neuronal circuits utilise glutamate for neurotransmission. We are continuing to measure these metabolites in a 1.5 cm^3 volume encompassing the left anterior cingulate with MRS at 4.0 Tesla in healthy volunteers and never-treated patients with schizophrenia. The comments by Feinberg & Guazzelli regarding dreaming prompted us to examine our 4.0 Tesla MRS data with respect to the sleep/wake state of the healthy volunteer control subjects. In total, 15 healthy volunteers have been scanned to date. Of these subjects, two could not be reached. Each of the remaining subjects were asked whether they had slept during the spectroscopy study. Of these subjects, six remembered sleeping at some point during the study, six did not sleep, one person could not remember. Comparison of metabolite levels (mean (s.d.)) between subjects who remembered sleeping (n=6) and those who did not sleep (n=6) showed a significant increase in glutamine levels (9.9 (2.7) *v.* 6.1 (1.7), P=0.02) in subjects who remembered sleeping, using a two-tailed t-test. The relative glutamine increase in the subjects who reported that they were asleep was comparable to the increase previously observed in patients with schizophrenia (Bartha *et al*, 1997).

Since it is highly likely that subjects who were sleeping during data acquisition were dreaming, the increased levels of glutamine in these subjects may be due to increased glutamatergic activity in the anterior cingulate, part of the basal ganglia–thalamocortical circuit implicated by Feinberg & Guazzelli in schizophrenia. The similarity of findings between healthy subjects who were likely to be dreaming and patients with schizophrenia highlights the need for further investigation of these regions in schizophrenia.

Bartha, R., Williamson, P. C., Drost, D. J., *et al* (1997) Measurement of glutamate and glutamine in the medial prefrontal cortex of never-treated schizophrenic patients and healthy controls using proton magnetic resonance spectroscopy. *Archives of General Psychiatry*, **95**, 464–473.

Feinberg, I. & Guazzelli, M. (1999) Schizophrenia – a disorder of the corollary discharge systems that integrate the motor systems of thought with the sensory systems of consciousness. *British Journal of Psychiatry*, **174**, 196–204.

R. Bartha Laboratory for Functional Magnetic Resonance Research, John P. Robarts Research Institute, Box 5015, 100 Perth Drive, London, Ontario, N6A 5K8, Canada

P. C. Williamson University Campus, London Health Sciences Centre, 339 Windermere Road, London, Ontario, N6A 5A5, Canada

D. J. Drost Lawson Research Institute, 268 Grosvenor Street, London, Ontario, N6A 4V2, Canada

A. K. Malla WMCH, London Health Sciences Centre, 392 South Street, London, Ontario, N6B 5C2, Canada.

R. W. J. Neufeld Psychology Department, Social Sciences Centre, University of Western Ontario, London, Ontario, N6A 5C2, Canada

Quick rating of depressed mood

Sir: I would like to comment on McKenzie & Marks' (1999) claims about some characteristics of their rating scale.

First, test–retest reliability was assessed by comparing self-rated single item (D1P) on its own with ratings presented in the Fear Questionnaire, one given immediately after the other. Although assessments were made at six different visits, repeated measures analysis was not used.

As the bias introduced seems substantial (e.g. lack of random order, recall, practice) the resulting distortion of estimates may be large. Additionally, the construct's time instability means that the range included the whole spectrum of values, resulting in a further inflation of Pearson's r. It is not surprising that Pearson's r was unity or near unity at each visit.

Second, they indicate that Pearson's $r=-0.54$ between D1P at discharge, and self-rated percentage improvement has predictive merit. However, 71% of the variability in self-rated percentage improvement is unexplained by D1P at discharge, and the reduction in errors of prediction is only 16% when using this information. They also conclude that pre-treatment D1P scores predict a binary category (drop-outs with low mood given as referral reason/otherwise) by referring to different means: 6.7 (drop-outs) and 3.7 (non-drop-outs). However, most pre-treatment D1P scores will not predict membership to either subset simply because of the substantial distribution overlap (s.d.=1.83 and 2.47, respectively). The positive skewness of non-drop-outs would not invalidate the latter statement as it did not seem to have seriously violated the t-test's assumptions. I conclude that D1P's predictive validity is rather low.

Third, concurrent validity was assessed by comparing D1P with D1C (same item completed by clinician), the 21-item Beck Depression Inventory (BDI), and a 'true' depression factor in the BDI (factor analysis). However, the authors used Pearson's r, a technique that does not measure agreement.

They also report a regression line with an intercept of 1.07 and a slope of 0.71, and claim interval correspondence between the BDI and D1P. Interestingly, as these values indicate no equivalence (i.e. intercept not 0, slope not 1) it could be concluded that there is lack of agreement between the scales. However, this argument is flawed because measurement error is not considered. As linear regression does not often give single values for error, it is difficult to assess comparability using this technique.

It is unfortunate that the authors have not used appropriate methods to assess agreement between measurement instruments, such as the well-known techniques proposed by Altman & Bland (1983).

Altman, D. G. & Bland, J. M. (1983) Measurement in medicine: the analysis of method comparison studies. *Statistician*, **32**, 307–317.

McKenzie, N. & Marks, I. (1999) Quick rating of depressed mood in patients with anxiety disorders. *British Journal of Psychiatry*, **174**, 266–269.

D. Marchevsky Department of Psychiatry, Campbell Centre, Hospitals Campus, Standing Way, Eaglestone, Milton Keynes MK6 5NG

Authors' reply: Marchevsky noted Altman & Bland's (1983) suggestions to compare two methods of measurement. These have some merit but do not change our conclusions.

Test–retest reliability was studied by comparing the D1P on its own and as one of 23 questions embedded within the Fear Questionnaire; presented at the same time were problems and targets, work/social adjustment and the Compulsion Checklist (Marks, 1986). Apart from the Compulsion Checklist all the items were rated on a nine-point scale. The order in which the two D1P modes and the other scales were rated varied (albeit non-randomly) across subjects. It thus seems unlikely that there were order effects, or that bias was introduced by practice or recall, since that would require recall of one among nearly 40 similar ratings. Not only Pearson's r but also Cohen's kappa, a chance-corrected measure of agreement, showed values close to unity.

Dr Marchevsky questions the use of Pearson's r and beta from linear regression. These equal one another (0.71 in the instance he quoted). One might debate how much an r measure of association also measures agreement. However, that r represents the degree to which variability in one