Table 2 Description of psychological placebo conditions in each study

Study	Description of PP
Besyner (1979)	Nonspecific group: "Therapist behavior was limited to reflection and clarification of verbal material and questioning to facilitate discussion. It may be argued that such procedures are akin to, if not identical with, those employed by Rogerian therapists. While the validity of this argument cannot be denied, it is the belief of this researcher that such procedures are considered to be minimally therapeutic." (page 70, line 10)
Fuchs and Rehm (1977)	Nonspecific therapy: "Session 1 began in the same way as the self-control procedure with introductions, collection of deposits, a review of confidentiality issues, and a 10-minute group interaction assessment procedure. As in the other groups, participants were given an information sheet and a general introduction to group therapy concepts, generally from a nondirective framework. From that point on and throughout the ensuing sessions, therapists in this condition attempted to elicit discussion of past and current problems, to encourage group interaction, and to reflect and clarify feelings in an empathic manner. Although therapists at times suggested simple exercises within the group to facilitate open discussion, they were specifically instructed neither to recommend out-of-therapy activity nor explicitly to teach behavioral principles. These sessions lasted approximately 2 h weekly, as did self-control therapy sessions." (page 209, left column, line 24)
Hegerl et al. (2010)	Guided self help group (GSG): "In the GSG, a supportive atmosphere was created, allowing the participants to communicate about their situation and daily life, but no psychotherapeutic intervention was allowed by the group leader." (page 33, right column, line 1)
Kelly (1982)	Nondirective group:"The nondirective group served as a control group and met for the same amount of time as the other groups, but did not undergo their treatment procedures. Outside of behavior change strategies and cognitive strategies, the group was free to discuss any topics (e.g., support, jobs, etc.). All sessions, with the exception of the first, consisted of a review of the previous meeting's topic and a discussion of issues the group members felt were important. The therapist behavior during all sessions was as consistent as possible. An attempt was made to provide all group members with maximum empathy and warmth." (page 41, line 10)
Propst (1980)	Therapist Contact plus Self-Monitoring: "Participants in this condition simply met for a discussion group and kept track of their daily mood. For homework they were to record items for group discussion on their mood cards. The content of the discussion was up to the participants, as the therapists participated as little as possible." (page 172, line 5)

Table 2 continued

Study	Description of PP
Serfaty et al. (2009)	Talking Control: "Clearly defined criteria for the TC group were used to prevent CBT from being delivered. Talking control therapy was developed during our feasibility work, and details are available from the authors. The therapists practiced delivering the TC in role plays with the supervisor so that difficult questions could be addressed. Dysfunctional beliefs were not challenged; however, the therapists were asked to show interest and warmth, encouraging participants to discuss neutral topics such as hobbies, sports, and current affairs. No advice or problem solving was given, and there was little focus on emotional issues. No suggestions for behavioral tasks were offered. So for example, if the patient said, "My daughter does not like me as she never comes to visit me," the therapist would ask, "How many children do you have?" (page 1334, right column, line 8)

than NT (OR 2.24, 1.32–3.88) and that CBT was nominally, but not significantly, superior to PP (OR 1.30, 0.53–2.94), which in turn was superior to NT (OR 1.73, 0.67–4.84). For all the comparisons, the credible intervals were relatively wide because of the lack of power. The specificity of CBT was estimated to constitute 35.0 % (−99.5 to 180.3 %) of its efficacy over NT.

Pooling all available evidence, the estimate for the specificity of CBT had an extremely wide credible interval. In other words, overall, the currently available best evidence was compatible with both the no specificity hypothesis, i.e., the Dodo bird verdict (Baardseth et al. 2013; Luborsky et al. 2002; Luborsky and Singer 1975; Smith and Glass 1977; Wampold et al. 1997), as well as all foregoing point estimates ranging between 25 and 72 %(Barker et al. 1988; Bowers and Clum 1988; Lambert and Barley 2001; Stevens et al. 2000). However, post hoc exploratory analyses revealed that CBT of adequate length had a specificity component of about 50 %, with a 95 % credible interval between 20 and 85 %. We may now assume, with some confidence, that CBT has a non-zero specific component in the treatment of depression in adults.

There is now corollary evidence to suggest that the Dodo bird verdict is not universally operative. Critical incident stress debriefing is a form of crisis counseling aimed at preventing the development of posttraumatic stress disorder. It is typically delivered to a group of trauma survivors in a single 1–3-h session that takes place within 1 week of the trauma event. Although it does contain many common factors, such as empathic listening by experts in the field with credible explanatory models, specific factors appear to be at work leading to null to harmful results

Table 3 Pair-wise meta-analyses and sensitivity analyses

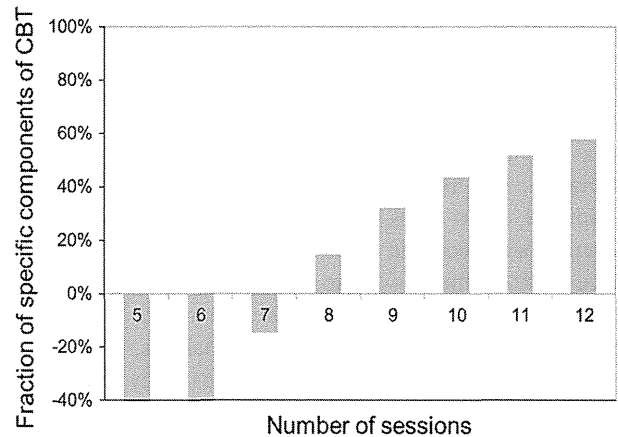|	Pair wise meta-analyses		Allocation concealment		Blinding of assessors		Treatment fidelity		Included disorders		Response imputed	
	OR (95 % CI)	n	OR (95 % CI)	n	OR (95 % CI)	n	OR (95 % CI)	n	OR (95 % CI)	n	OR (95 % CI)	n
CBT versus NT	2.07 (1.35–3.18)	13	1.79 (1.18–2.71)	10	1.31 (0.67–2.52)	1	7.00 (2.31–21.19)	1	1.49 (1.03–2.15)	9	1.31 (0.67–2.52)	1
CBT versus PP	1.74 (0.79–3.83)	6	1.55 (0.84–2.83)	1	NA	0	2.54 (1.34–4.82)	2	2.11 (1.16–3.83)	3	4.89 (1.53–15.66)	1
PP versus NT	2.04 (0.40–10.55)	1	NA	0	NA	0	NA	0	NA	0	NA	0

n number of included studies

Table 4 Odds ratios of response and specificity of CBT estimated in MTM and its meta-regression

	Overall MTM	Meta-regression MTM	
		<10sessions	≥10sessions
CBT versus NT	2.24 (1.32 to 3.88)	1.53 (1.02 to 2.28)	7.37 (3.74 to 15.15)
CBT versus PP	1.30 (0.53 to 2.94)	0.55 (0.27 to 1.20)	2.71 (1.42 to 5.33)
PP versus NT	1.73 (0.67 to 4.84)	2.72 (1.28 to 5.76)	2.72 (1.28 to 5.76)
CBT specific component	35.0 % (−99.5 % to 180.3 %)	−159.6 % (−958.4 % to 90.6 %)	50.4 % (19.7 % to 85.0 %)

Numbers in parentheses represent 95 % credible intervals

Fig. 3 Specific component of CBT for each number of sessions

(Rose et al. 2002; van Emmerik et al. 2002). Cottraux et al. (2001) demonstrated that cognitive therapy and exposure therapy may have differential degrees of effectiveness on obsessive–compulsive disorder (OCD), with the former having greater effects on depression and anxiety and the latter having greater effects on intrusive thoughts and OCD symptoms. They also reported some analyses showing that the amount of specific effects increases from post-treatment to follow-up, which could indicate that the post-treatment results are more strongly influenced by common factors, while follow-up assessments can reflect more specific components.

The number of included studies may appear limited in comparison with some recent systematic reviews of CBT for depression (Barth et al. 2013; Jakobsen et al. 2011), but our objective was not to perform a systematic review of CBT in general but to ask a focused question regarding the specificity of CBT by performing a network meta-analysis, for which the homogeneity and consistency of the included interventions and populations were more important than for traditional pairwise meta-analyses. We therefore focused on face-to-face CBT, with patients who were diagnosed as having acute depression according to operationalized diagnostic criteria or by scoring above the accepted threshold of a validated depression screening instrument. We also did not include behavior therapy or third-wave CBT in order to focus on narrowly defined CBT. We excluded studies if they employed protocolized pharmacotherapy in conjunction with CBT. Neither did we include the waiting list control, often used in psychotherapy research, as an NT control because there is a growing suspicion that the waiting list control may be differentiated from the NT condition (Watanabe et al. 2007). We further limited PP to interventions that were regarded as lacking an active component by researchers in the trial but that were explained as having an active component to the participants. We did not consider so-called counseling or supportive psychotherapy as PP because we believe these techniques have active components and should be classified as an active treatment. We adopted this narrow definition of PP in order to avoid bias due to researcher allegiance. All in all, out of the 128 studies found in the original study selection, we were only able to include 18 studies comparing CBT with PP and/or NT during the acute phase treatment of adults with depression (Fig. 1).

Several caveats are in order before we conclude. First, despite our systematic and comprehensive search of the literature, we were able to include only a relatively small number of studies. Thus, for example, although the network meta-regression revealed that the specific component of CBT may constitute half of its efficacy when CBT was given for ten or more sessions, it ought to be noted that only 5 of the 18 studies had ten or more sessions. Secondly, the evidence was not only quantitatively, but also qualitatively less than desirable. Allocation concealment was reported to be adequate in only three studies, and assessor blinding was reported in only one of the 18 studies. Furthermore, only three studies examined treatment fidelity in a satisfactory manner, and the response rates had to be imputed from the reported continuous outcomes in all but two studies. The results, however, were robust to sensitivity analyses. Thirdly, the heterogeneity of evidence network among CBT, PP, and NT, measured in terms of the median between-trial standard deviation, was relatively large when compared with the estimated effect sizes between the treatment arms. The heterogeneity coupled with the small sample size may have limited the power to detect relatively weak but important effect modifiers. We were not able to conduct many of the pre-planned sensitivity analyses, and where we were able to perform such analyses, they may have lacked an adequate power. However, when we included characteristics of the trials as effect modifiers and when the heterogeneity arising from the number of sessions was accounted for, the median between-trial standard deviations decreased. Last, but not least, our analytical model supposes a simple additive relationship between specific and non-specific components. However, it is imaginable that some interaction may exist between the two types of components: for example, if a treatment is very effective from its beginning, this would increase the patients' expectations for a positive outcome and hence would increase the placebo effect, but this can occur only in the treatment group. We would need better-designed studies, possibly with multiple control conditions with differential intensities, to detect such interactions.

On the other hand, the strengths of the present study may be as follows. First and foremost, we started with a well-formulated and well-focused clinical question to examine the specificity of a well-delineated intervention, i.e. CBT, for a specific clinical condition, i.e. acute phase treatment of depression in adults. Secondly, we followed the Cochrane review methodology. Comprehensive literature searches were conducted so as to minimize publication bias (Egger et al. 2003). Detailed manuals were prepared to guide the selection and data extraction of studies in duplicates. We also examined possible sources of bias and

conducted analyses following an intention-to-treatment principle as closely as possible. Thirdly, the use of MTM has enabled us to examine the consistency of the totality of evidence surrounding CBT, PP, and NT and to derive the most precise estimate of the specific component of CBT possible based on randomized evidence, while adjusting for possible effect modifiers. Thus, the main weaknesses of previous reviews, namely the unfocused inclusion of participants and interventions, the lack of systematic searches, and the small effect sizes with wide 95 % CI, have all been addressed in this study.

In conclusion, the present study represents the most up-to-date and comprehensive summary for the specificity hypothesis of CBT for depression. Despite the quantitatively and qualitatively limited body of randomized evidence examining this issue, the present study suggested a non-null specific component for one form of psychotherapy for one particular disorder. Future studies are needed to assess the specificity of CBT and other well-defined psychotherapies of adequate length and of satisfactory quality for various psychiatric disorders and psychological problems. Such psychotherapies, when they do exist, should be given preference in the provision and training of psychotherapies. The Dodo bird verdict is on the verge of extinction.

References

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology, 48,* 251–257.

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., et al. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review, 33*(3), 395–405.

Barker, S. L., Funk, S. C., & Houston, B. K. (1988). Psychological treatment versus nonspecific factors: A meta-analysis of conditions that engender comparable expectations for improvement. *Clinical Psychology Review, 8*(6), 579–594.

Barth, J., Munder, T., Gerger, H., Nuesch, E., Trelle, S., Znoj, H., et al. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med, 10*(5), e1001454.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Besyner, K. J. (1979). *The comparative efficacy of cognitive and behavioral treatments of depression: A multi-assessment approach*. A Thesis Submitted to the Graduate Faculty of Texas Tech University.

Bowers, T. G., & Clum, G. A. (1988). Relative contribution of specific and nonspecific treatment effects: Meta-analysis of placebo-controlled behavior-therapy research. *Psychological Bulletin, 103*(3), 315–323.

Chambless, D. (2002). Beware the dodo bird: The dangers of overgeneralization. *Clinical Psychology Science and Practice, 9,* 13–16.

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cottraux, J., Note, I., Yao, S. N., Lafont, S., Note, B., Mollard, E., et al. (2001). A randomized controlled trial of cognitive therapy versus intensive behavior therapy in obsessive compulsive disorder. *Psychotherapy and Psychosomatics, 70*(6), 288–297.

da Costa, B. R., Rutjes, A. W., Johnston, B. C., Reichenbach, S., Nüesch, E., Tonia, T., et al. (2012). Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: Meta-epidemiological study. *International Journal of Epidemiology, 41*(5), 1445–1459.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *JAMA, 263*(10), 1385–1389.

Dowrick, C., Dunn, G., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., Lehtinen, V., et al. (2000). Problem solving treatment and group psychoeducation for depression: multicentre randomised controlled trial: Outcomes of Depression International Network (ODIN) Group. *BMJ, 321*(7274), 1450–1454.

Dunn, G., Maracy, M., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., et al. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *British Journal of Psychiatry, 183,* 323–331.

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology, 78*(1), 107–126.

Egger, M., Juni, P., Bartlett, C., Holenstein, F., & Sterne, J. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment, 7*(1), 1–76.

Ellis, A. (1979). *Theoretical and empirical foundations of rational-emotive therapy*. Monterey: Brooks/Cole.

Faramarzi, M., Alipor, A., Esmaelzadeh, S., Kheirkhah, F., Poladi, K., & Pash, H. (2008). Treatment of depression and anxiety in infertile women: Cognitive behavioral therapy versus fluoxetine. *Journal of Affective Disorders, 108*(1–2), 159–164.

Fuchs, C. Z., & Rehm, L. P. (1977). A self-control behavior therapy program for depression. *Journal of Consulting and Clinical Psychology, 45*(2), 206–215.

Furukawa, T. (1999). From effect size into number needed to treat. *Lancet, 353*(9165), 1680.

Furukawa, T. A., Cipriani, A., Barbui, C., Brambilla, P., & Watanabe, N. (2005). Imputing response rates from means and standard deviations in meta-analyses. *International Clinical Psychopharmacology, 20*(1), 49–52.

Greenberg, L. S., & Newman, F. L. (1996). An approach to psychotherapy change process research: Introduction to the special section. *Journal of Consulting and Clinical Psychology, 64*(3), 435–438.

Hamamci, Z. (2006). Integrating psychodrama and cognitive behavioral therapy to treat moderate depression. *The Arts in Psychotherapy, 33*(3), 199–207.

Hamdan-Mansour, A. M., Puskar, K., & Bandak, A. G. (2009). Effectiveness of cognitive-behavioral therapy on depressive symptomatology, stress and coping strategies among Jordanian university students. *Issues in mental health nursing, 30*(3), 188–196.

Hegerl, U., Hautzinger, M., Mergl, R., Kohnen, R., Schutze, M., Scheunemann, W., et al. (2010). Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: A randomized, controlled trial including a patients' choice arm. *International Journal of Neuropsychopharmacology, 13*(1), 31–44.

Higgins, J. P., & Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine, 15*(24), 2733–2749.

Hollon, S. D., & DeRubeis, R. J. (1981). Placebo-psychotherapy combinations: Inappropriate representations of psychotherapy in drug-psychotherapy comparative trials. *Psychological Bulletin, 90*(3), 467–477.

Hunot, V., Moore Theresa, H. M., Caldwell Deborah, M., Furukawa Toshi, A., Davies, P., Jones, H., et al. (2013). 'Third wave' cognitive and behavioural therapies versus other psychological therapies for depression. *Cochrane Database of Systematic Reviews, 10.* http://onlinelibrary.wiley.com/doi/10.1002/146518 58.CD008704.pub2/abstract. doi:10.1002/14651858.CD008704. pub2.

Jakobsen, J. C., Hansen, J. L., Storebo, O. J., Simonsen, E., & Gluud, C. (2011). The effects of cognitive therapy versus 'no intervention' for major depressive disorder. *PLoS ONE, 6*(12), e28299.

Kelly, M. L. (1982). Rational emotive therapy versus Lewinsohnian based approaches to the treatment of depression. *Dissertation Abstracts International, 43*(6-B).

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy, 38*(4), 357–361.

Leucht, S., Hierl, S., Kissling, W., Dold, M., & Davis, J. M. (2012). Putting the efficacy of psychiatric and general medicine medication in perspective: A review of meta-analyses. *British Journal of Psychiatry, 200,* 97–106.

Lewinsohn, P. M., Antonuccio, D. O., Breckenridge, J. S., & Teri, L. (1984). *The coping with depression course: A psychoeducational intervention for unipolar depression*. Eugene, OR: Castalia.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology Science and Practice, 9,* 2–12.

Luborsky, L., & Singer, B. (1975). Comparative studies of psychotherapies: Is it true that "every won has one and all must have prizes"? *Archives of General Psychiatry, 32*(8), 995–1008.

Miranda, J., Chung, J. Y., Green, B. L., Krupnick, J., Siddique, J., Revicki, D. A., et al. (2003). Treating depression in predominantly low-income young minority women: A randomized controlled trial. *JAMA, 290*(1), 57–65.

Omer, H., & London, P. (1989). Signal and noise in psychotherapy: The role and control of non-specific factors. *British Journal of Psychiatry, 155*, 239–245.

Propst, L. R. (1980). The comparative efficacy of religious and nonreligious imagery for the treatment of mild depression in religious individuals. *Cognitive Therapy and Research, 4*(2), 167–178.

Rose, S., Bisson, J., Churchill, R., & Wessely, S. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). *Cochrane database of systematic reviews, 2*(2), CD000560.

Rosenthal, R., & Frank, J. D. (1956). Psychotherapy and the placebo effect. *Psychological Bulletin, 53*(4), 294–302.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6*, 412–415.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA, 273*(5), 408–412.

Serfaty, M. A., Haworth, D., Blanchard, M., Buszewicz, M., Murad, S., & King, M. (2009). Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: A randomized controlled trial. *Archives of General Psychiatry, 66*(12), 1332–1340.

Shinohara, K., Honyashiki, M., Imai, H., Hunot, V., Caldwell Deborah, M., Davies, P., et al. (2013). Behavioural therapies versus other psychological therapies for depression. *Cochrane Database of Systematic Reviews*, (10). http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008696.pub2/abstract. doi:10.1002/14651858.CD008696.pub2.

Siev, J., Huppert, J., & Chambless, D. (2010). Treatment specificity for panic disorder: A reply to Wampold, Imel, and Miller (2009). *Behavior Therapist, 33*, 12–14.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *The American psychologist, 32*(9), 752–760.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L., & Sutton, A. J. (2000). Publication and related biases. *Health Technology Assessment, 4*(10), 1–115.

Stevens, S. E., Hynan, M. T., & Allen, M. (2000). A meta-analysis of common factor and specific treatment effects across the outcome domains of the phase model of psychotherapy. *Clinical Psychology-Science and Practice, 7*(3), 273–290.

van Emmerik, A. A., Kamphuis, J. H., Hulsbosch, A. M., & Emmelkamp, P. M. (2002). Single session debriefing after psychological trauma: A meta-analysis. *Lancet, 360*(9335), 766–771.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes.". *Psychological Bulletin, 122*, 203–215.

Watanabe, N., Hunot, V., Omori, I. M., Churchill, R., & Furukawa, T. A. (2007). Psychotherapy for depression among children and adolescents: A systematic review. *Acta Psychiatrica Scandinavica, 116*(2), 84–95.

Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis

Tomofumi Miura, Hisashi Noma, Toshi A Furukawa, Hiroshi Mitsuyasu, Shiro Tanaka, Sarah Stockton, Georgia Salanti, Keisuke Motomura, Satomi Shimano-Katsuki, Stefan Leucht, Andrea Cipriani, John R Geddes, Shigenobu Kanba

Summary

Background Lithium is the established standard in the long-term treatment of bipolar disorder, but several new drugs have been assessed for this indication. We did a network meta-analysis to investigate the comparative efficacy and tolerability of available pharmacological treatment strategies for bipolar disorder.

Methods We systematically searched Embase, Medline, PreMedline, PsycINFO, and the Cochrane Central Register of Controlled Trials for randomised controlled trials published before June 28, 2013, that compared active treatments for bipolar disorder (or placebo), either as monotherapy or as add-on treatment, for at least 12 weeks. The primary outcomes were the number of participants with recurrence of any mood episode, and the number of participants who discontinued the trial because of adverse events. We assessed efficacy and tolerability of bipolar treatments using a random-effects network meta-analysis within a Bayesian framework.

Findings We screened 114 potentially eligible studies and identified 33 randomised controlled trials, published between 1970 and 2012, that examined 17 treatments for bipolar disorder (or placebo) in 6846 participants. Participants assigned to all assessed treatments had a significantly lower risk of any mood relapse or recurrence compared with placebo, except for those assigned to aripiprazole (risk ratio [RR] 0·62, 95% credible interval [CrI] 0·38–1·03), carbamazepine (RR 0·68, 0·44–1·06), imipramine (RR 0·95, 0·66–1·36), and paliperidone (RR 0·84, 0·56–1·24). Lamotrigine and placebo were significantly better tolerated than carbamazepine (lamotrigine, RR 5·24, 1·07–26·32; placebo, RR 3·60, 1·04–12·94), lithium (RR 3·76, 1·13–12·66; RR 2·58, 1·33–5·39), or lithium plus valproate (RR 5·95, 1·02–33·33; RR 4·09, 1·01–16·96).

Interpretation Although most of the drugs analysed were more efficacious than placebo and generally well tolerated, differences in the quality of evidence and the side-effect profiles should be taken into consideration by clinicians and patients. In view of the efficacy in prevention of both manic episode and depressive episode relapse or recurrence and the better quality of the supporting evidence, lithium should remain the first-line treatment when prescribing a relapse-prevention drug in patients with bipolar disorder, notwithstanding its tolerability profile.

Funding None.

Introduction

Bipolar disorder is a complex disorder characterised by recurrent episodes of depression and mania (bipolar I disorder) or hypomania (bipolar II disorder).[1,2] The lifetime prevalence of bipolar I and II disorders has been estimated at about 0·5% and 1·5%, respectively.[3] Bipolar disorder is often chronic: results of long-term prospective follow-up studies show that the proportions of bipolar I patients who remain in remission are very low: 28% for 4 years and about 10% for 5 years.[4–6]

Long-term treatment is usually needed to minimise the risk of serious relapse or recurrence and to stabilise mood. Pharmacotherapy is the standard therapeutic approach. Lithium has been the standard long-term therapy for 40 years, but antiepileptics, antipsychotics, and antidepressants are also recommended and widely used in clinical practice. As the number and variety of available drugs increase, uncertainty about their comparative efficacy and tolerability increases, and questions remain about which agent should be used for which patient.[7–9]

When several treatment options are available for a specific indication, having a reliable estimate of comparative efficacy (prevention of any mood episode, of manic, hypomanic, or mixed episode, and of depressive episode), tolerability, and acceptability is clinically useful. In the absence of direct comparisons between all available treatments, a network meta-analysis can be used to synthesise the available direct and indirect evidence. This method has been successfully applied to guide clinical practices in medicine and psychiatry.[10–12] We did a systematic review and network meta-analysis of the efficacy and tolerability of pharmacological treatments for bipolar disorder to provide the most up-to-date, methodologically sound summary of the available evidence and to inform decisions about long-term treatment.

Department of Neuropsychiatry Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan (T Miura MD, H Mitsuyasu MD, K Motomura MD, S Shimano-Katsuki MD, Prof S Kanba MD); Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan (H Noma PhD); Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine and School of Public Health, Kyoto, Japan (Prof T A Furukawa MD); Department of Pharmacoepidemiology, Kyoto University School of Public Health, Kyoto, Japan (S Tanaka PhD); Department of Psychiatry, University of Oxford, Oxford, UK (S Stockton BA, A Cipriani PhD, Prof J R Geddes MD); Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece (G Salanti PhD); Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany (Prof S Leucht MD); and Department of Public Health and Community Medicine, Section of Psychiatry and Clinical Psychology, University of Verona, Verona, Italy (A Cipriani PhD)

Correspondence to:
Dr Tomofumi Miura, Department of Neuropsychiatry Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
tmiura@npsych.med.kyushu-u.ac.jp

Methods

Search strategy and selection criteria

Before beginning the review, we registered the study protocol with the PROSPERO database of systematic reviews (number CRD42012002739; appendix pp 2–11), and we did our systematic review in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Subsequent changes to the protocol are shown in the appendix (p 12). The overall dataset is available online.

We searched Embase, Medline, PreMedline, PsycINFO, and the Cochrane Central Register of Controlled Trials (CENTRAL) to identify eligible studies published between the date of the databases' inception and July 26, 2012, and we updated the search on June 28, 2013. We also searched international trial registers via the WHO's International Clinical Trials Registry Platform (ICTRP) and the US Food and Drug Administration (FDA) website on July 4, 2013, and asked pharmaceutical companies to provide additional information about their studies. Full details of the search strategies are given in the appendix (pp 13–26).

We included all randomised controlled trials comparing any pharmacological agent with placebo or active comparator, with at least 12 weeks of follow-up, for the maintenance treatment of patients with a primary diagnosis of bipolar disorder, irrespective of whether the patients' subtypes were specified or not. We also included trials in which the investigators did not use operationalised criteria, but apparently discriminated between bipolar illness and unipolar depression and provided the data separately for bipolar patients. We excluded studies focusing on child or adolescent bipolar disorder. The eligible pharmacological agents included not only the so-called mood stabilisers, but also any antipsychotics, antidepressants, and antiepileptic drugs. We included combination or augmentation studies when the two drugs used were specified, but excluded studies whose treatment group allowed either lithium or valproate as the baseline treatment. We included open trials and those with any level of blinding. We included blinded drugs, open-label drugs, and also open-label drugs plus blinded placebo into the same drug node in the network meta-analysis, because these three treatment groups should not differ in their pharmacological activities. To investigate the effect of blinding, we did a sensitivity analysis restricted to trials using double blinding. We excluded studies in which participants were randomly assigned to a maintenance treatment regimen while in an acute mood episode (so-called continuation studies); however, we included prophylaxis design (euthymic participants were eligible) and relapse prevention design (only those who responded to the investigational drug during the acute-phase treatment were eligible to be randomly assigned to either remain on the drug or be switched to placebo or comparator).

Outcome measures and data extraction

The primary outcomes were the number of participants with any recurrent mood episode (depressive, manic, hypomanic, or mixed) as defined by the study investigators (treatment efficacy) and the number of participants who dropped out of treatment because of adverse events (treatment tolerability), both at the longest available follow-up. Secondary outcomes included the number of participants who had a depressive episode, those who had a manic, hypomanic, or mixed episode, and those who discontinued treatment for any reason including relapse (treatment acceptability). We also examined the number of participants who completed suicide and the social functioning of all patients.

At least two of three reviewers (TM, HM, and TAF) selected the studies, and TM and HM, independently, were responsible for data extraction. We contacted the corresponding author or sponsor of the original article for further information when necessary. Any disagreements were resolved through discussion within the review team. We assessed the risk of bias in the included studies using the Cochrane Collaboration method, with an additional item to assess whether definitions of the mood episode relapse or recurrence were explicit or operationalised, or not.[13]

Statistical analysis

Network meta-analysis combines direct and indirect evidence for all relative treatment effects and provides estimates with maximum power.[14–18] Although an odds ratio (OR) is a frequently used effect measure in network meta-analyses, it is not necessarily an approximation to a risk ratio (RR), which is generally easier to interpret for clinicians. We therefore used RRs in our network meta-analysis since event rates were not small in some trials.

First, we did pair-wise meta-analyses of direct evidence using the random-effects model, with R version 3.0.0 and the metafor package.[19,20] Second, we did a random-effects network meta-analysis within a Bayesian framework using Markov chain Monte Carlo in OpenBUGS 3.2.2.[21] Comparative RRs are reported with their 95% credible intervals (CrIs). The network meta-analysis model and the BUGS codes are shown in the appendix (pp 27–30).

The assumption of transitivity[17,22] in the network (a prime requisite of network meta-analysis) was first assessed by considering the distributions of major effect modifiers (publication year, subtypes of bipolar disorder, percentage of female participants, inclusion of rapid-cycling bipolar disorder, mood state at recruitment, and treatment before randomisation) for all the comparisons in the networks. Consistency between direct and indirect sources of evidence was then statistically assessed globally (by comparing the fit and parsimony of consistency and inconsistency models) and locally (by calculating the difference between direct and indirect estimates in all closed loops in the network).[23–25] We graphically presented

the data and evaluated inconsistency using computational and graphical tools with STATA version 13.0.[23]

The treatment network will consist of closed loops and single-standing nodes. Because transitivity of single-standing nodes cannot be assessed, and its effect size estimates do not benefit from the network (ie, they cannot borrow strength from the entire network), but are often based on only one trial, analyses mainly focused on the treatment nodes constituting the closed-loop network.

We assessed the quality of evidence contributing to each network estimate with the GRADE framework, which characterises the quality of a body of evidence on the basis of the study limitations, imprecision, heterogeneity or inconsistency, indirectness, and publication bias.[26] The starting point for confidence in each network estimate was high, but was downgraded according to the assessments of these five aspects. We quantified the limitation of studies contributing to each network estimate by calculating the contributions from studies with an enrichment design and secondly by calculating those from studies at high risk of bias. The judgment of precision was based on whether the CrI around the point estimate overlapped with the clinically meaningful threshold.

We did sensitivity analyses using publication year, subtypes of bipolar disorder, rapid-cycling course of illness, enrichment design, sponsorship bias, duration of follow-up, and blinding of the treatment group.

Role of the funding source

This study received no external funding. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

We identified 10815 references through the electronic searches and retrieved 114 potentially eligible studies to analyse in detail (figure 1). We excluded 83 reports that did not meet the eligibility criteria, and identified two further studies when we updated our search. We also found one candidate trial from the WHO ICTRP search; however, insufficient information was available and we therefore regarded the study as awaiting assessment. We found another candidate trial[27] from inquiries to pharmaceutical companies and requested detailed information about it, but the clinical data of the study were not available from the company. We did not find any unpublished trials from the FDA website.

In our network meta-analysis, we included 33 trials published between 1970 and 2012, including 6846 participants. Table 1 lists the included studies (for details and references, see appendix pp 31–46) and table 2 reports their summary characteristics. The mean age of
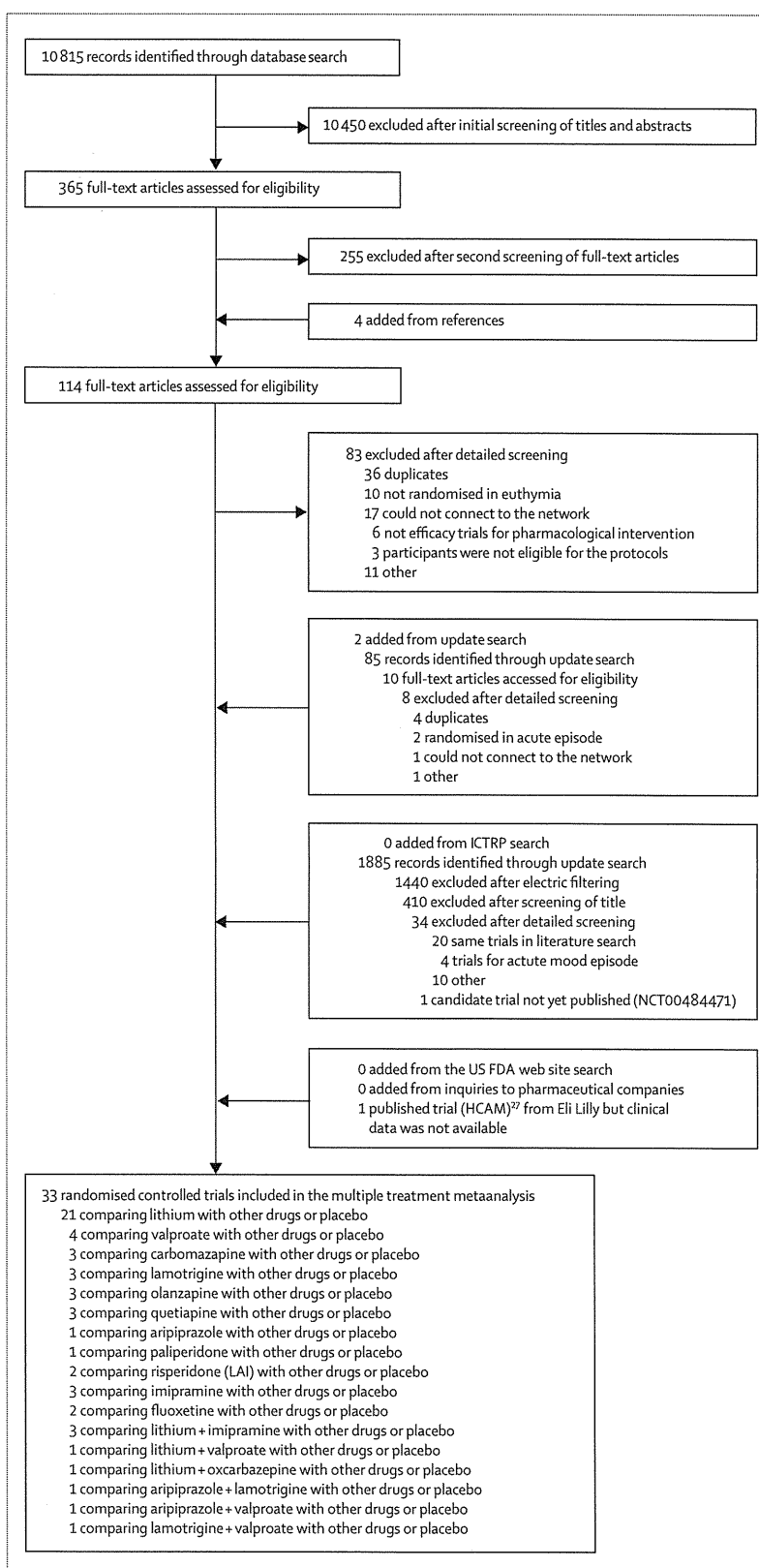

***Figure 1:* PRISMA flowchart**
ICTRP=WHO International Clinical Trials Registry Platform. FDA=Food and Drug Administration. LAI=longacting injection. PRISMA=Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

	Interventions (number of participants)	Included diagnosis	Mood status at recruitment	Blinding	Enrichment design
Melia, 1970	Lithium (5) vs placebo (6)	BP	Euthymia	Double-blind	No
Cundall, 1972	Lithium (8) vs placebo (5)	BP	Unknown	Double-blind	Yes
Prien, 1973a	Lithium (18) vs imipramine (13) vs placebo (13)	BP	Depressive episode	Double-blind	No
Prien, 1973b	Lithium (101) vs placebo (104)	BP	Manic episode/hypomanic episode	Double-blind	Yes
Dunner, 1976	Lithium (16) vs placebo (24)	BP-II, BP other	Euthymia	Double-blind	No
Fieve, 1976	Lithium (24) vs placebo (29)	BP-I, BP-II	Euthymia	Double-blind	No
Kane, 1981	Lithium + imipramine (37) vs lithium + placebo (38)	BP-I	Euthymia	Double-blind	No
Kane, 1982	Lithium + imipramine (6) vs lithium (4) vs imipramine (5) vs placebo (7)	BP-II	Euthymia	Double-blind	No
Prien, 1984	Lithium + imipramine (36) vs imipramine (36) vs lithium (42)	BP	Manic episode/hypomanic episode/mixed episode/depressive episode	Double-blind	Yes
Coxhead, 1992	Lithium (16) vs carbamazepine (15)	BP	Euthymia	Double-blind	No
Bowden, 2000	Valproate (187) vs lithium (91) vs placebo (94)	BP-I	Manic episode/mixed episode/euthymia	Double-blind	No
Calabrese, 2000	Lamotrigine (93) vs placebo (89)	BP-I, BP-II	Manic episode/hypomanic episode/mixed episode/depressive episode/euthymia	Double-blind	Yes
Kleindienst, 2000	Lithium (86) vs carbamazepine (85)	BP-I, BP-II, BP-NOS	Manic episode/hypomanic episode/mixed episode/depressive episode	Open	No
Bowden, 2003	Lamotrigine (59) vs lithium (46) vs placebo (70)	BP-I	Manic episode/hypomanic episode	Double-blind	Yes
Calabrese, 2003	Lamotrigine (171) vs lithium (121) vs placebo (121)	BP-I	Depressive episode	Double-blind	Yes
Hartong, 2003	Carbamazepine (30) vs lithium (23)	BP-I, BP-II	Euthymia	Double-blind	No
Amsterdam, 2005	Fluoxetine (8) vs placebo (4)	BP-II	Depressive episode	Double-blind	Yes
Calabrese, 2005	Lithium (32) vs valproate (28)	BP-I, BP-II	Manic episode/hypomanic episode/mixed episode/depressive episode/euthymia	Double-blind	No
Tohen, 2005	Olanzapine (217) vs lithium (214)	BP-I	Manic episode/mixed episode	Double-blind	No
Tohen, 2006	Olanzapine (225) vs placebo (136)	BP-I	Manic episode/mixed episode	Double-blind	Yes
Keck, 2007	Aripiprazole (78) vs placebo (83)	BP-I	Manic episode/mixed episode	Double-blind	Yes
Vieta, 2008	Lithium + oxcarbazepine (26) vs lithium (29)	BP-I, BP-II	Euthymia	Double-blind	No
Amsterdam, 2010	Fluoxetine (28) vs lithium (26) vs placebo (27)	BP-II	Depressive episode	Double-blind	Yes
Geddes, 2010	Lithium (110) vs valproate (110) vs lithium + valproate (110)	BP-I	Euthymia	Open	No
Quiroz, 2010	Risperidone LAI (140) vs placebo (135) for efficacy outcome; risperidone LAI (154) vs placebo (149) for safety outcome	BP-I	Manic episode/mixed episode/euthymia	Double-blind	Yes
Koyama, 2011	Lamotrigine (45) vs placebo (58)	BP-I	Manic episode/mixed episode/depressive episode/euthymia	Double-blind	Yes
Weisler, 2011	Quetiapine (404) vs lithium (364) vs placebo (404)	BP-I	Manic episode/mixed episode/depressive episode/euthymia	Double-blind	Yes
Woo, 2011	Valproate + aripiprazole (40) vs valproate (43)	BP-I	Manic episode/mixed episode	Double-blind	Yes
Carlson, 2012	Aripiprazole + lamotrigine (178) vs lamotrigine (173)	BP-I	Manic episode/mixed episode	Double-blind	Yes
Berwaerts, 2012	Paliperidone (152) vs placebo (148)	BP-I	Manic episode/mixed episode	Double-blind	Yes
Young, 2012	Quetiapine (291) vs placebo (294)	BP-I, BP-II	Depressive episode	Double-blind	Yes
Bowden, 2012	Lamotrigine (45) vs lamotrigine + valproate (41)	BP-I, BP-II	Depressive episode/euthymia	Double-blind	Yes
Vieta, 2012	Risperidone LAI (132) vs placebo (135) vs olanzapine (131)	BP-I	Manic episode/mixed episode/euthymia	Double-blind	Yes

See appendix (pp 31–46) for more details and references. BP=bipolar disorder. LAI= longacting injection.

Table 1: Summary of randomised controlled trials of treatments for bipolar disorder with at least 12 weeks' follow-up

participants was 40·2 years (SD 12·8) and 3633 (55%) of 6655 participants for whom data were reported were women. The eligible diagnoses in primary studies were bipolar I disorder (15 [45%] trials), bipolar II disorder (four [12%] trials), both bipolar I and II disorder (eight [24%] trials), and unspecified bipolar disorder (six [18%] trials). Rapid-cycling bipolar disorder was excluded in five (15%) studies and included in 12 (36%) studies; no mention of it was made in the remaining 16 (48%) trials.

Participants were assigned to placebo or to one of the following 17 treatment interventions: aripiprazole, carbamazepine, fluoxetine, imipramine, lithium, lithium plus imipramine, lithium plus oxcarbazepine, lithium plus valproate, lamotrigine, aripiprazole plus lamotrigine, valproate plus lamotrigine, olanzapine, paliperidone, quetiapine, risperidone longacting injection (LAI), valproate, and valproate plus aripiprazole. Two non-blinded randomised trials were included. The mean of the study durations of the included studies was 74·0 weeks (SD 37·6; range 17·3–171·4). We noted considerable differences across studies in mood states of the participants at study recruitment (table 2) and in treatments to stabilise

	Studies (N=33)
Recruitment area	
Cross-continental	11 (33%)
North America	14 (42%)
Europe	6 (18%)
Asia	2 (6%)
Number of treatment groups	
Two	23 (70%)
Three or more	10 (30%)
Blinding	
Open-label	2 (6%)
Single-blind	0
Double-blind	31 (94%)
Diagnostic criteria	
Not operationalised	4 (12%)
Feighner criteria	2 (6%)
Research Diagnostic Criteria	3 (9%)
DSM-III	1 (3%)
DSM-III-R	2 (6%)
DSM-IV	14 (42%)
DSM-IV-TR	7 (21%)
Included diagnosis	
Bipolar I disorder	15 (45%)
Bipolar II disorder	4 (12%)
Bipolar I and II disorder	8 (24%)
Bipolar disorder (subtype not specified)	6 (18%)
Inclusion of rapid cycling	
Included	12 (36%)
Excluded	5 (15%)
Unclear	16 (48%)
Mood statuses at recruitment	
Acute mood episode	16 (48%)
Depressive episode	5 (15%)
Manic/hypomanic/mixed episode	8 (24%)
Any acute mood episode	3 (9%)
Acute mood episode or euthymia	7 (21%)
Euthymia	6 (18%)
Unclear	4 (12%)
Mood statuses of most recent episode	
Reported*	23 (70%)
Not reported	10 (30%)
Enrichment design	
Yes	19 (58%)
No	14 (42%)
Sponsorship	
Unclear	3 (9%)
Yes	22 (67%)
No	8 (24%)

DSM=Diagnostic and Statistical Manual of Mental Disorders. *Depressive episode was reported for 1970 participants and a manic/hypomanic/mixed episode was reported for 3660 participants.

Table 2: Summary characteristics of the 33 included studies
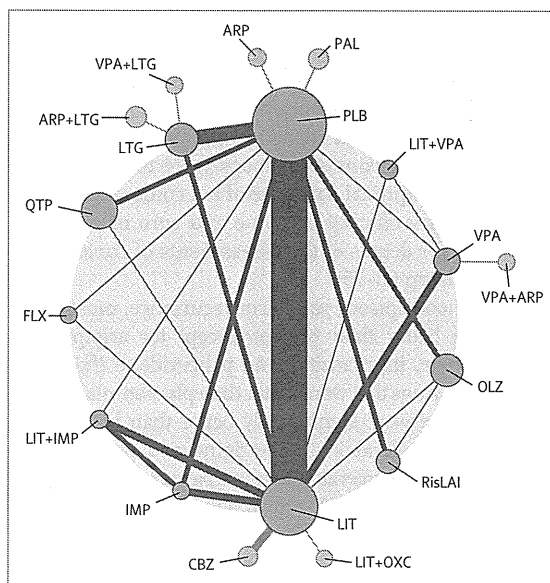
Figure 2: Network of all eligible comparisons for the network meta-analysis
Each node (circle) corresponds to a drug included in the analysis, with the size proportional to the number of participants randomly assigned to that drug. Each line represents direct comparisons between drugs, with the width of the lines proportional to the number of trials comparing each pair of treatments. The treatment nodes in the closed-loop network are purple, whereas single-standing nodes and their connections are light blue. All the monotherapies, except for ARP, PAL, and CBZ, were compared with at least two other treatment nodes (ie, were in the closed-loop network). 12 (50%) of 24 comparisons for the primary efficacy outcome and seven (29%) of 24 comparisons for tolerability were done in more than one trial. ARP=aripiprazole. CBZ=carbamazepine. FLX=fluoxetine. IMP=imipramine. LIT=lithium. LTG=lamotrigine. OLZ=olanzapine. OXC=oxcarbazepine. PAL=paliperidone. PLB=placebo. QTP=quetiapine. RisLAI=risperidone longacting injection. VPA=valproate.

trials, whereas treatment before randomisation was not restricted in six (18%) trials.[28] In eight (24%) trials, neither one of the treatment groups had an advantage from the active run-in design (any one of the study drugs or both of them were used to stabilise mood episodes) or participants were recruited in a euthymic mood. 22 (67%) studies were done, at least in part, under industry sponsorship. Other risks of bias of the included studies are presented in the appendix (pp 47–50).

Figure 2 shows the network of eligible comparisons for the network meta-analysis. Of 153 possible pair-wise comparisons among 18 interventions, 24 direct comparisons were made for our primary outcomes (the networks for each outcome are provided in the appendix pp 51–54). Distributions of the major effect modifiers in each comparison are shown in the appendix (pp 55–58). The summaries of pair-wise meta-analyses (primary and secondary outcomes, test of heterogeneity, and funnel plots in comparison with lithium and placebo) are shown in the appendix (pp 59–67).

Figure 3 presents the results of the network meta-analyses for the primary outcomes. The heterogeneity variances of the random-effects network meta-analysis models for primary outcomes were 0.147 for any mood episode relapse or recurrence and 0.366 for tolerability.

mood episodes before randomisation (appendix pp 55–58). An enrichment design—ie, selection of patients who responded acutely to treatment—was used in 19 (58%)

Also, the assumption of global consistency was supported by a better trade-off between model fit and complexity when consistency was assumed than when it was not. Tests of local inconsistency revealed that the percentages for inconsistent loops were to be expected according to empirical data (one of ten comparison loops for the primary efficacy outcome and zero of seven for tolerability; for details of the assessments of consistency, see appendix pp 68–75).
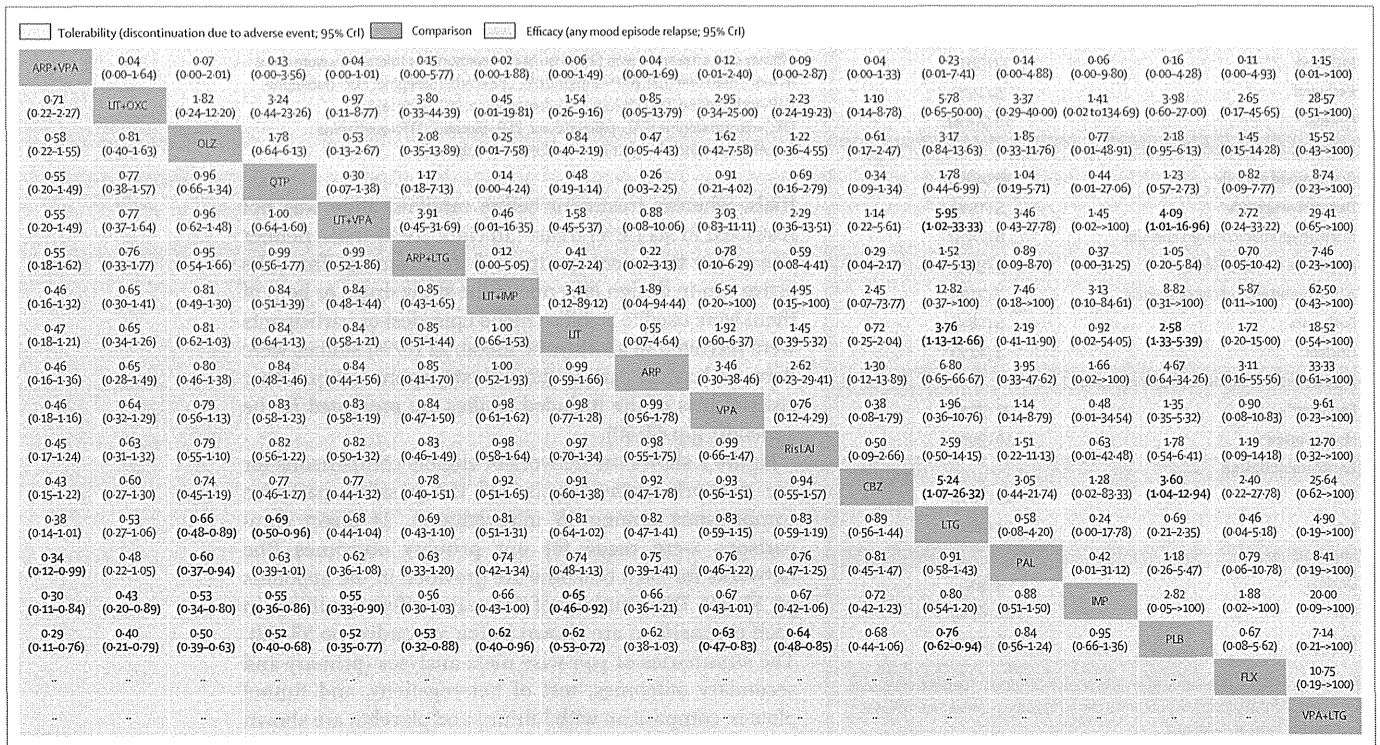
For any mood episode relapse or recurrence, most of the drugs were better than placebo except for aripiprazole, carbamazepine, imipramine, and paliperidone (figure 3). Of the active drugs that were better than placebo, olanzapine and quetiapine were significantly better than lamotrigine (figure 3). For tolerability, lamotrigine and placebo were significantly better tolerated than carbamazepine, lithium, or lithium plus valproate (figure 3). The results of secondary outcomes are presented in the appendix (pp 76–80).

Figure 4 presents ranked forest plots of RRs for compounds that are included in the closed-loop network in comparison with placebo. The quality of evidence for any mood episode relapse or recurrence was rated as moderate for lithium and olanzapine, very low for lithium plus imipramine, and low for all the others (for details of the estimation of the quality of the evidence, see appendix pp 81–106). Lithium was better than placebo in the prevention of both manic and depressive relapse or recurrence, but less well tolerated than placebo. Quetiapine was also better than placebo in the prevention of both manic and depressive relapse or recurrence. Olanzapine was significantly better than placebo in the prevention of manic but not depressive relapse or recurrence. In the other interventions, either one or both of the secondary efficacy outcomes were statistically non-significant.

We also presented results in a two-dimensional plot of RR of each drug in comparison with placebo for any mood relapse or recurrence versus tolerability, and depressive relapse or recurrence versus manic, hypomanic, or mixed relapse or recurrence (appendix pp 107–09). The cumulative probability plots and SUCRAs (surface under the cumulative ranking curve) for all the included treatment groups are presented in the appendix (pp 110–20).

Because the number of completed suicides was zero or one in most of the trials, we did not calculate their RRs, and showed the raw numbers in the appendix (pp 121–24). Only five trials reported social functioning as measured by the Global Assessment of Functioning scale or the Global Assessment Scale.

Tolerability (discontinuation due to adverse event; 95% CrI) — Comparison — Efficacy (any mood episode relapse; 95% CrI)

ARP+VPA	0.04 (0.00-1.64)	0.07 (0.00-2.01)	0.13 (0.00-3.56)	0.04 (0.00-1.01)	0.15 (0.00-5.77)	0.02 (0.00-1.88)	0.06 (0.00-1.49)	0.04 (0.00-1.69)	0.12 (0.01-2.40)	0.09 (0.00-2.87)	0.04 (0.00-1.33)	0.23 (0.01-7.41)	0.14 (0.00-4.88)	0.06 (0.00-9.80)	0.16 (0.00-4.28)	0.11 (0.00-4.93)	1.15 (0.01-›100)
0.71 (0.22-2.27)	**LIT+OXC**	1.82 (0.24-12.20)	3.24 (0.44-23.26)	0.97 (0.11-8.77)	3.80 (0.33-44.39)	0.45 (0.01-19.81)	1.54 (0.26-9.16)	0.85 (0.05-13.79)	2.95 (0.34-25.00)	2.23 (0.24-19.23)	1.10 (0.14-8.78)	5.78 (0.65-50.00)	3.37 (0.29-40.00)	1.41 (0.02 to134.69)	3.98 (0.60-27.00)	2.65 (0.17-45.65)	28.57 (0.51-›100)
0.58 (0.22-1.55)	0.81 (0.40-1.63)	**OLZ**	1.78 (0.64-6.13)	0.53 (0.13-2.67)	2.08 (0.35-13.89)	0.25 (0.01-7.58)	0.84 (0.40-2.19)	0.47 (0.05-4.43)	1.62 (0.42-7.58)	1.22 (0.36-4.55)	0.61 (0.17-2.47)	3.17 (0.84-13.63)	1.85 (0.33-11.76)	0.77 (0.01-48.91)	2.18 (0.95-6.13)	1.45 (0.15-14.28)	15.52 (0.43-›100)
0.55 (0.20-1.49)	0.77 (0.38-1.57)	0.96 (0.66-1.34)	**QTP**	0.30 (0.07-1.38)	1.17 (0.18-7.13)	0.14 (0.00-4.24)	0.48 (0.19-1.14)	0.26 (0.03-2.25)	0.91 (0.21-4.02)	0.69 (0.16-2.79)	0.34 (0.09-1.34)	1.78 (0.44-6.99)	1.04 (0.19-5.71)	0.44 (0.01-27.06)	1.23 (0.57-2.73)	0.82 (0.09-7.77)	8.74 (0.23-›100)
0.55 (0.20-1.49)	0.77 (0.37-1.64)	0.96 (0.62-1.48)	1.00 (0.64-1.60)	**LIT+VPA**	3.91 (0.45-31.69)	0.46 (0.01-16.35)	1.58 (0.45-5.37)	0.88 (0.08-10.06)	3.03 (0.83-11.11)	2.29 (0.36-13.51)	1.14 (0.22-5.61)	5.95 (1.02-33.33)	3.46 (0.43-27.78)	1.45 (0.02-›100)	4.09 (1.01-16.96)	2.72 (0.24-33.22)	29.41 (0.65-›100)
0.55 (0.18-1.62)	0.76 (0.33-1.77)	0.95 (0.54-1.66)	0.99 (0.56-1.77)	0.99 (0.52-1.86)	**ARP+LTG**	0.12 (0.00-5.05)	0.41 (0.07-2.24)	0.22 (0.02-3.13)	0.78 (0.10-6.29)	0.59 (0.08-4.41)	0.29 (0.04-2.17)	1.52 (0.47-5.13)	0.89 (0.09-8.70)	0.37 (0.00-31.25)	1.05 (0.20-5.84)	0.70 (0.05-10.42)	7.46 (0.23-›100)
0.46 (0.16-1.32)	0.65 (0.30-1.41)	0.81 (0.49-1.30)	0.84 (0.51-1.39)	0.84 (0.48-1.44)	0.85 (0.43-1.65)	**LIT+IMP**	3.41 (0.12-89.12)	1.89 (0.04-94.44)	6.54 (0.20-›100)	4.95 (0.15-›100)	2.45 (0.07-73.77)	12.82 (0.37-›100)	7.46 (0.18-›100)	3.13 (0.10-84.61)	8.82 (0.31-›100)	5.87 (0.11-›100)	62.50 (0.43-›100)
0.47 (0.18-1.21)	0.65 (0.34-1.26)	0.81 (0.62-1.03)	0.84 (0.64-1.13)	0.84 (0.58-1.21)	0.85 (0.51-1.44)	1.00 (0.66-1.53)	**LIT**	0.55 (0.07-4.64)	1.92 (0.60-6.37)	1.45 (0.39-5.32)	0.72 (0.25-2.04)	3.76 (1.13-12.66)	2.19 (0.41-11.90)	0.92 (0.02-54.05)	2.58 (1.33-5.39)	1.72 (0.20-15.00)	18.52 (0.54-›100)
0.46 (0.16-1.36)	0.65 (0.28-1.49)	0.80 (0.46-1.38)	0.84 (0.48-1.46)	0.84 (0.44-1.56)	0.85 (0.41-1.70)	1.00 (0.52-1.93)	0.99 (0.59-1.66)	**ARP**	3.46 (0.30-38.46)	2.62 (0.23-29.41)	1.30 (0.12-13.89)	6.80 (0.65-66.67)	3.95 (0.33-47.62)	1.66 (0.02-›100)	4.67 (0.64-34.26)	3.11 (0.16-55.56)	33.33 (0.61-›100)
0.46 (0.18-1.16)	0.64 (0.32-1.29)	0.79 (0.56-1.13)	0.83 (0.58-1.23)	0.83 (0.58-1.19)	0.84 (0.47-1.50)	0.98 (0.61-1.62)	0.98 (0.77-1.28)	0.99 (0.56-1.78)	**VPA**	0.76 (0.12-4.29)	0.38 (0.08-1.79)	1.96 (0.36-10.76)	1.14 (0.14-8.79)	0.48 (0.01-34.54)	1.35 (0.35-5.32)	0.90 (0.08-10.83)	9.61 (0.23-›100)
0.45 (0.17-1.24)	0.63 (0.31-1.32)	0.79 (0.55-1.10)	0.82 (0.56-1.22)	0.83 (0.50-1.32)	0.83 (0.46-1.49)	0.98 (0.58-1.64)	0.97 (0.70-1.34)	0.98 (0.55-1.75)	0.99 (0.66-1.47)	**RisLAI**	0.50 (0.09-2.66)	2.59 (0.50-14.15)	1.51 (0.22-11.13)	0.63 (0.01-42.48)	1.78 (0.54-6.41)	1.19 (0.09-14.18)	12.70 (0.32-›100)
0.43 (0.15-1.22)	0.60 (0.27-1.30)	0.74 (0.45-1.19)	0.77 (0.46-1.27)	0.77 (0.44-1.32)	0.78 (0.40-1.51)	0.92 (0.51-1.65)	0.91 (0.60-1.38)	0.92 (0.47-1.78)	0.93 (0.56-1.51)	0.94 (0.55-1.57)	**CBZ**	5.24 (1.07-26.32)	3.05 (0.44-21.74)	1.28 (0.04-12.94)	3.60 (1.04-12.94)	2.40 (0.22-27.78)	25.64 (0.62-›100)
0.38 (0.14-1.01)	0.53 (0.27-1.06)	0.66 (0.48-0.89)	0.69 (0.50-0.96)	0.68 (0.44-1.04)	0.69 (0.43-1.10)	0.81 (0.51-1.31)	0.81 (0.64-1.02)	0.82 (0.47-1.41)	0.83 (0.59-1.15)	0.83 (0.59-1.19)	0.89 (0.56-1.44)	**LTG**	0.58 (0.08-4.20)	0.24 (0.00-17.78)	0.69 (0.21-2.35)	0.46 (0.04-5.18)	4.90 (0.19-›100)
0.34 (0.12-0.99)	0.48 (0.22-1.05)	0.60 (0.37-0.94)	0.63 (0.39-1.01)	0.63 (0.36-1.08)	0.63 (0.33-1.20)	0.74 (0.42-1.34)	0.74 (0.48-1.13)	0.75 (0.39-1.41)	0.76 (0.46-1.22)	0.76 (0.47-1.25)	0.81 (0.45-1.47)	0.91 (0.58-1.43)	**PAL**	0.42 (0.01-31.12)	1.18 (0.26-5.47)	0.79 (0.06-10.78)	8.41 (0.19-›100)
0.30 (0.11-0.84)	0.43 (0.20-0.89)	0.53 (0.34-0.80)	0.55 (0.36-0.86)	0.55 (0.33-0.90)	0.56 (0.30-1.03)	0.66 (0.43-1.00)	0.65 (0.46-0.92)	0.66 (0.36-1.21)	0.67 (0.43-1.01)	0.67 (0.42-1.06)	0.72 (0.42-1.23)	0.80 (0.54-1.20)	0.88 (0.51-1.50)	**IMP**	2.82 (0.05-›100)	1.88 (0.02-›100)	20.00 (0.09-›100)
0.29 (0.11-0.76)	0.40 (0.21-0.79)	0.50 (0.39-0.63)	0.52 (0.40-0.68)	0.52 (0.35-0.77)	0.53 (0.32-0.88)	0.62 (0.40-0.96)	0.62 (0.53-0.72)	0.62 (0.38-1.03)	0.63 (0.47-0.83)	0.64 (0.48-0.85)	0.68 (0.44-1.06)	0.76 (0.62-0.94)	0.84 (0.56-1.24)	0.95 (0.66-1.36)	**PLB**	0.67 (0.08-5.62)	7.14 (0.21-›100)
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	**FLX**	10.75 (0.19-›100)
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	**VPA+LTG**

Figure 3: Efficacy (any mood episode relapse or recurrence) and tolerability (discontinuation due to adverse event) according to the network meta-analysis

Comparisons between treatments should be read from left to right and the estimates are in the cell in common between the column-defining treatment and the row-defining treatment. Drugs are reported in order of efficacy (any mood episode relapse or recurrence) ranking estimated by SUCRA (surface under the cumulative ranking curve). For tolerability, a risk ratio (RR) lower than 1·00 favours the row-defining treatment. For any mood episode relapse or recurrence, a RR lower than 1·00 favours the column-defining treatment. Significant results are in bold. The RR of drug B over drug A can be obtained by calculating the inverse of the RR of drug A over drug B. ARP=aripiprazole. CBZ=carbamazepine. CrI=credible interval. FLX=fluoxetine. IMP=imipramine. LIT=lithium. LTG=lamotrigine. OLZ=olanzapine. QTP=quetiapine. OXC=oxcarbazepine. PAL=paliperidone. PLB=placebo. RisLAI= risperidone longacting injection. VPA=valproate.

We did sensitivity analyses with respect to publication year, bipolar disorder subtype, rapid-cycling course of illness, enrichment design, sponsorship from pharmaceutical company, study duration, and blinding of the trial (appendix pp 125–31). When analyses were restricted to trials with bipolar I disorder, lithium plus imipramine seemed to increase manic relapse or recurrence. Exclusion of the studies without rapid-cycling bipolar disorder participants left 12 trials, and we noted no differences in the conclusions of primary and secondary outcomes when assessing these trials only. Giving less weight to studies with enrichment design, sponsorship from a pharmaceutical company had no or little effect on estimates of all the outcomes across the network. When the studies were restricted to those that had at least 52 weeks of follow-up or those with a double-blind design, the results showed little or no effect on estimates of any outcomes (appendix pp 125–31).

Discussion

Our comprehensive search for relevant trials identified 33 randomised controlled trials (6846 participants) of drug therapies in the maintenance treatment of bipolar disorder.

Within the main network consisting of closed loops (figure 2), all drugs or combinations, except for imipramine, were significantly more efficacious in the prevention of any mood episode relapse or recurrence than was placebo, by sizeable margins. With respect to the secondary outcomes of prophylactic efficacy, only quetiapine and lithium prevented relapse or recurrence of both polarities of the mood episode, compared with placebo (figure 4). However, we noted considerable differences in design features of the included trials (table 1). Lithium was the dominant node in the evidence network, and the evidence for lithium was well balanced in terms of mood states at recruitment, with small (or possibly null) contributions from enrichment design trials (despite its discovery about 60 years ago, most evidence about lithium has been produced in the past 15 years and lithium has often been the reference drug in registration studies about second-generation antipsychotics, ruling out the potential for sponsorship bias). In quetiapine and lamotrigine studies, the participants were more balanced in terms of mood states at study entry than were participants in olanzapine trials, but they were enriched; in olanzapine trials only participants with an acute or recent manic or mixed episode were recruited, but they were more balanced in terms of enrichment than were quetiapine and lamotrigine trials (table 1; appendix p 90). In risperidone longacting injection and fluoxetine studies, participants with specific polarity were recruited and only those responding to the investigational drug were eligible (table 1; appendix pp 55–58). Olanzapine, lithium plus valproate, and risperidone longacting injection seemed to be more prophylactic for manic episodes than for depressive episodes, whereas lamotrigine might be more prophylactic for depressive episodes (figure 4). These
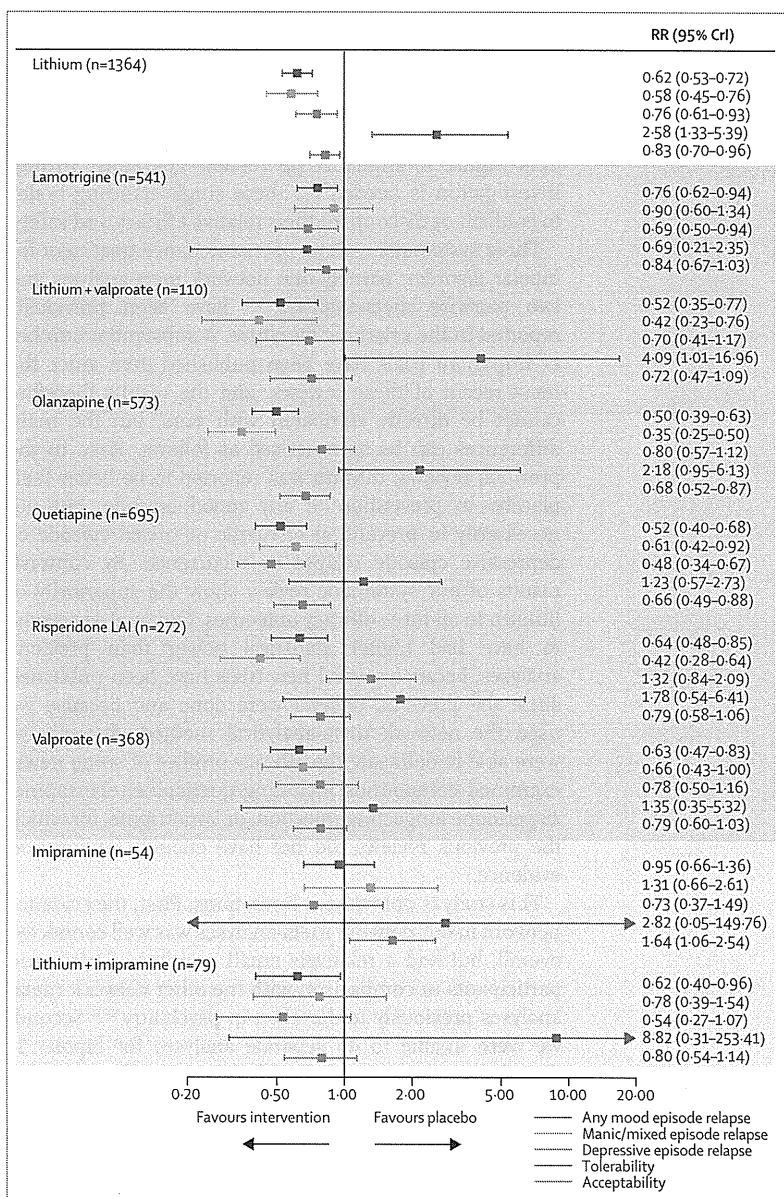
Figure 4: Efficacy according to type of mood episode recurrence or relapse, and tolerability and acceptability, compared with placebo

Results from the main closed-loop network are shown for any mood episode relapse or recurrence (dark blue line), manic, hypomanic, or mixed episode relapse or recurrence (green line), depressive episode relapse or recurrence (light blue line), tolerability (dark red line), and acceptability (red line). Fluoxetine is excluded from the plot because the result for manic, hypomanic, or mixed episode relapse or recurrence was not reported. The interventions are divided into three groups: the white background shows that all three efficacy outcomes are statistically significant and the confidence in estimate of RR to prevent any mood episode relapse is moderate; the light blue background shows that either one of three efficacy outcomes is statistically non-significant or the confidence in estimate is low; and the light green background shows that two or more of the efficacy outcomes are statistically non-significant or the confidence in estimates is low or very low. Treatments are presented in alphabetical order in each group. RR=risk ratio. CrI=credible interval. LAI=longacting injection.

drugs could be a second choice for a patient who has a specific dominant polarity.

We then examined the single-standing nodes, which do not form closed loops and are often connected to the

main network by one trial only with a relatively small sample size (figure 2). In our evidence network, several combination treatments seem to have favourable point estimates of RRs for efficacy, tolerability, or acceptability. However, they were not statistically significant, with wide CrIs (figure 3; appendix pp 76–80). Therefore, further investigation is needed for these single-standing nodes to confirm or disconfirm their relative efficacy and safety.

Three systematic reviews of maintenance treatment for bipolar disorder, namely one network meta-analysis and two pairwise meta-analyses,[29–31] have been previously reported in the scientific literature. A substantial number of important trials have been published even since the most recent of these reviews, and the results therefore cannot be directly compared with ours, but the main differences can be summarised as follows. First, in the previous reviews, lithium was reported to be better than placebo in prevention of any mood episode, but not necessarily in prevention of manic or mixed episode or depressive episode relapse or recurrence. By contrast, results of our systematic review show the superiority of lithium in all three efficacy outcomes. Our analysis seems to have had higher statistical power than previous analyses, because several new trials have been published since the previous reviews were done and because we used the network meta-analytical method. Second, we were able to delineate the efficacy profiles of some newly examined compounds including quetiapine, olanzapine, risperidone longacting injection, or lamotrigine, for which the previous reviews did not have enough randomised evidence.

This study is not without limitations. First, the evidence network in our network meta-analyses was well connected overall, but had a relatively small number of trials and participants in comparison with the other network meta-analyses previously undertaken in psychiatry.[10–12] Second, we were unable to do separate analyses for bipolar II disorder or for rapid-cycling bipolar disorder, and different drugs might have different efficacy profiles for different subtypes. However, exclusion of the few studies that focused on these disorders or inclusion of them in the total evidence network did not materially change the results. Third, many of the studies of maintenance treatment for bipolar disorder were funded by pharmaceutical companies and used the enrichment design to select patients who responded to treatment in the acute phase (tables 1, 2), which might give clear advantage to the investigational drug and cause a sponsorship bias. The effect of these study limitations were taken into account when we assessed the quality of evidence behind major comparisons. However, sensitivity analyses taking into account the effect of potentially favouring the newest treatments across the network did not produce materially different results.

In conclusion, even though the generalisation of our study's findings to real-world clinical practice will be difficult, some important clinical implications can be drawn. Lithium seems to be the most reasonable candidate for a first-line treatment option for the long-term treatment of bipolar disorder (it is one of the most effective treatments in the prevention of both manic and depressive episodes, with the most robust and unbiased evidence, with a higher rate of adverse events than placebo, but not substantially more dropout due to any cause). Quetiapine might also be a suitable choice, but because the quetiapine studies were heavily biased by enrichment design, the evidence supporting quetiapine should be interpreted with caution. Additionally, when a patient's dominant polarity is known, evidence suggests that olanzapine is more antimanic than is quetiapine and lithium, and lamotrigine is more effective than placebo in the prevention of depressive relapse or recurrence. The other drugs in the closed-loop network—except for imipramine and lithium plus imipramine—should be considered as third-line treatments even though they are all more effective than placebo in the prevention of any mood episode. All these drugs have very different side-effect profiles and this important clinical issue has to be taken into account at the individual patient level.

Two research implications follow. First, our results suggest that some drugs could be divided into two classes according to their relative efficacy of prophylactic activity against depressive episodes or manic, hypomanic, or mixed episodes. The relation between patients' polarity and drugs' characteristics should be more clearly recognised and researched in future trials. Second, because none of the examined and available mono-therapies is clearly effective for all required aspects of bipolar maintenance therapy, and because some of the trialled cotherapies provide hopeful leads (albeit with wide CrIs), future research in this domain should focus on the above-mentioned stronger candidates and their combinations.

References
1 American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th edn. Arlington, VA: American Psychiatric Association, 2013.
2 Phillips ML, Kupfer DJ. Bipolar disorder diagnosis: challenges and future directions. *Lancet* 2013; 381: 1663–71.
3 Merikangas KR, Jin R, He JP, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry* 2011; 68: 241–51.
4 Goodwin FK, Jamison KR. Manic-depressive illness, 2nd edn. Oxford, UK: Oxford University Press, 2007.
5 Keller MB, Lavori PW, Coryell W, Endicott J, Mueller TI. Bipolar I: a five-year prospective follow-up. *J Nerv Ment Dis* 1993; 181: 238–45.
6 Tohen M, Waternaux CM, Tsuang MT. Outcome in mania. A 4-year prospective follow-up of 75 patients utilizing survival analysis. *Arch Gen Psychiatry* 1990; 47: 1106–11.
7 National Institute for Health and Care Excellence. Bipolar disorder: the management of bipolar disorder in adults, children and adolescents, in primary and secondary care, July 2006. http://www.nice.org.uk/CG038 (accessed April 14, 2014).
8 Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) collaborative update of CANMAT guidelines for the management of patients with bipolar disorder: update 2013. *Bipolar Disord* 2013; 15: 1–44.
9 Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *Lancet* 2013; 381: 1672–82.
10 Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* 2011; 378: 1306–15.
11 Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746–58.
12 Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 2013; 382: 951–62.
13 Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions, version 5.1.0, updated March, 2011. www.cochrane-handbook.org (accessed Aug 5, 2013).
14 Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331: 897–900.
15 Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996; 15: 2733–49.
16 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23: 3105–24.
17 Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods* 2012; 3: 80–97.
18 Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17: 279–301.
19 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–88.
20 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010; 36: 1–48.
21 Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Stat Med* 2009; 28: 3049–67.
22 Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013; 159: 130–7.
23 Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PloS One* 2013; 8: e76654.
24 Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making* 2013; 33: 641–56.
25 Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; 62: 857–64.
26 Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011; 64: 401–06.
27 Amsterdam JD, Garcia-España F, Fawcett J, et al. Efficacy and safety of fluoxetine in treating bipolar II major depressive episode. *J Clin Psychopharmacol* 1998; 18: 435–40.
28 Cipriani A, Barbui C, Rendell J, Geddes JR. Clinical and regulatory implications of active run-in phases in long-term studies for bipolar disorder. *Acta Psychiatr Scand* 2014; 129: 328–42.
29 Soares-Weiser K, Bravo Vergel Y, Beynon S, et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder. *Health Technol Assess* 2007; 11: iii–iv, ix–206.
30 Smith LA, Cornelius V, Warnock A, Bell A, Young AH. Effectiveness of mood stabilizers and antipsychotics in the maintenance phase of bipolar disorder: a systematic review of randomized controlled trials. *Bipolar Diord* 2007; 9: 394–412.
31 Vieta E, Günther O, Locklear J, et al. Effectiveness of psychotropic medications in the maintenance phase of bipolar disorse: a meta-analysis of randomized controlled trials. *Int J Neuropsychopharmacol* 2011; 14: 1029–49.

– 89 –

妊産婦の保健を対象とした系統的レビューに携わる人材発掘の調整と育成

研究分担者 大槻　克文（昭和大学江東豊洲病院　准教授）

研究要旨

　臨床研究報告を系統的にレビューするコクラン計画は世界的には認知度が高く、その有用性は高い。また、その作成に携わる人材やその支援体制は確立している。一方で、日本におけるコクラン共同計画の認知度は低く、特に周産期領域での人材発掘とその育成は喫緊の課題である。本分担研究では、題目の通り『妊産婦の保健を対象とした系統的レビューに携わる人材発掘の調整と育成』を平成 25 年度に引き続き 26 年度の目的とした。

　今年度は、１．コクランレビューワークショップ参加者に対する周産期領域、特に産科領域からのサポート、２．周産期領域での学会等における「コクランレビューに関する説明会の開催、３．学会や医局でのロビー活動（啓発活動）、４．次年度の方策検討、について活動を行った。

A. 研究目的

　日本におけるコクラン共同計画の認知度は低く、特に周産期領域での人材発掘とその育成は喫緊の課題である。本分担研究では、題目の通り『妊産婦の保健を対象とした系統的レビューに携わる人材発掘の調整と育成』を平成 25 年度に引き続き平成 26 年度の目的とした。

B. 研究方法

　１　「コクランレビューワークショップ参加者に対する周産期領域、特に産科領域からのサポート」

　本研究の主任研究者である森　臨太郎　独立行政法人・国立成育医療研究センター・研究所　政策科学研究部・部長が複数回主催するコクランレビューワークショップに出席し、本研究への理解を深めるとともに、参加者とのコミュニケーションを介して、周産期領域、特に産科領域からのサポートを行う。さらに出席者の所属施設や背景を分析することで、人材の偏りの有無につき分析を行うこととした。

　２　「周産期領域での学会等における「コクランレビューに関する説明会の開催」

　本邦で開催される周産期領域、産婦人科領域での学会や研究会主催者に働きかけ、「周産期領域での学会等における「コクランレビューに関する説明会の開催」を試みることとした。

　３　「学会や医局でのロビー活動（啓発活動）」

　上記２．と共に周産期領域での各種学会や医局において、当分担研究者の知りうる限りの若手医師に対して、「日本におけるコクラン共同計画の認知度」を高めるべく、啓発活動を行うこととした。

　４　「次年度の方策検討」

上記1.～3.をふまえた上で、より効果的、効率的な啓発活動のあり方を、緻密に検討することとした。

（倫理面への配慮）
　本研究は人材の発掘と育成が目的であり、通常の臨床研究に求められる倫理面への配慮は前提としない。

C. 研究結果
　1　「コクランレビューワークショップ参加者に対する周産期領域、特に産科領域からのサポート」
　本ワークショップに出席し、本研究への理解を深めるとともに、参加者とのコミュニケーションを介して、周産期領域、特に産科領域からのサポートを行うこととした。

　2　「周産期領域での学会等における「コクランレビューに関する説明会の開催」
　平成27年4月に横浜で開催される第67回日本産科婦人科学会(学術集会会長：峰岸　敬教授（群馬大学）) 事務局に対して、日本におけるコクラン共同研究の主旨を説明し,当該学術集会内での「コクランレビューに関する説明会の開催」開催許可を依頼した。また、本邦で開催される周産期領域、産婦人科領域での学会や研究会主催者に働きかけ、「周産期領域での学会等における「コクランレビューに関する説明会の開催」を試みた。

　3　「学会や医局でのロビー活動（啓発活動）」
　上記2.と共に周産期領域での各種学会や医局において、当分担研究者の知りうる限りの若手医師に対して、「日本におけるコクラン共同計画の認知度」を高めるべく、啓発活動を実施した。

　4　「次年度の方策検討」

上記検討結果を踏まえて、主任研究者である森臨太郎先生と問題点の抽出と協議を行い。平成27年度の方策を検討した。

D. 考察
　今年度は、1.「コクランレビューワークショップ参加者に対する周産期領域、特に産科領域からのサポート」、2.「周産期領域での学会等における「コクランレビューに関する説明会」の開催」、3.「学会や医局でのロビー活動（啓発活動）」、4.「次年度の方策検討」、について活動を行った。
　平成25年7月に開催された日本周産期新生児医学会学術集会期間中でのワークショップにおいては、若手医師の出席者数が決して多い状況ではなかった。コクランレビューワークショップでは当初は産婦人科医師の出席者はほぼ皆無であったが、次第に参加者が増えてきた。しかしながらこのワークショップへの出席者には周産期医療の第一線で勤務質ている者は少なく、本邦の医療従事者の職務環境（多忙など）が影響している可能性が垣間見られた。学会や会合で若い先生へ声を掛け、コクラン共同研究の説明を行うも、多忙と英語力への不安あるが故に興味を有することができないという意見が大多数をしてめていた。
　以上より、本邦でのコクラン共同研究、特に『妊産婦の保健を対象とした系統的レビューに携わる人材発掘の調整と育成』には多大の労力、時間、臨床家の職務環境整備などが必要であることが改めて認識された。

E. 結論
　産科領域での人材発掘と育成に関しては、今一度今後の方策を緻密に考える必要はあることが明らかであった。

F. 研究発表
1. 論文発表

1. Tokunaka M, Hasegawa J, Oba T, Nakamura M, Matsuoka R, Ichizuka K, Otsuki K, Okai T, Sekizawa A. Decidual polyps are associated with preterm delivery in cases of attempted uterine cervical polypectomy during the first and second trimester. J Matern Fetal Neonatal Med. 2014 Jul 30:1-3.

2. Hayakawa M, Ito Y, Saito S, Mitsuda N, Hosono S, Yoda H, Cho K, Otsuki K, Ibara S, Terui K, Masumoto K, Murakoshi T, Nakai A, Tanaka M, Nakamura T Incidence and prediction of outcome in hypoxic-ischemic encephalopathy in Japan. ; Executive Committee, Symposium on Japan Society of Perinatal and Neonatal Medicine. Pediatr Int. 2014 Apr;56(2):215-21.

3. Otsuki K, Tokunaka M, Oba T, Nakamura M, Shirato N, Okai T. Administration of oral and vaginal prebiotic lactoferrin for a woman with a refractory vaginitis recurring preterm delivery: appearance of lactobacillus in vaginal flora followed by term delivery. J Obstet Gynaecol Res. 2014 Feb;40(2):583-5.

4. 大槻 克文. 妊娠後半期における妊娠維持機構とその破綻 日産婦データベースを用いた因子解析と多施設共同 RCT に基づく背景別早産予防対策.. 日本産科婦人科学会雑誌 66, 2499-2511(2014)

5. 苅部 瑞穂, 安藤 直子, 野村 奈央, 奥山 亜由美, 秋野 亮介, 宮上 哲, 安水 渚, 折坂 勝, 佐々木 康, 大槻 克文, 長塚 正晃. 腹腔鏡下手術にて間質部妊娠が確定できなかった１例. 神奈川産科婦人科学会誌 51 巻 1 号 Page44-47(2014.09)

6. 三科 美幸, 長谷川 潤一, 滝 元宏, 仲村 将光, 松岡 隆, 市塚 清健, 大槻 克文, 関沢 明彦. 腎出血により新生児死亡を来した後部尿道弁の１例. 日本周産期・新生児医学会雑誌 50 巻 3 号 Page1068-1072(2014.09)

7. 大槻克文. 婦人科診療ハンドブック「感染症 30. 細菌性腟症の診断と治療」198-202 杉山徹 編 中外医学社

8. 大槻克文. NICU マニュアル 第 5 版「1 総論 リスク評価 4. 胎児心拍モニタリング」新生児医療協議会 編 金原出版株式会社

9. 大槻克文 『たまごクラブ』1 月号「産院に行くべき"危険サイン"リスト」、風讃社 2014

10. 大槻 克文. Pre-mo［プレモ］春号『「切迫早産」乗り越えて無事産まれました』. p88-91

11. 大槻 克文. Pre-mo［プレモ］秋号.

12. 大槻 克文. たまごクラブの 2014/08/12 発売号 『おなかの赤ちゃんの不安 スッキリ解消クリニック』. P122-128

13. 大槻 克文. たまごクラブ 2014/11/15 発売号『妊娠 12~36 週のママ、必読！[張り・痛み]おなかの[出血][破水]流産・早産の兆候、見逃さないで!』. P36-51

14. 大槻 克文. 【感染症診療 update】 (II 章)主要な臓器感染症 産科感染症 絨毛膜羊膜炎. 日本医師会雑誌 143, S236-S239(2014)

15. 大槻 克文, 神保 正利, 太田 創. 【管理法はどう変わったか?:温故知新 産科編】頸管無力症. 周産期医学 44 巻 3 号 Page331-336(2014.03)

16. 太田 創, 大場 智洋, 大槻 克文, 徳中 真由美. 【ルチーンケアの根拠を答えられますか?ふりかえりの助産業務と「なぜ?」「どうして?」エビデンス】 前期破水で内診してはいけないのはなぜですか. ペリネイタルケア 33 巻 3 号 Page246-248(2014.03)

17. 太田 創, 大場 智洋, 大槻 克文, 徳中 真由美. 【ルチーンケアの根拠を答えられますか?ふりかえりの助産業務と「なぜ?」

「どうして?」エビデンス】 切迫早産で安静の指示が出るのはなぜですか.. ペリネイタルケア 33 巻 3 号 Page241-245(2014.03)

18. 宮上 哲, 大槻 克文.【読み方がわかる!説明できる!産科の臨床検査ディクショナリー これさえあれば妊婦健診で困らない! エコーもCTGも】 (第8章)ケーススタディ 検査はこう活用しよう! 妊婦が羊水流出感を自覚した. ペリネイタルケア 2014 新春増刊 Page277-281(2014.01)

19. 奥山 亜由美, 大槻 克文.【読み方がわかる!説明できる!産科の臨床検査ディクショナリー これさえあれば妊婦健診で困らない! エコーもCTGも】 (第6章)分娩時に必要な検査を理解しよう! Bishopスコア. ペリネイタルケア 2014 新春増刊 Page208-211(2014.01)

20. 秋野 亮介, 大槻 克文.読み方がわかる!説明できる!産科の臨床検査ディクショナリー これさえあれば妊婦健診で困らない! エコーもCTGも】 (第4章)特別なニードがある場合の検査を理解しよう! 早産マーカー. ペリネイタルケア 2014 新春増刊 Page193-195(2014.01)

21. 太田 創, 大場 智洋, 徳中 真由美, 大槻 克文.【読み方がわかる!説明できる!産科の臨床検査ディクショナリー これさえあれば妊婦健診で困らない! エコーもCTGも】 (第2章)超音波検査を理解しよう! 子宮頸管長・内子宮口の形態. ペリネイタルケア 2014 新春増刊 Page110-116(2014.01)

22. 折坂 勝, 大槻 克文.【読み方がわかる!説明できる!産科の臨床検査ディクショナリー これさえあれば妊婦健診で困らない! エコーもCTGも】 (第1章)妊娠中の基本検査を理解しよう! 細菌関連検査 腟分泌物培養検査. ペリネイタルケア 2014 新春増刊 Page060-063(2014.01)

23. 小出 容子, 大槻 克文, 山本 松男, 関沢 明彦.【新たな早産予防戦略】 歯周病と早産. 産科と婦人科 81 巻 1 号 Page51-54(2014.01)

24. 大槻 克文.【新たな早産予防戦略】 頸管長と早産. 産科と婦人科 81 巻 1 号 Page39-45(2014.01)

2. 学会発表

1. 大槻 克文.妊娠後半期における妊娠維持機構とその破綻 日産婦データベースを用いた因子解析と多施設共同 RCT に基づく背景別早産予防対策. 日本産科婦人科学会第66回学術講演会 東京 (2014.04)

2. Katsufumi OTSUKI. Trial for prevention of preterm birth in Japan. 46th International Conference on Pathophysiology of Pregnancy. Tokyo (2014.09)

3. 大槻 克文.ラクトフェリンによる早産抑制効果について. 第4回臨床ラクトフェリンシンポジウム 東京 (2014.03)

4. 大槻 克文.早産ハイリスク症例に対する腟内環境改善を目指したラクトフェリン (LF) による早産予防の試み. JAPAN2014(第12回ヘルスフードエキスポ) 日本ラクトフェリン学会セミナー 東京ビッグサイト (2014.05)

5. 大槻 克文.本邦における早産予防対策の現状と今後の方策. 江戸川区産婦人科医会 東京 江戸川区医師会館 (2014.09)

6. 大槻 克文.切迫早産予知のための管理と考え方.第2回江東豊洲産婦人科懇話会 東京 江東区 (2014.12)

7. 折坂 勝, 佐々木 康, 宮上 哲, 遠武 孝祐, 苅部 瑞穂, 野村 奈央, 奥山 亜由美, 小谷 美帆子, 安水 渚, 安藤 直子, 大槻 克文, 長塚 正晃, 高橋 諄.子宮体部扁平上皮癌の一例. 日本婦人科腫瘍学会第56回学術集会 (2014.06)

8. 大場 智洋, 大槻 克文, 徳中 真由美,

長谷川 潤一，松岡 隆，市塚 清健，関沢 明彦.本邦の周産期センターにおける早産予防管理の多様性と問題点．日本産科婦人科学会第66回学術集会（2014.04）

1．奥山 亜由美，大槻 克文，野村 奈央，秋野 亮介，宮上 哲，苅部 瑞穂，安水 渚，折坂 勝，安藤 直子，佐々木 康，長塚 正晃.妊娠中の子宮頸管ポリープ切除と病理診断及び妊娠予後に関する検討．日本産科婦人科学会第 66 回学術集会（2014.04）

2．秋野 亮介，安藤 直子，野村 奈央，奥山 亜由美，宮上 哲，安水 渚，苅部 瑞穂，折坂 勝，佐々木 康，大槻 克文，長塚 正晃.帝王切開分娩後多量出血に対する子宮動脈塞栓術施行により子宮筋層壊死・子宮留膿症をきたし子宮腟上部切断術を余儀なくされた重症妊娠高血圧腎症の一例．日本産科婦人科学会第 66 回学術集会（2014.04）

3．宮上 哲，安藤 直子，秋野 亮介，苅部 瑞穂，野村 奈央，奥山 亜由美，安水 渚，折坂 勝，佐々木 康，大槻 克文，長塚 正晃.腹腔鏡下子宮全摘出術の標準化と若手への教育．日本産科婦人科学会第66回学術集会（2014.04）

4．大場 智洋，大槻 克文，仲村 将光，長谷川 潤一，関沢 明彦.早産ハイリスク症例に対する腟内環境改善を目指したラクトフェリン（LF）による早産予防の試み．第 31 回日本産婦人科感染症研究会学術集会 神戸(2014.06)

5．神保正利、西 健、内山心美、小山恵子、太田 創、前田雄岳、岡崎美寿歩、荒木美智子、大山 香、大槻克文.新規開院した大学附属病院における医師勤務体制構築への新たな取り組み．第 128 回関東連合産科婦人科学会総会・学術集会 松本（2014.10）

6．西 健、内山心美、小山恵子、太田 創、前田雄岳、岡崎美寿歩、荒木美智子、大山 香、神保正利、大槻克文.クラウドシステムを用いた医師勤務管理と業務効率化の試み．第 128 回関東連合産科婦人科学会総会・学術集会 松本（2014.10）

G．知的財産権の出願・登録状況
1．特許取得　なし
2．実用新案登録　なし
3．その他　なし

国際蘇生法連絡委員会(International Liaison Committee on Resuscitation: ILCOR) ガイドライン策定におけるコクランレビュー活用の検討

研究分担者　田村正徳　埼玉医科大学総合医療センター

研究要旨

目的：国際蘇生法連絡委員会(International Liaison Committee on Resuscitation: ILCOR)では 2015 年のコンセンサスの改定にむけ GRADE(Grading of Recommendations Assessment, Development and Evaluation)システムを導入し、蘇生に関するガイドライン策定予定である。今回 ILCOR の旧論文評価法の改善のため、GRADE システムを採用した既存のコクランレビューを活用することが有用かを検討する。2012 年 12 月の ILCOR の会議においてコクランレビューを活用した GRADE evidence profile および GRADE finding table を例として発表、会議参加意見を収集した。その結果を踏まえさらにコクランレビューには含まれない非ランダム化比較試験 3 文献も加え 2014 年 12 月の ILCOR の会議において再度発表し、意見を収集した。

結果：コクランレビューを活用することにより、質の高い評価表を速やかに作成でき、その評価結果は ILCOR 会議において受け入れは良好であった。

考察：コクランレビューを活用することにより ILCOR ガイドライン作成において、その質を改善し、また作業をスムースとなり得る。

研究協力者:

研究協力者　　杉浦崇浩
静岡済生会総合病院　新生児科

A. 研究目的

国際蘇生法連絡委員会(International Liaison Committee on Resuscitation: ILCOR)では 2015 年のコンセンサスの改定に向け GRADE(Grading of Recommendations Assessment, Development and Evaluation)システムを導入し、蘇生に関するガイドラインを策定する予定である。今回 ILCOR の旧論文評価法の改善のため、GRADE システムを採用した既存のコクランレビューを活用することが有用かを検討する。

B. 研究方法

2012 年 4 月 28 日にアメリカ合衆国、ボストンにて開催された ILCOR 新生児部門会議で幾つかのクリニカルクエッション候補が選考され、その際同時に GRADE システムを採用したワークシートを作成し、例示する事が決定された。そこで既存のコクランレビュー(Rabe H, Cochrane Database Syst Rev. 2012 Aug 15;8:CD003248) を活用し GRADE evidence profile および GRADE finding table を作成し、2012 年 10 月 18 日のオーストリア、ウィーンでの ILCOR 全体会議および 2012 年 12 月 2 日のアメリカ合衆国、ワシントン D.C での ILCOR 新生児部門会議において発表した。その結果を踏まえ 2014 年 12 月 7 日のアメリカ合衆国、ワシントン D.C での ILCOR 新生児部門会議においてコクラン