AUTHOR'S REPLY: The curiously dogmatic tone of Lee & Chan's criticism might be easier to accept if they had read our paper more carefully, since many of the points they raise are discussed in our paper.

The rationale for our design is simple. Conventional cognitive–behavioural therapy (CBT) (19 sessions over 18 weeks) is undoubtedly an intensive form of treatment. The resources to supply this form of therapy to the 1% of young women who meet the diagnostic criteria for bulimia nervosa is unlikely to be available in most areas. The long-term outcome of this form of therapy is not nearly as optimistic as Lee & Chan suggest. Fewer than 40% of patients are in remission after one year (Fairburn *et al*, 1993). Our brief CBT (eight sessions) requires less therapist time, and can easily be taught to non-specialists. Drug treatment alone is associated with a high rate of non-compliance in bulimia nervosa. Our design aimed to test a model of treatment which would be applicable in ordinary clinical practice, reduce therapist time, improve compliance, and optimise response.

Of course there is a risk of reaching a ceiling effect with a combination of psychotherapy and pharmacotherapy. However, it stands to reason that a ceiling effect is less likely to occur with a less intensive, rather than more intensive form of therapy. This point is dealt with in detail in our paper.

The comparison of our design with a study comparing the combination of antidepressant or placebo with electroconvulsive therapy (ECT) is entirely inappropriate. The correct comparison would replace ECT with CBT. Such a study would be entirely justified, and when done, has shown that the combination of antidepressant and CBT is probably superior to each form of treatment alone (e.g. Hollan *et al*, 1992).

The authors suggest that, before performing this study, we should have compared d-fenfluramine with placebo. This study has already been done, and is discussed at length in our paper (Russell *et al*, 1988). In fact, the high drop-out rate from this study is one of the factors which led us to the study design, which succeeded in having an exceptionally low drop-out rate.

Lee & Chan's description of CBT does not do justice to the model used in our study (derived from Fairburn's model (Fairburn, 1985)). The educational component, outlining the interaction between attitudes, eating behaviour, and biology, was heavily emphasised. The effect of medication on biological processes is easily incorporated within this model, and does not, Lee & Chan assert without supportive evidence, negate the effects of one or other treatment.

Lee & Chan suggest that d-fenfluramine may have a role in the treatment of obese bulimics. It is unlikely that they will be able to test this hypothesis without including a psychological package in a treatment trial, since a trial of d-fenfluramine versus placebo will almost certainly be undermined by high drop-out rates. I look forward to reading how they will be able to learn by our mistakes.

FAIRBURN, C. (1985) Cognitive behavioural treatment for bulimia. In *Handbook of Psychotherapy for Anorexia Nervosa and Bulimia* (eds D. M. Garner & P. E. Garfinkel), pp. 160–192. New York: Guilford Press.

FAIRBURN, C. G., JONES, R., PEVELER, R. C., *et al* (1993) Psychotherapy and bulimia nervosa. Longer-term effects of interpersonal psychotherapy, behaviour therapy and cognitive behavior therapy. *Archives of General Psychiatry*, **50**, 419–428.

HOLLAN, S. D., DERUBEIS, R. J., EVANS, M. D., *et al* (1992) Cognitive therapy and pharmacotherapy for depression singly and in combination. *Archives of General Psychiatry*, **49**, 774–781.

RUSSELL, G. F. M., CHECKLEY, S. A., FELDMAN, J., *et al* (1988) A controlled trial of d-fenfluramine in bulimia nervosa. *Clinical Neuropharmacology*, **11**, S146–159.

THOMAS A. FAHY

*Institute of Psychiatry*
*Denmark Hill*
*London SE5 8AF*

**Diagnostic agreement in psychiatry**

SIR: The important study by Okasha *et al* (*Journal*, May 1993, **162**, 621–626) compared diagnostic reliability for ICD–9, ICD–10 and DSM–III–R. However, readers would have liked to have known whether the differences in overall reliability between the diagnostic systems reached statistical significance. The original description of kappa (Cohen, 1960) gave the simple arithmetic for testing for the significance of the difference between two independent kappas. Unfortunately, the reader cannot do this testing on the basis of Tables 1 and 2 in the Okasha *et al* paper because for inter-rater reliability one requires cross-tabulation of the two clinicians' allocation of cases to calculate both observed and chance agreement. Without significance testing the intriguing finding of higher overall inter-rater reliability for ICD–10 compared with both ICD–9 and DSM–III–R might be explained by chance.

Secondly, the authors correctly stated that kappa is base-rate dependent so that interpretation of results for disorders comprising less than 5% of the sample should be treated cautiously. However, in stating that this is "one of the main criticisms of kappa" they are, perhaps, unaware of the alternative view that this base-rate dependence is indeed one of kappa's strengths. Shrout *et al* (1987) have comprehensively rebutted the argument of Spitznagel &