Non-immunocompromised patients with invasive aspergillosis

Patient	Sex	Age	Clinical details
1	M	60	Multisystem failure following aortobifemoral grafts for peripheral vascular disease
2	F	53	Chronic rheumatic aortic and mitral valve disease; acute renal failure due to vasculitic-type glomerulonephritis
3	M	74	Panlobular emphysema and postmortem evidence of 1-antitrypsin deficiency

categories (table).[2] None of these patients was neutropenic during their hospital stay and none was diagnosed before necropsy as having invasive aspergillosis. Unlike the cases of Dr Hovenden and colleagues, two showed widely disseminated disease, indicating that systemic spread may also occur in such patients.

I agree with Dr Hovenden and colleagues in most respects and particularly welcome their emphasis on the importance of taking aspergillus isolates seriously in any patient with progressive pulmonary infiltrates. My research has shown, however, that several of their comments require some qualification. While it is true that profound neutropenia is common in many cases of invasive aspergillosis, this was a feature in only two out of 12 liver transplant patients with aspergillosis.[2] We have previously shown that a more important risk factor in this group is the use of high dose steroids to prevent rejection.[3] Furthermore, in our series there was no significant association between the presence of neutropenia and the incidence of systemic spread. In contrast, there was a highly significant association with use of high dose corticosteroids.

Finally, the association between aspergillus and building work may have been overemphasised. Despite the presence of a substantial building project on the hospital site, the source of infection in our liver transplant patients at the Queen Elizabeth Hospital, Birmingham, was eventually traced to a contaminated air shaft (T Elliot, personal communication).

A P BOON

St James's University Hospital,
Leeds LS9 7TF

1 Hovenden JL, Nicklason F, Barnes RA. Invasive pulmonary aspergillosis in non-immunocompromised patients. *BMJ* 1991;**302**:583-4. (9 March.)
2 Boon AP, O'Brien D, Adams DH. A ten year review of invasive aspergillosis detected at necropsy. *J Clin Pathol* (in press).
3 Boon AP, Adams DH, Buckels J, McMaster P. Cerebral aspergillosis in liver transplantation. *J Clin Pathol* 1990;**43**: 114-8.

Radiation workers and childhood leukaemia

SIR,—One thing that emerges from the article by Mr James Urquhart and colleagues is that neither paternal employment nor occupational radiation exposure at Dounreay can apparently explain the excess of childhood leukaemia within the Caithness area.[1] Because of the relations found earlier for Sellafield in West Cumbria,[2] however, the following comments are relevant.

Annual average doses of external ionising radiation among workers in the nuclear industry have fallen over the years. For example, mean annual doses to radiation workers at Sellafield decreased from around 15 mSv during the 1950s to around 5 mSv during the 1980s.[3,4] At Dounreay, which became operational later than Sellafield, overall doses have been lower—thus 13% of Dounreay radiation workers had accumulated over 100 mSv to 1979 (mean 47 mSv)[5] compared with 34% of Sellafield workers to 1983 (mean 124 mSv).[3] The results presented by Urquhart and coworkers therefore relate to a lower exposure working environment, and as they are based on a smaller study, as the authors say, they do not contradict the earlier findings.

For comparison, results from the studies in the Dounreay and Sellafield areas for workers at the nuclear installations only and their children are shown in the table, with non-Hodgkin's lymphoma included as it is not possible to separate leukaemia cases in the former paper.[1] If the Sellafield figures are proportionally scaled down to the size of the Dounreay study then little major difference is seen on the small numbers involved.

MARTIN J GARDNER

MRC Environmental Epidemiology Unit,
Southampton General Hospital,
Southampton SO9 4XY

1 Urquhart JD, Black RJ, Muirhead MJ, *et al.* Results of a case-control study of leukaemia and non-Hodgkin's lymphoma in children in Caithness near the Dounreay nuclear installation. *BMJ* 1991;**302**:687-92. (23 March.) [And see correction, *BMJ* 1991;**302**:818. (6 April.)]
2 Gardner MJ, Snee MP, Hall AJ, Powell CA, Downes S, Terrell JD. Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *BMJ* 1990;**300**:423-9.
3 Smith PG, Douglas AJ. Mortality of workers at the Sellafield plant of British Nuclear Fuels. *BMJ* 1986;**293**:845-54.
4 BNFL Health and Safety Directorate. *Annual report on occupational safety 1989.* Risley: British Nuclear Fuels, 1990.
5 Beral V, Inskip H, Fraser P, Booth M, Coleman D, Rose G. Mortality of employees of the United Kingdom Atomic Energy Authority, 1946-1979. *BMJ* 1985;**291**:440-7.

Preventing fraud

SIR,—Ms Jane Smith seems to concentrate on the teaching of research ethics and reduction of the need for publication to reduce fraud, asking for retention of data to help in investigation, thus seeming to say that prevention and investigation are important.[1] Detection of fraud, however, is not considered.

In this regard the advent of good clinical practice should play a part. Its elements are laid down in many documents and laws from the American Food and Drugs Administration[2] and are now brought to Europe by the guidelines issued by the Committee for Proprietary Medicinal Products.[3] The major tenets of good clinical practice are to assure the protection of the research subject and to confirm the integrity of the data.

At present good clinical practice relates only to the development of pharmaceuticals and medical devices. The job of assuring the validity of the data lies initially with the investigator and with monitors, who, among other obligations, have to validate data by comparing the research data written into case report forms with source documents such as patient and physician records and other sources of raw data such as laboratory and test reports.

Pharmaceutical companies that work to good clinical practice must then have at least one tier of auditors to audit the work of both monitor and investigator. In the United States further assurance of the data is provided by the Food and Drugs Administration, which has a large team of compliance officers to check data validity. It is this typically American attitude of checks and balances that maximises the chance of detecting fraud in American pharmaceutical research. The Food and Drugs Administration publishes a blacklist of investigators who have fallen foul of the audit. It also has the ability to bring criminal charges against investigators perpetrating fraud.

Our company has been operating to good clinical practice in the United Kingdom for several years, and we have strongly suspected fraud on several occasions. It is likely that the advent of good clinical practice in Britain and Europe will increase the detection of fraud. The long term effect should be to reduce the frequency of its occurrence.

At present in the United Kingdom, if fraud is detected in pharmaceutical research the only action available is to report the matter to the General Medical Council. Europe does not yet have specific laws on the subject. No government yet undertakes quality audits, with the exception of Greece, which has two auditors. Indeed, the only other country intending to audit is France, which has trained 10 auditors. These, however, have yet to undertake a definitive audit.

All these schemes may help detect fraud in pharmaceutical research, but they do not do anything to increase the detection of fraud in academic research. It is surely not enough to make fraud less likely to occur, nor enough to investigate it once suspicions arise and allegations are made. We contend that academic institutions should set up procedures for auditing research under their jurisdiction in parallel with the developments in the pharmaceutical industry.

The pharmaceutical industry seems to be doing so much to detect fraud, but many academics are still more suspicious of research sponsored by companies than they are of work sponsored from outside the industry. Perhaps academic and professional bodies have something to learn from the experience of the pharmaceutical industry regarding good clinical practice and assuring the validity and integrity of data.

I M DEWS
M J VANDENBURG

Medical and Clinical Research Consultants,
Romford, Essex RM7 7DA

1 Smith J. Preventing fraud. *BMJ* 1991;**302**:362-3. (16 February.)
2 United States Code of Federal Regulations. Title 21:Sections 50, 56 and 312.
3 CPMP Working Party on Efficacy of Medicinal Products. *Note for guidance: good clinical practice for trials on medicinal products in the European community.* Brussels: Commission of the European Communities, 1990.

Father's external radiation dose (mSv) before conception of child for leukaemia and non-Hodgkin's lymphoma cases and controls born to Dounreay and Sellafield workers

Nuclear installation	Child*	Total preconception dose		6 Months preconception dose	
		<100 mSv	≥100 mSv	<10 mSv	≥10 mSv
Dounreay	Case	3	0	2	1
	Control	14	1	14	1
Sellafield	Case	7	4	6	4
	Area control	42	5	31	8
	Local control	59	3	46	5

*See Gardner *et al* for definition of area and local controls (some children could be both). Area controls are more comparable with the controls selected by Urquhart *et al*.[1]

Confidence intervals and sample sizes

SIR,—Dr Leslie E Daly has argued that it is important to estimate appropriate sample sizes and that sample sizes based on the width of confidence intervals can result in studies too small to achieve meaningful results.[1] He suggests that calculations for sample sizes based on confidence intervals should specify the confidence level, the minimum size of the comparative measure that we wish to estimate unambiguously, and the chance of achieving this if the measure actually had this minimum value (power).

We fear that consideration merely of statistical factors in estimating sample sizes conveys a false respectable rigour, which is misleading. All the

factors on which Dr Daly suggests sample sizes should be calculated are arbitrary. The significance level and power should not be selected arbitrarily based on convention but be based on the clinical significance of obtaining false positive and false negative study results. This will then be a judgment based on relevance in real life rather than on statistical dogma.

Calculations of sample size also ignore biases that can be introduced by large studies simply because of the sheer practical difficulties in obtaining a high participation or response. Similarly, as Rothman argues, sample sizes based on technical considerations alone fail to take into account the value of the information obtained.[2] It is important to consider what additional benefit there is to be gained for each increase in sample size. This will amount to an implicit marginal analysis of costs and benefits. Other practical issues of a sociocultural and political nature will also need to be considered.

One point from Dr Daly's paper illustrates the absence of clinical considerations. He has compared sample sizes based on the hypothesis testing approach and those based on the confidence interval approach and has shown in his example that a study based on the latter would have only a 50% power of detecting the corresponding smallest worthwhile differences. What about the largest worthwhile difference? To any physician or patient in despair it is the largest worthwhile difference that will inspire hope, and let us not underestimate the "power" of such hope.

Statistical consideration is only one of many in deciding sample sizes and must not be allowed to dominate such decisions or lend spurious scientific respectability. As Rothman summarises, "In the final analysis, the question of the most appropriate study size is not a technical decision to be determined by calculation but a judgment to be determined by experience, intuition, and insight"[2] —and above all, by relating to the value of the information in bettering people's lives.

K M V NARAYAN
A BISSET
L MACDONALD
J CRESSWELL
B WILSON
V J DEV
A CADZOW

Department of Public Health Medicine,
Grampian Health Board,
Aberdeen AB9 8QP

1 Daly LE. Confidence intervals and sample sizes: don't throw out all your old sample size tables. *BMJ* 1991;**302**:333-6. (9 February.)
2 Rothman KJ. Objectives of epidemiologic study design. In: Rothman KJ. *Modern epidemiology*. Boston: Little, Brown, 1986.

SIR,—Many of Dr Leslie E Daly's arguments have been debated in medical and statistical journals but there is still much confusion and misunderstanding surrounding the problem of determining sample size in either a confidence interval or a hypothesis testing context.[1] The arguments put forward in favour of a hypothesis testing approach, however, may not be as convincing as the simple blood pressure example would have us believe.

Dr Daly suggests an example of a study in which a mean difference of 5 mm Hg is observed with a confidence interval for that difference of 0 mm Hg to 10 mm Hg. How much different is that information from the slightly different situation he describes of a mean difference of 4 mm Hg with a confidence interval from −1 mm Hg to 9 mm Hg? One is "statistically significant" by some arbitrary cut off rule, the other is not. In practical terms, the data we have say either that one treatment is the same as the other (0 mm Hg and −1 mm Hg are in practice the same thing) or that one treatment may have a beneficial effect of up to about 9 mm Hg or 10 mm Hg (these two values also being trivially different). Dr Daly claims to determine sample size to fit in with estimation "with only a slight change of wording," though it is not the wording that matters but the concepts and objectives behind them. Dr Daly's objective would seem to be that of ensuring adequate power for a significance test.

Decision making rarely yields a simple yes or no answer ($p<0{\cdot}05$ or $p \geq 0{\cdot}05$). Clinical trials and other comparative studies can rarely be assessed on the basis of a single efficacy parameter. We all strive to have as few efficacy parameters as possible, but the judgment of whether to prescribe (or register) any given treatment is a balance between how much good it might do, how much harm it might do, and, possibly, how much it costs. What should we ask if one treatment is significantly better than another but has significantly more side effects? Surely, "How much better and how much more?"

None of this is to say that significance testing is irrelevant. The size of an effect, however, is the crux of the problem, and appropriate clinical decisions may often be made in the absence of statistical significance but in the presence of sufficiently accurate estimates. Probably neither approach to sample size determination is appropriate to all studies, so I too would urge researchers to keep their old sample size tables and nomograms. They should not use them, however, without careful regard to the objectives of any particular study.

SIMON DAY

Lilly Research Centre Limited,
Windlesham,
Surrey GU20 6PH

1 Daly LE. Confidence intervals and sample sizes: don't throw out all your sample size tables. *BMJ* 1991;**302**:333-6. (9 February.)

AUTHOR'S REPLY,—The letters from Dr Narayan and colleagues and Mr Day raise some interesting issues relating to the place of sample size calculations in the planning of clinical research projects. I too think that it is wrong to base sample sizes solely on statistical criteria, and I would be the first to argue that the various parameters that go into such a calculation must be based on relevance to real life and sound clinical judgment. Mr Day's comment on the necessity to take account of treatment costs and side effects in addition to treatment efficacy are also germane.

I fail to see, however, the relevance of "the largest worthwhile difference" mentioned by Dr Narayan and colleagues. Usually, sample size calculations are determined in a hypothesis testing framework and are based on a requirement that there be a good chance (the power) of detecting, with a statistically significant result, a treatment effect as large as or larger than a specified minimum worthwhile difference. Of course if the treatment effect is in reality much larger than this minimum the study will be much more likely to detect it. Sample size calculations, however, must be based on a realistic judgment of treatment effects, and if these are set too high the study may in fact fail to detect worthwhile differences that are smaller. A study that misses a real effect can inspire no hope whatsoever—in either patient or physician.

The main intention of my paper was in fact to warn against some recent methods of estimating sample size that seem clinically more relevant than the traditional techniques based on hypothesis testing but have serious shortcomings. The alternative methods, based on a confidence interval philosophy, suggest that the sample size be determined by prespecifying the width of confidence intervals (a measure of precision) without any consideration of power. I argued that such approaches paradoxically give unacceptably small study numbers and proposed a change of wording that would make the use of established methods compatible with a confidence interval analysis. Mr Day suggests that my arguments are not as convincing as they might seem.

In my paper I used the example of a study to detect a blood pressure difference between treatment and control groups of at least 5 mm Hg. A sample size of 251 in each group was determined on the basis of an 80% chance that the 95% confidence interval would exclude zero—equivalent to a 20% chance that zero would be included in the interval. Prespecifying the width of the 95% confidence interval as 10 mm Hg resulted in the smaller figure of 123—which gave a 50% chance of the interval including zero.

Mr Day seems to base his argument on the fact that including zero in the confidence interval does not really matter and that the interpretation of a confidence interval running from −1 to 9 mm Hg is essentially the same as that of an interval running from 0 to 10 mm Hg. Although technically the former would be interpreted as a non-significant result, I agree that the two intervals are equivalent from the point of view of interpretation.

It is instructive, however, to examine the results that might be obtained from studies with the two different sample sizes if the true treatment effect were in fact 5 mm Hg. The smaller sample size gives a 34·8% chance that a confidence interval resulting from the study would include −1 mm Hg, whereas the corresponding chance with the larger sample size is only 8·1%. In more extreme situations the inadequacy of the smaller sample size becomes even clearer. The smaller study has a 16·4% chance of ending up with a confidence interval including −2·5 mm Hg and an 11·9% chance of the interval including −3 mm Hg. The corresponding figures with the larger sample size are 1·3% and 0·6%. There is a considerable difference of interpretation between a confidence interval running from −3 to 7 mm Hg and one running from 0 to 10 mm Hg, and such confidence intervals, compatible with the treatment increasing blood pressure by a clinically important amount, are quite likely with the smaller sample size. The smaller study could easily miss the treatment effect it was designed to detect.

Though it may not be the only criterion for determining the numbers required, a formal estimation of sample size is important in planning a clinical trial. It is vital, however, that such estimations are appropriate, and to this end I urge that we stay with the standard methods, reworded if necessary to fit into a confidence interval framework.

LESLIE DALY

Department of Community Medicine
and Epidemiology,
University College Dublin,
Dublin 2,
Ireland

Rectal examination in patients with abdominal pain

SIR,—We would like to make three comments on the paper of Mr J M Dixon and colleagues[1] from the viewpoint of the general practitioner. Firstly, there is no acknowledgment or discussion of the filtering effect of an initial assessment by a general practitioner on the sample of patients they studied. It is not clear from the paper how the sample of patients was obtained. We were informed by the authors that most patients had been seen and referred to the hospital by their general practitioner. The sample used in this study is therefore a highly selected one, and how it was obtained should have been stated.

Secondly, the authors have chosen to describe the value of the symptoms and signs by using odds ratios. This is only one of several methods that can be used to assess symptoms and signs, some of which will contribute different information and may be used in different situations to provide an overall picture of how useful a test is. For example, the specificity of a test gives an indication of how useful a test is at ruling out disease. By reworking