Clinicians are familiar with the audit cycle. Medical practice is examined, improvements are suggested, and the practice is re-examined after a suitable interval, to allow the improvements to be made. There are difficulties with audit, but all departments of anaesthesia can point to aspects of practice that audit has improved. In 1992, this journal published what was really an audit of statistical practice at the Anaesthetic Research Society. All abstracts in the precirculated booklets for the three meetings in 1990 were examined. Simple statistical errors were grouped as errors in study design, or in choice of statistical method; errors in the presentation of variability; and errors in the presentation of probability. There had been studies of statistics in papers published in anaesthesia journals (for example, reference 3), but by the time a full paper is published, fellow workers, editors and referees have influenced the final version. The statistics presented in precirculated abstracts give a chance to assess the statistical knowledge of researchers in anaesthesia, which is otherwise impossible to assess: researchers are unlikely to answer multiple choice question papers sent out by the curious, even if they could spare the time. In the survey, 61 of 94 abstracts from 1990 presenting numerical information contained errors.

Five years on, we repeated part of this survey, looking at three booklets from 1995–1996. We did not repeat the whole survey, but looked only at the more common errors reported in the earlier one: no mention of randomization (in clinical trials); lack of description of statistical methods; failure to show variability; standard error given on samples of undefined or small (n < 10) size; statistical significance claimed without data; and values of probability given without data. Our findings were disappointing: 34 abstracts of 80 that needed statistics contained these simple errors. This proportion is not directly comparable with the 61 of 94 in the earlier survey because we have included fewer errors this time; Table 1 gives a better comparison, showing the number of errors, rather than the number of abstracts.

There may have been some improvement. There were fewer abstracts in which significance was claimed or probabilities given without data to support them (although this error occurred in 10 of 80 abstracts), but overall there were a disappointing number of errors. This is especially disappointing because two of the faults, failure to show variability or to describe the statistics used, are mentioned in the guidelines on presentation of papers to the Anaesthetic Research Society.

Over the years, medical statisticians have been critical of medical researchers, and it is from the publications of one of those critics, D. G. Altman, that the original list of errors was drawn. Error is the word that Altman uses and, with its implication of not keeping to recognized good practice, is more appropriate than other possible words, such as mistake or fault. A mistake implies a wrong choice or carelessness. A fault has implications of responsibility, which an error does not necessarily have. Anyone using statistics as an experimental tool should know how to use them — just as they should know how to use cell cultures, neuromuscular monitoring, liquid chromatography or whatever other experimental tool they use in their research. Error is also a useful word in this context because of the familiar expressions “error of omission” and “error of commission, and most of the errors from our repeated survey are errors of omission: it seems that investigators are not aware of the good practice required for statistical description.

The errors described by Altman, leaving aside for the moment misuse of the standard error of the mean, are not defendable. Randomization is necessary to reduce bias and because statistical theory is based on the idea of random sampling. Randomization is too important for it to be assumed; it should be stated explicitly. The statistical tests used must be described, to enable the reader to identify statistical errors and to determine whether the conclusions of a study are valid; in addition, and of fundamental importance, readers should not have to guess or presume what the investigators did. A mean or median with no indication of variability is simply uninterpretable. A statement of probability or of statistical significance without data is meaningless. Overall, if there are statistical errors, the conclusion of a study may be incorrect and misleading. There is no authority in statistics who would make allowances for these simple errors, so why do they occur in abstracts at the Anaesthetic Research Society? Why do they occur in spite of an explicit audit published in this journal asking that they should not occur and, for two of the errors, an explicit instruction in the guide to presenters? This is not just a domestic problem for the Anaesthetic Research Society, but a problem for the whole specialty. If the Obstetric Anaesthetists’ Association speaks for its subspecialty, and the other sub-

Table 1 Statistical errors, as absolute number of errors, and as percentage of abstracts needing statistics, identified in two surveys of precirculated Anaesthetic Research Society abstracts. Errors occurred in study design (1 and 2); presentation of variability (3 and 4); and presentation of probability (5 and 6). see = standard error of the mean. P= probability

<table>
<thead>
<tr>
<th>Errors</th>
<th>Abstracts 1990</th>
<th>As %</th>
<th>Abstracts '95–'96</th>
<th>Number</th>
<th>As %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total abstracts</td>
<td>115</td>
<td>—</td>
<td>91</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Needing statistics (n)</td>
<td>94</td>
<td>—</td>
<td>80</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>1 No mention of randomization</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2 Lack of statistical description</td>
<td>29</td>
<td>31</td>
<td>17</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>3 No variability shown</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4 see, no n or n &lt; 10</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>5 Significance claimed, no data</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>6 P given, no data</td>
<td>11</td>
<td>12</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total errors</td>
<td>74</td>
<td>—</td>
<td>49</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>
specialty groups likewise, then the Anaesthetic Research Society speaks for research and research method, and is a window on that.

Abstracts cannot, of course, contain every detail of a study. One could argue that the investigators would not make these errors of omission in their full reports. That may be so, but abstracts are scientific presentations, even if not as polished as the final reports. They are presented to peers for scientific appraisal, which is less easy if detail is lacking. Investigators do not have the excuse that space is limited: 20 of the 34 abstracts had at least two lines spare; nine had more than five lines spare. There is a view that the abstracts of the Society should not be published in the *British Journal of Anaesthesia*. Whether or not they are published, the precirculated abstracts are still available to members of the Society. A future decision not to publish them would not affect the validity of future surveys, although if we take the statistical criticism seriously it throws doubt on the scientific validity of the abstracts.

The errors described in these surveys are basic. A survey by Avram and co-workers was much more comprehensive. They were greatly concerned with the misuse of “elementary” hypothesis tests, making examples particularly of repeat testing and the use of parametrically based tests such as the $t$ test on data suitable only for non-parametric analysis. They considered their study used “only simple criteria” our survey is even simpler. Nor have we considered another important statistical problem: studies that are too small to detect important differences — so-called beta, or Type II, error.

There are three broad questions. Why are these errors made? Do they matter? How can statistical practice be improved? The first is easy to answer. Anaesthetists in general are not comfortable with statistics. Candidates for the Fellowship examinations approach cardiorespiratory tutorials with enthusiasm; they approach statistics tutorials with dread and resignation. Yet the amount of statistics that those candidates may later do, is small. The necessary degree of understanding is far less than in cardiorespiratory physiology.

There is published evidence of anaesthetists’ dislike of statistics. In a survey reported by Burnstein and colleagues, a quarter of respondents in the unprompted section thought knowledge of statistics was irrelevant to the practice of anaesthesia. Half the respondents felt that the only knowledge needed was of the various statistical tests, implying that a “recipe-book” approach is all that is necessary. Almost half the respondents felt unable to give a tutorial on statistics. We suspect that anaesthetists actively involved in research would give different answers, but our surveys, and the lack of statistical insight of many of the presenters at the Anaesthetic Research Society, suggest that statistics is still not taken seriously. Presenters are often trainee anaesthetists, but while the error may be theirs, most of the responsibility lies elsewhere, with their supervisors.

Do the errors matter? We would not be writing this editorial if we did not think so, and there are enough published commentaries on medical statistics to suggest that there are many others who think likewise. So how can statistics be improved? We concluded our earlier survey by suggesting that by using statistical checklists, and with the help of local statisticians, the Anaesthetic Research Society had the opportunity to play its part in improving the general standard of the reporting of medical research. In the intervening 5 years that has not happened. Statistical checklists and local statisticians are not as helpful as a proper understanding of statistical principles, which boils down really to a proper understanding of probability and of the spread of data. These principles underlie the ideas of taking samples from a population and of the valid estimation of measures from those samples. No knowledge of the mechanics of any statistical test is needed, as long as there is some knowledge of which test is appropriate — for which a statistician is helpful; and of what the result means — which is where the understanding is essential. A researcher with a proper understanding of statistics would not present a mean without variability, or fail to state the statistical test used.

A researcher with a proper understanding of statistics would also know when the standard error of the mean is appropriate. The misuse of the standard error, especially its use for small samples, is not as simple an error as failing to show variability, but it is a good indication of statistical understanding. The standard error does not show the spread of data; the standard deviation does that. The standard error enables a calculation of confidence in the mean, and it is this confidence interval that is the statistic with real meaning. A 95% confidence interval of 10 mm Hg around a mean arterial pressure of 125 mm Hg implies 95% probability that the true mean arterial pressure of the population from which the sample was drawn lies between 115 and 135 mm Hg. For large samples, the 95% confidence interval is twice the standard error. For smaller samples, not only is the standard error larger, but the 95% confidence interval becomes progressively more than twice the standard error. The size of sample below which the standard error (and, by implication, the confidence interval) becomes less useful is a matter of convention, in just the same way that the cut-off of $P<0.05$ is convention. It just seems a matter of common sense that it is not wise to try to fix the mean of the population from a sample of five or six.

Are we right to criticise research in anaesthesia and bring these criticisms to possibly wider attention, especially when anaesthesia research is under enough strain? We have not compared anaesthesia with other specialties, nor anaesthesia in Britain with that in other countries, though there is no reason to suppose anaesthesia, or Britain, is any better or worse than other specialties or countries. When we complete the clinical audit cycle in our hospitals we compare now with then, rather than compare here with elsewhere, and that is what our surveys have done. There is little change. Whatever damage publication of our findings might do now, it will be worse in the future if nothing happens. If other specialties wish to take us to task, then they must do so on evidence of better statistical practice in their specialties — and comparisons of our surveys with their published work will not do that.

Since Altman complained that “The general standard of statistics in medical journals is poor” there have been great improvements. In that time
many journals, of which the British Journal of Anaesthesia is one, have introduced statistical checklists. Otherwise-accepted manuscripts are now far more likely to be sent to a statistician or at least to someone with a statistical interest. The system is much improved, but our survey, and experience at research meetings, such as those of the Anaesthetic Research Society, suggests that better statistical knowledge among researchers would ease the workload of referees. It is a shame when discussion of an interesting project becomes bogged down because of the misuse of simple statistics.

More statistical emphasis in the anaesthesia examinations is not the answer. Candidates soon forget. Perhaps the emphasis should change from statistics in the narrow sense to the wider aspects of good study design. All practising clinicians must have some knowledge of this to enable any critical appraisal of what is published. But it is those actively involved in research who need to have an immediate feel for statistics: supervisors have a responsibility to teach their researchers, or to ensure there is someone who can do so; researchers must realize the importance of statistics. Researchers also have a responsibility: to ask for teaching in statistics relevant to their projects. We need to get away from the all-too-common feeling among researchers that statistics is just an awkward encumbrance to the real business of research, a feeling often paradoxically coupled with an undue reliance on the magic figure of a probability of less than 0.05.

Altman and many others have criticised medical journals. We can take some comfort that one of the errors that they and we have looked at — the misuse of standard error — is rife in one of the foremost scientific journals, Nature. Nature publishes little of relevance to anaesthesia but this is changing with the recent interest in the possibility that specific protein sites are involved in the mode of action of anaesthetics. In their paper, Mihic and co-workers show data with a statistical interest. The system is much better today than it was 20 years ago, but its improvement is not because they are particularly informative. Our specialty’s response to this should not be that what Nature does is good enough for us, but that even a journal with the highest scientific reputation can get it wrong, and in some ways we can better them. The widespread impression that clinical research is less sound, statistically and otherwise, than basic scientific research may be false: Muers wrote of respiratory physiology that whereas “therapeutics needs tens of thousands of patients, valid and persuasive physiological evidence continues to be drawn from studies of a mere one or two subjects.”

The statistical sensibility of anaesthesia journals is higher than it used to be, but it needs to filter down to those doing the research in anaesthesia. There is no instant solution, but those who have this sensibility must take every opportunity to impart it to those who do not. This means that they must not be afraid to speak statistically whenever they can, and researchers must expect to be questioned on their basic understanding of statistics if that understanding appears to be shaky.

N. W. G. thanks Doug Altman for helpful comments and suggestions.

References