# SOME ERRORS IN GAS ANALYSIS USING THE HALDANE APPARATUS

By E. T. RENBOURN, M.D., M.R.C.P., B.Sc. (Lond.) and J. McK. ELLISON, B.A. (Cantab.)
*Department of Applied Physiology, London School of Hygiene and Tropical Medicine*

With the Technical Assistance of L. M. Croton

(With 5 Figures in the Text)

## INTRODUCTION

A number of physical methods have been introduced during the last few years for the analysis of gases. Amongst these we have those involving interferometry (Lowell-Olsen, 1948), paramagnetism (Rein, 1943), thermal conductivity (Daynes, 1920), infra-red absorption (Dingle & Pryce, 1940), and absolute gas-flow analysis (Grove-White & Sander, 1949). Nevertheless, they are not yet suitable for accurate routine work and their errors are not yet completely known.

In physiological work the Haldane gas analysis apparatus and its modifications (Krogh, 1920; Carpenter, Fox & Sereque, 1929) remain standard, and the accuracy has been laid down by Haldane (1912) and repeated by such recent authorities as Douglas & Priestley (1924), Gemmill (1931), Peters & Van Slyke (1932), Thorpe & Whiteley (1938), MacLeod (1941), and Hawk, Oser & Summerson (1947). Although the accuracy obtained by an analysis is a fundamental feature of the method, there is a paucity in the literature of data concerned with error of volumetric analysis, and limits of error are sometimes quoted without reference to the methods by which they are derived.

The present work is concerned with the standard 10 ml. Haldane gas analysis apparatus, used particularly for measuring the logarithmic decrement of carbon dioxide concentration in domestic ventilation studies (Renbourn, Angus, Ellison & Jones, 1949), where accuracy is required at concentrations between 5 and 0·5 % $CO_2$ if large errors in air change are to be avoided. While the portable Haldane apparatus gives far higher accuracy it is considerably more difficult to use in routine work and does not measure concentrations of $CO_2$ above 1·0 %. The larger 20 ml. apparatus gives double the accuracy of the 10 ml. apparatus, but apart from this is liable to the same errors. However, the considerations raised by us appear in general applicable to any method of volumetric analysis. A study has been made of the magnitude and sources of errors encountered in routine work, of the variation in errors between individuals and of secular trends in error. It may be assumed that in this paper all technical precautions as detailed by previous authors have been taken. Since technical and instrumental errors are dealt with adequately by Douglas & Priestley (1924) and Peters & Van Slyke (1932), they are in principle not discussed in this paper.

https://doi.org/10.1017/S0022172400015035 Published online by Cambridge University Press

## EXPERIMENTAL METHOD AND RESULTS

All workers concerned in the gas analysis quoted in this paper were experienced in use of the apparatus, and the technique detailed by Haldane. Adjustments and readings were taken with a hand lens, and sampling tubes were tested for leaks prior to use. Estimates of error of method have been derived from the following sources.

(1) *Differences between duplicate mercury calibrations of the Haldane burette*

The accuracy achieved in mercury calibration may be regarded as the limiting accuracy of making readings similar to those involved in gas analysis with the instrument under optimal conditions. Such calibration eliminates the error of judgement in subdividing the smallest engraved division. All calibrations were done in the manner laid down by Haldane with the tap sealed to the bottom of the burette and without a telescope. Parallax was minimized by changing eye level so that the engraved calibration immediately below that to be read coincided with its reflexion in the mercury. Such a precaution may not always be used in routine work.

(A) *Duplicate calibration of dry burette*

In this case one instrument was calibrated dry with mercury by the same worker under similar conditions on two consecutive days. The standard deviation (S.D.) of the difference in volume between corresponding points on the two calibrations has been used to estimate the error.

The corrections required according to the two dry calibrations and also to the wet calibration of the same burette are shown in the graph in Fig. 1. Comparisons of derived volumes of the dry calibrations are available for 16 points over the range of the burette scale. The mean difference between the two calibrations is 0·0023 c.c. and the S.D. of the difference is 0·0014 ml. Using the '$t$' test for the significance of the difference,

$$t = 0{\cdot}0023 \times \frac{\sqrt{16}}{0{\cdot}0014} = 6{\cdot}57 \quad (15 \text{ D.F.}; P < 0{\cdot}1\%; \text{Sig.}).$$

Thus under optimum conditions there is a possible bias error in the same worker on different days, and this bias may alter by as much as 0·002 ml. within 24 hr. The mean laboratory temperature difference between the two days was less than 1·0° C., and the range of variation on each day less than 1·0° C., so that thermal expansion of the burette cannot account for this. Constant bias, however, does not alter the estimate of gas concentration, since it is derived from a difference in volume in which bias cancels out. Only if the bias alters considerably during the course of a determination will the calculated concentration be affected. In this set of readings there does not appear to be any reason for distinguishing between an altering bias and random error, since there is no significant correlation between consecutive difference between the two calibrations ($r = -0{\cdot}18$, not significant).

(B) *Difference between wet and dry calibrations*

Haldane recommended that the burette be calibrated under conditions of use, viz. moistened with dilute acid, but this is possibly not always done in practice. Owing to the thin film of moisture on the walls of the wet burette the two calibrations are substantially different and the difference increases with volume. It is reasonable to assume that this increase is linear, and if allowance be made for a linear component as well as a constant difference an estimate of standard deviation of the difference may be obtained from the residual variance. This may be regarded as a measure of the error of difference between the calibrations.

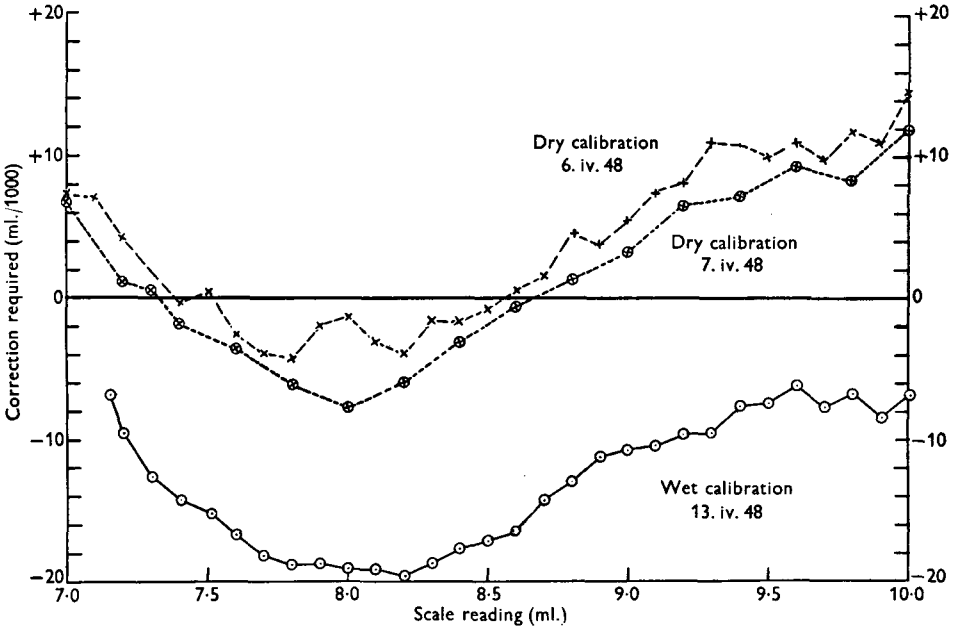


Fig. 1. Haldane gas analysis apparatus: No. 1. Upper, dry calibration. Lower, wet calibration.

The graph in Fig. 1 shows the calibration curves for two dry and one wet calibration of the same burette. The mean difference between the earlier (i.e. upper curve) dry calibration and the wet calibration was 0·0167 ml.; the difference increased by 0·0018 ml./ml., and the S.D. obtained from the residual variance was 0·00115 ml. (26 D.F.). If the difference between wet and dry calibration is extrapolated to zero volume, this corresponds (within the error of estimation) to zero difference between calibrations. This result is not self-evident, since the errors due to moisture in the bulb of the burette are not necessarily comparable to those in the narrow tubing of the calibrated scale. The consequence is, however, that the difference between calibrations is proportional to the volume. Gas concentration is calculated as a ratio of two volumes, and in this case by chance no error is introduced by using the dry calibration instead of the wet.

(C) *Duplicate calibrations of wet burette*

These were carried out before and after breaking and replacing of a tap at the top of a burette. Since this introduces a constant difference the standard deviation of difference gives an estimate of the error of calibration precisely analogous to that in (A).

The S.D. of the difference between the wet calibrations is 0·0012 ml. (31 D.F.). The serial correlation coefficient between successive volume differences is less than 2%, and not significant.

Each of A, B and C provides an estimate of error of measurement of volume difference. Gas concentrations are derived from measured volume differences divided by initial volumes, and since initial volumes of gas are usually about 10 ml. the S.D. of volume difference may be expressed in terms of percentage gas by multiplying by 10.

The agreement between the three estimates is extremely good; the variance ratio between greatest and least is 1·53 and the probability only slightly less than 20%. A combined estimate of variance gives a S.D. of difference corresponding to about 0·013% gas.

### (2) *Error of replicate determinations of combined carbon dioxide and oxygen and of carbon dioxide*

On a number of occasions replicate determinations of $CO_2$ were made and on one occasion of $CO_2$ and $O_2$. In each case the basis of comparison for estimating error has been the replicate determination of gas concentration from the same sampling tube.

The tubes have a relatively small volume (about 75 ml.), and the turbulence produced in taking samples by evacuated tubes allows a reasonable assumption that all samples from one tube have in fact the same composition. The volume of the tubes, however, restricts the number of determinations to three.

(A) *Combined oxygen and carbon dioxide determinations*

These were done on atmospheric air in order to check the calibration of the instrument as suggested by Haldane. Three determinations were made from each of three tubes by an exceptionally careful worker with many years experience with this apparatus. In this instance it has not been possible to apply a calibration correction, since the instrument involved was destroyed through enemy action. All nine estimations were performed on the same instrument, however, and the part of the scale used was roughly the same in each instance; consequently, it is unlikely that the difference in calibration corrections would substantially modify the differences between estimations, though possibly modifying the means considerably. At the low levels of $CO_2$ found in external air there is reason to believe that unconscious bias may play a very much more important part (see below, § (3)), and therefore these figures have not been included with the data.

The combined estimations have been subjected to analysis of variance (Fisher, 1941). The analysis is summarized in Table 1.

Table 1. *Analysis of variance*

| Source of variation | 9 × sum of squares | D.F. | 9 × mean square | Variance ratio | Significance |
|---|---|---|---|---|---|
| Between samples | 0·5078 | 2 | 0·2539 | 3·6 | Not sig. |
| Within samples | 0·4188 | 6 | 0·0708 | — | — |
| Total | 0·9266 | 8 | — | — | — |

S.D. from residual 'within samples' = 0·088 %.

The S.D. is very much higher than the total error of ± 0·02 % suggested by Haldane for $O_2$, in spite of the fact that these estimations were for checking calibration and carried out with special care. In contrast may be noted the figures for atmospheric $CO_2$, where bias is suspected. In a series of seven triplicate estimations done by the same worker the range was 0·03–0·05 % $CO_2$ ,and for any sampling tube the range was never more than 0·01 %.

(B) *Carbon dioxide estimations. Designed experiments*

These more recent data on the error of the method were obtained from designed experiments, and the estimations carried out under routine conditions. The data have been obtained from the work of one technician and are therefore strictly comparable to one another. Only estimations of $CO_2$ were done, since a large number of $O_2$ or combined $CO_2$ and $O_2$ determinations would have been too time-consuming to be completed in a day's work. Two sets of such experiments have been carried out, separated in time by about 10 months. In each set the design of the experiment was subject to the basic limitation that only three estimations could be done on the gas in one tube, and also to the condition in our particular case that only three instruments were in normal use.

(i) *Preliminary experiments.* These consisted of two groups of nine estimations of $CO_2$ of concentrations ranging from atmospheric air to 21·0 %. These nine estimations consisted of three from each of three sampling tubes, one estimation from each tube being carried out on each instrument. The technician was not aware of the source of air in the sampling tubes. The arrangement of estimates was a 3 × 3 Latin square, the rows corresponding to each of three instruments, the columns to each of three analyses (1st, 2nd, 3rd) and the letters to each of three tubes. In every case, the volume readings were recorded by an independent observer in order to minimize any bias due to knowledge of earlier results. Since three instruments were operated simultaneously, there is little likelihood of bias due to memorizing the initial volume. Calibration corrections were applied, and any remaining differences between instruments or between the orders of estimation eliminated by analysis of variance.

The analyses are summarized in Tables 2 and 3.

(ii) *Later experiments.* Subsequently a further set of three experiments was carried out in order to obtain an estimate of error for the instrument based on a larger number of observations and to determine whether there was any variation of error with time. The total range of $CO_2$ concentration covered was atmospheric to 14·0 %. In each case three different sources of air were used, three samples from each source were taken and three estimations done on each sample. Thus in each

experiment there were nine tubes and twenty-seven estimations in all. The first two experiments were designed to minimize bias of replication—'prejudice error'—on the part of the technician. The third experiment was done under routine conditions and bias given free play. The technician and observers were aware of the nature of all the experiments, and in this sense the last experiment was not 'normal'; but it is not considered that the technician was under any obvious sense of stress.

Table 2. $3 \times 3$ *Latin square* (1); *analysis of variance*

| Source of variation | $9 \times$ sum of squares | D.F. | $9 \times$ mean square | Variance ratio |
|---|---|---|---|---|
| Between instruments | 0·2474 | 2 | 0·1237 | $<1$ Not sig. |
| Between estimations | 0·3662 | 2 | 0·1831 | $\simeq 1{\cdot}3$ Not sig. |
| Between tubes | 7106·2478 | 2 | 3553·1239 | $>10{,}000$ Sig. 0·1 % |
| Residual | 0·2786 | 2 | 0·1393 | |
| Total | 7107·1400 | 8 | | |

From residual S.D. = 0·13 % approx.; 2 D.F.
If all non-significant variances are combined S.D. = 0·13 % approx.; 6 D.F.

Table 3. $3 \times 3$ *Latin square* (2); *analysis of variance*

| Source of variation | $9 \times$ sum of squares | D.F. | Mean square | Variance ratio |
|---|---|---|---|---|
| Between instruments | 0·3744 | 2 | 0·1872 | $\simeq 1{\cdot}4$ Not sig. |
| Between estimations | 0·2808 | 2 | 0·1404 | $\simeq 1{\cdot}05$ Not sig. |
| Between tubes | 760·5414 | 2 | 380·2707 | $>1000$ Sig. 0·1 % |
| Residual | 0·2664 | 2 | 0·1332 | |
| Total | 761·4630 | 8 | | |

From residual S.D. = 0·12 % approx.; 2 D.F.
If all non-significant variances are combined S.D. = 0·13 % approx.; 6 D.F.

*Sampling.* When external air was used, three samples were taken by three people standing several yards apart across the wind outdoors in the country and well away from any building. The sampling tubes were held overhead and contamination by expired air minimized by holding the breath for 15 sec. before sampling. For higher concentrations of $CO_2$, the mixture was prepared in a Douglas bag, which was thoroughly mixed. Samples were taken from a side tube after this had been thoroughly flushed out by the mixture.

*Order of sampling and elimination of bias.* In the first two experiments it was desired to reduce bias to a minimum and at the same time to employ a design of experiment so that the results would be suitable for treatment by variance analysis. As in the earlier experiments bias of replication was minimized by the recording of all volume readings by an independent observer. The nine tubes in each experiment were numbered at random and the technician given a table indicating the order in which estimations were to be done and the instruments to be used in each case. This table was arranged so that it could be split up in a number of different ways. If it is to be regarded as a 'two-way' table consisting of three columns corresponding to the three instruments and nine rows corresponding to the order of the estimation, then each column contains one estimation

from each tube, and each group of three rows contains one estimation from each tube. Furthermore, within any of these groups of three rows, each column contains one estimation from each bag. Subject to these restrictions assignment was random. Each group of three rows and three columns therefore constitutes a $3 \times 3$ Latin square in which rows correspond to the order of estimation, columns to instruments, and letters to bags.

In variance analysis of these data the differences between the three Latin squares ('between squares' in Tables 4 and 5) indicates any tendency for the $CO_2$ in the tubes to change with time. Differences between bags, between tubes and between instruments are assessed directly. The 'bags × instruments interaction' measures the extent to which the differences between the bags are characteristic of the instruments on which estimations are made. Since the different bags correspond to different concentrations, a significant 'bags × instruments interaction' would indicate an instrument error varying with concentration, due perhaps to calibration or to differences in lighting.

*Order of analysis in experiment where bias was given full scope.* In the third experiment samples were obtained in a similar way, but the technician knew the source of the samples and was responsible for recording all figures. In his arrangement of the order of the analyses he was subject only to the restriction that the gas in each tube was estimated once and once only in each of the three instruments. Adequate scope for the normal bias of replication in routine work was present.

Variance analyses are given in Tables 4–6. The S.D. from the three analyses, after the various effects have been eliminated by the analysis, are in every instance considerably greater than that quoted by Haldane. Moreover, the S.D. of the estimation on day 2 is significantly greater than that of either day 1 or day 3, although days 1 and 3 agree with one another tolerably well. These results may be summarized by means of the variance ratios:

$$\frac{\text{Day 1}}{\text{Day 3}},\ F = 1{\cdot}86;\ n_1 = 10,\ n_2 = 16;\ \text{Not sig.}$$

$$\frac{\text{Day 2}}{\text{Day 1}},\ F = 3{\cdot}34;\ n_1 = n_2 = 10;\ \text{Sig. } P = 5\,\%.$$

$$\frac{\text{Day 2}}{\text{Day 3}},\ F = 6{\cdot}20;\ n_1 = 10,\ n_2 = 16;\ \text{Sig. } P = 0{\cdot}1\,\%.$$

From these data it appears that a number of conclusions can be drawn:

(*a*) In no instance is the 'between squares' mean square significantly greater than the residual, which indicates that there is no evidence for a systematic change with time of measured concentration of $CO_2$ from a tube.

(*b*) In one instance, the difference between instruments is significant in spite of the use of calibration corrections, indicating that characteristic instrument errors may sometimes creep in after applying the most careful corrections. Such errors might be due to differences in, say, lighting, but any such mechanism attributed to them is purely speculative. On the other hand, the difference may not be real, and the statistical significance is perhaps due to an 'Error of the second kind' (Neyman & Pearson, 1928).

Table 4. *Analysis of variance: day* 1

| Source of variation | 27 × sum of squares | D.F. | 27 × mean square | Variance ratio |
|---|---|---|---|---|
| Between squares | 0·0344 | 2 | 0·0172 | < 1 Not sig. |
| Between instruments | 0·6098 | 2 | 0·3049 | 5·06 Sig. 5 % |
| Between bags | 1561·7258 | 2 | 780·8629 | > 12,000 Sig. 0·1 % |
| Between tubes within bags | 0·6576 | 6 | 0·1096 | 1·82 Not sig. |
| Bags × instruments interaction | 0·4564 | 4 | 0·1141 | 1·89 Not sig. |
| Residual by subtraction | 0·6022 | 10 | 0·0602 | |
| Total | 1564·0862 | 26 | | |

Variance based on residual = 0·00233; hence S.D. = 0·047 %, 10 D.F.
Variance based on residual and non-significant items
= 0·00295; hence S.D. = 0·054 %, 22 D.F.

Table 5. *Analysis of variance: day* 2

| Source of variation | 27 × sum of squares | D.F. | 27 × mean square | Variance ratio |
|---|---|---|---|---|
| Between squares | 0·4146 | 2 | 0·2073 | 1·03 Not sig. |
| Between instruments | 0·5634 | 2 | 0·2817 | 1·40 Not sig. |
| Between bags | 22632·4872 | 2 | 11316·2436 | > 55,000 Sig. 0·1 % |
| Between tubes within bags | 0·4266 | 6 | 0·0711 | < 1 Not sig. |
| Bags × instruments interaction | 0·3168 | 4 | 0·0792 | < 1 Not sig. |
| Residual by subtraction | 2·0100 | 10 | 0·2010 | |
| Total | 22636·2186 | 26 | | |

Variance based on residual = 0·00744; hence S.D. = 0·086 %, 10 D.F.
Variance based on residual and non-significant items
= 0·00576; hence S.D. = 0·076 %, 24 D.F.

Table 6. *Analysis of variance: day* 3

| Source of variation | 27 × sum of squares | D.F. | 27 × mean square | Variance ratio |
|---|---|---|---|---|
| Between instruments | 0·0344 | 2 | 0·0172 | < 1 Not sig. |
| Between bags | 16828·5314 | 2 | 8414·2657 | > 270,000 Sig. 0·1 % |
| Between tubes within bags | 0·2370 | 6 | 0·0395 | 1·22 Not sig. |
| Residual by subtraction | 0·5164 | 16 | 0·0323 | |
| Total | 16829·3192 | 26 | | |

Variance based on residual = 0·00120; hence S.D. = 0·035 %, 16 D.F.
Variance based on residual and non-significant items
= 0·00122; hence S.D. = 0·035 %, 24 D.F.

(*c*) In no instance is the 'between tubes within bags' mean square significantly greater than the residual mean square. This appears to indicate that the technique of sampling whether from bags or more directly as from the atmosphere is satisfactory, within the error of estimation.

(*d*) In no instance is the 'bags × instruments interaction' mean square significantly greater than the residual mean square. This indicates that there are no

systematic differences between calibrations of the instruments varying between different parts of the scale.

(*e*) The ratios of the standard deviations estimated either from the residuals or from the sums of all the non-significant items in the analyses indicate marked and significant variability in error from day to day. It appears possible, though it is not statistically proven, that in both instances in which bias tendency was eliminated the residual error was greater than in the instance in which it was not.

(*f*) In every instance the residual error expressed as one standard deviation was very much greater than the limits of error tolerated by Haldane (from 2·5 to 13 times as great).

### (3) *Examination of terminal digits in readings of gas volume*

The Haldane gas analysis apparatus has a burette of 10 ml. graduated from 7 to 10 ml. in units of 0·01 ml. Consequently one division corresponds to 0·1 % gas, and if the instrument is to measure gas concentrations to 0·01 % the technician must divide these smallest divisions corresponding to about 1 mm. in width into ten parts by eye. If, in fact, this is being done accurately, the distribution of terminal digits of his readings of volume should be equally divided (apart from random statistical variation) between the possibilities 0–9 (Myers, 1906; Yule, 1927).

The departmental laboratory note-books of three experienced workers were examined, and large numbers of figures spread over a considerable period obtained. In all instances the work involved was routine, and the workers were not aware that their figures were to be subject to scrutiny. Random statistical variation was tested for significance by means of the $\chi^2$ test.

In the case of two workers, the figures have been divided up according to the time and circumstances of the estimations involved (see Figs. 2–4). The individual G (Fig. 2) has a tolerably even distribution of digits among the possibilities, although even in this case the deviation from the expectation of equal divisions is very highly significant ($\chi^2 = 62{\cdot}5$, 9 D.F.; $P < 0{\cdot}1\,\%$). Operator F has a marked (significant) preference for 0 and 5 (see Fig. 3). Moreover, the predominance of these figures is much greater in routine estimations than in triplicate estimations on atmospheric air, in which there is an attempt at special accuracy and also a tendency towards bias.

If the 0's and 5's are grouped together in each case, the results may be tabulated as follows:

| | Routine estimations | Triplicate estimations on external air | Total |
|---|---|---|---|
| 0 or 5 | 339 | 90 | 429 |
| Other digits | 71 | 63 | 134 |
| Total | 410 | 153 | 563 |

Using 'tests of significance for 2 × 2 contingency tables' (Yates, 1934; and Fisher & Yates, 1948), we find

$$\chi_c = 5{\cdot}8 \text{ (approx.)}; \; m = 36{\cdot}5; \; p = 0{\cdot}27; \; P < 1\,\%, \text{ Sig.}$$

There is hence a strong indication that the accuracy of this worker varies from time to time, possibly with the nature of the work involved. Moreover, in estimations of $CO_2$ in atmospheric air there appears to be evidence of bias by expectation

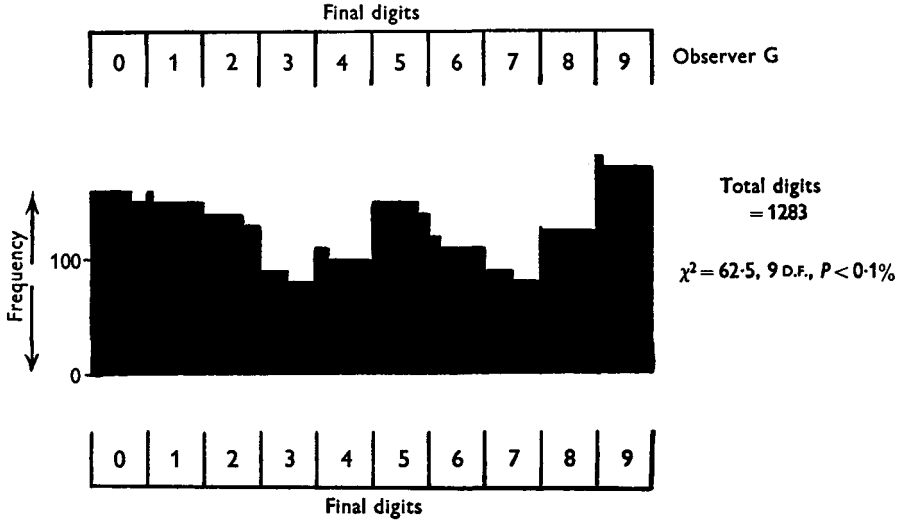


Fig. 2. Readings of Haldane burette frequency distribution of final digits (1).

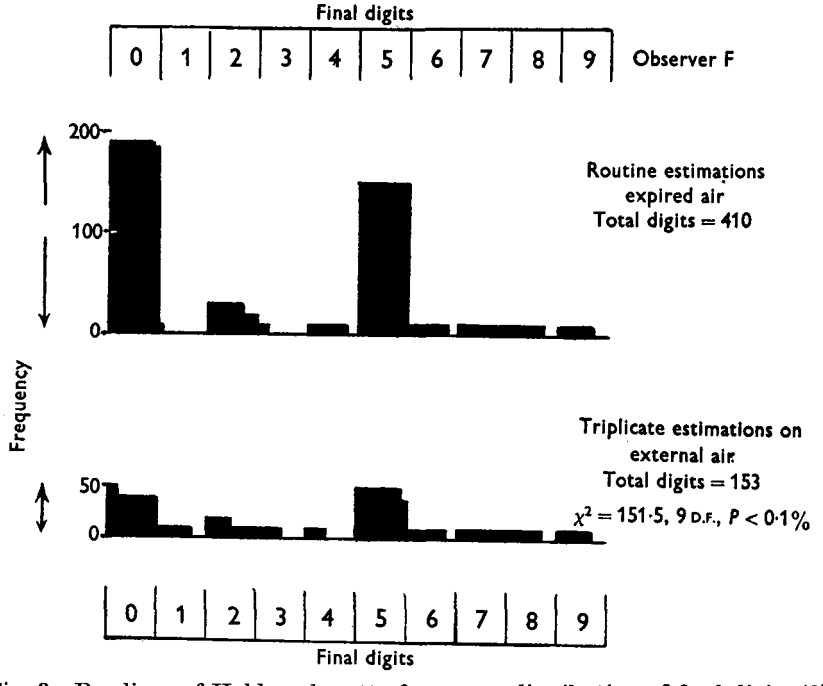


Fig. 3. Readings of Haldane burette frequency distribution of final digits (2).

of particular results. In this case the difference between the initial and final volume is so small that division of the initial volume (of approximately 10 ml.) and subsequent multiplication by 100 to convert to percentage is equivalent to multiplying by ten only. Consequently since the 'expected result' of about 0·03 % $CO_2$ is well

known there is a bias towards reading a change in volume of about 0·003 ml. In the reading of initial volume, and of final volume after $O_2$ absorption (in which the volume difference is much greater and the relationship is not so simple unless the initial volume is exactly 10 ml.), there is not the same likelihood of bias. To illustrate

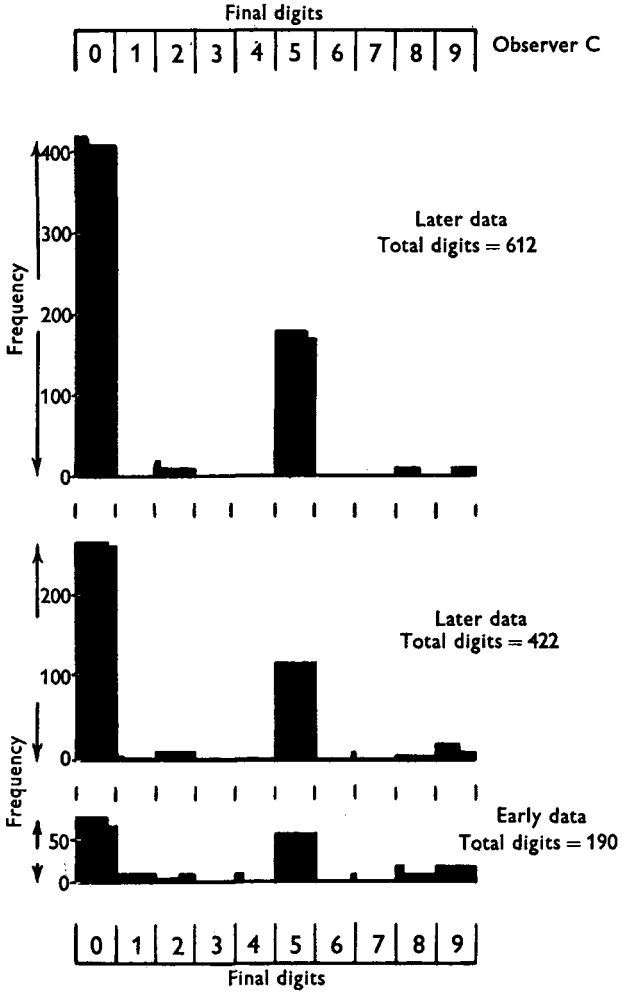


Fig. 4. Readings of Haldane burette frequency distribution of final digits (3). Readings of volume obtained from laboratory note-books have been classified according to their final digits. The histograms have been derived from this classification for various groups of data from three individuals.

this, twenty-one estimations of combined $CO_2$ and $O_2$ in atmospheric air were examined. The $CO_2$ concentrations ranged from 0·03 to 0·05 %, a very much narrower range than would be expected from data with as high a standard deviation as indicated in § (2) above. The distribution of final digits was:

| | $CO_2$ only | $O_2$ and $CO_2$ | Total |
|---|---|---|---|
| 0 or 5 | 6 | 25 | 31 |
| Other digits | 15 | 17 | 32 |
| | 21 | 42 | 63 |



https://doi.org/10.1017/S0022172400015035 Published online by Cambridge University Press

Applying the 'tests of significance for $2 \times 2$ contingency tables' (see above)

$$\chi_c = 2{\cdot}04;\ m = 10{\cdot}3;\ p = 0{\cdot}49;\ P < 5\%;\ \text{Sig.}$$

It would seem, therefore, that there is bias towards 'expected' results.

Operator C has an even more marked preference for 0 and 5, and it increases with time. In the later data his preference for 0 rather than 5 is apparent. It is clear that none of these workers can have attained the accuracy claimed by Haldane. There are marked differences in accuracy between workers, and for any one worker there are variations with time and with the accuracy which is regarded as necessary.

### (4) *Additional minor sources of data*

(A) Six observers made the necessary settings and readings for measuring the volume of a fixed quantity of gas in a Haldane gas burette. The settings were grossly maladjusted between each setting, and the observers were unaware of each other's results. This was done on six different volumes of gas and the thirty-six readings subjected to analysis of variance to determine the mean differences between observers and the random error.

The analysis is summarized in Table 7. The standard deviation of a single volume reading corresponds to 0·024 % gas. Since gas concentrations are estimated from differences in volume, the S.D. of gas estimation is $0{\cdot}024\ \sqrt{2} = 0{\cdot}034\%$ gas.

Table 7. *Analysis of variance: readings of Haldane burette by different observers*

| Source of variation | 36 × sum of squares | D.F. | 36 × mean square | Variance ratio |
|---|---|---|---|---|
| Between observations | 29·301648 | 5 | 5·8603296 | > 25,000 Sig. 0·1 % |
| Between observers | 0·000936 | 5 | 0·0001872 | < 1 Not sig. |
| Residual | 0·005088 | 25 | 0·0002035 | — |
| Total | 29·307672 | 35 | — | — |

Hence variance from residual = 0·00000565; S.D. = 0·0024 ml.; 25 D.F.

(B) Two observers went through the process of routine analysis for $CO_2$ of the same three samples of air simultaneously. Both took the readings of volume in every case, but were unaware of each other's figures. After the absorption of $CO_2$ appeared to be complete the process of passing over to fresh potash was repeated fifteen times for about three minutes each time, in order to detect whether there was any error due to slow absorption of the final traces of $CO_2$.

The results are illustrated in Fig. 5. In the case of instrument no. 3, there is no clearly defined trend during the course of the measurements, but the range of variation in the case of one observer is 0·006 ml. and in the case of the other 0·009 ml. This would have a marked effect on calculated $CO_2$ concentration. In the case of instruments nos. 1 and 2 a trend appears after a time. This trend is certainly due to a leak of some sort, but it might go undetected in routine work. In the case of instrument no. 2 the trend is slow and consecutive readings agree well. They may even agree exactly owing to experimental error or to subconscious bias of replication. It is difficult to state at what point the random error was superseded by the systematic 'leak' error.

(C) The line thickness of the engraved scale on the burette introduces an error whose magnitude is indicated by the thickness of the lines expressed as a fraction

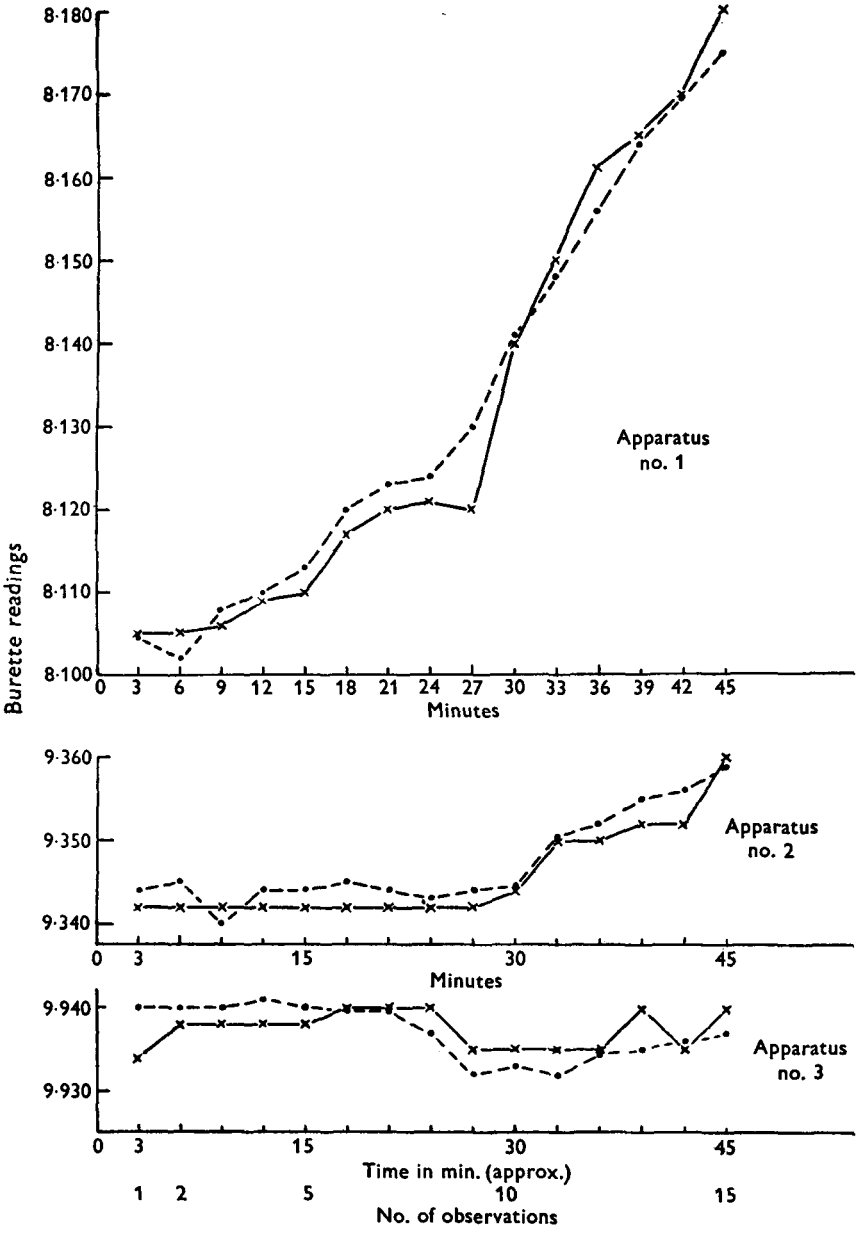


Fig. 5. Changes in reading Haldane burette with two observers. Consecutive readings of gas volume by two observers after repeated absorption of $CO_2$ at 3 min. intervals. The absorption was considered complete at time 0.

of the scale length. The thickness of forty adjacent calibration lines on each of two burettes were measured. In one case the line thickness was 10% of the total length of scale, and in the other case 14%, corresponding to 0·010 and 0·014% gas.

https://doi.org/10.1017/S0022172400015035 Published online by Cambridge University Press

## DISCUSSION

The errors to which any measurement is subject may be roughly divided into three classes:

(1) Errors in design and construction of the instrument.

(2) Errors in the technique of the worker including systematic errors.

(3) Subjective and psychological factors, including a bias towards a particular result, learning and 'warming up' factors, fatigue, haste, boredom and the physiological state of the individual.

In any statement of error of a method it is important that all of these be taken into account, and consequently that the estimate of error be derived from measurements obtained under conditions as similar to those under discussion as possible. Clearly the ideal is to obtain the estimate of error from the very figures involved, but this is not always possible; moreover, this method may not enable the detection or elimination of bias error and consequently may conceal a very important factor in certain types of experiment. The next best alternative is to obtain an estimate of error from experiments designed to eliminate it under conditions of routine use after all controllable factors have been minimized or eliminated. However, the very complexity of such a planned experiment may be another psychological factor influencing the accuracy of the technician and so defeat its own ends. Whilst the use of automatic recording instruments may reduce subjective and psychological types of error, it would be of interest to know whether such forms of measurement introduce errors of their own.

The experimental work in this paper was intended to provide such an estimate of error for the 10 ml. Haldane gas analysis apparatus, and to detect the principal sources of error. Haldane (1898, 1912) himself attributed to this apparatus an error of $\pm 0{\cdot}01\,\%$ for $CO_2$ and $\pm 0{\cdot}02\,\%$ for $O_2$ (the figures represent a range and not a standard deviation), and these figures are quoted by most recent authorities. Gemmill (1931) gives a frequency histogram of differences between duplicates for $CO_2$ and $O_2$ estimations which are somewhat larger than those quoted by Haldane, but the differences are appreciably less than those found by us in designed experiments.

If calibration errors are neglected the main errors may be classified as follows:

| Nature of error | Source of data | Error if expressed as S.D. (% gas) | Error if expressed otherwise (% gas) |
|---|---|---|---|
| Instrumental | Difference between calibrations | 0·013 % | — |
| | Mean line thickness | — | 0·010–0·014 % |
| Technique + instrumental | Sampling: Mixing | — | Negligible |
| | Undetected leaks in system | — | Not known |
| | Error replicate setting and reading burette | 0·034 % | — |
| Subjective + technique + instrumental | Replicate determination $CO_2$ and $O_2$ | 0·027–0·13 % | — |
| | Final digits | — | Up to 0·05 % |

https://doi.org/10.1017/S0022172400015035 Published online by Cambridge University Press

This subdivision of errors is to a considerable extent arbitrary, but not grotesquely so, and it provides a basis for discussion. It makes clear that any improvement of the instrument, particularly in finer engraving of lines, is not likely to improve the performance of the instrument as used. In view of the use of the compensating thermo-barometer in the Haldane apparatus the total error seems to be far larger than can be adequately explained by changes of vapour pressure or of gas absorption with temperature. It is probable, however, that a considerable part of the technical and subjective error is due to parallax, the observer subconsciously moving the head to a position which gives the desired result, and clearly an improvement in the instrument to the extent of providing a telescope will greatly reduce errors of this nature.

The largest and most variable component of error appears to be psychological in nature. In some cases, as in the experiment of day 2 summarized in § (2)(B) (ii), this factor does not appear to be present at all, and the accuracy obtained is substantially that of technical and instrumental error alone. In other cases, as in § (2) (B) (i), it is far larger than the other errors combined, giving a total S.D. of 0·13 % gas. A number of features stand out, however:

(1) The accuracy of an observer varies markedly from time to time under similar conditions of work. This is shown by data obtained from designed experiments (§ (2)(B)(i) and (ii)) and from examination of terminal digits (§ (3)), and variation may be present or absent over a period as short as one day § (2)(B)(ii), days 1 and 2).

(2) The changes in accuracy are not necessarily associated with improvement with time (vide § (3)).

(3) There is a tendency to be more accurate with work believed to be of greater importance (vide § (3)).

(4) There is a tendency to read the scale so as to obtain an 'expected result' by subconscious bias (vide § (2)(A) and § (3)). The latter may also influence the readings of volume after gas absorption and thereby mask errors due to leaks (vide § (4)(B)).

The fact that even experienced individuals vary between themselves in their experimental measurements was pointed out by Maskelyne for astronomical observations as early as 1799, and it is now common knowledge that such variation occurs in all fields of measurement. Although secular or intra-individual trends in error are apparently less well known, a reference to them was made by Maskelyne (1799) and they are fully described by Pearson (1901), Pearson (1922) and by Tocher (1926) for chemical analysis. Such trends may be diurnal, day to day, seasonal or even extend over a period of years, and both Tocher (1926) and Pearson (1922) noted a change in error during the period of a day. It is possible that physiological rhythms over these periods of time (Renbourn, 1947) may play some part. Kleitman (1939) has shown that the efficiency curves for various tasks runs parallel to the diurnal curve of body temperature for a particular individual. MacFarlane (1945) and Wiehl (1946) mention intra-individual variations of error in haematological work, and such variations are discussed more fully by Jones (1938) for clinical nutritional assessment of children and by Birkelo,

Chamberlain, Phelps, Schools, Zachs & Yerushalmy (1947) for clinical radiological diagnosis.

The accuracy of the Haldane apparatus depends in great part upon the supposed ability to divide a scale division of about 1 mm. width into ten equal parts. This may be possible with a telescope, but our results show that with experienced workers using a hand lens as suggested by Haldane such a division is not accurate. The explanation lies partly in subconscious bias towards certain numbers which varies from individual to individual and which may itself show secular trends. The thickness of the calibration line produces a small error corresponding to that found by Haldane for the whole method. It agrees tolerably well with the error of burette calibration (vide § (1)).

In the routine estimation of error of a method, duplicates or replicates are commonly used. The former are frequently taken as routine and hence over a longer period of time. They may for this reason be a better index of error in routine analysis. MacFarlane (1945) and Biggs & MacMillan (1948) have both shown that unconscious bias plays a part in the high level of consistency shown by some experienced technicians in haematological work. However, if the technician is not aware of his previous results there is usually an appreciable fall in accuracy. Our own work fits in well with such a finding, and it is possible that errors of this type occur in most forms of measurement, particularly where these are subjective.

The evidence in this paper suggests that in our hands the 10 ml. Haldane gas analysis apparatus has not the sensitivity of measuring $CO_2$ concentrations below 0·5 % to within an error of 10 %, and neither the $CO_2$ level nor the combined $CO_2$ and $O_2$ level in atmospheric air can be regarded as a suitable index of the calibration of the burette or the accuracy of the technician. Nevertheless, Haldane, an extremely careful worker, laid down the value of such estimations, and the statement is quoted in all standard texts dealing with the method.

The inability of most recent workers to repeat the accuracy of Haldane in the use of the CO haemoglobinometer (MacFarlane, 1945), and ourselves in the use of the gas analysis apparatus, may be due to less refined apparatus or techniques in present-day laboratory work, but we agree with MacFarlane that a possible explanation lies in the lack in the past of appreciation of error due to bias. Biggs & MacMillan suggest that bias of replication is due in part to the nature of training of the laboratory worker. A standard authority suggests a range of error not to be exceeded by a good technician. With experience this is reached by many workers. A bias or 'prejudice' towards consistency or 'accuracy' has been introduced, and only designed experiments will show its presence.

This paper is concerned with the errors of the Haldane apparatus as a precision instrument. For physiological work and for the estimation of $CO_2$ and $O_2$ levels as found in expired and alveolar air samples the error is nevertheless well within the order found in most standard laboratory methods.

SUMMARY

1. The paper represents an attempt to estimate the error of the standard 10 ml. Haldane gas analysis apparatus under routine conditions of use.

2. The ultimate error of the instrument under optimum conditions has been obtained from duplicate mercury calibrations of the same instrument and corresponds to S.D. = 0·013 % gas. A significant day-to-day variation in calibration was found.

3. The accuracy in routine use has been determined by replicate estimations of $CO_2$ and $O_2$ in air and from estimation of $CO_2$ in designed experiments. The data show an error corresponding to S.D. = 0·027 to 0·13 %. This is far larger than that described by Haldane.

4. Examination of the distribution of terminal digits in burette readings shows that the smallest scale division is in fact not being divided into ten parts by the worker. Preference for certain digits varies with the individual, with time, and with the accuracy required.

5. There is a marked tendency for a worker to obtain an 'expected result', whether this is derived from a previous estimation of the same sample or from knowledge of the expected result; such bias tends to reduce the apparent error of estimation. Bias of replication—'prejudice error'—may play a part in many forms of measurement.

6. The results and conclusions are discussed in relation to other published data on errors of measurement in general.

We wish to thank Mr W. R. Luxton for advice in a number of the experiments; Mr W. R. Hindes for the preparation of the figures and tables. We are indebted to Prof. G. P. Crowden, in whose Department the experimental work was done, for constant encouragement and for the provision of a considerable part of the data from departmental files.

REFERENCES

Biggs, R. & MacMillan, R. L. (1948). *J. clin. Path.* **1**, 269, 288.
Birkelo, C. C., Chamberlain, W. E., Phelps, P. S., Schools, P. E., Zachs, D. and Yerushalmy, J. (1947). *Jour. Amer. med. Assoc.* **133**, 359.
Carpenter, T. M., Fox, E. L. & Sereque, A. F. (1929). *J. biol. Chem.* **83**, 211.
Daynes, H. A. (1920). *Proc. Roy. Soc.* A, **97**, 273.
Dingle, E. H. & Pryce, A. W. (1940). *Proc. Roy. Soc.* B, **129**, 468.
Douglas, C. G. & Priestley, J. G. (1924). *Human Physiology.* Oxford.
Fisher, R. A. (1941). *Statistical Methods for Research Workers.* Edinburgh.
Fisher, R. A. (1942). *The Design of Experiments.* Edinburgh.
Fisher, R. A. & Yates, F. (1948). *Statistical Tables*, p. 5 and table 8.
Gemmill, C. L. (1931). *Amer. J. Physiol.* **98**, 135.
Grove-White, C. W. & Sander, A. G. (1949). (Personal communication.)
Haldane, J. S. (1898). *J. Physiol.* **22**, 465.
Haldane, J. S. (1912). *Methods of Gas Analysis.* London.
Hawk, P. B., Oser, B. C. & Summerson, W. H. (1947). *Practical Physiological Chemistry*, London.
Jones, H. R. (1938). *J. R. Statist. Soc.* **101**, 1.
Kleitman, N. (1939). *Sleep and Wakefulness.* Chicago.

KROGH, A. (1920). *Biochem. J.* **14**, 267.
LOWELL-OLSEN, H. (1948). *Rep. Univ. Wisconsin. C.M.* 514.
MACFARLANE, R. G. (1945). *Spec. Rep. Ser. med. Res. Coun., Lond.*, no. 252, p. 16.
MACLEOD, J. J. R. (1941). *Physiology in Modern Medicine.* London.
MASKELYNE, N. (1799). *Astronomical Observations*, p. 333.
MYERS, C. S. (1906). *J. R. anthrop. Inst.* **36**, 255.
NEYMAN, J. & PEARSON, E. S. (1928). *Biometrika*, **20** A, 175.
PEARSON, E. S. (1922). *Biometrika*, **14**, 23.
PEARSON, K. (1901). *Philos. Trans.* **198**, 253.
PETERS, J. P. & VAN SLYKE, D. D. (1932). *Quantitative Clinical Chemistry*, **2**. *Methods.* London.
REIN, H. (1943). *Schr. dtsch. Akad. LuftfahrtForsch.* **7**, 73.
RENBOURN, E. T. (1947). *J. Hyg., Camb.*, **45**, 456.
RENBOURN, E. T., ANGUS, T. C., ELLISON, J. MCK. & JONES, M. S. (1949). *J. Hyg., Camb.*, **47**, 1.
THORPE, J. F. & WHITELEY, M. A. (1938). *Thorpe's Dictionary of Applied Chemistry.* London.
TOCHER, J. F. (1926). *Analyst*, **51**, 338.
WIEHL, D. C. (1946). *Millbank Mem. Fund Quart.* **24**, 5.
YATES, F. (1934). *J. R. Statist. Soc.* Suppl. **1**, 217.
YULE, G. U. (1927). *J. R. Statist. Soc.* **90**, 570.

(*MS. received for publication* 25. III. 50.)