tion in 500 general practices in England and Wales. The questionnaire is short — one side of A4 with 10 questions — and takes approximately two minutes to complete. The introductory letter, signed by the researchers, was addressed to the practice manager requesting him or her to give it to the general practitioner most involved in cervical smear testing. For the follow up of non-respondents, the practice manager was asked to give the questionnaire to the general practitioner most likely to complete it; the follow-up letter to the general practitioner was signed by the head of department. The response rate after the first questionnaire was 50%, and after the second questionnaire the response rate had increased to 76%.

Although Sibbald and colleagues gained a good response from the telephone questionnaire, the costs of obtaining a completed questionnaire by telephone were approximately four times higher than by post. In addition, telephone surveys of general practitioners can be frustrating and time consuming for the researcher who finds the general practitioner's telephone number engaged, on answerphone or that the general practitioner is unavailable in surgery or out on visits.

DIANA MASON
PIPPA OAKESHOTT

Division of General Practice and Primary Care
St George's Hospital Medical School
Hunter Wing
Cranmer Terrace
London SW17 0RE

References
1. Heberlein TA, Baumgartner R. Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *Am Soc Rev* 1978; 4: 447-462.
2. Goyder JC. Further evidence on factors affecting response rates to mailed questionnaires. *Am Soc Rev* 1982; 47: 550-553.

Leicester assessment package

Sir,
I am concerned that the authors of two papers[1,2] conclude that the Leicester assessment package can be recommended for assessment of consultation competence in general practice. However, the results presented lead to the conclusion that the Leicester assessment package cannot (yet) be recommended.

Any procedure passing or failing individuals will suffer from misclassification. That is, a proportion of individuals who should pass will be failed, and a proportion of individuals who should fail will

pass. The question of how to establish exactly who should pass and fail is difficult — hence the need for assessment packages. A gold standard assessment might be approximated by a number of experienced assessors (say four) each assessing a large number of consultations (say 20) several times over a period of months (say three), and deciding finally in conference whether to pass or fail. Less intensive methods, such as the Leicester package, should be validated against such a gold standard (criterion validation) to establish sensitivity and specificity — a process which clearly requires detailed specification of how the pass/fail decision is to be made. Deciding how high sensitivity and specificity need to be, taking into account the different costs associated with low sensitivity (prejudicial to careers) and low specificity (prejudicial to the health of the public) is essentially a political decision, and requires an open debate. The face validation reported[1] in no way guarantees that the Leicester assessment package will be able to approximate the results of a gold standard satisfactorily.

Variability owing to temporal factors (that is repeatability) was not assessed, and could affect both trainees and assessors. If temporal variability is large (good days and bad days occur frequently and have a considerable effect on performance) achieving adequate reliability will require assessments to be made on several well-spaced occasions.

The authors attempted to assess the variability due to differing assessors, patients and doctors. The estimates of components of variance due to these sources were not presented, however, and are likely to be imprecise owing to the small sample sizes. Brennan emphasizes that the 'study design be as large as possible. Doing so helps ensure that the resulting estimated variance components will be as stable as possible'.[3] Moreover, samples of five or six are unlikely to be representative of parent distributions, and very unlikely to include significant representation from the tails of the distributions. But this is where likely causes for concern are: difficult patients, borderline doctors, eccentric assessors. As a result, there can be little confidence that the results obtained will be stable (that is, a repeat study might produce very different estimates).

The authors used (Cronbach's) alpha coefficients to assess internal consistency. Unfortunately, although often called a measure of internal consistency, alpha increases not only with consistency, but also with number of items. A matrix of correlation coefficients gives a much bet-

ter idea of consistency. For a given value of alpha, and a given number of items, it is possible to estimate the mean correlation coefficient. For an assessor with alpha = 0.8 (intra-assessor consistency) the mean correlation between scores given for pairs of patients will be about 0.4.

One assessor was recognized as 'inconsistent' (with an estimated mean correlation coefficient of 0.04 corresponding to alpha = 0.22), and it was suggested that 'all assessors should be trained and calibrated before being sanctioned to assess real candidates...'. This may have the desired effect, but is hardly a firm basis for recommending the Leicester package. Evidence that training of assessors does improve consistency (and hence overall reliability) is needed.

The correct interpretation of alpha, and the generalizability coefficient, is as an indicator of the internal reliability of the mean of several items. One definition of alpha is the ratio of the true (with perfect assessment) subject (in this case, doctor) variance to the expected observed (with variation due to assessors and cases) subject variance. If the observed variance σ_o^2 is taken as equal to the true variance σ_t^2 plus a random error variance σ_e^2, then it is simple to see that $\sigma_e^2/\sigma_t^2 = (1/\alpha) - 1$ so that $\sigma_e/\sigma_t = \sqrt{((1/\alpha)-1)}$. Suppose that the Leicester assessment package were actually measuring exactly the same qualities as an idealized gold standard assessment (this needs establishing by criterion validation) — but with random measurement error added. If alpha = 0.8 then from above we have $\sigma_e/\sigma_t = 0.5$ that is the (random) error standard deviation is half the true subject standard deviation. Sensitivity and/or specificity will be relatively low. Suppose the true cut off (for pass/fail) were two true subject standard deviations below the true mean — so that the worst 2.3% of doctors should fail (assuming perfect assessment, and normality). A particular doctor whose true score is 0.5 standard deviation below the pass mark (that is in the worst 0.6%) will have a 16% chance of passing. Fairly difficult calculations show that the sensitivity will be 98%, but the specificity only 73% (that is more than one in four of those who should fail will in fact pass). The negative predictive value would be just 45% (that is only 45% of those who failed would actually deserve to fail).

This is why Nunnally[4] suggests that while for basic research purposes a reliability of 0.7 or higher is sufficient, for 'many applied settings a reliability of 0.8 is not nearly high enough... In those applied settings where important decisions are made with respect to specific test scores, a reliability of 0.90 is the min-

imum that should be tolerated, and a reliability of 0.95 should be considered the desirable standard.' If it is 'generally accepted' that a generalizability coefficient of 0.8 is sufficient for assessing clinical competence, then one can only suppose that such assessments (and the decisions based on them) are not deemed particularly important.

DAVID BRAUNHOLTZ

Division of General Practice and Public
  Health Medicine
Leeds University
30 Hyde Terrace
Leeds LS2 9LN

References
1. Fraser RC, McKinley RK, Mulholland H. Consultation competence in general practice: establishing the face validity of prioritized criteria in the Leicester assessment package. Br J Gen Pract 1994; 44: 109-113.
2. Fraser RC, McKinley RK, Mulholland H. Consultation competence in general practice: testing the reliability of the Leicester assessment package. Br J Gen Pract 1994; 44: 293-296.
3. Brennan RL. Elements of generalisability theory. Iowa City, IA: American College Testing Program, 1983: 3, 30, 101.
4. Nunally JC. Psychometric theory. 2nd edition. New York, NY: McGraw Hill, 1978: 245.

Arranging emergency hospital admission

Sir,
In the month when the Secretary of State for Health floated the idea of closing 40% of the National Health Service's remaining acute hospital beds (speech to the National Association of Health Authorities and Trusts , 22 June 1994), the results of a study of problems encountered by general practitioners arranging hospital admission have assumed an even greater importance (June Journal, p.251). The study found that problems were experienced by general practitioners during the hospital admissions procedure in 35% of cases, and 21% of telephone calls resulted in a refusal to admit a patient to a particular hospital. The Secretary of State seems to think that the care currently being delivered in these beds can be relocated either to the private sector or to primary care in the community. The balance between these two in the Secretary of State's vision, like much else, is not yet clear.

Those of us who struggle to provide a high standard of primary care against a background of a falling number of hospital beds are fearful about the future. We are told that there are too many hospital beds and yet our regular experience of difficulties in securing a bed for emergency admission contradicts what we are told. In our bewilderment, it is reassuring to find that our experience is validated by research. Now we must hope that the future planning of the NHS will be based on scientific research rather than political rhetoric.

STEPHEN AMIEL
JUDY BENNETT
IONA HEATH
AZHAR MALIK
DANIEL TOEG

Caversham Group Practice
Kentish Town Health Centre
2 Bartholomew Road
London NW5 2AJ

Identifying the agenda in the consultation

Sir,
Middleton's interesting paper on the attitudes of general practitioners to lists and the patients who bring them (July Journal, p.309) highlights the possible barriers that doctors may have to making full use of patients' written lists in the consultation. He argues that encouraging patients to bring lists might help solve a common communication problem in the consultation, namely that the patient's agenda is not fully identified and addressed.

However, embracing the written lists of those few patients who do bring them is only one way of tackling this fundamental area. The wider issue here is how to help doctors understand the importance of identifying and confirming early on in the consultation as many as possible of the problems that the patient wishes to discuss, whether he or she brings a written list or not. Here, we can learn a lot from North American research and teaching about the medical interview and communication skills which place considerable emphasis on this initial survey or screening of problems and on agenda setting.

Stewart and colleagues have shown that 54% of patients' complaints and 45% of their concerns are not elicited[1] while Starfield and colleagues record that in 50% of visits, the patient and the doctor do not agree on what is the main problem.[2] Burack and Carpenter found that patients and doctors agreed on the chief complaint in only 76% of somatic problems and in only 6% of psychosocial problems.[3] Several investigators have shown that patients often have more than one concern to discuss and the mean number of concerns ranged from 1.2 to 3.9 in both new and return visits.[2,4-6] These studies warn of the danger of premature and limited hypothesis testing before a wider spectrum of concerns has been identified.

In a key piece of research, Beckman and Frankel have shown that doctors frequently interrupt patients before they have completed their opening statement — after a mean time of only 18 seconds — and that this behaviour both limits the number of complaints elicited and increases the number of complaints arising late in the consultation.[7,8] They have also shown that the order in which patients present their problems does not correlate with their clinical importance. Therefore, the apparent assumption of many doctors that the first complaint mentioned is the chief one may considerably reduce the accuracy and efficiency of the consultation.

Beckman and Frankel have also shown which specific communication skills help doctors to identify as many as possible of the patient's complaints and which skills known to be helpful later on in the consultation, such as clarifying, echoing and repetition, are in fact counterproductive early on in the interview. Several North American teaching texts now propose the following sequence for the early part of the consultation:[9-11]

● encouraging the patient to discuss his/her main concerns by attentive listening without interruption or premature closure;
● confirming the list identified so far by summarizing;
● checking repeatedly for additional concerns, 'is there anything else you wished to discuss today?', until the patient indicates that there is none;
● negotiating an agenda for the consultation.

In the teaching of trainees and trainers in the East Anglian region, explaining that most patients can be expected to bring more than one problem on any one occasion, and that a survey of problems and agenda setting should be part of the structure of all consultations, helps doctors to experience less conflict during consultations, to be more patient-centred and to use time more effectively. Accuracy and efficiency are increased and uncertainty is reduced for both the patient and the doctor. As patients are often unaware of the time allocated to them by the appointment system, and how long it might take to explore any problem with the doctor, early identification of problems allows priorities to be negotiated. Such an open approach at the beginning of the consultation means that the patients are usually agreeable to