

important enough to detect nor the power of the trial to detect that reduction are reported. If the power of the study is low (which is likely given the small numbers in each group) then lack of significance is not proof of lack of effect. If the size of the effects found and their confidence intervals had been reported we would have been able to make a sensible judgement.²

Finally, the authors in their reply state that 'significant differences are not in themselves enough to reject the null hypothesis'. It is not clear what they mean. If they mean that in a study with 54 significance tests we should interpret the odd significant results with caution, we could not agree more. We oppose the slavish use of *P*-values.³

It would have been better if (say) three endpoints had been used, for example a measure of morbidity, asthma remedy use and health service resource use. These three endpoints would have had a clear interpretation and would have been measuring different outcomes.

2. Trials with repeated measurements of response over time require a prespecified policy for statistical analysis, aimed at a single specific hypothesis of interest—repeated significance tests at each time point should be avoided. White and colleagues report nine significance tests at half-yearly intervals, 54 in all. The analysis of variance for each morbidity measure at the end of each time period tests a different hypothesis. Each relates to a different clinical effect and a different time after which one expects these effects to be observed.

If one is interested in the effect on morbidity over time, then the presence of a time trend can be tested using a multivariate analysis (as we suggested) which allows for the correlation between general practitioners' scores over time and which more efficiently uses the repeated measures.

An alternative approach is to decide *a priori* how long it is likely to take for the intervention to produce a clinically meaningful effect. Once that time has been decided an appropriate test of the difference in outcome between the general practitioner groups at that time point should be performed with the confidence intervals reported.

3. The magnitude of the clinical effects for the primary endpoints should be stated along with the confidence limits. Nowhere in White and colleagues' paper are the general practitioner scores reported for the intervention and control groups separately. They test the significance of the differences in the scores achieved by the general practitioner groups but do not

present the size of difference or clinical effect being tested. Thus we have no idea of the clinical importance of the effects being tested.

This is all the more surprising since they state that 'one has to interpret the importance of results from a clinical point of view'. We could not agree more.

4. The intended size of the trial and the power calculations should be specified in advance. When a study is being planned the researchers need to decide what is the smallest size of clinical effect they consider to be worth detecting and at what level of statistical significance. The appropriate size of the study (in this case the number of general practitioners) is determined by the decision as to the power of the study to detect such an effect at that level of significance if indeed the intervention does produce that effect.

Results must be evaluated in the context of prior knowledge, corroborative studies, dose-response relationships and their reproducibility.⁴ However, taking the above into account the finding of a statistically significant result is, as far as we are aware, the only basis for rejecting a study's null hypothesis.

TREVOR A SHELDON
PHILIP MONK

Department of Community Health
University of Leicester
Leicester Royal Infirmary
PO Box 65
Leicester LE2 7LX

References

1. Pocock S, Hughes M, Lee R, *et al*. Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987; 317: 426-432.
2. Rose G. Beta blockers in the treatment of myocardial infarction. *Br Med J* 1980; 2: 259.
3. Doll R, Peto R. When is significant not significant? *Br Med J* 1977; 2: 259.
4. Gore S, Altman D. *Statistics in practice*. London: British Medical Association, 1982: 71.

Workload of full-time women GPs

Sir,

The paper by Judith Cooper and colleagues on workload and remuneration for part-time women in general practice (*October Journal*, p.400) has failed to emphasize that the most important factor in deciding the profit sharing ratio for part-time partners is the amount of out of hours and weekend work that they do.

Although the study shows how much on-call work is done by the two groups as a whole, this is not related individually to their daytime commitment. As a wife, mother and full-time general prac-

titioner, I am in no doubt at all that the nights and weekends on call are by far the most stressful, tiring and intrusive part of a general practitioner's workload. It is also the most dangerous. The review body's figure of 13.5% of gross remuneration for out of hours work bears no resemblance at all to a realistic payment for the degree of disruption, wear and tear and fatigue that a full part in the average general practitioner's rota brings. It is not only the hours of night work either, but a busy night on call makes the following day's work far harder to cope with.

The amount of money earned for the number of nights on call should not be in a linear relationship because the more nights done the harder the load is to bear. The reward should accordingly be increased or decreased exponentially.

Therefore, full-time general practitioners, working at nights and weekends, may feel that their part-time partners, who do not share this load fully, should share significantly less of the profits regardless of their daytime working hours.

J D BILLINGHAM

French Weir Health Centre
French Weir Avenue
Taunton TA1 1NW

Trainee collaborative research in the Essex faculty

Sir,

The paper by Timmins and colleagues (*October Journal*, p.423) demonstrates that a trainee collaborative study provides a healthy symbiosis: trainees participate in audit and encounter the College at a local level while the College benefits from a novel research tool. Clearly, as the authors point out, other faculties might consider involving local trainees in similar studies, to mutual benefit.

Given this pioneering approach, it seems ironic that the decision of the Joint Committee on Postgraduate Training for General Practice, in February 1988, to withdraw its recognition for training from the North East Thames region would have affected the trainees cooperating with the Essex faculty during the period of this study. Presumably, the training environment was so poor as to warrant withdrawal of recognition, rendering trainees ineligible for the College examination, yet adequate enough to stimulate the 'high standard' of work described by the authors which provided the basis of a paper deemed suitable for publication in the *College Journal*. It will be interesting to see how rapidly and capably trainees in less blighted regions are encouraged to