SIGNIFICANT OR IMPORTANT?

The letter in this issue of the Journal from Professor Vere, Dr de Saintonge and Dr Sharman (de Saintonge, Vere & Sharman, 1977) raises an important question that is too seldom asked by medical practitioners. When a clinical trial has shown that one treatment is significantly more effective than another, is this sufficient to convince us that we should use it?

The answer is that further examination of the data is required to make a decision because a difference that is statistically significant may not be important clinically.

With this in mind the authors examined the tests of significance that we would normally use in comparing the effectiveness of two hypnotic drugs, diazepam and nitrazepam, in inducing and maintaining sleep. Because nitrazepam is more expensive than diazepam they modified the test of significance in such a way that their experiment would need to show with a high degree of certainty that nitrazepam was three times as effective as diazepam before the difference would be significant; but if diazepam were better than nitrazepam a very much smaller degree of certainty for a smaller difference would be statistically significant. They are in effect converting a test of significance (that an observed difference could not have occurred by chance) into a test of importance. For those with an interest in statistics, using the usual null hypothesis, a difference between two means is said to be significant, $P < 0.05$, when one mean differs by more than 2 standard errors from the first. Diagrammatically Figure 1 shows the distribution of their difference about a mean of zero (as prescribed by the null hypothesis); if the result is significant ($P < 0.05$), then the observed difference is found to be in one of the two tail areas (shaded), each of which has area 0.025 of the total area under the curve, so that together they amount to 0.05 ($=\alpha$). We use both tails because the difference is essentially ambiguous in its sign, and we are not to know in advance of the experiment which mean will be the greater.

de Saintonge, Vere & Sharman (1977) however have set asymmetric limits, as shown in Figure 2, so that diazepam will be described as significantly better than nitrazepam when the difference in favour of diazepam is in the first shaded segment, α_1 , with area 0.04, but that a difference in favour of nitrazepam will not be described as significant until it is in the segment α_2 , of area 0.01. The combined risk of error if the null hypothesis is true thus remains at 0.05.

de Saintonge, Vere & Sharman (1977) have carried the analysis a stage further, however. They are keen not to run any serious risk of failing to detect a sizable difference between these drugs because of the play of

chance in the experiment. They decided what was the size (D) of the minimal difference of importance between them, and decided also what risk they were prepared to accept ($\beta = 0.10$) of failure to detect it.

In Figure 3 there appear two possible distributions for the difference between nitrazepam and diazepam: the first, centred on a mean of zero, is that of the null hypothesis; the second, with mean D , is the distribution which has the minimal difference which should not fail to be detected, if it exists – that is, if

Figure 1 Distribution of difference between means of two hypnotic drugs (nitrazepam and diazepam) with symmetrical limits where the two tail areas (shaded) are equal.

Figure 2 Distribution of difference between means of two hypnotic drugs (nitrazepam and diazepam) with asymmetrical limits where the two tail areas (shaded) are unequal.

Figure 3 Two possible distributions for the difference between means of two hypnotic drugs. For explanation see text.

that hypothesis is true. The null hypothesis will be rejected if the observed positive difference is greater than the value k , or accepted if the observed difference is less than k . But if the second hypothesis is true, this acceptance of the null hypothesis equates with a rejection of the real difference D : this situation will occur when the observed difference is found in the shaded area labelled ' β ', and if this area is 0.10, then the conditions are satisfied that the real difference of D will be rejected not more often than 0.10 or once in ten times.

de Saintonge, Vere & Sharman (1977) are to be congratulated on this attempt to convert a test of significance into a test of importance. It has the great merit that they decided in advance by how much and with what degree of certainty their experiment would have to show that nitrazepam was better than diazepam.

It has two advantages. Firstly most doctors find statistics difficult and find it difficult enough to understand the niceties of a test of significance based on the conventional form of the null hypothesis, let alone when the basis of decision depends on a deliberate introduction of a skewed test. Secondly if the price ratio of nitrazepam and diazepam changes the validity of their result will change.

It would seem to us to be easier for those who design clinical trials to continue to use the usual form

of tests of significance based on the null hypothesis. But it is vital that a *statistically significant* difference should not necessarily be assumed to be an *important* difference. It is not going to be worthwhile using nitrazepam to get on average an extra 15 min sleep at three times the cost of diazepam. It might be worthwhile if for instance it got on average an extra hour and 15 min.

It is extremely important that doctors give thought to these matters and that they are not persuaded by advertisers or others to accept statistically significant differences in the performance of drugs as necessarily indicating a difference of practical importance of value.

O.L. WADE

*Department of Therapeutics and Clinical
Pharmacology,
Queen Elizabeth Hospital,
Birmingham B15 2TH*

J.A.H. WATERHOUSE

*Department of Medical Statistics,
University of Birmingham,
Birmingham B15 2TH*

Reference

- CHAPUT-DE SAINTONGE, D.M., VERE, D.W. & SHARMAN, V.L. (1977). Diazepam or nitrazepam? A question of cost. *Br. J. clin. Pharmac.*, **4**, 492-494.