Using the Grubbs and Cochran tests to identify outliers

Analytical Methods Committee, AMCTB No. 69

[illegible] (TB N . 39) [illegible]

Significance tests can be used with care and caution to help decide whether suspect results can be rejected as genuine outliers (assuming that there is no obvious explanation for them such as equipment or data recording errors), or must retained in the data set and included in its later applications. Such tests for outliers are used in the conventional way, by establishing a *null hypothesis*, H_0, that the suspect results are *not* significantly different from the other members of the data set, and then rejecting it if the probability of obtaining the experimental results turns out to be very low (*e.g.* $p < 0.05$). If H_0 can be rejected the suspect result[s] can also be rejected as outliers. If H_0 is retained the suspect results must be retained in the data set. These situations are often distinguished by converting the experimental data into a *test statistic* and comparing the latter with critical values from statistical tables. If the two are very similar, *i.e.* the suspect result is close to the boundary of rejection or acceptance, the test outcome must be treated with great circumspection.

[illegible]

Fig. 1 shows as a dot plot the results obtained when the cholesterol level in a single blood serum sample was measured seven times. The individual measurements are 4.9, 5.1, 5.6, 5.0, 4.8, 4.8 and 4.6 mM. The dot plot suggests that the result 5.6 mM is noticeably higher than the others: can it be rejected as an outlier? This decision might have a significant effect on the clinical interpretation of the data. Several tests are available in this situation. For years the most popular was the Dixon or *Q*-test, introduced in 1951. It has the advantage that (as in this example) the test statistic can often be calculated mentally. It has been superseded as the recommended method (ISO 17025) by the Grubbs test (1950), which compares the difference between the suspect result and the mean of all the data with the sample standard deviation. The test statistic, *G*, in this simplest case is thus given by:

$$G = |\text{suspect value} - \bar{x}|/s \qquad (1)$$

where the sample mean and standard deviation, $\bar{x}$ and s, are calculated with the suspect value *included*. In our example $\bar{x}$ and s are 4.97 and 0.32 respectively, so $G = 0.63/0.32 = 1.97$, which is less than the two-tailed critical value ($p = 0.05$) of 2.02. We conclude that the suspect value, 5.6 mM, is not an outlier, and must be included in any subsequent application of the data (the Dixon test leads to the same conclusion). One important lesson of this example is that in a small data sample a single suspect measurement must be *very* different from the rest before it qualifies for rejection as outlier. In this instance the Grubbs and Dixon tests agree that only if the suspect measurement is as