teril (Taro)	X	X		
Atypical Antipsychotics					
Aripiprazole*◊	Abilify (Otsuka)	X	X	X	
Ziprasidone*◊	Geodon (Pfizer)	X	X		
Risperidone*◊	Risperdal (Janssen Pharm)	X	X		
Asenapine◊	Saphris (Organon Sub Merck)	X	X		
Quetiapine*◊	Seroquel (Astrazeneca)	X			X
Olanzapine*◊	Zyprexa (Lilly)	X	X	X	
Olanzapine/fluoxetine combination*◊	Symbyax (Lilly)				X
Lurasidone◊	Latuda (Sunovion Pharms)				X
Offset drug side-effects					
Metformin*				X	
Verapamil*				X	

*=generic forms available ◊=black box warnings

II. The Key Questions

Key questions, PICOT, and analytic framework were posted for public comment from December 19, 2013 to January 10, 2014. In response to comments provided, we made several changes. We separated physically-based somatic treatments, such as electroconvulsive therapy, into a third category of treatments. There was disagreement whether somatic treatments best fit with pharmacologic treatment, as suggested by the commenter, or nonpharmacologic treatments, as suggested by several team members. We also reduced key question 2 to simple statements of harms without regard to comparators. Finally, we made adjustments to patient characteristics in key question 3 by removing types of mania (this is well-covered by including bipolar types), and adding age, race/ethnicity, and SES to address one commenters request for the addition of cultural factors and likely to have been examined in the literature.

Key Question 1: What is the efficacy and comparative effectiveness of pharmacologic and nonpharmacologic treatments for adults with bipolar disorder?

    a. How do pharmacologic treatments (monotherapy or combination therapies) affect patient centered outcomes when compared with placebo?
    b. How do pharmacologic treatments (monotherapy or combination therapies) affect patient centered outcomes when compared with other active pharmacologic treatment?
    c. How do behavioral health treatments (psychotherapy, psychosocial interventions) affect patient centered outcomes when compared with usual care?
    d. How do behavioral health treatments (psychotherapy, psychosocial interventions, chronotherapy) affect patient centered outcomes when compared with other active treatment?
    e. How do somatic treatments (electroconvulsive therapy (ECT), transcranial magnetic stimulation (TMS)) affect patient-centered outcomes when compared with other active treatment?
    f. How do comprehensive programs affect patient centered outcomes when compared with usual care?

Key Question 2: What are the harms from pharmacologic and nonpharmacologic treatments for adults with bipolar disorder?

    a. What are the harms from pharmacologic treatments?
    b. What are the harms from behavioral health treatments?
    c. What are the harms from somatic treatments?
    d. What are the harms from comprehensive programs?

Key Question 3: What is the effectiveness of treatments to reduce the metabolic change (metabolic syndrome, glucose dysregulation, weight gain) side effects of first line pharmacologic treatments?

Key Question 4: Which patient characteristics predict the effectiveness and harms of pharmacologic and nonpharmacologic treatments for people with bipolar disorder, including disease-specific characteristics such as bipolar type, phase severity, pediatric onset, new onset, treatment resistant, types of depression, and other comorbidities and patient characteristics such as substance use, other psychiatric comorbidities, medical comorbidities, age, sex, race/ethnicity, socioeconomic status?

Key Questions 1- 3 will be examined within the context of Key Question 3. Table 2 shows, at the categorical level, acceptable comparisons.

Table 2. Includable comparisons

	Placebo/ Waitlist/ Usual Care	Pharma-cologic	Behavioral	Somatic	Comprehensive	Interventions to Reduce Side Effects
Pharmacologic	X	X				
Behavioral	X	X	X			
Somatic	X	X		X		

Comprehensive	X	X	X		X	
Interventions to Reduce Side-Effects	X					X

PICOTS

Table 3. PICOTS

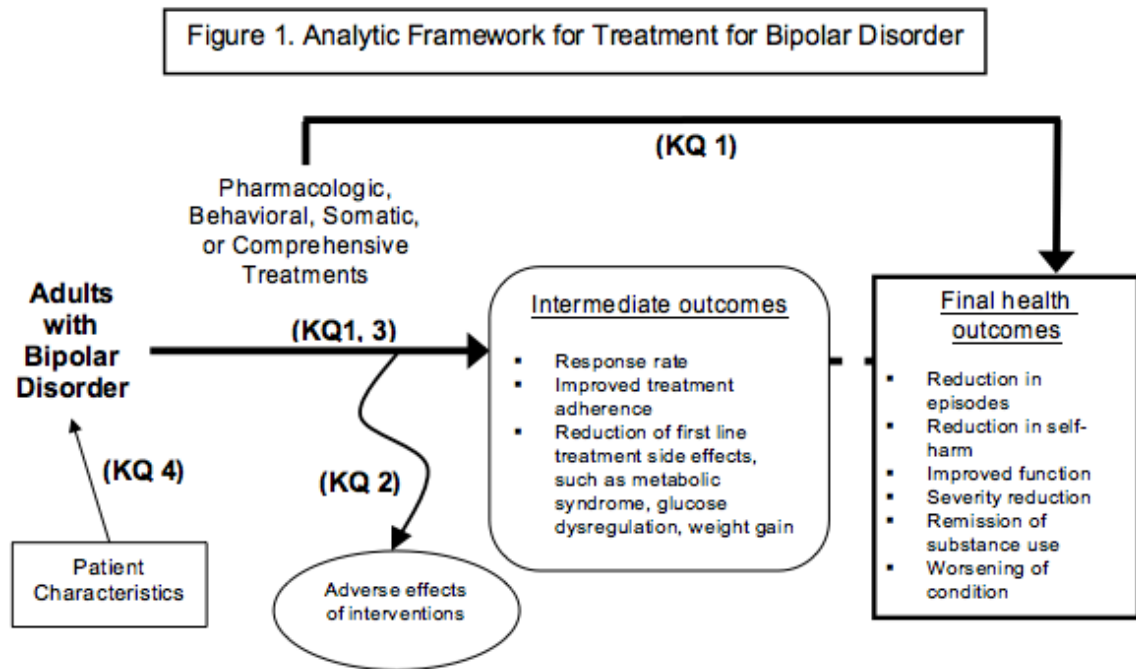| PICOTS	Included	Excluded
Population	Adults, 18+ years old, with any bipolar disorder. Includes pregnant women	Pediatric bipolar patients Studies with samples of greater than 25% identified as schizoaffective with bipolar symptoms. Schizoaffective disorder is distinguished by higher psychotic symptoms than bipolar disorder.

Intervention	Pharmacologic treatment
	• Manic episodes – lithium, anticonvulsants, antipsychotics
	• Depressive or mixed episodes – lithium, anticonvulsants, antipsychotics, antidepressants
	• Maintenance phase – lithium, anticonvulsants, antipsychotics, antidepressants
	• Combination therapy –
	   o Two or more medications begun simultaneously with similar therapeutic goal;
	   o Augmentation with a second medication to boost response when patient's symptoms have only partially remitted
	   o Two medications with different goals
	
	Behavioral treatment
	• Psychotherapy, such as cognitive behavioral therapy (CBT)
	• Family-focused therapy
	• Interpersonal and social rhythm therapy
	• Psychoeducation
	• Chronotherapy
	
	Somatic treatment treatment
	• Electroconvulsive therapy (ECT)
	• Transcranial magnetic stimulation (TMS)
	
	Comprehensive programs – multicomponent programs incorporating pharmacological, psychological, and social components in an integrated fashion.
	
	Interventions to reduce side effects of medications given for prolonged periods (metabolic syndrome, glucose dysregulation, weight gain) (such as verapamil, metformin)
		Over-the-counter botanicals, nutritional supplements, dietary approaches (including omega 3)
		
		Programs designed only as treatment plus adherence only, where it is not therapeutic but only adherent.

Comparator groups	Pharmacologic treatment – placebo, active control Behavioral or Somatic treatment – placebo/sham, usual care, or active control Comprehensive treatment – placebo or active	

	control	
	Interventions – placebo, waitlist, active control, usual care	
Outcomes	Final health or patient-centered outcomes: Reduction of episodes family • Remission/Prevention of episodes • Increased time between episodes/Time to remission • Reduced hospitalization • Prevention of episodes Reduction in self-harm • Reduction in suicide • Reduction in suicidal thoughts or self-harming behaviors Improved function • Improved social and occupational functioning • Change in disability • Health related quality of life Severity reduction Remission of co-occurring substance use disorder Worsening of condition Intermediate outcomes • Treatment response • Improved treatment adherence • Reduction of first line treatment side effects (metabolic syndrome, glucose dysregulation, weight gain) Adverse effects of interventions • Switching phases • Increase metabolic syndrome, glucose dysregulations, weight gain • Reported adverse effects	Time to drug effect Drug tolerance studies; phase II studies All other intermediate outcomes, such changes in physiologic conditions
Timing	Acute mania/mixed episode: at least 3 weeks Acute depression: at least 3 months Maintenance or prevention: at least 6 months	
Setting	Inpatient and outpatient for mania or mixed episodes. Outpatient for depression, maintenance, and prevention.	

III. Analytic Framework

Figure 1. Analytic Framework for Treatment for Bipolar Disorder

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

Studies will be included in the review based on the PICOTS framework outlined in Table 3 and the study-specific inclusion criteria described in Table 4.

Table 4. Study inclusion criteria

Category	Criteria for Inclusion
Study Enrollment	Studies that enroll adults with any form of bipolar disorder (Bipolar I, Bipolar II, Bipolar otherwise specified, Bipolar not otherwise specified, rapid cycling) using any diagnostic process. Studies that enroll bipolar disorder patients along with other patients with DSM-V diagnoses will be included if the bipolar patients are analyzed separately.
Study Design and Quality	Systematic reviews, RCTs, nonrandomized controlled trials, and prospective cohort studies will be included for each population and treatment option. Prospective studies must include a comparator and appropriate methods to correct for selection bias. Studies specifically addressing treatment harms may also include retrospective and case series designs. Systematic reviews must include risk of bias assessment with validated tools. Observational studies that do not adequately report study information to allow the abstraction of time sequences for treatment and followup duration or have indeterminable numerators and denominators for outcomes and adverse event

	rates will be excluded at the abstraction phase.
Time of Publication	1970 forward for trials of pharmacologic and somatic treatments. Lithium was FDA approved in 1970. 1994 forward for all other literature, including systematic reviews. This corresponds with the period during which systematic reviews and evidence-based research approaches have been applied to behavioral health.
Publication type	Published in peer reviewed journals
Language of Publication	English

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions

We will search Ovid Medline, Ovid PsycInfo, Ovid Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL) to identify previous systematic reviews, randomized controlled trials, and prospective cohort studies published and indexed in bibliographic databases. Our search strategy, which appears in Appendix A, was created by staff and a biomedical librarian, and reviewed by a second independent librarian. Our search strategy included relevant medical subject headings and natural language terms for the concept of bipolar disorder. This concept was combined with filters to select RCTs, observational studies, and systematic reviews.

We will search for systematic reviews published since 1994. We anticipate that older, established treatments will be covered by prior reviews. (See Risk of Bias section below for discussion of quality assessments of systematic reviews.) We will also search for RCTs and prospective cohort studies published since 1994. For those older, established treatments without a prior high quality review, we will perform targeted searches for studies published since 1970. We will supplement these systematic reviews with backward citation searches of relevant systematic reviews. For newer treatments, which are less likely to have a prior systematic review, we will search for RCTs and prospective cohort studies published since 1994. We will update searches while the draft report is under public/peer review.

Studies reporting treatment for side effects or treatment harms will be retained if they report measures of widely recognized clinical relevance for primary diagnostic or monitoring, such as HbA1c levels or fasting glucose. The measures should have utility, powerful enough for the individual patient. Studies attempting to develop or establish new etiologic pathways are outside of scope.

We will review bibliographic database search results for studies relevant to our PICOTS framework and study-specific criteria. Search results will be downloaded to EndNote. Titles and abstracts will be reviewed by two independent investigators to identify studies meeting PICOTS framework and inclusion/exclusion criteria. All studies identified as relevant by either investigator will undergo full-text screening. Two investigators will independently screen full text to determine if inclusion criteria are met. Differences in screening decisions will be resolved by consultation between investigators, and, if necessary, consultation with a third investigator. We will document the inclusion and exclusion status of citations undergoing full-text screening. Throughout the screening process, team members will meet regularly to discuss training material and issues as they arise to ensure consistency of inclusion criteria application.

We will conduct additional grey literature searching to identify relevant completed and ongoing studies. Relevant grey literature resources include trial registries and FDA databases. We will search ClinicalTrials.gov and the International Controlled Trials Registry Platform (ICTRP) for ongoing studies. We will also review Scientific Information Packets (SIPs) sent by manufacturers of relevant interventions. Grey literature search results will be used to identify studies, outcomes, and analyses not reported in the published literature to assess publication and reporting bias and inform future research needs.

C. Data Abstraction and Data Management

Studies meeting inclusion criteria will be distributed among investigators for data extraction. One investigator will extract relevant study, population demographic, and outcomes data. Data fields to be extracted will be determined based upon proposed summary analysis. These fields will include author, year of publication; setting, subject inclusion and exclusion criteria, intervention and control characteristics (intervention components, timing, frequency, duration), followup duration, participant baseline demographics, comorbidities; method of diagnosis, enrollment, and severity, descriptions and results of primary outcomes and adverse effects, and study funding source. Relevant data will be extracted into web-based extraction forms created in Xcel. Data will be exported into Excel spreadsheets for descriptive analysis. Data will be analyzed in RevMan 5.21[12] software. Evidence tables will be reviewed and verified for accuracy by a second investigator.

We will catalogue one-off studies of interventions (only one study examining a particular intervention) with low sample size. These studies will be made available in an appendix, indexed by intervention type. However, they will not be abstracted due to the inability to determine a strength of evidence beyond insufficient. The report text will provide a simple map of the literature in this category.

Systematic reviews determined to have fair or good methodology (see Section D below) will be used to replace de novo data extraction processes for specific population/treatment/outcome comparisons that are sufficiently relevant. Systematic reviews of fair or good quality that are deemed to have potential author conflict of interest, such as due to reviewing a body of literature to which the authors had substantially contributed, will be subjected to random quality checks of 10% of included study data abstraction. Individual studies in included systematic reviews will be tracked for contribution to unique population/treatment/outcome comparisons to avoid double-counting study results.

D. Assessment of Methodological Risk of Bias of Individual Studies

Risk of bias of eligible studies will be assessed using instruments specific to study design. For RCTs, questionnaires developed from the Cochrane Risk of Bias tool will be used. The seven domains included in this tool include sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data (i.e., was incomplete outcome data adequately addressed), selective reporting, and other sources of bias (i.e., problems not covered by other domains). For behavioral health trials, the presence of treatment fidelity, that is, treatment

definition and implementation, will also be evaluated. Outcome measurement issues inherent in the psychometric properties of the questionnaires used to measure outcomes and assessment methods used to detect change in those questionnaire results will be specifically evaluated for detection bias. Specific study methodology or conduct will be used to judge potential risk of bias with respect to each domain following guidance in the Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0.[13]

We developed an instrument for assessing risk of bias for observational studies based on the RTI Observational Studies Risk of Bias and Precision Item Bank.[14] We selected items most relevant in assessing risk of bias for this topic, including participant selection; attrition, ascertainment, and appropriateness of analytic methods. The preliminary risk of bias assessment form is provided in Appendix B. The form will be tested by investigators using an initial sample of included studies and will be finalized by full team input.

Two investigators will independently assess risk of bias for all included studies. Investigators will consult to reconcile any discrepancies in overall risk of bias assessments. Overall summary risk of bias assessments for each study will be classified as low, moderate, or high based upon the collective risk of bias inherent in each domain and confidence that the results are believable given the study's limitations. When the two investigators disagree, a third party will be consulted to reconcile the summary judgment.

Systematic reviews that assess risk of bias for included individual studies will be assessed for review quality.[15] Study-level risk of bias must be assessed using validated risk of bias tools appropriate to study design. Systematic review quality and risk of bias will be assessed using modified AMSTAR criteria. An additional question regarding the appropriateness of the review findings given the contributing studies will be added.

E. Data Synthesis

We will summarize the results into evidence tables and synthesize evidence for each unique population, comparison, and outcome combination. When a comparison is adequately addressed by a previous systematic review of acceptable quality and no new studies are available, we will reiterate the conclusions drawn from that review. When new trials are available, previous systematic review data will be synthesized with data from the additional trials.

We will summarize included study characteristics and outcomes in evidence tables. If available, observational literature examining treatment benefits will be used for treatments or subgroups not adequately addressed by published RCTs. Using a random effects model, we will calculate risk ratios (RR) and absolute risk differences (RD) with the corresponding 95 percent confidence intervals (CI) for binary primary outcomes. Weighted mean differences (WMD) and/or standardized mean differences (SMD) with the corresponding 95 percent CIs will be calculated for continuous outcomes. We will assess the clinical and methodological heterogeneity and variation in effect size to determine appropriateness of pooling data.[16] If data are appropriate for pooling, meta-analysis will be performed.

We will assess statistical heterogeneity with Cochran's Q test and measure magnitude with I^2 statistic.[16] When direct evidence on certain comparisons is not available, indirect comparison will be explored.[16] When pooling is not appropriate due to lack of comparable studies or heterogeneity, qualitative synthesis will be conducted. Decisions for pooling will be based on the homogeneity of study populations based on inclusion

criteria, likely match of diagnostic processes, specific interventions, and the ability to treat outcome measures as similar.

We will use minimum important differences (MIDs) to assess the efficacy and comparative effectiveness of outcomes with well-established MIDs. When standard MIDs for a particular outcome is not available, we will use a statistical difference to assess efficacy and comparative effectiveness and calculate the minimum detectable difference that the data allowed (β=.8, α=.05). We will calculate MID for outcomes for which strength of evidence is assessed.

Results will be organized by bipolar type, phase, and patient characteristics. Tables 5-7 provide examples of this framework. Each of the tables will be created separately for Bipolar I, Bipolar II, Rapid Cycling, Bipolar Not-Otherwise-Specified, and mixed populations (where appropriate) and populated with relevant outcome information. Further levels of detail for intersections of different patient characteristics will be determined as the literature is examined. Some columns or cells may be exploded into their own matrix, for example phase severity, again depending on the literature available.

Table 5. Outcomes by age of onset and bipolar symptom severity

	New Onset (i.e., 1st episode)	Pediatric onset	Late Onset (BPVI)	Types of Depression	Phase severity
Acute Mania episode				--	
Acute Depression episode	--				
Maintenance					
Prevention of episodes					
Mixed Episodes				--	

Note: shaded cells not valid group.

Table 6. Outcomes by patient demographic characteristics

	Adult 18-64	Adult 65+	Sex	Adult Pregnant Women	Race/ Ethnicity	SES
Acute Mania episode						
Acute Depression episode						
Maintenance						
Prevention of episodes						
Mixed Episodes						

Table 7. Outcomes by comorbidity and treatment resistance

	Lithium Treatment Resistant	Comorbid Substance Abuse	Other Comorbid psychiatric	Comorbid medical
Acute Mania episode				
Acute Depression episode				
Maintenance				
Prevention of episodes				
Mixed Episodes				

We will also conduct several sensitivity analyses. Outcomes in studies assessed as having a high risk of bias will be compared to synthesized evidence as a means of sensitivity analysis. Contradictions will be investigated in further depth.

Similar to sensitivity analysis by risk of bias score, we will conduct sensitivity analyses by the level of diagnostic accuracy of bipolar disorders. Diagnosing bipolar disorder is challenging because of the lack of specificity of many of the symptoms, the need to understand the course of the illness symptoms over time, and the lack of objective measures to confirm diagnosis.[17] Further, several secondary symptoms of mania are shared by other psychiatric diagnoses, and bipolar is commonly both over and under-diagnosed.[17] Patients themselves can be biased toward a bipolar disorder diagnosis.[17] Therefore, a detailed understanding of the diagnostic processes used to establish study samples have implications for the ability to find a signal for effectiveness, if one such exists, as well as the applicability of the results.

In order to accomplish this sensitivity analysis, we developed a method of categorizing study diagnostic assessment processes, resulting in overall summary scores of Excellent, Good, Fair, and Poor. (See Appendix C for the assessment tool). The diagnostic process assessment tool incorporates information on what diagnostic tools the studies used, the reliability of the study diagnostic assessment process, the diagnostic criteria, evidence for the ability of the diagnostic raters, and the sources of information used for establishing the diagnosis. Using a set of 18 examples drawn from includable primary trials, the tool was piloted by the TEP members for face validity, usability, and evidence for the ability to discriminate between assessment processes. The tool was then revised based on feedback from the TEP.

We will also check for heterogeneity in results based on diagnostic criteria, that is, whether diagnoses were based on DSM-IIIR or DSM-IV.

We will assess harms as dichotomous variables to acknowledge the inherent difficulties of assessing harms, and also to simplify analysis. There are various ways to assess harms; each has problems. One can use RCT and controlled cohort data, but they generally have small samples, short follow-ups. One can use case series, but they have no controls and the rate of "adverse events" among persons getting placebos is high.[18] One can use case-control studies, but they are subject to recall bias. One can examine the general experience with the intervention, but this does not exclude the possibility that persons with the target condition have different susceptibilities. We will use reported harms from RCTs, prospective cohort, retrospective case-control, and case series.

F. Grading the Strength of Evidence (SOE) for Major Comparisons and Outcomes

The overall strength of evidence for primary outcomes of KQ1 within each comparison will be evaluated based on four required domains: (1) study limitations (risk of bias); (2) directness (single, direct link between intervention and outcome); (3) consistency (similarity of effect direction and size); and (4) precision (degree of certainty around an estimate).[19] A fifth domain, reporting bias, will be assessed when SOE based upon the first four domains is moderate or high.[19] Based on study design and conduct, risk of bias will be rated as low, medium, or high. Consistency will be rated as consistent, inconsistent, or unknown/not applicable (e.g., single study) based on the direction, magnitude, and statistical significance of all studies. Directness will be rated as either

direct or indirect based on the need for indirect comparisons when inference requires observations across studies. That is, more than one step is needed to reach the conclusion. Precision will be rated as precise or imprecise based on the degree of certainty surrounding each effect estimate or qualitative finding. An imprecise estimate is one for which the confidence interval is wide enough to include clinically distinct conclusions. The potential for reporting bias, when assessed, will be evaluated by the potential for publication bias, selective outcome reporting bias, and selective analysis reporting bias. Other factors that may be considered in assessing strength of evidence include dose-response relationship, the presence of confounders, and strength of association.

Based on these factors, the overall strength of evidence for each outcome will be rated as:[19]

- **High:** Very confident that estimate of effect lies close to true effect. Few or no deficiencies in body of evidence, findings believed to be stable.
- **Moderate:** Moderately confidence that estimate of effect lies close to true effect. Some deficiencies in body of evidence; findings likely to be stable, but some doubt.
- **Low:** Limited confidence that estimate of effect lies close to true effect; major or numerous deficiencies in body of evidence. Additional evidence necessary before concluding that findings are stable or that estimate of effect is close to true effect.
- **Insufficient:** No evidence, unable to estimate an effect, or no confidence in estimate of effect. No evidence is available or the body of evidence precludes judgment.

We will assess strength of evidence for published systematic reviews replacing de novo review processes that did not provide a strength of evidence assessment based on a GRADE or GRADE-equivalent method and incorporating all relevant articles, including new articles identified in bridge searches. For prior systematic reviews that did provide acceptable strength of evidence, the impact of new articles on the overall body of evidence will take into consideration the differences in strength of evidence domains and the relative contributions of the prior review and the new articles.

We will assess strength of evidence for validated scales (such as the Beck Depression Inventory, Young Mania Rating Scale, Hamilton Depression Rating Scale, Clinical Global Improvement Scale) and commonly used items that examine improved function (such as the WHOQOL-BREF or the Functional Assessment Short Test). We will not assess strength of evidence for less commonly measured items such as increased time between episodes or hospitalizations. Attempted suicide and other self-harming behaviors will also not be assessed for strength of evidence due to the difficulty of defining and measuring such behaviors.

G. Assessing Applicability

Bipolar research generally draws from highly defined populations, resulting in samples that are often drawn from subpopulations rather than the bipolar populations at large. Thus, the ability to infer generalizability can be compromised. Applicability of studies will be determined according to the PICOTS framework. Study characteristics that may affect applicability include, but are not limited to, the population from which the study participants are enrolled, diagnostic assessment processes, narrow eligibility

criteria, and patient and intervention characteristics different than those described by population studies of bipolar disorder.[20] These applicability issues are present in the synthesis frameworks and sensitivity analyses described in more detail in the data synthesis section.

V. References

1. Gum, A.M., King-Kallimanis, B., Kohn, R., 2009. Prevalence of mood, anxiety, and substance-abuse disorders for older Americans in the national comorbidity survey-replication. The American Journal of Geriatric Psychiatry 17, 769–781.
2. Ferrari, A. J., et al. (2011). "A systematic review of the global distribution and availability of prevalence data for bipolar disorder." Journal of Affective Disorders **134**(1-3): 1-13.
3. Samame, C., et al. (2012). "Social cognition in euthymic bipolar disorder: systematic review and meta-analytic approach." Acta Psychiatrica Scandinavica **125**(4): 266-280.
4. Sole, B., et al. (2011). "Are bipolar II patients cognitively impaired? A systematic review." Psychological Medicine **41**(9): 1791-1803
5. Valtonen H, Suominen K, Partonen T, Ostamo A, Lonnqvist J. Time patterns of attempted suicide. J Affect Disord 2006; 90: 201–207
6. Brady, KT., Sonne, SC. (1995) "The relationship between substance abuse and bipolar disorder." Journal of Clinical Psychiatry 56 Suppl 3: 19-24
7. Bipolar Disorder. National Institute of Mental Health, retrieved Oct 10, 2013. Available from: http://www.nimh.nih.gov/health/publications/bipolar-disorder/complete-index.shtml
8. DSM-5 Development, American Psychiatric Association retrieved Oct 15, 2013. Available from: http://www.dsm5.org/Documents/changes%20from%20dsm-iv-tr%20to%20dsm-5.pdf
9. Amann, B., C. Born, et al. (2011). "Lamotrigine: when and where does it act in affective disorders? A systematic review." Journal of Psychopharmacology **25**(10): 1289-1294.
10. Nierenberg AA, Burt T, Matthews J, Weiss AP. Mania associated with St. John's wort. Biol Psychiatry. 1999 Dec 15;46(12):1707-1708.
11. Perlick DA, Rosenheck RA, Clarkin JF, Maciejewski PK, et. al. Impact of family burden and affective response on clinical outcome among patients with bipolar disorder. Psychiatr Serv. 2004 Sep;55(9):1029-1035
12. Review Manager (RevMan) [Computer program]. Version 5.2. Copenhagen: The Nordic Cochrane Centre TCC, 2012.
13. Higgins JPT, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions: Version 5.1.0. The Cochrane Collaboration; 2011.
14. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. Journal of Clinical Epidemiology. 2011.
15. White C, Ip S, McPheeters M, et al. Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. Agency for Healthcare Research and Quality. Rockville, MD: 2009. http://effectivehealthcare.ahrq.gov/repFiles/methodsguide/systematicreviewsreplacedenovo.pdf.

16. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. Journal of Clinical Epidemiology. 2011 Nov;64(11):1187-97. PMID 21477993.

17. Donohue AW. Diagnosing Bipolar Disorder in the Community Setting. Journal of Psychiatric Practice. 2012 Nov;18(6):395-407

18. Weihrauch TR, Gauler TC. Placebo--efficacy and adverse effects in controlled clinical trials. Arzneimittelforschung. 1999 May;49(5):385-93.

19. Agency for Healthcare Research and Quality. Grading the strength of a body of evidence when assessing health care interventions--AHRQ and the effective health-care program: An Update Draft Report. Rockville, MD: June 2012. http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=1163.

20. Atkins D, Chang S, Gartlehner G, et al. Assessing the Applicability of Studies When Comparing Medical Interventions. 2010. http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=603. Accessed on AHRQ Publication No. 11-EHC019-EF.

VI. Definition of Terms

Not applicable.

VII. Summary of Protocol Amendments

If we need to amend this protocol, we will give the date of each amendment, describe the change and give the rationale in this section. Changes will not be incorporated into the protocol. Example table below:

Date	Section	Original Protocol	Revised Protocol	Rationale
This should be the effective date of the change in protocol	Specify where the change would be found in the protocol	Describe the language of the original protocol.	Describe the change in protocol.	Justify why the change will improve the report. If necessary, describe why the change does not introduce bias. Do not use justification as "because the AE/TOO/TEP/Peer reviewer told us to" but explain what the change hopes to accomplish.

VIII. Review of Key Questions

AHRQ posted the key questions on the Effective Health Care Website for public comment. The EPC refined and finalized the key questions after review of the public comments, and input from Key Informants and the Technical Expert Panel (TEP). This input is intended to ensure that the key questions are specific and relevant.

IX. Key Informants

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts constitute a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes and identify particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicting opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design, and methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor do they contribute to the writing of the report. They have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. The EPC considers all peer review comments on the draft report in preparation of the final report. Peer reviewers do not participate in writing or editing of the final report or other products. The final report does not necessarily represent the views of individual reviewers. The EPC will complete a disposition of all peer review comments. The disposition of comments for systematic reviews and technical briefs will be published three months after the publication of the evidence report.

Potential Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

EPC core team members must disclose any financial conflicts of interest greater than $1,000 and any other relevant business or professional conflicts of interest. Related financial conflicts of interest that cumulatively total greater than $1,000 will usually disqualify EPC core team investigators.

XIII. Role of the Funder

This project was funded under Contract No. xxx-xxx from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviewed contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

Appendix A. Search algorithm for treatments for bipolar disorder

Database: Ovid MEDLINE(R) <1946 to February Week 4 2014> Search
>Strategy:
----

1   meta analysis as topic/
2   meta-analy$.tw
3   metaanaly$.tw.
4   meta-analysis/
5   (systematic adj (review$1 or overview$1)).tw.
6   exp Review Literature as Topic/ (7347)
7   or/1-6
8   cochrane.ab.
9   embase.ab.
10   (psychlit or psyclit).ab.
11   (psychinfo or psycinfo).ab.
12   or/8-11
13   reference list$.ab.
14   bibliograph$.ab.
15   hand search.ab.
16   relevant journals.ab.
17   manual search$.ab.
18   or/13-17
19   selection criteria.ab.
20   (data adj2 (extract* or abstract*)).ab.
21   1   or 20
22   review/
23   2   and 22
24   Comment/
25   Letter/
26   editorial/
27   animal/
28   human/
29   2   not (28 and 27)
30   or/24-26,29
31   or 1   or 1   or 23
32   3   not 30
33   randomized controlled trials as topic/
34   randomized controlled trial/
35   random allocation/
36   double blind method/
37   single blind method/
38   clinical trial/
39   clinical trial, phase i.pt.
40   clinical trial, phase ii.pt.
41   clinical trial, phase iii.pt.
42   clinical trial, phase iv.pt.
43   controlled clinical trial.pt.

44   randomized controlled trial.pt.
45   multicenter study.pt.
46   clinical trial.pt. (484436)
47   exp clinical trials as topic/
48   or/33-47
49   (clinical adj trial$).tw.
50   ((singl$ or doubl$ or treb$ or tripl$) adj (blind$3 or mask$3)).tw.
51   placebos/
52   placebo$.tw.
53   randomly allocated.tw.
54   (allocated adj2 random$).tw.
55   or/49-54
56   4   or 55
57   case report.tw.
58   case report.tw.
59   letter/
60   historical article/
61   or/57-60
62   5   not 61
63   exp cohort studies/
64   cohort$.tw.
65   controlled clinical trial.pt.
66   epidemiological methods/
67   limit 6   to yr=1971-1983
68   or/63-65,67
69   (ae or to or po or co).fs.
70   side effect$.ti,ab.
71   side effect$.ti,ab.
7   ((adverse or undesireable or harm$ or serious or toxic) adj3 (effect$ or reaction$ or event$ or outcome$)).ti,ab.
73   exp product surveillance, postmarketing/
74   exp adverse drug reaction reporting systems/
75   exp clinical trials, phase iv/
76   exp poisoning/
77   exp substance-related disorders/
78   exp drug toxicity/
79   exp abnormalities, drug induced/
80   exp drug monitoring/
81   exp drug hypersensitivity/
82   (toxicity or complication$ or noxious or tolerability).ti,ab.
83   exp postoperative complication/
84   exp intraoperative complications/
85   or/69-84
86   exp Bipolar disorder/
87   bipolar*.ti.
88   cyclothymia.ti.
89   (rapid adj cycl*).ti.
90   (mania or hypomania or manic or hypomanic).ti.

91   or/86-90
92   3 and 91
93   6 and 91
94   6 and 8 and 91
95   limit 92 to yr="1994-Current"
96   limit 9 to yr="1994-Current" (3286) RCTs
97   limit 9 to yr="1994-Current" (1518)
98   9 not 9 (522) Systematic reviews
99   9 not (95 or 96) (989) Cohort harms
100   (68 and 91) not 8 (3060)
101   limit 10 to yr="1994-Current" (2188) Cohort benefits
102   10 not (95 or 9 or 97)
103   limit 9 to "all child (0 to 18 years)"
104   limit 9 to "all adult (19 plus years)"
105   10 and 104
106   10 not 105
107   9 not 10 (3034) RCTs without pediatric-only
108   limit 98 to "all child (0 to 18 years)"
109   limit 9 to "all adult (19 plus years)"
110   10 and 109
111   10 not 110
112   9 not 11 (498) Systematic reviews without pediatric-only
113   limit 9 to "all child (0 to 18 years)"
114   limit 99 to "all adult (19 plus years)"
115   11 and 114
116   11 not 115
117   9 not 11 (923) Cohort harms without pediatric-only
118   limit 10 to "all child (0 to 18 years)"
119   limit 10 to "all adult (19 plus years)"
120   11 and 119
121   11 not 120
122   10 not 12 (1556) Cohort benefits without pediatric-only
123   (resection or prostate or radiofrequency or sealer or ablation or hip or fibrillation).ab.
124   10 not 12 (2855) RCTs without pediatric-only or resection, ablation, etc.
125   11 not 12 (495) Systematic reviews without pediatric-only or
resection, ablation, etc.
126   11 not 12 (800) Cohort harms without pediatric-only or resection, ablation, etc.
127   12 not 12 (1515) Cohort benefits without pediatric-only or resection, ablation, etc.

Appendix B. Risk of Bias Assessment Form for Observational Studies

_____ **_Author_** _____ **_Year_** _____ **[PMID]** _____ **Reviewer** _____

Question	Response		Criteria	Justification
			Internal Validity	
1. Is the study design prospective, retrospective, or mixed?	Prospective	☐	Outcome has not occurred at the time the study is initiated and information is collected over time to assess relationships with the outcome.	
	Mixed	☐	Studies in which one group is studied prospectively and the other retrospectively.	
	Retrospective	☐	Analyzes data from past records.	
2. Are inclusion/exclusion criteria clearly stated?	Yes	☐		
	Partially	☐	Some, but not all, criteria stated or some not clearly stated.	
	No	☐		
3. Are baseline characteristics measured using valid and reliable measures and equivalent in both groups?	Yes	☐		
	No	☐		
	Uncertain	☐	Could not be ascertained.	
4. Is the level of detail describing the intervention adequate?	Yes	☐	Intervention described included adequate service details	
	Partially	☐	Some of the above features.	
	No	☐	None of the above features.	
5. Is the selection of the comparison group appropriate?	Yes	☐	Considering bipolar type, diagnostic assessment, other patient characteristics	
6. Did researchers isolate the impact from a concurrent intervention or an unintended exposure that might bias results?	Yes	☐	Accounted for concurrent informal care.	
	Partially	☐		
	No	☐		
7. Any attempt to balance the allocation between the groups (e.g. stratification, matching, propensity scores)?	Yes	☐	(if yes, what was used?)	
	No	☐		
	Uncertain	☐	Could not be ascertained.	
8. Were outcomes assessors blinded?		☐	Who were outcome assessors?	
9. Are outcomes assessed using valid and reliable measures, implemented consistently across all study participants?	Yes	☐	Measure valid and reliable (i.e. objective measures, well validated scale, provider report); and equivalent across groups.	
	Partially	☐	Some of the above features (partially validated scale)	
	No	☐	None of the above features. (self-report, scales with lower validity, reliability); not equivalent across groups	
	Uncertain	☐	Could not be ascertained.	
10. Is the length of follow-up the same for all groups?	Yes	☐		
	No	☐		

	Uncertain ☐	Could not be ascertained.	
11. Did attrition result in a difference in group characteristics between baseline and follow-up?	Yes ☐	(measurement period of interest if repeated measures)	
	No ☐		
	Uncertain ☐	Could not be ascertained (i.e. retrospective designs where eligible at baseline could not be determined)	
12. If baseline characteristics are not similar, does the analysis control for baseline differences between groups?	Yes ☐		
	No ☐		
	Uncertain ☐	Could not be ascertained (i.e. retrospective designs where eligible at baseline could not be determined)	
13. Are confounding and/or effect modifying variables assessed using valid and reliable measures across all study participants?	Yes ☐		
	No ☐		
	Uncertain ☐	Could not be ascertained (i.e. retrospective designs where eligible at baseline could not be determined)	
	NA ☐	No confounders or effect modifiers included in the study.	
14. Were the important confounding and effect modifying variables taken into account in the design and/or analysis (e.g. through matching, stratification, interaction terms, multivariate analysis, or other statistical adjustment)?	Yes ☐		
	Partially ☐	Some variables taken into account or adjustment achieved to some extent.	
	No ☐	Not accounted for or not identified.	
	Uncertain ☐	Could not be ascertained	
15. Are the statistical methods used to assess the primary outcomes appropriate to the data?	Yes ☐	Statistical techniques used must be appropriate to the data.	
	Partially ☐		
	No ☐		
	Uncertain ☐	Could not be ascertained	
16. Are reports of the study free of suggestion of selective outcome reporting?	Yes ☐		
	No ☐	Not all prespecified outcomes reported, subscales not prespecified reported, outcomes reported incompletely.	
	Uncertain ☐	Could not be ascertained.	
17. Funding source identified	No ☐		Industry, government, university, Foundation (funded by what money source?)
	Yes ☐	Who provided funding?	
	Uncertain ☐		

	Overall Assessment			
18. Overall Risk of Bias assessment	Low	☐	Results are believable taking study limitations into consideration	
	Moderate	☐	Results are probably believable taking study limitations into consideration	
	High	☐	Results are uncertain taking study limitations into consideration	

Appendix C. Draft Diagnostic Assessment Rubric

Criteria	Poor	Fair	Good	Excellent
Assessment Tool	Not Reported/ No	CIDI, other simple screening tools, like MDQ, Young Mania rating scale/Beck or Hamilton for current symptoms	MINI, SADS-not modified to meet DSM-IV	SCID, SADS-modified to meet DSM-IV
Process Reliability	Not Reported/ No	Back-end quality check of interview and diagnosis (e.g., 2nd rater reviews taped interviews)	Consensus, no details	Consensus with reliability measures
Diagnostic Criteria	Not Reported/ No	Mention DSM	DSM confirmed (states confirmed but no indication that all criteria were assessed or how confirmed)	Meet DSM (adequately met all criteria available at the time; if SCID, presume met)
If applicable: **BP-NOS**	Not Reported/ No	Details of definition not provided	Details of definition provided	Specific definition met
Raters	Not Reported/ No	Trained, no quantitative measures of quality, no certification process	MD, PhD, Resident, MA clinician with no quantitative measures of quality, no certification process	Clinicians or non-clinicians with standardized training and evaluation process
Source of Information	Patient self-report of diagnosis/ No source indicated	Medical record, only ICD9 or other summary statement, lacks detail	Interview with participant OR <u>detailed</u> medical record review	Interview with patient and second source (close relative or spouse, or detailed medical record review)
OVERALL RATING	Poor	Fair	Good	Excellent