British Journal of Psychiatry (1995), **166**, 378–381

Comparison of Carbamazepine and Lithium in the Prophylaxis of Bipolar Disorders

A Meta-analysis

R. DARDENNES, C. EVEN, F. BANGE and A. HEIM

Background. This meta-analysis assessed the equipotency of carbamazepine and lithium prophylaxis in bipolar disorder.
Method. We selected only randomised, double-blind, controlled studies comparing carbamazepine with lithium from a manual and computerised search, and subjected them to a quality inventory. Their statistical results were weighted by their quality score and combined.
Results. Four studies met our criteria, yielding, overall, $P=0.15$. This result is not straightforward because the studies showed significant heterogeneity ($P<0.01$).
Conclusion. Differences in statistical power and sensitivity of outcome measure explain this heterogeneity and the conflicting results of the studies. Therefore, the prophylactic efficacy of carbamazepine remains questionable.

The purpose of this work is to examine, by means of a meta-analysis, the validity of the opinion stating that carbamazepine (CBZ) equals lithium in prophylactic efficacy (Post *et al*, 1991).

A meta-analysis seemed to be timely inasmuch as other studies on this subject are unlikely to be undertaken in the near future. Investigations have indeed switched toward other compounds such as valproate.

Method

Selection of studies

In a manual and computerised bibliographic search covering the years 1970–93, we included only randomised, double-blind, controlled studies allowing statistical treatment (Placidi *et al*, 1986; Watkins *et al*, 1987; Lusznat *et al*, 1988; Coxhead *et al*, 1992).

The study by Okuma *et al* (1981) compared CBZ with placebo (and therefore could not enter the meta-analysis) and failed to demonstrate CBZ's superiority statistically ($P<0.10$). For obvious ethical reasons, it is likely to remain the only study comparing CBZ with placebo.

Qualitative analysis

We subjected the studies to a quality inventory formed of 16 explicit, previously defined, and uniformly applied criteria. This inventory resulted in an overall quality score summing the scores attributed to each item (0 or 1 point, depending on whether the study possessed the characteristic under investigation). These criteria, derived from those of King (1990), are defined as follows:

(1) Description of patients: at least age, sex, and number of previous episodes.
(2) Use of diagnosis criteria.
(3) Diagnosis confirmed by two independent assessors.
(4) Homogeneity of diagnosis (only bipolar disorders = 1, everything else = 0).
(5) No selection bias favouring treatment responders or non-responders.
(6) Concomitant psychotropic treatment (antidepressant or neuroleptic allowed = 0; not allowed = 1).
(7) Description of side-effects.
(8) Relevance of the dependent variable (failure rate = 0; length in remission = 1).
(9) At least one dependent variable directly reflecting patients' psychiatric conditions.
(10) Patients followed for at least five months.
(11) Patients assessed at least monthly.
(12) Compliance checking (e.g. plasmatic dosage or count of tablets).
(13) Drop-out rate ($\leqslant 15\% = 1$; $> 15\% = 0$).
(14) Statistical analysis of major hypotheses.
(15) Sample size based on statistical power calculation.
(16) Multivariate methods for two or more dependent variables.

It is difficult to interpret the results of a comparison of lithium and CBZ when data referring to previous treatments are not reported. Patients included have often had their illness for a long time

and presumably received lithium prophylaxis. When this is not specified, the lithium group may comprise a higher proportion of lithium non-responders, an obvious bias. Likewise, lithium non-responders may respond differently to CBZ.

The outcome measure is a critical point; we gave 1 point to the studies using the length in remission as an outcome measure and points to those using the failure rate, a method which puts on the same level the early and late recurrences and results in a loss of information and sensitivity. Because none of the other authors distinguished between 'relapse' and 'recurrence' we also avoided this issue.

The overall quality score (Table 1) allowed us to ascertain whether the studies were homogeneous in quality. At the quantitative stage of the meta-analysis, the statistical result of each study was weighted by its quality score.

Quantitative meta-analysis

We combined the statistical results (P levels) of the four studies to ascertain whether there was a difference in efficacy between CBZ and lithium. The methods of adding the standard normal deviation (Z) and of adding weighted Zs were used (Rosenthal, 1991). The latter enabled us to include the quality score of each study. The legitimacy of this calculation was ascertained by testing the statistical homogeneity of the studies for statistical significance (P) and effect size, r and its transformation Z_r being used as effect size estimates (Rosenthal, 1991):

$$r = \frac{Z}{\sqrt{N}} \qquad Z_r = \frac{1}{2} \log_e \frac{1+r}{1-r}$$

We computed the exact P levels by the χ^2 method when the proportion of failures was the outcome measure or the only usable variable. As Overall & Rhoades recommend (1987), when P was less than 0.001, we retained this value instead of the exact P.

Results

Qualitative results

Homogeneity of diagnosis (item 4). Coxhead *et al* (1992) included only bipolar patients. The other studies also included either schizoaffective (Lusznat *et al*, 1988), schizoaffective and schizophreniform (Placidi *et al*, 1986), or unipolar disorders (Watkins *et al*, 1987). For the last-named disorder, lithium does not overtly hold the reference position as it does for the prophylaxis of bipolar disorder. To prove that a new treatment is as effective as the reference drug, one must use the latter in its optimal conditions. Therefore, the studies assessing heterogeneous samples had no point.

Previous response to treatment as a selection bias. Watkins *et al* (1987) selected subjects who did not receive preventive treatment or had stopped it, thinking they no longer needed it. In Placidi *et al*'s sample (1986), there was a significant intergroup difference of the proportion of lithium non-responders (lithium group = 5/27 and CBZ group = 13/29; $\chi^2 = 4.44$; d.f. = 1; $P < 0.05$). Lusznat *et al* (1988) did not report the previous prophylactic treatments. Coxhead *et al* (1992) studied patients currently receiving lithium for whom it was thought "medically and ethically appropriate" to change treatment. Many of them probably responded poorly.

Relevance of the outcome measure. Watkins *et al* (1987) did not refer to the length of remission but to the additional time in remission from the previous time in remission (which had not been covered by

Table 1
Quality score, characteristics, and results of studies

Study	Placidi *et al* (1986)	Watkins *et al* (1987)	Lusznat *et al* (1988)	Coxhead *et al* (1992)
Quality score (out of 16 pts)	6	8	7	11
Diagnostic criteria	DSM-III*	DSM-III	DSM-III	DSM-III
Diagnosis	BP+SA+SE	BP+UP	BP+SA	BP
Follow-up duration	2-36 months	>prior cycle length	12 months	12 months or until relapse
Sample size	83	52	40	31
Drop-outs (%)	27 (32.5%) 68% after 12 months	15 (29%)	11 (27.5%)	3 (10%)
Outcome measure	Proportion of failures	Additional time in remission	Proportion of failures	Proportion of failures
Statistical results	$P=0.89$	$P<0.001$	$P=0.10$	$P=0.70$
Interpretation	CBZ = lithium	CBZ < lithium	CBZ = lithium	CBZ = lithium

BP: bipolar disorder; UP: unipolar disorder; SE: schizophreniform episodes; SA: schizoaffective disorder.
*American Psychiatric Association, 1980.

https://doi.org/10.1192/bjp.166.3.378 Published online by Cambridge University Press

any prophylactic agent). This study receives 1 point because, although the duration of remission since the episode preceding the index episode was collected retrospectively, it conveys additional information about individual patients and the specificity of the course of their disorders. Placidi *et al* (1986) did a survival analysis without displaying its numeric results. They also used a 'relapse index'; i.e. the number of relapses divided by the number of months under treatment of all the subjects of each group. This variable is as relevant and sensitive as the length in remission, for it also takes the time factor into account. Unfortunately, it is not usable because data necessary to calculate an exact *P* level were not provided. Therefore, the proportion of failures remained the only usable variable for a meta-analysis, and no point was given to this study. Lusznat *et al* (1988) and Coxhead *et al* (1992) both used the proportion of patients who relapsed.

Follow-up duration >5 months. Watkins *et al* (1987) gave only the average follow-up duration (lithium group: 20 months, CBZ group: 16 months). Nevertheless, because the follow-up exceeded the previous length in remission, this study receives 1 point. In Coxhead *et al*'s work (1992), the patients who did not relapse were followed up for 12 months, while those who relapsed fell into the failure category, so that it was no longer necessary to continue the follow-up. Therefore, this study gets 1 point. Placidi *et al* (1986) indicate a follow-up duration of 2–36 months, in some cases less than the five months required.

Quantitative results

The results of the four studies appear to be in clear conflict: Watkins *et al* (1987) found a significant difference in favour of lithium. The other three did not find any difference.

Significance levels were combined, yielding overall *P* values of 0.12 (combined $Z = 1.165$) and 0.15 (combined weighted $Z = 1.04$) when weighted with the quality scores. This could be read as showing an equal efficacy of the two compounds. But it states only that the observed difference in favour of lithium bore a 12% risk of resulting from sample variation under the null hypothesis. The studies were significantly heterogeneous ($\chi^2 = 15.98$; d.f. = 3; $P < 0.01$ for significance tests; $\chi^2 = 19.92$; d.f. = 3; $P < 0.001$ for effect size estimates). This heterogeneity prevents us from unequivocally interpreting the overall *P* value (0.15). The studies are few and bear unique characteristics; therefore, it was impossible to particularise homogeneous subsets of studies in order to find the grounds of this heterogeneity.

Discussion

Placidi *et al* (1986) used variables allowing for the time factor that were unusable for a meta-analysis; we then referred to a less sensitive variable (proportion of failures). This might have modified the final results.

However, there are very few randomised, double-blind controlled studies, and their results conflict. Disentangling these conflicting results chiefly affects the sensitivity of the outcome measure and the issue of statistical power.

To examine the issue of statistical power, one can calculate the smallest difference that could have been demonstrated to be significant by the negative studies (which failed to attain statistical significance), considering the sample sizes. For this purpose, we applied Casagrande *et al*'s (1978) formula to the negative study assessing the largest sample (Placidi *et al*, 1986), and we found a minimal difference (Δ) of 45%, meaning that it could have detected only a difference of efficacy greater than 45% (of failures among patients) between lithium and CBZ, with a type-two error (β) of 5%. In other words, the most powerful of the negative studies could, at best, suggest that the difference of failure rates lay between 0% and 45%. Such a high Δ precludes us from stating that CBZ is equivalent in effect to lithium. To reduce Δ to 25%, a level more consistent with the hypothesis of equality, we would require 80 subjects in each group (with $\beta = 5\%$). Nevertheless, Placidi *et al* (1986) concluded that "CBZ and lithium appear to possess comparable prophylactic profiles". Likewise, Coxhead *et al* (1992) concluded: "CBZ is equal in efficacy and tolerability to lithium in the prophylaxis of bipolar disorder". But the gap between "lack of significant difference" and "equal efficacy" may be bridged only when the type-two error is weak (when statistical power is high); this was hardly the case in any of the negative studies.

The other important point is the outcome measure: when it does not allow for a time factor, it puts early and late recurrences on the same level; therefore, it greatly loses sensitivity. The only study (Watkins *et al*, 1987) which found a significant difference (in favour of lithium) used an outcome measure that allowed for the time factor. The differences between the studies in sensitivity stem from whether or not this factor was allowed for. They might partly explain, together with differences in statistical power, the contradictory results recorded among the studies. The heterogeneity of these studies is also related to other differences in design: sample homogeneity, previous treatment response, follow-up duration, and drop-out rate. Their effect on the discrepancy of the results, although probable, could neither be

identified nor quantified because of the small number of examined studies. Nevertheless, the difference in outcome measure segregates the studies in the same way as their results do (Watkins *et al*'s work is opposite to that of the three others).

Because null results are as interesting as positive ones (whatever their direction), this meta-analysis is unlikely to be flawed by a bias of publication. Nevertheless, its results could be reversed by future studies: adding a single study yielding findings similar to Watkins *et al*'s in the computations would result in the superiority of lithium (combined $Z=2.52$; $P=0.006$).

In conclusion, if we examine the data arising from controlled studies using either placebo or lithium as control drugs, it appears that CBZ has not yet conclusively demonstrated its prophylactic efficacy with samples of affective patients not selected for any particular characteristic such as cycle length, or lithium non-response or response. Thus, lithium should remain the preferred treatment for prophylaxis.

However, open studies and 'ABA' studies suggest that at least some patients or subsets of patients might respond to CBZ. Therefore, additional investigations should be undertaken to (a) verify, with non-selected samples, the results of the only study (Watkins *et al*, 1987) meeting both a satisfactory statistical power and a sensitive outcome measure; and (b) determine whether CBZ affects clearly defined clinical patterns or stages of the course of affective disorders. Such investigations would help to clarify the indications of CBZ.

References

American Psychiatric Association (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM-III). Washington, DC: APA.

Casagrande, J. T., Pike, M. C. & Smith, P. G. (1978) An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics*, **34**, 483–486.

Coxhead, N., Silverstone, T. & Cookson, J. (1992) Carbamazepine versus lithium in the prophylaxis of bipolar affective disorder. *Acta Psychiatrica Scandinavica*, **85**, 114–118.

King, P. R. (1990) Appraising the quality of drug evaluation research. II. A methodological review procedure for prophylactic treatment trials. *Journal of Clinical Psychopharmacology*, **10**, 22–26.

Lusznat, R. M., Murphy, D. P. & Nunn, C. M. (1988) Carbamazepine vs lithium in the treatment and prophylaxis of mania. *British Journal of Psychiatry*, **153**, 198–204.

Okuma, T., Inanaga, K., Otsuki, S., *et al* (1981) A preliminary double-blind study on the efficacy of carbamazepine in prophylaxis of manic–depressive illness. *Psychopharmacology*, **73**, 95–96.

Overall, J. E. & Rhoades, H. M. (1987) Adjusting *p* values for multiple tests of significance. In *Psychopharmacology: the Third Generation of Progress* (ed. H. Y. Meltzer), pp. 1013–1018. New York: Raven Press.

Placidi, G. F., Lenzi, A., Lazzerini, F., *et al* (1986) The comparative efficacy and safety of carbamazepine versus lithium: a randomised, double-blind 3-year trial in 83 patients. *Journal of Clinical Psychiatry*, **47**, 490–494.

Post, R. M., Altshuler, L. L., Ketter, T. A., *et al* (1991) Antiepileptic drugs in affective illness. Clinical and theoretical implications. In *Advances in Neurology*, Vol. 55 (eds D. Smith, D. Treitman & M. Trimble), pp. 239–277. New York: Raven Press.

Rosenthal, R. (1991) *Meta-Analytic Procedures for Social Research*. Newbury Park: Sage.

Watkins, S. E., Callender, K., Thomas, D. R., *et al* (1987) The effect of carbamazepine and lithium on remission from affective illness. *British Journal of Psychiatry*, **150**, 180–182.

R. Dardennes, MD, **C. Even**, MD, **F. Bange**, MD, **A. Heim**, MD, Clinique des maladies mentales et de l'encéphale, Hôpital Sainte Anne, Paris

Correspondence: Dr Dardennes, CMME, Hôpital Sainte Anne, 1 rue Cabanis, 75014 Paris, France

(First received 28 January 1994, final revision 16 May 1994, accepted 6 June 1994)