patients with panic disorder with no, mild, or moderate agoraphobic avoidance, cognitive therapy was superior to applied relaxation not only on measures of panic but also on measures of agoraphobic avoidance. In contrast, the LT study of exposure treatment found no specific effect of situational exposure on panic at *any* level of agoraphobic avoidance.

Corrections. Two statements in Marks *et al*'s letters require correction. In their first letter (Marks *et al*, 1994) they state that situational exposure "eliminated 96% of panics" in the LT study. This is misleading as the 96% figure seems to come from an assessment carried out 35 weeks after the end of exposure therapy (43 week assessment) and based on only 61% of the initial sample. At the end of exposure therapy (8 weeks) the group receiving exposure and placebo showed a more modest 63% reduction in panic frequency. When commenting on our exclusion of severe agoraphobics in the current letter, Marks *et al* state that "most LT cases (were) severe". This is also misleading as their use of the term "severe" is different and less stringent than ours. They appear to be referring to the DSM–III category of "extensive phobic avoidance" whereas we were referring to the DSM–III–R category of "severe agoraphobia". Many cases meeting criteria for the former would be categorised in DSM–III–R as "moderate agoraphobia" not as "severe agoraphobia" and would therefore have been included in our study.

CLARK, D. M. (1993) Cognitive mediation of panic attacks induced by biological challenge tests. *Advances in Behaviour Research and Therapy*, **15**, 75–84.

MARKS, I. M., SWINSON, R. P., BASOGLU, M., *et al* (1993) Alprazolam and exposure alone and combined in panic disorder with agoraphobia: a controlled study in London and Toronto. *British Journal of Psychiatry*, **162**, 776–787.

MARKS, I. M., BASOGLU, M. & NOSHIRVANI, H. (1994) Cognitive therapy for panic disorder (letter). *British Journal of Psychiatry*, **165**, 557–559.

D.M. CLARK
P.M. SALKOVSKIS
A. HACKMANN
H. MIDDLETON
A. WELLS
M. GELDER

Department of Psychiatry,
Warneford Hospital,
Oxford OX3 7JX

Home-based versus in/out-patient care for people with serious mental illness

SIR: Marks *et al* (*BJP*, August 1994, **165**, 179–194) mention in their discussion that "outcome was rather better with home care than in/out-patient care", whereas in the abstract they report that "outcome was superior with home-based care". There appears to be some uncertainty about the interpretation of the results, and we would like to offer some suggestions that we feel would facilitate interpretation of this important and carefully conducted study.

Analyses. The authors have identified 19 endpoints in the trial (number of admissions, days in hospital, GAS, BPRS, PSEtotal, PSEdah, PSEbso, PSEnsr, PSEnsn, patient's satisfaction, relative's satisfaction, SAS global, SAS social, SAS extended family, SAS parents, SAS daily living skills, SAS economic, SAS work and SAS marital). As no primary, prespecified outcome has been identified, all 19 are of equal importance from the statistical point of view. For each outcome, separate significance testing is carried out at each of the three measurement points, so that the reader is presented with 55 tests in the main article, followed by another 255 in the appendix. *P*-values at the 10% significance level and over are reported "non-significant", and *P*-values <0.1 are apparently considered significant. This approach is problematic. First of all, presenting multiple *P*-values can only lead to a subjective conclusion, because it is impossible to decide how many of the 55 tests in the main article need to be "significant" in order for the trial to show superiority of home-based treatment over standard care. Fifteen (27%) of the tests were significant at the 5% level, but this still leaves 40 measurements where the confidence intervals include values that indicate that home-based care actually makes patients worse. Secondly, multiple conventional significance testing can seriously inflate the overall Type I error rate (false positive results), and the study protocol should specify in advance how the problem of multiple testing is going to be dealt with. One way to control *P*-values when all endpoints are analysed on equal terms is the Bonferroni correction. However, the Bonferroni correction becomes too conservative if the endpoints are correlated (as is clearly the case in this trial), and the modification proposed by Simes (1986) is more appropriate, as realistically the Type I error becomes $<\alpha$ (Pocock *et al*, 1987*a*). Applying the modified procedure by Simes to the data of Marks *et al* (only taking into account that there are 19 endpoints and ignoring the fact that further separate tests were carried out at each measurement point, and in five subgroups), reveals that of the 15 results significant at the 5% level, only 8 survive (15% of the total of 55). Thirdly, and most importantly, the plethora of significance testing is

based on the false premise that each time point is of separate interest in its own right. In reality, the primary hypothesis is more global (across all measurements over the follow-up period, is there a tendency for superior functioning in the group with home-based treatment?). While the descriptive data shown by the authors would tend to favour this hypothesis, their analyses appear to be biased towards the null hypothesis of no difference. We therefore think that it is important that the authors use a simple summary statistic approach or global test as described by Frison & Pocock (1992) and Pocock *et al* (1987*b*). Furthermore, instead of presenting many tests involving subgroups, they should test for interaction with some relevant variables. Statistical tests for interaction assess directly whether a prognostic factor affects the difference in treatments (Pocock, 1983).

Another issue that the authors could address is the fact that in both the intervention group and the control group, there appears to be a significant *trend* to improve on clinical and social measures with each subsequent time point, but the trend appears to be stronger in the intervention group. The authors could therefore use a test which allows for the examination of a treatment time interaction. This would reduce the number of tests, and be compatible with what the data tell us.

Time in hospital. The authors' decision to include baseline measures for endpoints as a covariate in some of the analyses is to be commended. For example, if there are large differences on a measure at the pre-treatment baseline assessment, then these differences need to be adjusted for in assessing the effect of treatment at follow-up; the results are not readily interpretable if such baseline differences are not taken into account in assessing the treatment effect. As time in hospital is arguably the most important outcome measure, we were disappointed to note that the authors did not discuss the fact that important baseline differences between the two groups on first admission status may have affected the results (73% of DLP patients were first admissions, versus only 57% of controls; $\chi^2=5.4$; $P=0.02$). Because readmission risk is increased with each subsequent admission (Mortensen & Eaton, 1994), the reported results may be confounded, as controls are more likely to be readmitted regardless of any treatment difference. Furthermore, the higher proportion of prevalent cases with established chronicity in the control group may have confounded other treatment associations as well.

The data provided by Marks and colleagues reveal large differences between the median and mean time spent as in-patients in both groups, indicating that this measure is skewed. As the reported *t*-tests are invalidated in the presence of significant skewness, we would like to suggest an appropriate transformation or the use of non-parametric methods, taking into account confounding by baseline admission status.

FRISON, L. & POCOCK, S. J. (1992) Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, **11**, 1685–1704.

MORTENSEN, P. B. & EATON, W.W. (1994) Predictors for readmission risk in schizophrenia. *Psychological Medicine*, **24**, 223–232.

POCOCK, S. J. (1983) *Clinical Trials, A Practical Approach*. Chichester: John Wiley.

—, S. J., GELLER, N. & TSIATIS, A. (1987*a*) The analysis of multiple endpoints in clinical trials. *Biometrics*, **43**, 487–498.

—, S. J., HUGHES, M. D. & LEE, R. J. (1987*b*) Statistical problems in the reporting of clinical trials. A survey of three medical journals. *New England Journal of Medicine*, **317**, 426–432.

SIMES, R. J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751–754.

J. VAN OS
K. MCKENZIE
K. GILVARRY
T. FAHY

Brixton Case Management Project
Institute of Psychiatry,
London SE5 8AF

CORRIGENDA

BJP, **166**, 80–86. In authors names and details, Robert A. Burns should read Roger A. Burns, MRCPsych.

BJP, **166**, 263. In the letter *ECT seizure threshold and fluoxetine*, the correct initials for the authors are R. I. Tobiansky and G. G. Lloyd.