XXII. AN INVESTIGATION OF THE ACCURACY OF THE BARCROFT DIFFERENTIAL MANOMETER IN RESPIRATION STUDIES.

BY ALEXANDER STEVEN CORBET AND
WALTER REGINALD WOOLDRIDGE.

From the London School of Hygiene and Tropical Medicine.

(Received November 30th, 1935.)

ALTHOUGH the Barcroft apparatus is used extensively for the study of respiration in biological fluids, we have found no published work in which the accuracy of the results obtained has been submitted to a critical examination. This is somewhat surprising when it is considered that the Barcroft apparatus is employed so frequently for comparative work where the interpretation is often founded on small differences.

In view of these considerations and from a knowledge of a certain variability in results from experiments made in triplicate as recommended by Wooldridge and Standfast [1936], it appeared desirable to submit the results obtained by use of the Barcroft technique to a statistical examination. Our work has shown that the use of four, or even of three, manometers usually suffices to give a mean with a reasonably low standard error. For qualitative work serious objection can hardly be raised against duplicate experiments, but a single experiment cannot be recommended, for it may give a dubious result which cannot be checked effectively.

The results from the following five series of experiments formed the basis for the statistical work, but use has been made also of the extensive data collected by one of us (W. R. W.) in collaboration with Mr A. F. B. Standfast. In these experiments the usual technique was followed, the manometric cups containing a total volume of liquid of 3 ml. In the following brief description only the active reagents present in the cups are mentioned.

Series I. Left-hand cup: *Pseudomonas fluorescens*; right-hand cup: *Ps. fluorescens* in *M*/40 glucose solutions. All apparatus and reagents sterilised before use.

Series II. As in Series I, but apparatus and reagents not sterilised before use.

Series III. Left-hand cup: *Pseudomonas fluorescens* + *Bacterium coli*; right-hand cup: *Ps. fluorescens* + *Bact. coli* in *M*/40 glucose solution.

Series IV. Left-hand cup: tap water; right-hand cup: crude sewage.

Series V. Left-hand cup: activated sludge; right-hand cup: sludge + crude sewage (Seitz-filtered).

Table I.

Series	Barcroft manometers used	Results rejected*	Duration of experiment in hours	Mean at end of experiment in μ l.	% standard error at end of experiment
I	14	1	9.0	175 (standard error 4.71)	2.7
II	17	2	5.0	179 (" 4.42)	2.5
III	16	2	9.5	177 (" 5.75)	3.2
IV	16	2	8.0	189 (" 6.16)	3.3
V	15	1	8.0	570 (" 10.67)	1.9

* Rejections were made on the basis of anomalous pressure difference-time curves, as explained below.

The mixtures or liquids in all cups were buffered to p_{H} 7.4 and temperature variations during any one experiment were slight, rarely exceeding 1° .

Readings were taken at half-hourly intervals, and experiments were continued for at least 8 hours, except those in Series II, which were terminated after 5 hours. A summary of the results is given in Table I.

A criterion for the acceptance or rejection of results.

Occasionally results from replicate experiments with the Barcroft apparatus may be so divergent that it appears advisable to reject one or more of them when calculating the mean. Sometimes, investigation at the conclusion of an experiment reveals the cause of such aberrant results (*e.g.* a leak, condensation of fluid in the manometer tube *etc.*), but in many instances no explanation is forthcoming and the rejection of a result suspected of being erroneous becomes an arbitrary procedure. We have therefore sought a criterion whereby the acceptance or rejection of a result is freed from personal bias.

Fig. 1a. Plot of pressure differences against time on double logarithmic paper (Series II).

Fig. 1b. Plot of pressure differences against time on double logarithmic paper (Series IV).

In a general way it may be said that provided that the temperature remains reasonably constant, the course of oxygen absorption in the Barcroft cups tends to approximate to the equation $Y = Ft^m$, where Y is the pressure difference between the left- and right-hand cups after time t , and F and m are constants [*cf.* Corbet, 1934, 1]. Thus the plot of the pressure difference against the time elapsed on double logarithmic paper approaches a straight line. Under some conditions however, particularly when gas uptake is rapid, the values of F and

m change so rapidly that the resulting curves tend to be parabolic rather than linear: nevertheless, the curves from replicate experiments are more or less parallel although, in general, they tend to converge with increasing time. We found that the pressure difference-time curves of experiments known to be faulty did not follow their fellows, but usually differed so radically in character that such experiments could be detected immediately by this means and legitimately rejected. Series of curves from two sets of identical experiments are depicted in Figs. 1 *a* and 1 *b*: it will be seen that both series contain a faulty experiment (dotted curve) which can be omitted justifiably from further consideration. Occasionally, the plot on double logarithmic paper reveals parallel but widely spaced curves. This implies a high standard error and one may feel tempted to reject the divergent results: we do not consider that such a procedure can be upheld, however, for then the question which results have to be rejected becomes a matter of personal opinion. The occasional appearance of these presumably normal results demonstrates the importance of not drawing conclusions from a single experiment.

The question of the acceptance or rejection of results suspected of being faulty is one of considerable importance. In fact we might go so far as to say that the degree of precision possible in experiments with these differential manometers depends on the detection and elimination of such faulty experiments. On the other hand, rejection of readings suspected of being too high or low for insufficient reasons will produce results whose accuracy is fictitious rather than real.

Calculation of the mean.

The usual statistical methods of assessing the accuracy of experimental data are based on the assumption that these observations are normally distributed about their mean and, if such is the case with results derived from experiments with the Barcroft apparatus, it is correct to express the mean of the results by dividing their sum by the total number of experiments. On the other hand, if the logarithms of the values, and not the values themselves, are normally distributed (as is the case with parallel plate counts of soil and milk samples [Corbet, 1934, 2]), then it is preferable to determine the geometric mean.

To decide conclusively whether the values from a series of replicate Barcroft experiments are normally distributed requires more results than are usually available but we have examined the question in the light of the results obtained from the five series of experiments mentioned above. The normality of the distributions was tested by calculation of Pearson's constants β_1 and β_2 (see Table II). For a symmetrical distribution β_1 is zero whilst β_2 approximates to 3, β_2 being greater or less according to whether the curve is leptokurtic or platykurtic. In Table II are included results found by Cook and Stephenson [1928] for oxygen absorption by bacteria in $M/200$ glucose solution; this experiment differs from those of Series I–V in that the oxidation was carried to completion.

Table II.

Series	Number	Actual values		Logarithms of actual values	
		β_1	β_2	β_1	β_2
I	13	0.62	1.96	0.21	1.94
II	15	0.16	1.85	0.092	2.42
III	14	0.21	2.02	0.17	2.12
IV	14	0.24	1.76	0.83	1.98
V	14	0.01	2.18	1.33	2.24
Cook and Stephenson [1928]	28	0.23	2.45	0.23	2.45

Deviations of β_1 and β_2 from zero and 3 respectively can only be regarded as significant if the respective values of $\sqrt{\beta_1}$ and of β_2 are $> 2 \times 0.65$ and 2×1.31 for 14 results, and 2×0.46 and 2×0.93 for 28 results [Tippett, 1931]. It will be seen from the table that, although the distributions of the results of Series I–III are not significantly asymmetrical, the logarithms of the actual values suggest a more normal distribution; on the other hand, the actual values from Series IV and V, rather than their logarithms, are normally distributed. In no case, however, is the divergence from zero and 3 of the β constants sufficient to justify working with logarithmic rather than arithmetic values.

The standard error in Barcroft experiments.

The standard deviation is a measure of the distribution of the individual results about the mean value: it is usual to express this deviation in terms of the standard error and to regard as statistically significant any results which differ from the mean by more than twice the standard error. The standard error and standard deviation are related in the manner shown by the following expression:

$$\text{standard error} = \frac{\sigma}{\sqrt{N}} = \sqrt{\frac{\Sigma (x - \bar{x})^2}{N(N-1)}},$$

where σ is the standard deviation, $\Sigma (x - \bar{x})^2$ is the sum of the squares of the individual deviations from the mean and N the number of results.

It is desirable to ascertain how the standard error varies with duration of the experiment, the degree of replication and the magnitude of the mean. The standard error remains constant during the course of a single experiment only if the curves representing the oxygen absorption with time are parallel to one another; when these curves converge, diverge or cross, the standard error changes accordingly. Although occasionally instances of divergence or intersection are encountered, results from a large number of experiments show that these curves tend to converge as time proceeds, and so, as a general rule, the standard error diminishes as the experiment progresses.

The plot of the percentage standard error against time in the five series of experiments referred to (Table III and Fig. 2) shows that the standard error approaches a constant value which, for all practical purposes, is attained within 5 or 6 hours. The result of summing the curves I, III, IV and V (II was continued for 5 hours only) and taking the mean percentage standard errors is depicted also in Fig. 2. The curves for Series I–III (pure cultures of bacteria) show initially a high standard error which falls rapidly; in Series IV and V (sewage and sludge) the standard errors are almost constant throughout. The explanation of the difference in character between these two types of curves may be that, whilst the introduction of an active strain of bacteria into a sterile medium with an abundance of a readily oxidisable compound was associated with initial growth phases of varying duration in different manometers, with material containing a micro-population already in a condition approaching equilibrium, no very abrupt changes occurred and the gas absorption in the manometers was much more uniform. On the other hand, in a series of experiments in which the oxygen absorption by soil was measured in presence of KOH, the curve representing the decrease in percentage standard error with time was almost identical with that for Series III.

It is evident from Fig. 2 that, as a general rule, a consistently low percentage standard error can be expected only after an experiment has been in progress for not less than 5 hours.

Table III. *Variation of standard error with duration of experiment.*

Series	% standard error after			
	1 hour	3 hours	4.5 hours	6.5 hours
I	8.58	4.10	3.12	2.76
II	7.32	3.54	2.68	—
III	26.06	9.40	5.35	3.71
IV	3.70	3.18	3.24	3.22
V	2.48	1.78	1.47	1.75

Fig. 2. Variation in standard error with duration of experiment. The dotted curve represents the mean values for the curves of Series I, III, IV and V.

The calculated variation in the percentage standard error with the degree of replication found in our five series of experiments is shown in Table IV and Fig. 3.

The figures and graphs show that a standard error of about 5% would be obtained with 4 manometers for $6\frac{1}{2}$ hours and with 5 manometers for $4\frac{1}{2}$ hours; duplicate results gave a standard error under 10% for periods over 4 hours. It will be recalled that faulty results were eliminated, so as a general practice it is advisable to prepare one more replicate than the number considered to give a sufficiently low standard error.

There is some evidence suggesting that the percentage standard error may increase slightly as the magnitude of the mean decreases; if this is so, it appears to be too slight to have any practical significance. It is possible, too, that the percentage standard error may decrease somewhat as the experimental temperature is raised. Most of our data refer to work carried out at 20° , so we are unable to test this point adequately.

Table IV. *Variation of standard error with degree of replication.*

(a) After 4½ hours			No. of replicate experiments					
	15	14	13	8	6	4	3	2
	% standard error found			% standard error calculated				
Series								
I	—	—	3.12	3.83	4.42	5.41	6.25	7.65
II	2.68	—	—	3.55	4.09	5.02	5.79	7.09
III	—	5.35	—	6.82	7.87	9.62	11.13	13.64
IV	—	3.24	—	4.13	4.76	5.83	6.74	8.25
V	—	1.47	—	1.88	2.17	2.66	3.07	3.75
Mean	3.17			4.04	4.66	5.71	6.60	8.08

(b) After 6½ hours			No. of replicate experiments					
	14	13	8	6	4	3	2	
	% standard error found		% standard error calculated					
Series								
I	—	2.76	3.38	3.90	4.78	5.52	6.76	
III	3.71	—	4.72	5.45	6.67	7.70	9.44	
IV	3.22	—	4.11	4.74	5.81	6.70	8.21	
V	1.75	—	2.23	2.57	3.15	3.64	4.46	
Mean	2.86		3.61	4.17	5.10	5.89	7.22	

(c)						
	No. of replicate experiments					
	28	14	7	4	3	2
	% standard error found	% standard error calculated				
Cook and Stephenson [1928] (to completion)	1.73	2.39	3.24	4.48	5.18	6.34

Fig. 3. Variation in standard error with replication. The curves A, B and C represent the mean percentage standard errors for Series I-V after 3, 4½ and 6½ hours respectively. Results from the data of Cook and Stephenson (experiments carried to completion) are expressed by the curve D.

Examination of the significance of differences between means.

Often in experiments with the Barcroft apparatus it is necessary to compare results of two or more series of experiments: with large differences the interpretation to be placed upon them usually presents no difficulty, but the precise significance to be attached to small differences can be decided only after statistical examination of the data.

When the standard errors of two means are known, the probability of the difference between them arising by chance can be estimated and, if the odds against are sufficiently great, this difference is accepted as real. Usually, it is satisfactory to work on the $P=0.05$ level of significance; that is, to regard as significant all differences which would occur by chance less often than once in 20 times.

Dixon and Elliott [1930] concluded that the absorption of carbon dioxide in the Barcroft manometer is carried out more efficiently in presence of a roll of filter-paper soaked in KOH than by KOH solution alone. We have tested this conclusion in a replicated experiment and find that, under the particular conditions of our experiments, there is no significant difference between the two methods.

Two parallel sets of experiments, in which oxidation of the sterilised crude sewage was effected by the addition of mixed sewage bacteria, and in which carbon dioxide absorption was effected either by KOH solution or by KOH-saturated papers placed in the central tube of the Barcroft cups, gave the following results:

(a) KOH solution: mean of 6 experiments = 84.3 (s.e. 1.18) μl .

(b) KOH-saturated papers: mean of 8 experiments = 80.0 (s.e. 3.36) μl .

Fisher's t [1934] is calculated from the formula

$$t = \frac{(\bar{x} - \bar{x}')}{s \sqrt{\frac{1}{N} + \frac{1}{N'}}},$$

where $\bar{x}$ and $\bar{x}'$ are the respective means derived from N and N' observations and s is obtained from the relation

$$s^2 = \frac{\Sigma (x - \bar{x})^2 + \Sigma (x' - \bar{x}')^2}{(N - 1) + (N' - 1)},$$

the terms $\Sigma (x - \bar{x})^2$ and $\Sigma (x' - \bar{x}')^2$ representing the respective sums of the squares of the individual deviations from the means.

For the results of the experiments (a) and (b) we have

$$s^2 = \frac{(42 + 632)}{(5 + 7)} = 56.17,$$

and

$$t = \frac{(84.3 - 80.0)}{\sqrt{56.17} \cdot \sqrt{\frac{1}{6} + \frac{1}{8}}} = 1.063.$$

Entering the table for t at $n = [(N - 1) + (N' - 1)]$ (that is, at $n = 12$), it is found that P is greater than 0.3 for $t = 1.063$, implying that the difference between the means of (a) and (b) can arise by chance about one in 3 times and so is without significance.

A second example involving differences between three sets of experiments is of some importance from the viewpoint of sewage purification. A crude sewage was sterilised by autoclaving and was subsequently aerated in equal quantities with washed activated sludge and with the same sludge previously autoclaved.

After 3 hours' aeration the mixtures were centrifuged and the supernatants examined in Barcroft manometers to see whether their rates of oxidation in presence of activated sludge were different from that of the original autoclaved sewage; in other words, to test whether 3 hours' aeration with either of the sludges effected any reduction in the content of oxidisable matter of the original sewage. The following results were obtained:

(a) Autoclaved sewage: mean of 4 experiments 234 (s.e. 25.2) μ l.

(b) Autoclaved sewage after aeration with activated sludge: mean of 4 experiments 81 (s.e. 17.4) μ l.

(c) Autoclaved sewage after aeration with autoclaved sludge: mean of 4 experiments 179 (s.e. 4.79) μ l.

From the results of experiments (a) and (b) we have $t=5.00$ and $n=6$, and P is found to be considerably less than 0.01 and so the difference between the means of experiments (a) and (b) is highly significant. That is to say, aeration of the sewage with activated sludge has significantly reduced its content of oxidisable matter.

For the means of (a) and (c), however, $t=2.147$ corresponding to a value of P between 0.1 and 0.05 and showing that the difference between these means can arise by chance once in 10 to 20 times. Thus, although aeration of a sewage with an autoclaved sludge appears to reduce its oxidisable matter somewhat, the results obtained hardly permit definite conclusions to be drawn. But, in this example, the standard error of (a) is unusually high, thus precluding any precise deductions from being made regarding a comparatively small difference between (a) and (c). The standard error might have been reduced by higher replication or continuance of the experiment for a longer period of time.

In the same way, with the means of (b) and (c), t is 5.425 for which, at $n=6$, P is much less than 0.01 and so the difference is highly significant. Thus, even if the autoclaved sludge had produced some purification of the sewage such purification is very considerably increased by aeration with a normal activated sludge.

These two examples should suffice to illustrate the value of the use of the Barcroft technique in the way which we suggest.

SUMMARY.

A statistical study of results obtained by the use of the Barcroft apparatus for respiration experiments has shown that deductions from single, or even duplicate, experiments should be accepted with caution and that a higher degree of replication is desirable.

On the basis of our conclusions we have adopted the following procedure.

(a) Experiments are planned in triplicate, quadruplicate, or even with higher replication, according to the standard error expected (see Fig. 3) and the probable magnitude of the differences to be examined.

(b) Experiments are continued for a period appropriate to the number of replicates used so as to give a standard error not appreciably greater than 5% (see Fig. 2).

(c) The actual values of the pressure differences found at hourly intervals are plotted on double logarithmic paper and results of any experiments giving anomalous curves (see text) are rejected without further examination.

(d) After conversion of the pressure differences into volumes (by multiplication of the manometer factors), the arithmetic means are found, the standard errors calculated, and the difference between the means investigated as shown on p. 138.

If such a procedure be adopted the Barcroft differential manometer can be employed usefully in the investigation of small or large differences in many biological oxidations.

This investigation was carried out as part of the programme of the Water Pollution Board of the Department of Scientific and Industrial Research. Our thanks are due to our colleague, Dr J. O. Irwin, for advice on certain statistical matters.

REFERENCES.

- Cook and Stephenson (1928). *Biochem. J.* **22**, 1370.
Corbet (1934, 1). *Soil Sci.* **37**, 109.
— (1934, 2). *Soil Sci.* **38**, 407.
Dixon and Elliott (1930). *Biochem. J.* **24**, 820.
Fisher (1934). Statistical methods for research workers. 5th ed., Table IV..
Tippett (1931). The methods of statistics, p. 59.
Wooldridge and Standfast (1936). *Biochem. J.* **30**, 141.