VOLUME 43, No. 5 APRIL 1944

A STATISTICAL EXAMINATION OF THE ACCURACY OF VITAMIN A ASSAYS

AN ANALYSIS OF THREE CO-OPERATIVE EXPERIMENTS DESIGNED TO ASCERTAIN THE VALUE OF THE CONVERSION FACTOR FOR TRANSFORMING SPECTROPHOTOMETRIC VALUES INTO INTERNATIONAL UNITS

BY J. O. IRWIN, *Of the Statistical Staff, Medical Research Council*

INTRODUCTION

The three experiments with which this report deals were undertaken with the object of re-examining the conversion factor 1600 which had been provisionally allotted by the Second International Conference on Vitamin Standardization (1934) for converting the results of spectrophotometric tests for vitamin A into international units.

The participants were A. L. Bacharach and E. L. Smith, Glaxo Laboratories, Ltd.; R. G. Booth and W. Stott, British Cod-Liver Oil Producers; W. A. Broom, Boots Pure Drug Co., Ltd.; H. M. Bruce and K. H. Coward, Pharmaceutical Society; A. D. Emmett, O. D. Bird, R. A. Brown, and M. Sturtevant, Messrs Parke Davis and Co., Detroit Laboratories; E. M. Hume, Lister Institute; K. M. Henry and S. K. Kon, National Institute for Research in Dairying; T. Moore, Dunn Nutritional Laboratory; R. A. Morton and J. R. Edisbury, Liverpool University; S. W. F. Underhill, Messrs British Drug Houses, Ltd.; H. Wilkinson, Messrs Lever Brothers, Ltd.

The result of the first experiment was very shortly reported by the secretary of the Vitamin A Sub-Committee (Hume, 1937). It may be described briefly as a collaborative experiment in which workers in nine different laboratories determined the vitamin A content of (*a*) a sample of halibut-liver oil, and (*b*) a concentrate prepared from the same oil, by simultaneous biological tests on these and the International Standard.

Determinations of the spectrophotometric values of the oil and concentrate at $E^{1\%}_{1\,cm.}$ 325 mμ* were made by R. A. Morton and J. R. Edisbury of the University of Liverpool on the original materials, and, at the end of each worker's experiment, on the remainders of the samples which had actually been used in the feeding test. Variation in technique in the biological tests was eliminated as much as possible, but workers were asked to use their usual technique in regard to strain of animals, basal diet, temperature of rat room, etc.† To facilitate calculations, two doses each of oil, concentrate and standard in the ratio 2 : 1 were tested.

The result of this experiment indicated a value of 1470 for the factor to be used to convert into international units vitamin A values of potent oils determined spectrophotometrically from the standard. Later a recalculation of the results by Irwin gave a conversion factor of 1570. A lower value for concentrates was obtained, but the spectrophotometric values of the residues were found to have fallen seriously; it is probable that a true average E value of the concentrate during the period of dosing in each laboratory would have yielded, with the corresponding biological value, a conversion factor similar to that above. Accordingly only the statistical analysis of the halibut-liver oil is included in this report.

The analysis was performed by the best methods available at the time (Irwin, 1937). These methods were adequate for determining an average estimate of the conversion factor and its error by combining the data from all laboratories. This is shown by the fact that the error of the final result was found to be substantially the same when calculated by two different methods, first by averaging the errors obtained from the internal evidence of the experiments in the different laboratories and secondly by considering the variation of the results from one laboratory to another. When the data of the halibut-liver oil experiments were first examined statistically, it was found that the workers in the different laboratories did not all use the same method of calculating their results. Part I of this report, therefore, contains sections dealing with the influence of method of calculation on the result and with unusually aberrant tests. It was unnecessary to repeat these sections when dealing with the two later experiments in Parts II and III.

The second experiment was arranged to investigate a particular oil which had already been much examined both biologically and spectrophotometrically. This was the U.S.P. reference oil for which a conversion factor of 1900–2000 had been reported. The same precautions were taken as in the first experiment. The mean value for the vitamin A potency calculated from the biological results by the present author was 2619 i.u./g., and the value for $E^{1\%}_{1\,cm.}$ 325 mμ was 1·44. Therefore the conversion factor was 1820. Irwin found that the odds against the possibility of the difference between this factor and the factor 1570 being due to chance were about 30 to 1 (Hume, 1939).

A third experiment was therefore arranged when a

* For the benefit of the non-technical reader: $E^{1\%}_{1\,cm.}$ is the extinction coefficient when a 1% solution is used with a cell of 1 cm. thickness. The extinction coefficient is the logarithm of the ratio of the intensity of the incident to that of the emergent light. The E value is multiplied by the conversion factor to obtain the result in international units per gram.

† Details of technique can be found in papers published by the participants. For the general plan, see p. 312.

vitamin A ester became available—solid vitamin A 2-β-naphthoate being used. The conversion factor found in this experiment was 1770 (Hume, 1943).

The third value lies between the first two, and a collective significance test for the three factors 1570, 1820 and 1770 shows that they do not differ significantly, the pooled estimate being 1740 with limits of error ($P=0{\cdot}99$) of 93–107%. This is the answer to the question for which the experiments were designed. Part IV deals with the conversion factor.

It seemed, however, that advantage should be taken of this unique collection of data to make a thorough statistical analysis of the variability of the animals and of the estimates of slope of the dosage-response curves obtained from the different laboratories. An important conclusion can be drawn from this analysis. In calculating the error of a result in an individual laboratory, proper attention must be paid to the error of the slope. Some workers in the general field of biological assay have been in the habit of using a slope from a dosage-response curve constructed in advance on the basis of a good deal of past experience, modifying it from time to time as experience accumulates. Others have used an estimate of slope obtained from the individual test itself. There is no doubt that the latter is the correct procedure. The trouble with the former is that large changes of slope may occur with time, and that these changes do not occur in a random fashion but exhibit a secular or quasiperiodic tendency. The use of a fixed slope therefore introduces an uncontrolled error for which it is impossible to allow. Even if regular records of slope are kept, and their standard deviation over a considerable period of time calculated, this does not get over the difficulty. For the variations are not random, and so this standard deviation gives no clue to the error of the slope at any particular time.*

Unfortunately, if a test be performed with a limited number of animals, for example twenty, the estimate of slope obtained will have a large error and this will introduce a large error into the result.† In such cases the approximate formula for the error of the result which, at any rate until very recently, it has been customary to use and which purports to take account of the error of the slope, itself greatly underestimates the error. It is also biased unless the average responses to the test and standard preparations are the same. The first point is obvious when we consider that the approximate formula still gives finite limits of error when the slope is *not* significantly different from zero. Yet a zero slope implies no increasing response to increased dosage, and nothing can be inferred from the sole consideration of such an assay. This difficulty can be overcome by the calculation of what are known as exact fiducial limits for the estimate of error. These become, as they should, $0-\infty$ when the slope is not significant. These fiducial limits are explained in Part V of this report, while Part VI deals with the correct method of analysing data statistically and obtaining fiducial limits when account has to be taken of the use of litter mates on corresponding doses.

* In a laboratory where routine tests are carried out regularly perhaps a compromise is possible. The slope of assays carried out within a comparatively short period prior to the assay under examination might be averaged to provide an estimate. This would reduce the error of estimate by providing more data while eliminating long-term variations in slope.

† For a given number of animals the error of the slope would be diminished by using three doses of test and three of standard in the ratio 1 : 2 : 4 instead of two in the ratio 1 : 2. This would reduce the error of the slope in the ratio 1 : 1·6 and lead to a smaller error in the final result. The 2 : 1 ratio could usually be used for the three doses without the slope departing significantly from linearity.

PART I

Experiment 1. The assay of a sample of halibut-liver oil by workers in nine different laboratories

1. *Introductory*

Data relating to the assay of a sample of halibut-liver oil in nine different laboratories have been analysed. Table 1* shows the results as reported by the workers themselves and as recalculated. The recalculation was performed separately for males and females, by the method explained in a paper on statistical method applied to biological assays (Irwin, 1937). The standard error of the logarithm of the result (σ_M) was calculated separately for males and females, and to get a final value the logarithms of the results for males and females were weighted inversely as σ_M^2.

An inspection of Table 1 shows: (i) that there may be considerable differences between the results as reported and the results as recalculated, and (ii) that the reported results appear to be less variable than the recalculated. The reason for both these discrepancies is to be found in the influence of the method of calculation on the result obtained.

2. *Influence of method of calculation on the result*

How all the workers arrived at their results has not been reported. A common method of calculation, however, seems to be to take the position and slope of the response curves given by the two doses of standard as correct, and to use this curve to obtain the number of units corresponding to the responses made to the two doses of the test preparation. The results of worker 3 may be used to show that this is a faulty procedure.

	Dose	log dose (x)	Response (y†)	No. of rats
		Males		
Standard	2 units	0·301	9·2	5
	4 units	0·602	33·6	5
Test	10 mg.	1·000	3·8	5
	20 mg.	1·301	7·5	4
		Females		
Standard	2 units	0·301	10·3	7
	4 units	0·602	26·2	6
Test	10 mg.	1·000	11·3	6
	20 mg.	1·301	15·5	6

Using these data, the equation to the straight lines

* The order of the laboratories in this and subsequent tables is *not* the same as that given on p. 291.

† Average increase in wt. (g.) in 5 weeks.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

through the two points provided by the two doses of the standard are

Males: $y - 9{\cdot}2 = 81{\cdot}1\,(x - 0{\cdot}3010)$, (1)
Females: $y - 10{\cdot}3 = 52{\cdot}8\,(x - 0{\cdot}3010)$. (2)

Putting $y = 3{\cdot}8$ and $7{\cdot}5$ in (1) and $11{\cdot}3$ and $15{\cdot}5$ in (2), calculating the corresponding values of x and then taking the antilogarithms, gives the following results for the number of units corresponding to 10 mg. halibut-liver oil (H.L.O.):

Males	No. of rats	Females	No. of rats	Total	No. of rats
1·71	5	2·09	6		
0·955	4	1·26	6		
Weighted mean					
1·37	9	1·67	12	1·54	21

So the final result is 154 units/g. which agrees closely with the result stated by worker 3.

points provided by the two doses of the test preparation are:

Males: $y - 3{\cdot}8 = 12{\cdot}3\,(x - 1)$, (3)
Females: $y - 11{\cdot}3 = 13{\cdot}8\,(x - 1)$. (4)

Putting $y = 9{\cdot}2$ and $33{\cdot}6$ in (3) and $10{\cdot}3$ and $26{\cdot}2$ in (4), calculating the corresponding values of x and then taking antilogarithms, gives the following results for the number of mg. of H.L.O. corresponding to 2 units of the standard preparation:

Males mg.	No. of rats	Females mg.	No. of rats	Total mg.	No. of rats
28	5	8	7		
1324	5	59	6		
Weighted mean					
676	10	32	13	312	23

Thus the same method of calculation would lead us to conclude that there were 2 units in 312 mg. of H.L.O.

Table 1. *Halibut-liver oil experiment. Results as stated by worker and as recalculated*

	As stated by worker		As recalculated: Males		As recalculated: Females		As recalculated: Combined		
	Result	M*	M	Result	M	Result	M (weighted mean)	Result	No. of animals used
1	119	$\bar{1}{\cdot}0755$	$\bar{1}{\cdot}0910$	123	$\bar{5}{\cdot}4092$	0†	$\bar{1}{\cdot}0910$	123	51
2	150	$\bar{1}{\cdot}1761$	$\bar{1}{\cdot}2502$	178	$\bar{1}{\cdot}0889$	123	$\bar{1}{\cdot}1562$	143	110
3	152	$\bar{1}{\cdot}1818$	$\bar{2}{\cdot}9901$	98	$\bar{1}{\cdot}1660$	147	$\bar{1}{\cdot}1063$	128	44
4	142	$\bar{1}{\cdot}1523$	$\bar{1}{\cdot}2156$	164	$\bar{1}{\cdot}0648$	116	$\bar{1}{\cdot}1743$	149	98
5	148	$\bar{1}{\cdot}1703$	$\bar{1}{\cdot}7947$	623†	$\bar{1}{\cdot}1216$	132	$\bar{1}{\cdot}4246$	266†	19
6	160	$\bar{1}{\cdot}2041$	$\bar{1}{\cdot}2996$	199	$\bar{1}{\cdot}0510$	112	$\bar{1}{\cdot}2873$	194	24
7	170	$\bar{1}{\cdot}2304$	$\bar{3}{\cdot}4949$	3†	$\bar{2}{\cdot}8339$	68†	$\bar{2}{\cdot}8245$	67†	19
8	143	$\bar{1}{\cdot}1553$	$\bar{1}{\cdot}2173$	165	$\bar{1}{\cdot}1553$	143	$\bar{1}{\cdot}1797$	151	66
9	147	$\bar{1}{\cdot}1673$	$\bar{1}{\cdot}1464$ $\bar{1}{\cdot}2256$	140 168	—	—	$\bar{1}{\cdot}1930$	156	49
Weighted mean	144	$\bar{1}{\cdot}1585$					$\bar{1}{\cdot}1947$	157	Mean 53
Weighted S.D.		0·0372					0·0729		
Mean limits of error per exp. ($P = 0{\cdot}99$)		80–125 %					65–154 %		

* M = logarithm of result in units per milligram. † Unusually aberrant tests.

Now in this particular case the slopes of the straight lines provided by the two doses of the test preparation were considerably flatter than those provided by the doses of the standard, being 12·3 for males and 13·8 for females as against 81·1 and 52·8. It therefore becomes of interest to inquire what happens if the procedure is reversed, that is to say if the position and slope of the response curve given by the two doses of the test substance are assumed to be correct, and this curve is used to obtain the number of mg. of H.L.O. corresponding to the responses made to the two doses of the standard preparation.

The equations to the straight lines through the two

or the potency would be $(2000/312) = 6{\cdot}4$ units/g. The difference between 154 and 6 is more than enough to show the need for a standard (and correct) method of calculation.

Large as are these differences in slope provided by the doses of the standard and test substances, they are not significantly different (see Table 3). The correct method of calculation finds (by the method of least squares) the two parallel straight lines best fitting the observed data; that is to say, it assumes that apart from sampling error the slopes provided by the standard and test substances are the same. The horizontal distance between the parallel straight lines then provides an

estimate of the potency ratio. In this case the results were 98 for males and 147 for females with a mean (correctly weighted) of 128. (An example of the correct method of calculation is given in the appendix to Part I, where worker 9's second experiment is used.

3. *Unusually aberrant tests*

In view of these considerations, the only discrepancies between the results as reported and as recalculated which call for further comment are those of workers 1, 5 and 7. In worker 5 this is due to the peculiar result obtained for the males (see data below on pp. 294 and 295). An examination of the test at once shows that something peculiar has occurred; indeed, from time to time tests do occur where it is obvious on inspection without doing any calculation at all that this is so. There are three further examples in this data, worker 1's result for females and worker 7's results for males and females. Nevertheless, exactly the same method of recalculation has been used in these cases because it was desired to show how a correct method of calculation, including the calculation of the error of the test from internal evidence, would itself reveal deficient tests.

Table 2 repeats the values of M given in Table 1, and shows, against each, the value of σ_M (the error of the test calculated from internal evidence) and the weight ($W_M = (10/\sigma_M^2)$ to be given to the value of M obtained. In the case of worker 1 (females) and worker 7 the error of the test calculated from internal evidence is so large that the result is given practically no weight, and the deficiency of the test is at once revealed. In the case of worker 5 (males) this is not so to the same extent, and an examination of the actual data soon reveals the reason. Doses of 0·6 and 1·2 units of the standard were given. There were two animals on the former dose, both of which died before the end of the test period and two on the latter dose. Doses of 3·2 and 6·4 mg. of H.L.O. were given and only one animal survived on each dose. The data are therefore as follows:

	Dose	log dose (x)	Response (y)	No. of rats
Standard	0·6 unit	$\bar{1}$·7782	—	—
	1·2 units	0·0792	72	2
Test	3·2 mg.	0·5052	50	1
	6·4 mg.	0·8062	20	1

Thus the standard gives no evidence about slope, and the test substance shows a large reversal. It is evident that something is wrong, and probably the test should be omitted altogether. Nevertheless, if we take the data at its face value we find a large negative slope ($b = -99{\cdot}7$).

Now the approximate formula for the error of the slope is given by

$$\sigma_M^2 = \Sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)+\frac{(\bar{y}_2-\bar{y}_1)^2\sigma_b^2}{b^4},$$

where

Σ^* = S.D. of animals on same dose,
σ_b = S.E. of slope,
n_1 = no. of animals on standard substance,
n_2 = no. of animals on test substance,
b = slope,
$\bar{y}_1$ = average response to doses of standard substance,
$\bar{y}_2$ = average response to doses of test substance.

If b is too large, as here, the portion of the error due to error in slope tends to be small and the calculation of the error from internal evidence does not reveal the weakness of the test (though, in this case, the reversal of the slope does). The weakness, in this case, and equally in cases where the slope is positive and too large, is, however, revealed when the result of the test is compared with the result of other tests on the same

* Where litter-mates have been used on corresponding doses Σ may be replaced by σ the S.D. of litter-mates on the same dose.

Table 2. *Halibut-liver oil. Values of M* with their standard errors and weights*

No.	Males M	Males σ_M	Males W_M	Females M	Females σ_M	Females W_M	Combined M	Combined σ_M	Combined W_M	Result Units/g.	Limits of error ($P=0{\cdot}99$) %
1	$\bar{1}$·0910	0·0602	2763	$\bar{5}$·4092	91·3	0	$\bar{1}$·0910	0·0602	2763	123	70–143
2	$\bar{1}$·2502	0·1640	372	$\bar{1}$·0889	0·1387	520	$\bar{1}$·1562	0·1059	891	143	53–187
3	$\bar{2}$·9901	0·2133	220	$\bar{1}$·1660	0·1528	428	$\bar{1}$·1063	0·1242	648	128	48–209
4	$\bar{1}$·2156	0·0601	2773	$\bar{1}$·0648	0·0975	1052	$\bar{1}$·1741	0·0511	3825	149	74–135
5	$\bar{1}$·7947	0·1712	341	$\bar{1}$·1216	0·1549	417	$\bar{1}$·4246	0·1148	758	266†	51–198
6	$\bar{1}$·2996	0·0503	3945	$\bar{1}$·0510	0·2204	206	$\bar{1}$·2873	0·0491	4151	194	75–134
7	$\bar{3}$·4949	12·09	0	$\bar{2}$·8339	0·8426	14	$\bar{2}$·8245	0·8392	14	67†	1–14500
8	$\bar{1}$·2173	0·0659	2300	$\bar{1}$·5553	0·0531	3551	$\bar{1}$·1797	0·0413	5851	151	78–128
9	$\bar{1}$·1464 $\bar{1}$·2256	0·0647 0·0541	2391 3418				$\bar{1}$·1930	0·0415	5808	156	78–128
Average σ_M								0·0604			
Mean limits of error per experiment ($P=0{\cdot}99$)											70–143

* M = logarithm of results in units per milligram.

† Unusually aberrant tests.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 3. *Halibut-liver oil. Quantities used in calculating the error from internal evidence*

Laboratory	Σ^2	D.F.	σ^2	D.F.	b	σ_b	b_1	b_2	b_1-b_2	$\sigma_{(b_1-b_2)}$	M	σ_M	Result
						Males							
1	78·68	21	50·14	7	55·9	12·1	58·8	52·3	+ 6·5	24·4	$\bar{1}$·0910	0·0602	123
2	187·31	49	115·25	7	18·5	12·5	10·7	25·5	−14·8	25·1	$\bar{1}$·2502	0·1640	178
3	233·32	15	42·50	2	48·7	23·4	81·1	12·3	+68·8	46·8	$\bar{2}$·9901	0·2133	98
4	370·34	36	469·46	3	104·1	20·3	95·6	113·7	−18·1	40·8	$\bar{1}$·2156	0·0601	164
5	72·00	1	—	—	−99·7	39·9	—	−99·7	—	—	$\bar{1}$·7947	0·1712	623*
6	32·48	7	—	—	68·6	11·5	57·6	77·5	−19·9	23·1	$\bar{1}$·2996	0·0503	199
7	261·00	2	—	—	6·6	46·5	0	13·3	−13·3	93·0	$\bar{3}$·4949	12·09	3*
8	139·88	28	112·95	11	59·5	14·2	54·1	66·2	−12·1	28·5	$\bar{1}$·2173	0·0659	165
9 (1st)	154·23	15	234·62	7	116·0	19·0	91·4	143·6	−52·2	38·0	$\bar{1}$·1464	0·0647	140
(2nd)	178·53	26	73·08	12	85·9	16·3	102·7	67·0	+35·7	32·6	$\bar{1}$·2256	0·0541	168
						Females							
1	178·91	21	56·33	9	0·75	17·5	−11·3	12·8	−24·1	24·7	$\bar{5}$·4092	91·3	0*
2	201·97	53	72·58	6	24·8	12·5	26·8	22·5	+ 4·3	25·1	$\bar{1}$·0889	0·1387	123
3	144·24	21	182·25	1	34·0	16·0	52·8	13·8	+39·0	32·0	$\bar{1}$·1660	0·1528	147
4	151·56	54	13·50	2	43·1	10·7	10·6	67·6	−57·0	21·7	$\bar{1}$·0648	0·0975	116
5	89·29	11	25·54	5	21·7	16·3	36·5	44·3	− 7·8	32·7	$\bar{1}$·1216	0·1549	132
6	57·70	9	70·75	2	26·4	14·1	1·1	55·4	−54·3	28·2	$\bar{1}$·0510	0·2204	112
7	106·97	9	67·11	3	13·4	19·7	5·8	23·3	−17·5	39·7	$\bar{2}$·8339	0·8392	67*
8	45·89	30	31·60	10	40·9	7·8	27·5	54·6	−27·1	15·6	$\bar{1}$·1553	0·0531	143

Σ^2 = variance of animals (not litter-mates) on same dose.
σ^2 = variance of animals (litter-mates) on same dose.
D.F. = no. degrees of freedom used in calculating corresponding variance.
b = best fitting slope for test and standard combined.
b_1 = slope from two doses of standard.
b_2 = slope from two doses of test.
M = log of result in units/mg.
σ_M = standard error of M.

* Unusually aberrant tests.

oil, from which (owing to its relatively small error) it differs significantly.

The data for worker 1 (females) are as follows:

	Dose	log dose (x)	Response (y)	No. of rats
Standard	1·93 units	0·2856	40·8	6
	3·86 units	0·5866	37·4	7
Test	9·64 mg.	0·9841	34·0	6
	19·28 mg.	1·2852	37·9	7

Here there is a reversal of slope for the standard, so that when a common slope is fitted to the data, it is practically flat ($b=0·7$). The result is that the error of the test becomes very large and it receives no weight in determining the final result.

The data for worker 7 (males) show a similar effect:

	Dose	log dose (x)	Response (y)	No. of rats
Standard	1·6 units	0·2041	53	2
	3·2 units	0·5052	53	1
Test	8 mg.	0·9031	39	1
	16 mg.	1·2041	43	2

There is practically no increased response to increasing dosage, the slope is very flat ($b=6·6$), and the error of the test consequently very large and its weight small.

The data for worker 7 (females) are as follows:

	Dose	log dose (x)	Response (y)	No. of rats
Standard	1·6 units	0·2041	37·25	4
	3·2 units	0·5052	39	3
Test	8 mg.	0·9031	29	4
	16 mg.	1·2041	36	2

Here there is practically no increased response to an increasing dose of standard, while the increase in response to increasing dosage of test is small. The slope ($b=13·4$) is still too flat and the error of the test consequently large.*

4. *Differences in variability of the reported and of the recalculated results*

We have to account for the fact that the reported results are less variable than the recalculated results. In arriving at the reported results, different laboratories may have used different methods of calculation, but it seems likely that the position and slope of the response curve given by the two doses of standard was taken as correct and the responses to the two doses of the test substance were interpreted in terms of this curve.

Now the slopes provided by the two doses of standard in this data were less variable than those provided by the two doses of the test substance.* Thus in using the slope from the former pair of points, instead of the best-fitting common slope part of the error of the test is neglected. The workers were using slopes which were too like one another.

There is another possible circumstance which may have contributed to this result. The workers had an idea of what the result ought to be. It was generally expected that the pairs of doses of test and standard would yield equivalent responses, and it was also known how many units per gram there were in the standard preparation. Now if the test on males or females yielded either for the standard or the test substance a slope which looked wildly wrong, what did the workers do about it? Did they reject the test on the males and use that on the females or vice versa? (In one case it is known that this was done.) Any such process of selection would tend to diminish the apparent error of the reported results.

5. *The error of the results*

The weighted standard deviation of the logarithms of the nine results is 0·0729, and the corresponding limits of error of the value assigned to the H.L.O. ($P=0·99$), 65–154 %. The weighted mean of the nine error variances σ^2_M calculated from internal evidence is 0·003642, and the corresponding standard deviation 0·0604. The corresponding limits of error ($P=0·99$) are 70–143 %. The former estimate of error is naturally somewhat greater because it must include any discrepancies between the different laboratories. But in point of fact the two values do not differ significantly.†

The standard error for a test with fifty-three animals, calculated from Coward's data (which were used for the Pharmacopoeia Addendum), is 0·1038 (limits of error ($P=0·99$), 54–185 %). The value 0·0729 obtained from the present results is just significantly‡ lower than the error from Coward's data (obtained from over 2000 animals), if we assume the latter to be correct. However, the latter calculation ignored the influence of error in

* The approximate nature of the formula hitherto in use for the error of the test is now more fully realized than when the above section was written. Where the slope is not significant, the 'fiducial limits' correctly calculated are $(0-\infty)$, and no information can legitimately be obtained from the test. All the aberrant results discussed here would now undoubtedly be omitted.

* The unweighted standard deviations of the slopes from the nine laboratories were as follows:

	Males	Females	Sexes combined
Standard	36·4	20·9	37·7
Test	66·5	21·2	52·1
Standard and test combined	61·6	14·1	48·4

Note that when the sexes are combined the common slope is more variable than the slope obtained from the standard preparation only.

† If we assume the value 0·0604 to be correct and remember 0·0729 is a standard deviation based on 8 degrees of freedom, we find

$$\chi^2=8\left(\frac{0·0729}{0·0604}\right)^2=11·6,$$

quite a normal value with 8 degrees of freedom (5 % point = 15·5).

‡ If we assume 0·1038 to be correct the corresponding weight is $10/(0·1038)^2=928$. The mean observed weight per experiment is 2745 with S.E. 758, $t=2·4$, 5 % point = 2·3. This points to the error being significantly less than Coward's.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

slope which is often large, so there does seem some evidence to show that the error of this series of results is somewhat lower than that given in the Pharmacopoeia Addendum. It is only reasonable to expect this as improvements in technique gradually evolve.

6. *Further points: heterogeneity in variability and slope*

It is of interest to inquire whether the variance in increase of weight of animals on the same dose was different in the different laboratories. In the first place it must be mentioned that in laboratories 2 and 4, increase of weight in 4 weeks was the criterion chosen, in all other laboratories increase of weight in 5 weeks. The variances from the tests in laboratories 2 and 4 are therefore not quite comparable with the others. Other things being equal the variance for a 4-week test should be less than that for a 5-week test (Coward, 1933), but Table 3 shows that the values of Σ^2 in laboratories 2 and 4 were greater than the average of the remainder. If therefore the variances in different laboratories differ significantly when results 2 and 4 are included, they would actually have differed still more significantly had 5-week tests been used in these two laboratories, since the variances would have been raised still further. When results 2 and 4 were included it was found that the variances of animals on the same dose differed significantly from laboratory to laboratory both in males and females. When they were excluded the variances differed significantly in the females only. In the males at any rate, therefore, the heterogeneity seems to be due to laboratories 2 and 4. The data do not show significant differences in the variances of *litter-mates* on the same dose whether nos. 2 and 4 are included or not.* Bartlett's test gave the following values of χ^2:

	Males	Females
	Including 2 and 4	
Σ^2	23·4 (D.F. 9), $P=0{\cdot}005$	21·7 (D.F. 7), $P=0{\cdot}004$
σ^2	10·6 (D.F. 6)	5·1 (D.F. 7)
	Excluding 2 and 4	
Σ^2	11·5 (D.F. 7)	13·8 (D.F. 5), $P=0{\cdot}02$
σ^2	5·9 (D.F. 4)	3·5 (D.F. 5)

Significant values are underlined.

The bucks were more variable than the does, the average variances being:

	Males	Females
	Including 2 and 4	
Σ^2	196·8 (D.F. 200)	141·9 (D.F. 208)
σ^2	130·9 (D.F. 49)	51·0 (D.F. 38)
	Excluding 2 and 4	
Σ^2	146·5 (D.F. 115)	105·2 (D.F. 101)
σ^2	107·6 (D.F. 39)	49·2 (D.F. 39)

The differences are significant and in the same direction as the results from Coward's data (Pharm. Comm. Rep. 1936).

We may also examine differences in slope from laboratory to laboratory (Table 3). These are occasionally significant. Looking at the males, but leaving out nos. 5 and 7, which are obviously anomalous, differences between nos. 1 and 2, nos. 1 and 4, nos. 1 and 9, nos. 2 and 4, nos. 8 and 9 (1) are significant or 5 pairs out of 28, whereas we should expect only 1 or 2. If nos. 2 and 4 are also omitted we have 2 pairs out of 15, whereas we should expect less than 1. In the females no. 1 is obviously anomalous, otherwise there are no significant differences.

Slopes obtained from the standard and from the H.L.O. do not differ significantly on the same test.

The average slope for bucks is somewhat greater than that for does, the figures being

	Including 2 and 4		Excluding 2 and 4	
	Males	Females	Males	Females
Mean	59·6	31·7	64·9	30·2
S.E. of mean	11·0	4·3	11·2	5·3

This agrees with the result from Coward's data (Coward, 1933; and Pharm. Comm. Rep. 1936).

APPENDIX TO PART I (METHOD OF CALCULATION)

(i) *Formulae*

The formulae for calculating the assay and its error for any number of doses of each substance have been given (Irwin, 1937). When, however, there are only two doses of the standard and two doses of the test preparation, and the two doses of the test preparation bear the same ratio to one another as the two doses of the standard preparation, these formulae are considerably simplified.

Let x_{11}, x_{12} be the logarithms of the two doses of the standard preparation, and x_{21}, x_{22} the logarithms of the two doses of the test preparation, and let

$$x_{12}-x_{11}=x_{22}-x_{21}=d. \tag{1}$$

Let the number of animals on the above four doses be $n_{11}, n_{12}, n_{21}, n_{22}$. Let $y_{11}, y_{12}, y_{21}, y_{22}$ be the corresponding responses, in this case average increase in weight in 5 weeks.

Let

$$\left.\begin{aligned} \bar{x}_1&=\frac{n_{11}x_{11}+n_{12}x_{12}}{n_{11}+n_{12}}, & \bar{x}_2&=\frac{n_{21}x_{21}+n_{22}x_{22}}{n_{21}+n_{22}},\\ \bar{y}_1&=\frac{n_{11}y_{11}+n_{12}y_{12}}{n_{11}+n_{12}}, & \bar{y}_2&=\frac{n_{21}y_{21}+n_{22}y_{22}}{n_{21}+n_{22}}. \end{aligned}\right\} \tag{2}$$

That is, $\bar{x}_1$, $\bar{x}_2$ are respectively the weighted means of the two log doses of the standard and the two log doses of the test preparations, and $\bar{y}_1$, $\bar{y}_2$ the weighted means of the corresponding responses.

Let $$v_1=\frac{n_{11}n_{12}}{n_{11}+n_{12}}, \quad v_2=\frac{n_{21}n_{22}}{n_{21}+n_{22}}. \tag{3}$$

* M. S. Bartlett's (1937*a*, *b*) test for homogeneity of variance was used. To test the homogeneity of a set of k estimated variances s_r^2 with n_r degrees of freedom a 'crude' value of χ^2 is calculated by computing $2{\cdot}3026\,(n\log_{10}s^2-\Sigma n_r\log_{10}s_r^2)$, where s^2 is the pooled variance with $n=\Sigma n_r$ degrees of freedom. If this value of χ^2 (with $(k-1)$ degrees of freedom) appears significant, it is corrected by dividing by a factor C, where

$$C=1+\frac{1}{3\,(k-1)}\left(\Sigma\frac{1}{n_r}-\frac{1}{n}\right).$$

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Then the best-fitting common slope to the four points is given by

$$b=\frac{\nu_1(y_{12}-y_{11})+\nu_2(y_{22}-y_{21})}{(\nu_1+\nu_2)\,d}, \qquad (4)$$

and M, the logarithm of the potency of the test preparation, will be given by

$$M=\bar{x}_1-\bar{x}_2+\frac{(\bar{y}_2-\bar{y}_1)}{b}, \qquad (5)$$

while the standard error of M is given by

$$\sigma_M^2=\left(\frac{1}{n_{11}+n_{12}}+\frac{1}{n_{21}+n_{22}}\right)\frac{\Sigma^2}{b^2}+\frac{(\bar{y}_2-\bar{y}_1)^2\,\sigma_b^2}{b^4}, \qquad (6)$$

with

$$\sigma_b^2=\frac{\Sigma^2}{(\nu_1+\nu_2)\,d^2}. \qquad (7)$$

Here Σ^2 is the variance in response of animals on the same dose, which must be estimated from the data. If litter-mates have been so used that each animal on each dose of the standard preparation has a litter-mate on the corresponding dose of the test preparation, Σ^2 in (6) but not in (7) may be replaced by σ^2, the variance of litter-mates on the same dose. This will usually be smaller than Σ^2 and the error of the test consequently reduced.

If all four doses are expressed in the same units by weight or volume, then M gives the logarithm of the ratio of the potency of the test preparation to that of the standard. But if the two doses of the standard preparation are given in 'units', while the two doses of the test preparation are given in milligrams, M gives the logarithm of the potency in units per milligram.

(ii) *Numerical example*

To illustrate the numerical work required the following data may be used (worker 9's second experiment):

	Dose	log dose (x)	Response (y)	No. of rats (n)
Standard	1 unit	0·0000	27·8556	9
	2 units	0·3010	58·7714	7
				16
Test	9·26 mg.	0·9666	49·4286	7
	18·52 mg.	1·2676	69·5857	7
				14

The calculation proceeds as follows:

Standard			Test
$y_{12}-y_{11}$	30·9158	$y_{22}-y_{21}$	20·1571
$\bar{y}_1$	41·3813	$\bar{y}_2$	59·5071
$\bar{x}_1$	0·1317	$\bar{x}_2$	1·1171
$\nu_1=\frac{63}{16}=$	3·9375	$\nu_2=\frac{49}{14}=$	3·5

$$b=\frac{3\cdot9375\,(30\cdot9158)+3\cdot5\,(20\cdot1571)}{0\cdot3010\,(7\cdot4375)}=\frac{192\cdot2808}{2\cdot2387}=85\cdot889.$$

$$M=-(1\cdot1171-0\cdot1317)+\left(\frac{59\cdot5071-41\cdot3813}{85\cdot889}\right)$$

$$=-0\cdot9854+\left(\frac{18\cdot1258}{85\cdot889}\right)=-0\cdot9854+0\cdot2110$$

$$=-0\cdot7744=\bar{1}\cdot2256$$

antilog $M=0\cdot168$.

Therefore potency of test substance is 168 units/g.

Notes

$$\bar{y}_1=\frac{9\,(27\cdot8556)+7\,(58\cdot7714)}{16}=41\cdot3813,$$

$$\bar{y}_2=\frac{7\,(49\cdot4286)+7\,(69\cdot5857)}{14}=59\cdot5071,$$

$$\bar{x}_1=\frac{9\,(0\cdot0000)+7\,(0\cdot3010)}{16}=0\cdot1317,$$

$$\bar{x}_2=\frac{7\,(0\cdot9666)+7\,(1\cdot2676)}{14}=1\cdot1171.$$

Calculation of the error

First it is necessary to estimate the variance of the increase in weight of animals on the same dose. This is usually independent of the dosage level. Therefore the information provided by the four doses may be pooled. The rule for forming the estimate is as follows:

For the animals on each dose calculate the sum of the squares of the deviations from the mean. Add these together and divide by $S\,(n-1)$ summed over the four doses, that is, by (total number of animals − 4). It is often convenient to use the relation

$$S\,(u-\bar{u})^2=S\,(u^2)-\bar{u}\,(nu)$$

or (sum of squares of deviations) = (sum of squares − (mean × total)).

In this particular case litter-mates were used on corresponding doses. It is therefore necessary to estimate the variance of litter-mates on the same dose. This quantity (σ^2) is usually independent of the dosage level, so it may be estimated from the data provided by the test, because the difference between litter-mates on corresponding doses (e.g. 1 unit of standard and 9·26 mg. of H.L.O.) will have a variance $2\sigma^2$. Thus the differences between litter mates on corresponding doses are formed, and their variance calculated as above. There will be two sets of differences, the sum of the squares of the deviations from the mean is calculated for each set, these are added together and divided by $S\,(n-1)$ summed over the two sets, that is by (total number of differences − 2). The result so obtained is halved and this gives the estimate of σ^2.

For the present test the data are as follows:

1 i.u.st.	9·26 mg. H.L.O.	2 i.u.st.	18·52 mg. H.L.O.
20·0	—	—	—
38·7	—	—	—.
41·8	54·5	50·0	59·4
38·0	60·3	78·3	72·7
13·5	26·8	41·0	60·3
39·0	56·5	77·7	92·4
22·0	52·7	34·3	64·5
11·7	50·6	54·6	72·0
26·0	44·6	75·5	65·8
Total:			
250·7	346·0	411·4	487·1
Mean:			
27·85	49·42857	58·771429	69·585714
Sum of squares:			
8089·07	17843·64	26207·08	34661·39

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Sum of squares of deviations:

	1105·6821	741·3458	2028·5141	766·1887
$n-1$:	8	6	6	6

Total sum of squares of deviations = 4641·7397.

Estimate of $\Sigma^2 = \frac{1}{26}(4641{\cdot}7397) = 178{\cdot}5284$.
The calculation of Σ^2 is shown immediately underneath.

The calculation of σ^2 proceeds as follows:

	9·26 mg. H.L.O. −1 i.u.st.	18·52 mg. H.L.O. −2 i.u.st.
	12·7	+ 9·4
	22·3	− 5·6
	13·3	+19·3
	17·5	+14·7
	30·7	+30·2
	38·9	+17·4
	18·6	− 9·7
Total	154·0	75·7
Mean	22·0	10·814286
Sum of squares	3943·38	2017·19
Sum of squares of deviations	555·38	1198·5485
$n-1$	6	6

Total = 1753·9285.
Estimate of $\sigma^2 = \frac{1}{2}\{\frac{1}{12}(1753{\cdot}9285)\} = 73{\cdot}0803$.

Finally, the estimates of Σ^2 and σ^2 so obtained are substituted in equations (6) and (7). In this case we find

$$\sigma_b^2 = \frac{178{\cdot}5284}{(0{\cdot}30103)^2\,(7{\cdot}4375)} = \frac{178{\cdot}5284}{0{\cdot}6739793} = 264{\cdot}8871,$$

$$\sigma_M^2 = \frac{73{\cdot}0803}{7376{\cdot}9203}\left(\frac{1}{16}+\frac{1}{14}\right) + \frac{(18{\cdot}1258)^2\,(264{\cdot}8871)}{54418953}$$

$$= 0{\cdot}0013268 + 0{\cdot}0015992 = 0{\cdot}0029260,$$

$$\sigma_M = 0{\cdot}05409,$$
$$1{\cdot}960\sigma_M = 0{\cdot}1060,$$
$$2{\cdot}576\sigma_M = 0{\cdot}1393,$$
$$\text{antilog } 1{\cdot}960\sigma_M = 1{\cdot}28,$$
$$\text{antilog } 2{\cdot}576\sigma_M = 1{\cdot}38.$$

Limits of error ($P = 0{\cdot}95$) are (78–128 %), ($P = 0{\cdot}99$) are (73–138 %).

PART II

EXPERIMENT 2. THE ASSAY OF A SAMPLE OF U.S.P. REFERENCE OIL BY WORKERS IN TEN DIFFERENT LABORATORIES

1. *The results and their errors*

Table 4 shows the results as reported by the workers themselves and as recalculated. Table 5 gives these results with their errors. The discrepancies between the results as stated and as recalculated are not relatively so large as for the H.L.O. Only one result, no. 9, appears unusually aberrant and that has a very large error. There is no need therefore to repeat §§ 2, 3 and 4 of Part I for the U.S.P. reference oil.

The weighted standard deviation of the logarithm of the ten results is 0·0559, and the corresponding limits of error ($P = 0{\cdot}99$) are 72–139 %. The weighted mean of the ten error variances is 0·002454, and the corresponding standard deviation 0·0495. The corresponding limits of error ($P = 0{\cdot}99$) are 75–134 %. The former estimate of error is again somewhat greater but not significantly greater than the latter ($\chi^2 = 11{\cdot}5$ with 9 degrees of freedom). Thus there are no significant differences between the results of the different laboratories.

The mean limits of error for the U.S.P. reference oil are of the same order of magnitude as for the H.L.O. and do not differ significantly from them. This was tested by comparing the average weights $\overline{W}$ for the two series. These are

H.L.O. $\overline{W} = 2745$
U.S.P.R.O. $\overline{W} = 4074$ } difference 1329, S.E. 1172.

The error of these results is again significantly lower than that given in the Pharmacopoeia Addendum.*

2. *Further points: heterogeneity in variability and slope*

The variances of increase in weight and slopes in the different laboratories were again examined for heterogeneity (Table 6). In laboratory 4 increase of weight in 4 weeks was the criterion chosen, in all other laboratories increase of weight in 5 weeks. The result of M. S. Bartlett's test for homogeneity of variance were as follows:

Values of χ^2

	Males	Females
	Including 4	
Σ^2	24·6 (D.F. 8), $P = 0{\cdot}002$	15·2 (D.F. 5), $P = 0{\cdot}010$
	Excluding 4	
Σ^2	23·2 (D.F. 7), $P = 0{\cdot}002$	11·1 (D.F. 4), $P = 0{\cdot}03$
	Excluding 2 and 4	
Σ^2	8·03 (D.F. 6)	1·06 (D.F. 3)
	All available data 2, 5, 8, 9	
σ^2	2·4 (D.F. 3)	5·6 (D.F. 3)

Significant values are underlined.

There is significant heterogeneity in the variance Σ^2 of animals on the same dose, and this whether laboratory 4 is included or not. Laboratory 2 has the highest variance in the males and laboratory 4 in the females. When both 2 and 4 are excluded, there is no significant difference in variance. This result was the same for the H.L.O. There are again no significant differences in the variances of *litter-mates* on the same dose.

The bucks are again more variable than the does, the average variances being:

	Males	Females
	Including 4	
Σ^2	162·4 (254 D.F.)	96·8 (172 D.F.)
	Excluding 4	
Σ^2	166·8 (235 D.F.)	86·1 (143 D.F.)
	Excluding 2 and 4	
Σ^2	133·6 (193 D.F.)	63·1 (100 D.F.)
	All available data 2, 5, 8, 9	
σ^2	142·0 (47 D.F.)	76·8 (57 D.F.)

* Weight from Pharmacopoeia Addendum = 910. The mean observed weight per experiment is 4074 with S.E. 857, $t = 3{\cdot}7$, 5 % point = 2·3.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 4. *U.S.P. reference oil. Results as stated by worker and as recalculated*

	As stated by worker		As calculated: Males		As calculated: Females		As calculated: Combined		
No.	Result	M*	M	Result	M	Result	M (weighted mean)	Final result	No. of animals used
1	2280	Ī·9600	—	—	Ī·9594	2277	Ī·9594	2277	40
2	2240	Ī·9523	Ī·9311	2133	Ī·9912	2450	Ī·9596	2278	93
3	2667	0·0281	0·0094	2552	0·0189	2611	0·0160	2595	93
4	2910–3020	0·0660–0·0821	Ī·9991	2495	0·1961	3927	0·0143	2584	56
5	2650	0·0253	0·0652	2905	Ī·9471	2213	0·0043	2525	32
6	3000	0·0791	0·0904	3079	0·1066	3196	0·0973	3128	39
7	3125	0·0969	0·1166	3270	—	—	0·1166	3270	32
8	2250	Ī·9542	Ī·9607	2284	Ī·9380	2167	Ī·9488	2222	58
9	Not stated		Ī·7271	1334	—	—	Ī·7271	1334	20
10	3120	—	0·0762	2980	—	—	0·0762	2980	54
Weighted mean							0·02015	2619	Mean 52
Weighted S.D.							0·0559		
Mean limits of error per exp. ($P=0·99$)							72–139 %		

* M = logarithm of potency ratio. The standard contained 2500 units/g.

Table 5. *U.S.P. reference oil. Values of M* with their standard errors and weights*

	Males			Females			Combined				Limits of error ($P=0·99$)
No.	M	σ_M	W_M	M	σ_M	W_M	M	σ_M	W_M	Result Units/g.	%
1	—	—	—	Ī·9594	0·0692	2091	Ī·9594	0·0692	2091	2277	66–151
2	Ī·9311	0·0620	2604	Ī·9912	0·0653	2345	Ī·9596	0·0450	4949	2278	77–131
3	0·0094	0·0778	1652	0·0189	0·0512	3810	0·0160	0·0428	5462	2595	78–129
4	Ī·9991	0·0541	3422	0·1961	0·1867	287	0·0143	0·0519	3709	2584	73–136
5	0·0652	0·0515	3770	Ī·9471	0·0499	4013	0·0043	0·0359	7783	2525	81–124
6	0·0904	0·0732	1867	0·1066	0·0845	1401	0·0973	0·0553	3268	3128	72–139
7	0·1166	0·0924	1172	—	—	—	0·1166	0·0924	1172	3270	58–173
8	Ī·9607	0·0805	1545	Ī·9380	0·0765	1710	Ī·9488	0·0554	3255	2222	72–139
9	Ī·7271	0·2739	133	—	—	—	Ī·7271	0·2739	133	1334	20–508
10	0·0762	0·0335	8921	—	—	—	0·0762	0·0335	8921	2980	82–122
Average σ_M								0·0495			
Mean limits of error per exp.											75–134

* M = logarithm of potency ratio. The standard contained 2500 units/g.

In the males (but not in the females) there are significant differences in slope from laboratory to laboratory (Table 6). These are almost entirely due to the difference between laboratory 10 and the others, 10 differing significantly from 3, 6, 7, 8 and 9. Apart from this, there are only three of the remaining thirty-three pairs of differences which are significant at the 5 % level against 1·5 expected. These are:

	(diff. in slope)/S.E.
3 and 5	2·02 (56 D.F.)
5 and 6	2·24 (34 D.F.)
5 and 9	2·17 (30 D.F.)

Slopes obtained from the standard and from the U.S.P. reference oil do not differ significantly in the same test.

The average slope for bucks is again greater than that for does, the figures being

	Including 4		Excluding 4	
	Males	Females	Males	Females
Mean	62·9	41·8	61·1	43·1
S.E.	7·9	4·1	8·4	4·2

The results do not differ significantly from those obtained for the H.L.O.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 6. *U.S.P. reference oil. Quantities used in calculating the errors from internal evidence*

Laboratory	Period weeks	Σ^2	D.F.	σ^2	D.F.	b	σ_b	b_1	b_2	b_2-b_1	$\sigma_{(b_2-b_1)}$	M	σ_M	Result
							Males							
2	5	319·14	42	119·27	19	54·311	17·53	54·15	54·49	− 0·34	35·11	Ī·9311	0·06197	2133
3	5	153·67	42	—	—	47·010	12·15	40·53	53·49	−11·96	25·39	0·0094	0·07780	2552
4	4	108·30	19	—	—	81·149	14·57	106·31	62·46	+43·85	29·46	Ī·9991	0·05406	2495
5	5	82·08	14	102·77	7	84·880	14·28	94·35	75·42	+18·93	28·55	0·0652	0·05150	2905
6	5	60·01	20	—	—	45·116	10·51	64·22	26·01	+38·21	21·02	0·0904	0·07319	3079
7	5	128·44	28	—	—	45·687	13·34	56·81	35·88	+20·93	26·73	0·1166	0·09238	3270
8	5	149·53	23	126·59	13	53·874	15·74	39·87	68·74	−28·87	31·50	Ī·9607	0·08046	2284
9	5	185·50	16	255·32	8	30·312	20·65	19·93	40·66	−20·73	41·30	Ī·7271	0·27386	1334
10	5	139·56	50	—	—	99·070	10·69	97·71	100·43	− 2·72	21·38	0·0762	0·03348	2980
							Females							
1	5	75·45	36	59·64	18	35·714	9·12	30·90	40·53	− 9·63	18·25	Ī·9594	0·06916	2227
2	5	139·49	43	74·84	21	38·745	11·45	36·88	40·53	− 3·65	22·91	Ī·9912	0·06531	2450
3	5	56·96	43	—	—	43·024	7·33	60·47	24·92	+35·55	14·66	0·0189	0·05123	2611
4	4	149·99	29	—	—	27·650	14·18	18·94	36·88	−17·94	28·27	0·1961	0·18667	3927
5	5	51·48	10	19·46	5	49·169	12·87	39·87	58·47	−18·00	25·75	Ī·9471	0·04992	2213
6	5	57·02	11	—	—	50·286	13·02	59·93	42·36	+17·57	23·68	0·1066	0·08450	3196
8	5	108·49	27	125·91	13	49·676	12·44	45·95	53·16	− 7·21	24·90	Ī·9380	0·07648	2167

Σ^2 = variance of animals (not litter-mates) on same dose.
σ^2 = variance of animals (litter-mates) on same dose.
D.F. = no. degrees of freedom used in calculating corresponding variance.
b = best fitting slope for test and standard combined.
b_1 = slope from two doses of standard.
b_2 = slope from two doses of test.
M = log of result in units/mg.
σ_M = standard error of M.

PART III

Experiment 3. The assay of a solution of the solid vitamin A 2-naphthoate against the international standard carotene, by workers in nine different laboratories

1. *The results and their errors*

Table 7 shows the results as reported by the workers themselves and as recalculated. Table 8 gives these results with their errors. There are no exceptionally aberrant results.

	$\overline{W}$	S.E.*	Mean no. of animals
H.L.O.	2745	746	53
U.S.P. reference oil	4074	707	52
Naphthoate: 5 weeks	2420	746	50
3 weeks	894	1120	55

The differences are not statistically significant, though there is a suggestion that the 3-week tests are less accurate.

The error given in the Pharmacopoeia Addendum yields, for the standard error of the logarithm of the result, 0·1241 for a 3-week test with fifty-five animals and 0·1069 for a 5-week test with fifty animals. Thus

Table 7. *Vitamin A β-naphthoate. Results as stated by worker and as recalculated*

		As stated by worker		As calculated						
				Males		Females		Combined		
No.	Period weeks	Result	M*	M	Result	M	Result	M (weighted mean)	Final result	No. of animals used
1 (*a*)	3	106	Ī·7243	Ī·7256	106	—	—	Ī·7256	106	72
(*b*)	5	96	Ī·6812	Ī·6845	97	—	—	Ī·6845	97	76
(*c*)	5 (select)	136·5	Ī·8341	Ī·8342	137	—	—	Ī·8342	137	40
2 (*a*)	3	200	0	—	—	0·0034	202	0·0034	202	20
(*b*)	5	200	0	—	—	0·1146	260	0·1146	260	20
3 (*a*)	3	140	Ī·8451	Ī·7122	103	Ī·9003	159	Ī·8484	141	53
(*b*)	5	102	Ī·7076	Ī·6280	85	Ī·7768	120	Ī·7593	115	51
4	5	177	Ī·9469	Ī·9941	197	Ī·8528	143	Ī·9440	176	125
5	5	Not stated		Ī·8100	126	Ī·8565	144	Ī·8237	133	44
7	5	Not stated		Ī·8679	148	—	—	Ī·8679	148	41
8	5	145	Ī·8603	Ī·9017	159	Ī·8087	129	Ī·8779	151	75
9 (*a*)	3	160	Ī·9031	Ī·9298	170	—	—	Ī·9298	170	56
(*b*)	5	160	Ī·9031	Ī·9539	180	—	—	Ī·9539	180	58
10	5	183	Ī·9614	Ī·9637	184	—	—	Ī·9637	184	44
Weighted mean 3 weeks (4 labs.)								Ī·9151, Ī·9185 } Ī·9180	164, 166 } 166	Mean 50, 57
5 weeks (9 labs.)										
Weighted S.D. 3 weeks								0·0897, 0·0695 } 0·0702		
5 weeks										
Mean limits of error per exp. (P=0·99) 3 weeks									59–170 % } 66–152	
5 weeks									66–151 %	

* M = logarithm of potency ratio. The standard contained 200 units/g.

For 3-week and 5-week tests respectively the weighted standard deviations of the logarithms of the results are 0·0897 and 0·0695, and the corresponding limits of error are 59–170 and 66–151 %. The weighted mean of the ten error variances is 0·011192 for the four 3-week tests and 0·004132 for the nine 5-week tests; the corresponding standard deviations are 0·1058 and 0·0643, giving limits of error of 53–187 and 68–146 % respectively. There are no significant differences between the two sets of estimates (χ^2 = 2·0 with 3 D.F. and 9·4 with 8 D.F. respectively). Thus there are no significant differences between the results of the different laboratories.

The mean limits of error for the naphthoate are of the same order of magnitude as for the earlier experiments. The average weights $\overline{W}$ are:

the average errors of the naphthoate test are again smaller than those given in the Pharmacopoeia Addendum. The differences are in the same direction as in the earlier experiments,† suggesting that the improvement in technique has been maintained.

2. *Further points: heterogeneity in variability and slope*

The variances of increase in weight of animals on the same dose and the slopes in the different laboratories were examined for heterogeneity as before.

* Pooled estimate of variance used. Cf. p. 296, col. 2, where an individual estimate is used.

† 3-week test weight from Pharmacopoeia Addendum = 649, whence t = 0·5; 5-week test weight from Pharmacopoeia Addendum = 875, whence t = 2·5. The latter value of t is significant at the 5 % level.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 8. *Naphthoate experiment. Values of M* with their standard error and weights*

No.	Period weeks	Males			Females			Combined			Result units/g.	Limits of error ($P=0{\cdot}99$) %
		M	σ_M	W_M	M	σ_M	W_M	M	σ_M	W_M		
1 (*a*)	3 ♂	Ī·7256	0·1578	402	—	—	—	Ī·7256	0·1578	402	106	39–255
(*b*)	5 ♂	Ī·6845	0·1112	809	—	—	—	Ī·6845	0·1112	809	97	52–193
(*c*)	5 ♂ (selected)	Ī·8342	0·0602	2762	—	—	—	Ī·8342	0·0602	2762	137	70–143
2 (*a*)	3 ♀	—	—	—	0·0034	0·1197	698	0·0034	0·1197	698	202	49–203
(*b*)	5 ♀	—	—	—	0·1146	0·1247	643	0·1146	0·1247	643	260	48–209
3 (*a*)	3	Ī·7122	0·3677	74	Ī·9003	0·2269	194	Ī·8484	0·1932	268	141	32–315
(*b*)	5	Ī·6280	0·3914	65	Ī·7768	0·1433	487	Ī·7593	0·1346	552	115	45–222
4	5	Ī·9941	0·0536	3479	Ī·8528	0·0723	1911	Ī·9440	0·0430	5390	176	77–129
5	5	Ī·8100	0·0951	1106	Ī·8565	0·1470	463	Ī·8237	0·0798	1569	133	62–161
7	5 ♂	Ī·8679	0·1360	541	—	—	—	Ī·8679	0·1360	541	148	45–224
8	5	Ī·9017	0·0793	1590	Ī·8087	0·1351	548	Ī·8779	0·0684	2138	151	67–150
9 (*a*)	3 ♂	Ī·9298	0·0673	2206	—	—	—	Ī·9298	0·0673	2206	170	67–149
(*b*)	5 ♂	Ī·9539	0·0445	5040	—	—	—	Ī·9539	0·0445	5040	180	77–130
10	5 ♂	Ī·9637	0·0564	3142	—	—	—	Ī·9637	0·0564	3142	184	72–140
Average σ_M	3 weeks								0·1058			
	5 weeks								0·0643			0·0715
Mean limits of error per exp.	3 weeks										53–187 %	
	5 weeks										68–146 %	66–152

* M = logarithm of potency ratio. The standard contained 200 units/g.

Table 9. *Vitamin A β-naphthoate. Quantities used in calculating the error from internal evidence*

No.	Period weeks	Σ^2	D.F.	σ^2	D.F.	b	σ_b	b_1	b_2	b_1-b_2	$\sigma_{(b_1-b_2)}$	M	σ_M	Result
						Males								
1 (*a*)	3	86·20	68	56·12	34	16·9	7·3	6·8	26·9	−20·1	14·6	Ī·7256	0·1578	106
(*b*)	5	186·65	72	125·54	36	25·1	10·4	19·4	30·8	−11·4	20·8	Ī·6845	0·1112	97
(*c*)	5	124·25	36	43·45	18	47·3	11·7	46·2	48·5	− 2·3	23·4	Ī·8342	0·0602	137
3 (*a*)	3	100·32	9	—	—	18·8	19·1	18·1	19·9	− 1·8	38·4	Ī·7122	0·3677	103
(*b*)	5	89·16	8	—	—	20·7	18·6	− 1·3	59·0	−60·3	37·3	Ī·6280	0·3914	85
4	5	330·71	60	—	—	84·8	15·1	79·3	90·3	−11·0	30·2	Ī·9941	0·0536	197
5	5	122·58	26	119·42	10	49·8	13·5	45·5	54·2	− 8·7	26·9	Ī·8100	0·0951	129
7	5	146·63	37	167·17	18	31·9	11·5	11·6	18·0	− 6·4	25·1	Ī·8679	0·1360	148
8	5	104·55	36	99·24	14	41·4	10·8	55·1	26·2	+28·9	21·8	Ī·9017	0·0793	160
9 (*a*)	3	59·87	52	45·24	16	27·6	6·9	21·4	33·9	−12·5	13·7	Ī·9298	0·0673	170
(*b*)	5	90·55	54	71·29	18	50·5	8·3	49·7	51·4	− 1·7	16·6	Ī·9539	0·0445	180
10	5	247·36	40	49·79	18	39·1	15·8	33·5	44·5	−11·0	31·6	Ī·9637	0·0564	184
						Females								
2 (*a*)	3	80·61	16	48·51	8	26·3	13·3	30·1	22·5	+ 7·6	26·7	0·0034	0·1197	202
(*b*)	5	142·35	16	80·34	8	36·1	17·7	43·0	29·1	+13·9	35·5	0·1146	0·1247	260
3 (*a*)	3	126·90	36	—	—	16·6	11·8	22·3	10·0	+11·3	23·8	Ī·9003	0·2269	159
(*b*)	5	169·73	35	—	—	36·8	13·9	34·5	39·5	− 5·0	27·8	Ī·7768	0·1433	120
4	5	152·55	57	—	—	49·0	10·5	56·5	41·2	+15·3	21·0	Ī·7125	0·0723	143
5	5	58·50	10	—	—	30·9	13·7	41·2	20·5	+20·7	27·4	Ī·8565	0·1470	144
8	5 (selected)	58·32	31	18·89	12	18·7	8·6	25·5	11·6	+13·9	17·2	Ī·8087	0·1351	129

Σ^2 = variance of animals (not litter-mates) on same dose.
σ^2 = variance of animals (litter-mates) on same dose.
D.F. = no. degrees of freedom used in calculating corresponding variance.
b = best fitting slope for test and standard combined.
b_1 = slope from two doses of standard.
b_2 = slope from two doses of test.
M = log of result in units/mg.
σ_M = standard error of M.

In the 3-week assays there are no significant differences in variance from laboratory to laboratory.

In the 5-week assays the results of M. S. Bartlett's test for homogeneity of variance Σ^2 were as follows:

Values of χ^2	
Males	Females
36·84 (D.F. 7), $P=0{\cdot}0002$	12·25 (D.F. 4), $P=0{\cdot}02$

In the males the differences between the variances (σ^2) of litter-mates on the same dose are just significant ($\chi^2=11{\cdot}43$ with 5 D.F.), and the two values for the females in laboratories 2 and 8 differ significantly (variance ratio $=4{\cdot}25$, 5 % point $=2{\cdot}85$).

In the 5-week tests, the bucks are, as before, more variable than the does, the average variances being:

	Males	Females
Σ^2	175·72 (D.F. 297)	129·57 (D.F. 149)
σ^2	89·11 (D.F. 96)	43·47 (D.F. 20)

In the 3-week tests, the results are as follows:

	Males	Females
Σ^2	76·57 (D.F. 129)	112·64 (D.F. 52)
σ^2	52·64 (D.F. 50)	48·51 (D.F. 8)

Here in non-litter-mates the does are somewhat more variable; there is no significant difference between the sexes in the variability of litter-mates.

There are no significant differences in slope between the different laboratories. The values of

$$\chi^2=S\{w_b(b-\bar{b})^2\}$$

are as follows:

	Males	Females
3-week tests	0·22 (D.F. 2)	0·30 (D.F. 1)
5-week tests	10·7 (D.F. 7)	5·19 (D.F. 4)

This being so there is no need to examine the difference between individual pairs.

Slopes obtained from the standard and from the U.S.P. reference oil do not differ significantly in the same test. There are no significant differences between the average slopes for bucks and does in this experiment; the figures being:

	3-week tests		5-week tests	
	Males	Females	Males	Females
Mean	18·1	20·9	46·3	32·2
S.E.	6·5	8·8	5·3	6·0

The average slopes for the naphthoate 5-week tests do not differ significantly from the corresponding results for the H.L.O. and the U.S.P. reference oil.

PART IV

THE CONVERSION FACTOR

The main object of all these experiments was to re-examine the conversion factor of 1600 which had been provisionally allotted by the Second International Conference on Vitamin Standardization (1934) for converting the results of the spectrophotometric test for vitamin A into international units. Table 10 gives the results. For the H.L.O. the conversion factor is estimated to be 1570, for the U.S.P. reference oil 1820, for the naphthoate 1770. The logarithms of the conversion factors and their standard errors are shown in the table. From these it may be concluded that the three values do not differ significantly.* Thus they are consistent with the hypothesis that the same conversion factor holds for all three substances. Accepting this we may pool the results and reach a conversion factor of 1740 with limits of error ($P=0{\cdot}99$) of 93–107 % or 1620–1860.

The error of the spectrophotometric determinations is too small to affect the error of the conversion factors. Thus for the naphthoate the mean of the logarithms of the E values is $\bar{2}{\cdot}971$ with a standard error of 0·004, and this includes any effect of deterioration. The effect of this error would be to raise the error variance of the logarithm of the conversion factor by $(0{\cdot}004)^2$ which would not alter the value given in Table 10.

PART V

EXACT FIDUCIAL LIMITS

The method of calculation it has been customary to use hitherto, and which has been described in the Appendix to Part I, gives the standard error of the logarithm of the result. From this standard error, which is an approximation, percentage limits of error corresponding to a probability level P (usually 0·95 or 0·99) are obtained. The interpretation to be given to these limits of error is that in the long run in a proportion P of experiments similar to the one actually performed, with samples of the same material, the result will lie between the corresponding percentages of its true value.

We may also invert this statement and say that, in the long run, in a proportion P of experiments the true value will lie between the corresponding percentages of the observed result. We then have limits which vary from experiment to experiment, because the observed result varies, but which are calculable by a definite rule and which are such that, in the long run, in a proportion P of experiments the true value will lie between them. Such limits are called 'fiducial limits'.

The customary approximation and the corresponding limits of error may be used whenever the slope is well above the level at which it would be just significant at the significance level chosen. If the slope is not significant or only just significant the approximation becomes useless, the true limits of error becoming zero and infinity. This is only common sense, since if the slope does not differ significantly from zero, no differentiation of the different doses is possible and no estimate of the median effective dose can be made. Recent work of Bliss (1935), Fieller (1941) and others has shown how such 'fiducial limits' may be calculated. This has been done and the results compared with the approximate

* Writing u for the logarithm of the conversion factor and w for its weight (the reciprocal of its error variance) the value of $\chi^2=S\{w(u-\bar{u})^2\}$ is 4·76 with 2 D.F. The 5 % point is 5·99.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

limits already obtained. The exact fiducial limits are given by

$$\bar{x}_1-\bar{x}_2+\frac{b(\bar{y}_2-\bar{y}_1)}{b^2-t^2B}\pm\frac{t}{(b^2-t^2B)}\sqrt{\{A(b^2-t^2B)+B(\bar{y}_2-\bar{y}_1)^2\}}, \quad (1)$$

where

$$A=s^2\left(\frac{1}{n_{11}+n_{12}}+\frac{1}{n_{21}+n_{22}}\right), \quad B=\frac{s^2}{(\nu_1+\nu_2)\,d^2}.$$

These are the estimated variances of $\bar{y}_2-\bar{y}_1$ and b respectively.

All these symbols except s^2 and t have been defined in the Appendix to Part I, pp. 297 and 298. Apart from complications due to using litter-mates, s^2 must be replaced by our estimate of Σ^2 the variance of animals on the same dose, while t is the value of Fisher's 't' corresponding to the level of significance chosen (the 5 % level for $P=0{\cdot}95$ and the 1 % value for $P=0{\cdot}99$) with the number of degrees of freedom on which the estimate of Σ^2 is based.

The formula (1) can also be written

$$\bar{x}_1-\bar{x}_2+\frac{\bar{y}_2-\bar{y}_1}{b}+\frac{t^2B(\bar{y}_2-\bar{y}_1)}{b(b^2-t^2B)}\pm t\sqrt{\left[\frac{A}{b^2}+\frac{B(\bar{y}_2-\bar{y}_1)^2}{b^4}+\frac{At^2B}{b^2(b^2-t^2B)}+B(\bar{y}_2-\bar{y}_1)^2\left\{\frac{1}{(b^2-t^2)^2B}-\frac{1}{b^4}\right\}\right]}, \quad (2)$$

where the relation between the exact and approximate formulae is clearly shown, for the first two members in each term give the limits hitherto used.

If the slope is significant $b^2-t^2B>0$, and if it is only just significant $b^2-t^2B=0$. In this case the limits for the logarithm of the potency ratio become $-\infty$ and ∞ and for the ratio itself 0 to ∞. This is in accordance with common sense, since nothing can be learnt from a slope which does not differ significantly from zero.

It may also be noted that if $b\bar{y}_2>b\bar{y}_1$, the approximate formula places the middle of the 'fiducial range' too low, and if $b\bar{y}_2<b\bar{y}_1$ too high. Thus the approximate formula in addition to underestimating the width of the fiducial range may bias its position.

Table 10. *Results, conversion factors and their errors*

	H.L.O.	U.S.P. ref. oil	Naphthoate (all tests)
M	2·195	3·418	2·219
Results	157 units/g.	2619 units/g.	166 units/g.
Average for one laboratory:			
σ_M	0·0729	0·0559	0·0715
Limits of error ($P=0{\cdot}99$)	65–154 %	72–139 %	66–152 %
Mean for all laboratories:			
σ_M	0·0243	0·0177	0·0198
Limits of error ($P=0{\cdot}99$)	87–116 %	90–111 %	89–112 %
$E^{1\,\%}_{1\,\text{c.m.}}$	0·10	1·44	0·094
Conversion factor	1570	1820	1770
log C.F.	3·196	3·260	3·248
$\sigma_{\text{log. C.F.}}$	0·0243	0·0177	0·0198
No. of tests	9	10	13
Average no. of animals	53	58	55

If complete sets of four litter-mates, one on each dose, had always been used, s^2 could have been correctly replaced by our estimate of σ^2, the variance of litter-mates on the same dose. In these experiments litter-mates were nearly always used on corresponding doses of test and standard, but it was by no means always the case that each animal on the lower dose of the standard or the test substance had a litter-mate on the higher dose. Thus in A, s^2 could legitimately be replaced by the estimate of σ^2, but in B in general it could not. Yet for the purpose of calculating fiducial limits it is necessary to use the same estimate in both cases.

It has been decided to use Σ^2 throughout. This will, as a rule, tend somewhat to overestimate the error while the use of σ^2 would underestimate it, but not necessarily always. For the estimate of σ^2 is based on a smaller number of degrees of freedom than that of Σ^2, and the corresponding value of b^2-t^2B, at a given level of significance, may therefore be smaller if σ^2 is used in place of Σ^2. This has the effect of tending to increase the error when σ^2 is used.

The difficulty could be avoided by using only complete sets of four litter-mates, and this point is taken up again in Part VI.

The results are shown in Tables 11, 12 and 13. They show that the approximate limits are very considerably in error unless the slope is more than 5 times its standard error with at least 15 D.F. Even then bias may result unless $\bar{y}_2-\bar{y}_1$ is small, for example, note Exp. 9 (b) in the H.L.O. series. It will therefore be desirable in future to calculate exact fiducial limits for individual tests.

Where however it is desired to obtain an average error for a number of results from different laboratories, there seems no alternative but to use approximate weights, as has been described.

PART VI

METHOD OF STATISTICAL ANALYSIS NOW RECOMMENDED, WHEN A NUMBER OF COMPLETE SETS OF FOUR LITTER MATES ARE AVAILABLE

1. *Analysis when there are no incomplete sets*

The object of this section is to describe the best method of statistical analysis now available when the data

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 11. *Halibut-liver oil. Comparison of exact and approximate fiducial limits*

No.	$\bar{y}_2 - \bar{y}_1$	b/S.E. of b	D.F.*	Fiducial limits % Approx. $P=0.95$	Approx. $P=0.99$	Exact $P=0.95$	Exact $P=0.99$
				Males			
1	− 8·31	4·62	21	76–131	70–143	62–136	44–152
2	− 1·08	1·48	49	48–210	38–265	0–∞	0–∞
3	−15·96	2·08	15	38–262	28–354	0–∞	0–∞
4	− 7·16	5·13	36	76–131	70–143	72–132	60–147
5	−37·00	−2·50	1	46–217	36–276	0–∞	0–∞
6	+ 1·07	5·97	7	80–125	74–135	75–136	62–168
7	−11·33	0·14	2	0–∞†	0–∞†	0–∞	0–∞
8	− 4·27	4·19	28	74–135	68–148	64–141	47–165
9 (*a*)	−15·99	6·11	15	75–134	68–147	72–128	59–141
9 (*b*)	+18·13	5·29	26	78–128	73–138	75–158	69–210
				Females			
1	− 2·92	0·04	21	0–∞	0–∞	0–∞	0–∞
2	− 5·15	1·98	53	53–187	44–228	0–∞	0–∞
3	− 4·20	2·13	21	50–199	40–248	0–250	0–∞
4	−10·58	4·03	54	64–155	56–178	48–144	25–161
5	− 2·82	1·33	11	50–201	40–251	0–∞	0–∞
6	− 7·52	1·87	9	37–270	27–370	0–∞	0–∞
7	− 6·67	0·68	9	2·2–4480	0·7–14800	0–∞	0–∞
8	− 5·24	5·24	30	79–127	73–137	69–131	56–144

* Number of degrees of freedom on which the S.E. of b is based.
† Upper limit $10^{23\cdot7}$% ($P=0.95$), $10^{33\cdot1}$% ($P=0.99$).

Table 12. *U.S.P. reference oil. Comparison of exact and approximate fiducial limits*

No.	$\bar{y}_2 - \bar{y}_1$	b/S.E. of b	D.F.*	Fiducial limits % Approx. $P=0.95$	Approx. $P=0.99$	Exact $P=0.95$	Exact $P=0.99$
				Males			
2	−3·00	3·10	42	76–132	69–144	49–168	18–248
3	+0·44	3·87	42	70–142	63–159	66–154	52–201
4	−1·58	5·57	19	78–128	73–138	75–131	67–155
5	+5·53	5·94	14	79–126	74–136	78–134	70–158
6	+4·08	4·29	20	72–139	65–154	71–160	61–228
7	+4·46	3·42	28	66–152	58–173	65–198	52–447
8	−0·33	3·42	23	70–144	62–161	59–167	36–261
9	−8·27	1·47	16	29–344	20–508	0–∞	0–∞
10	+7·55	9·27	50	86–116	82–122	86–118	82–126
				Females			
1	−1·45	3·92	36	73–137	66–151	63–148	46–181
2	−0·59	3·38	43	74–134	68–147	59–164	38–235
3	+0·55	5·87	43	79–126	74–136	78–129	71–144
4	+5·67	1·95	29	43–232	33–303	0–∞	0–∞
5	−2·60	3·82	10	80–125	74–134	57–156	27–221
6	+6·45	3·86	11	68–146	61–165	67–198	54–547
8	−2·58	3·99	27	71–141	64–157	63–146	45–178

* Number of degrees of freedom on which the S.E. of b is based.

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

consist mainly of observations on sets of four littermates, one member of each set being on each dose. (The two doses of the test substance and the two of standard are taken, of course, to be in the same ratio.) The method meets the difficulty noticed in the last section, because the slope (b) and the average difference response ($\bar{y}_2 - \bar{y}_1$) are always estimated from differences between isogenic pairs.

We will begin with a case when there are no incomplete sets of four.

Table 14 gives the results (laboratory 1 c) of a 5-week test on selected normally reacting litters, together with the complete calculation of the result and its error. The estimate of error is obtained from the analysis of variance of the results. The analysis of variance pro-

degrees of freedom, which are equal to the numbers of independent squares in the corresponding components, certain 'mean squares' are obtained. It has been shown that if there are no real dose or litter effects all these three mean squares would be statistically independent and have the same average value, only differing in a particular case owing to experimental error. If, on the other hand, there are real dose or litter effects the corresponding mean squares will exceed the third 'mean square' which provides the estimate of error. The significance of the difference between doses is tested by calculating the ratio ('variance ratio') of the mean square due to doses and that due to error. If the value of the variance ratio so obtained exceeds the 5% point, which may be obtained from a table, the effect of doses

Table 13. *Vitamin A β-naphthoate. Comparison of exact and approximate fiducial limits*

				Fiducial limits %			
				Approx.		Exact	
No.	$\bar{y}_2 - \bar{y}_1$	b/S.E. of b	D.F.*	$P = 0{\cdot}95$	$P = 0{\cdot}99$	$P = 0{\cdot}95$	$P = 0{\cdot}99$
				Males			
1 (*a*)	−4·64	2·32	68	49–203	39–255	2–181	0–∞
(*b*)	−7·92	2·41	72	61–165	52–193	2–179	0–∞
(*c*)	−7·85	4·04	36	76–131	70–143	55–142	33–160
3 (*a*)	−4·07	0·98	9	19–526	11–886	0–∞	0–∞
(*b*)	−6·67	1·11	8	17–585	8–1261	0–∞	0–∞
4	−0·47	5·62	60	79–127	73–137	77–130	69–145
5	−9·46	3·69	26	65–154	57–176	47–150	20–172
7	−3·98	2·77	37	54–185	45–224	20–184	0–∞
8	−3·44	3·83	36	70–143	62–160	59–146	39–174
9 (*a*)	−1·94	4·00	52	74–135	67–149	63–143	46–168
(*b*)	−2·33	6·08	54	82–122	77–130	77–126	69–137
10	−1·42	2·47	40	78–129	72–140	31–231	0–∞
				Females			
2 (*a*)	+0·90	1·98	16	58–172	49–203	0–∞	0–∞
(*b*)	+4·13	2·04	16	57–176	48–210	0–∞	0–∞
3 (*a*)	−1·66	1·41	36	36–278	26–384	0–∞	0–∞
(*b*)	−8·47	2·65	35	52–191	43–234	13–167	0–∞
4	−7·45	4·67	57	72–139	65–154	63–135	48–149
5	−4·43	2·26	10	52–194	42–239	0–235	0–∞
8	−3·75	2·17	31	54–184	45–223	0–197	0–∞

* Number of degrees of freedom on which the S.E. of b is based.

cedure is now familiar to most biological statisticians, but not to all workers in this field. The calculations will therefore be explained in detail.

The total sum of squares of the deviations of the observations (increases in weight) from their mean can be split up into three portions, the first depending on the differences between doses, the second on the differences between litters and the third solely on the differences from one litter to another of the differences between litter-mates. The first component where there are k letters is k times the sum of the squares of the deviation of the dose means from the general mean ($k = 10$ here). The second is 4 times the sum of the squares of the deviations of the litter means from the general mean. The third is obtained by subtraction but can also be derived explicitly as shown below. When these are divided by the corresponding numbers of

is judged to be significant. The 5% point is the value which would only be exceeded by chance once in 20 times if doses had no differential effect. It depends only on the number of degrees of freedom for doses and for error. The degrees of freedom as well as the sum of squares are additive. The total number of degrees of freedom is one less than the total number of observations (39), that for doses is one less than the number of doses (3), that for litters is one less than the number of litters (9), while that for error is $(3 \times 9 = 27)$. The sum of these $(3 + 9 + 27)$ is equal to the total number, 39.

Thus we are able to tell whether the growth response differs significantly from dose to dose or from litter to litter.

The numerical procedure is as follows. The sum of all the observations is 1729 the mean 43·225. The observations are summed in rows and columns, the sum of the squares of all the observations is 81857, the sum

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 14. *Calculation of the result and its error from a test based on isogenic sets only. Lab.* 1c. *Five-week test on selected normally reacting litters. Complete Litters*

	Standard		Test		
	y_{11} Dose 1	y_{12} Dose 2	y_{21} Dose 1	y_{22} Dose 2	Total
	38	44	31	42	155
	57	62	50	64	233
	29	45	31	52	157
	23	38	16	34	111
	26	59	29	44	158
	32	40	24	47	143
	43	51	30	40	164
	39	62	29	42	172
	51	74	35	49	209
	64	66	45	52	227
$S(y)$	402	541	320	466	1729
$\bar{y}$	40·2	54·1	32·0	46·6	43·225
$S(y^2)$	17830	30607	11086	22334	81857

		Sum of squares		D.F.		Mean square	
Doses	Slope	2030·625		1			
	Test *v.* standard	616·225	2648·075	1	3	882·692	
	Between slopes	1·225		1			
Litters			3360·725	9		373·414	V.R. = 9·07, sign.
Error	Slope	330·125		9			
	Test *v.* standard	558·525	1112·175	9	27	41·192	(s^2)
	Between slopes	223·525		9			
Total			7120·975	39			

s^2 is the estimate of σ^2, the variance of litter-mates on the same dose

	$u_1=(-y_{11}+y_{12}-y_{21}+y_{22})$	$u_2=(-y_{11}-y_{12}+y_{21}+y_{22})$	$u_3=(y_{11}-y_{12}-y_{21}+y_{22})$
	17	− 9	+ 5
	19	− 5	+ 9
	37	+ 9	+ 5
	33	−11	+ 3
	48	−12	−18
	31	− 1	+15
	18	−24	+ 2
	36	−30	−10
	37	−41	− 9
	9	−33	+ 5
Sum	+285	−157	+ 7
Sum of squares	9443	4699	899
Mean	+28·5	−15·7	+ 0·7
$\frac{1}{4}S(u-\bar{u})^2$	330·125	558·525	223·525

$$\tfrac{1}{40}\{(285)^2+(-157)^2+(7)^2\}=2648\cdot075.$$

of squares of deviations from the mean is

$$81857-(1729)^2/40=7120\cdot975.$$

The sum of squares of deviations due to doses is

$$\tfrac{1}{10}\{(402)^2+(541)^2+(320)^2+(466)^2\}-\frac{(1729)^2}{40}=2648\cdot075.$$

That due to litters is

$$\tfrac{1}{4}\{(155)^2+(233)^2+\ldots+(227)^2\}-\frac{(1729)^2}{40}=3360\cdot725.$$

The error term 1112·175 is then obtained by subtraction. Only the error term is actually required in the subsequent calculations to obtain the errors of the slope and of the result; but considerable light is thrown on the experiment by carrying the analysis of variance still further.

Let us call the responses to doses 1 and 2 of standard y_{11}, y_{12} that to doses 1 and 2 of test y_{21}, y_{22}.

The quantities $u_1=(-y_{11}+y_{12}-y_{21}+y_{22})$ provide a measure of the slope, $u_2=(-y_{11}-y_{12}+y_{21}+y_{22})$ a measure of the difference in response to test and standard, $u_3=(y_{11}-y_{12}-y_{21}+y_{22})$ a measure of the difference between the slopes for test and standard respectively.

These quantities are calculated for each litter, set down in columns and totalled. The totals are respectively

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

Table 14 (*continued*)

All available data

$y_{12}-y_{11}$	$y_{22}-y_{21}$	$y_{21}-y_{11}$	$y_{22}-y_{12}$
6	11	− 7	− 2
5	14	− 7	+ 2
16	21	+ 2	+ 7
15	18	− 7	− 4
33	15	+ 3	−15
8	23	− 8	+ 7
8	10	−13	−11
23	13	−10	−20
23	14	−16	−25
2	7	−19	−14
139	146	−82	−75

$b=\dfrac{285}{20\times 0\cdot 30103}=47\cdot 337$

$b^2=2240\cdot 792$

$\bar{y}_2-\bar{y}_1=\dfrac{-157}{20}=-7\cdot 85$

$(\bar{y}_2-\bar{y}_1)^2=61\cdot 6225$

$A=4\cdot 1192$

$B=45\cdot 456$

$\sigma_b=6\cdot 7421$

$M=-0\cdot 1658$

$=\bar{1}\cdot 8342$

Ratio $=0\cdot 6827$

Result 137 units/g.

Approx. $\sigma_M{}^2=0\cdot 0018383$

$0\cdot 0005579$

$0\cdot 0023962$

$\sigma_M=0\cdot 0490$

$1\cdot 96\ \sigma_M=0\cdot 0960$

$2\cdot 576\sigma_M=0\cdot 1262.$

	t	t^2	b^2-t^2B	(1) $b(\bar{y}_2-\bar{y}_1)/(b^2-t^2B)$	(2) $A(b^2-t^2B)+B(\bar{y}_2-\bar{y}_1)^2$
$P=0\cdot 95$	2·052	4·2107	2049·390	−0·1813	11242·960
$P=0\cdot 99$	2·771	7·6784	1891·763	−0·1964	10593·663

	$\sqrt{(2)}$	(3) $t\sqrt{(2)}/(b^2-t^2B)$	(1)−(3)	(1)+(3)	Limits in logs	
$P=0\cdot 95$	106·033	0·1062	−0·2875	−0·0751	$\bar{1}\cdot 7125$	$\bar{1}\cdot 9249$
$P=0\cdot 99$	102·926	0·1508	−0·3472	−0·0456	$\bar{1}\cdot 6528$	$\bar{1}\cdot 9544$

	Limits		% of actual
$P=0\cdot 95$	0·5158	0·8412	76–123 %
$P=0\cdot 99$	0·4496	0·9003	66–132 %

Limits of error %	
Approx.	Exact (fiducial)
80–125	76–123
75–134	66–132

+285, −157 and +7. $(285)^2/40$ (the divisor 40 occurs because each total is a linear function of all 40 observations) is the contribution of the slope to the sum of squares 'between doses'. Similarly $(157)^2/40$ is the contribution of the difference in response to test and standard and $(7^2)/40$ is the contribution of the difference between slopes. The sum of these three contributions is equal to 2648·075 and so agree with the sum of squares 'between' doses already calculated. Each of these components can be tested for significance against the error term; the slope is of course significant, so is the average difference in response between test and standard; the difference between slopes for test and standard is not significant. The last result justifies the use of an average slope. An experiment should aim at equalizing the average responses to test and standard; the second result shows that this was not completely achieved in this case, the error of the assay will be increased in consequence. To get a significant slope is of course essential if the experiment is to be interpreted at all.

The quantities u_1, u_2, u_3 can also be used to subdivide the error term and thus throw considerable light on its structure. They each depend only on differences between litter-mates, hence their variation from litter to litter clearly provides an estimate of the variance of litter-mates on the same dose. If for the ten litters the quantities $\frac{1}{4}S(u_1-\bar{u}_1)^2$, $\frac{1}{4}S(u_2-\bar{u}_2)^2$, $\frac{1}{4}S(u_3-\bar{u}_3)^2$ are calculated, their sum is found to be equal to the error term already obtained. They can be regarded as the contributions to the error term made by the variation

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

of (1) slope, (2) differences in response to test and standard, (3) the difference between slopes. It now remains to calculate the result of the assay, and its error. The differences $y_{12}-y_{11}, y_{22}-y_{21}, y_{21}-y_{11}, y_{22}-y_{12}$ are written down for each litter and summed. We then have for the slope

$$b=\frac{1}{2kd}\{S(y_{12}-y_{11})+S(y_{22}-y_{21})\}, \quad (1)$$

where k is the number of litters (10) and d the logarithm of the dose interval (0·30103). For the average difference in response to test and standard we have

$$\bar{y}_2-\bar{y}_1=\frac{1}{2k}\{S(y_{21}-y_{11})+S(y_{22}-y_{12})\}, \quad (2)$$

while the logarithm of the potency ratio

$$M=(\bar{y}_2-\bar{y}_1)/b.$$

Thus we find the potency ratio to be 0·6827, and since the standard solution contained 200 units/g. the result is 137 units/g.

equations which can only be solved by successive approximation and the computations become laborious.

Little will be lost, as a rule, by estimating the variance of litter-mates on the same dose from the complete litters, as before, and then using all the available isogenic pairs for the calculation of b and $\bar{y}_2-\bar{y}_1$. The procedure is exactly the same as before, except that $2k$ in equations (1), (2) and (3) above must be replaced by the number of pairs actually used. This number need not be the same for b and $\bar{y}_2-\bar{y}_1$, nor does it matter if different numbers of pairs are used from the test and standard substances or from the upper and lower doses. The value of $2k$ in A will be the same as that for $\bar{y}_2-\bar{y}_1$, while that in B will be the same as for b.

3. *Application of the method to the vitamin A 2-naphthoate*

This method has been applied, in the cases where it is appropriate, to the assays of vitamin A 2-naphthoate. Table 15 shows the results.

Table 15. *Errors of certain assays of vitamin A 2-naphthoate when full use is made of the fact that litter-mates have been used*

Assay	$\bar{y}_2-\bar{y}_1$	b/S.E. of b	P.F.	Approx. limits of error % ($P=0·95$)	Approx. limits of error % ($P=0·99$)	Exact fiducial limits % ($P=0·95$)	Exact fiducial limits % ($P=0·99$)	Comparison with previous method
				Males				
1 (*a*)	−4·64	2·90	51	53–188	43–231	19–163	0–190	Slight improvement
1 (*c*)	−7·85	7·02	27	80–125	75–134	76–123	66–132	Improvement
9 (*a*)	−1·94	2·25	24	58–172	49–204	2–180	0–∞	Worse
9 (*b*)	−2·34	5·81	27	82–122	77–130	79–120	77–130	Slight improvement
10	−1·43	6·09	27	81–124	75–133	78–126	70–138	Improvement
				Females				
2 (*a*)	0·90	2·81	12	62–162	53–190	48–267	.0–∞	Improvement
2 (*b*)	4·13	3·06	12	62–161	54–187	59–291	36–∞	Improvement

It remains to calculate the error of the result. We first calculate

$$A=\frac{2s^2}{2k} \quad \text{and} \quad B=\frac{2s^2}{2kd^2}=\sigma_b^2. \quad (3)$$

The approximate standard error of M is then given by

$$\sigma_M^2=\frac{A}{b^2}+\frac{(\bar{y}_2-\bar{y}_1)^2B}{b^4}, \quad (4)$$

and the approximate limits of error are as usual 100 antilog $1·960\sigma_M$ and 100 antilog $2·576\sigma_M$ for ($P=0·95$) and ($P=0·99$) respectively. The exact fiducial limits will now be given by formula (1) on p. 306 with $\bar{x}_1-\bar{x}_2=0$. The 5 and 1 % values of t for 27 D.F. must be obtained from the appropriate table (Fisher, 1941; Fisher & Yates, 1943), and then the calculation proceeds as is shown in Table 14.

2. *Analysis when there are mainly complete sets of litter-mates, with a few incomplete sets*

The problem of dealing with analysis of variance when a number of observations are missing has been studied in detail by Yates (1933). He has shown that the theoretically correct procedure is to estimate the missing observations by minimizing the error term in the analysis of variance. However, this results in a system of

Comparison with the method previously used (see Table 12) shows an improvement in all cases except 9 (*a*). The improvement in the approximate limits is only slight, that in the exact fiducial limits much more marked. The relation between the exact and the approximate limits is much the same as before.

Assay 9 (*a*) requires special examination. Using the original method 52 D.F. were available for the estimate of Σ^2 which has a value of 59·9; using the present method the difference between litters is not significant, and the estimate of σ^2 is 42·7 with only 24 D.F. Then there is a negligible gain in accuracy in using σ^2 and there is a loss in precision owing to the greater sampling error of the latter estimate. Further, the original method using all fifty-six animals gave a slope b of 27·6 with a standard error of 6·9. The present method which uses all the available isogenic pairs of which there are 18 gives $b=16·2$ with a standard error of 7·2. These two values are not significantly different, but while the former is 4·0 times its standard error the latter is only 2·25 times. This again increases the error of the result. Thus the fact that the latter method uses less data together with an unlucky slope more than outweighs the increased accuracy due to taking litter-mates into account. This, however, is exceptional and there is no

https://doi.org/10.1017/S0022172400013000 Published online by Cambridge University Press

doubt that as a rule the present method will give greater accuracy.

I must acknowledge my indebtedness to Mr E. A. G. Shrimpton, Mr O. Kempthorne and Mrs I. Mathison, who at one time or another have helped in the many numerical calculations involved. Dr K. H. Coward and Dr S. K. Kon gave valuable help in writing the introduction to this paper. Miss Hume's kindness and patience in co-operation deserves a special note.

APPENDIX. GENERAL PLAN OF EXPERIMENTS

I. The halibut-liver oil

Vitamin A sub-committee

General scheme. Spectrophotometric estimation of the absorption at 325 mμ, and biological estimation by as many laboratories as possible, are to be made on specially selected materials. The spectrophotometric observations are to be conducted by Dr R. A. Morton, with the help of Dr J. R. Edisbury.

Material. The material selected is a commercial sample of halibut-liver oil of which one gallon, of blue value about 2000, has been obtained through Dr Levern of the Torry Research Station, Aberdeen, and one half gallon, of blue value about 3200, has been presented to the Medical Research Council through Mr J. A. Cathcart, by the generosity of Messrs Parke, Davis. These two samples will be mixed and the portion, not immediately used for this research will be stored in the low temperature storage chamber at the Torry Research Station. The portion to be utilized in this investigation will be sent to Dr R. A. Morton of Liverpool University, who will prepare from some of it, after saponification, a sterol-free concentrate.

Biological tests will be made to compare the halibut-liver oil and the concentrate derived from it with the international β-carotene standard.

Spectroscopic tests will be made to determine $E^{1\%}_{1\,cm.}$ 325 mμ for the oil and the concentrate at the beginning of the experiment. They will also be made on the diluted solutions of those two substances, which are actually fed to the experimental animals, at the beginning and end of the biological tests, to ascertain whether the stability of the feeding solutions has been maintained throughout. Colorimetric and spectroscopic tests will be made in the same way to test the stability of the standard β-carotone solution used in the biological tests.

It is hoped that the test materials will be ready about the third week in April. Experimental rats should therefore be prepared for this date, as the materials should be tested biologically as soon as possible after preparation, otherwise a spectroscopic re-test for stability would be necessary before the biological test could begin and even so it would not be very satisfactory.

Diluent. A peroxide-free coconut oil will be selected by Mr MacLennan of Messrs Levers and will be saturated with hydroquinone to produce a 0·05% solution, as described by Morgan (*Biochem. J.* 28, 1180). This will be used for all dilutions, and a supply will be sent to all participants for equalizing the amount of solvent administered to test animals receiving the smaller doses.

Dosage. The dosage of oil and concentrate shall be based on the value of $E^{1\%}_{1\,cm.}$ 325 mμ reported by Dr Morton. Two doses of each material shall be administered to provide 2 and 4 international units (i.u.) of vitamin A daily, as calculated by the use of a conversion factor of 2000 for both oil and concentrate.

The daily doses of β-carotene standard per rat shall also be 2 and 4 i.u.

Dilution. The dilutions of the oil and the concentrate shall be carried out by Dr Morton, and of the β-carotene standard by Dr Hartley. The actual solutions to be fed will thus be identical for all the participants, and each will receive two bottles of each solution, one of which is for use and the other to be opened only in emergency, apart from which it is to be retained unopened for reference. Dr Morton will also retain a certain amount of each material, stored at 0° C., as a check.

Strength of the feeding solution. The strength of all three feeding solutions shall be such that 0·020 g. (a conveniently sized drop) of coconut oil contains 4 i.u. (2·4 μg. of standard β-carotene, or an amount of the oil or of the concentrate, calculated by means of a 2000 conversion factor to contain 4 i.u.).

Bottling of the feeding solution. The feeding solution shall be bottled in wide-mouthed bottles of 25 c.c. capacity, stoppered without an air space, of which each participant shall receive two for each solution. The total amount of each test material issued to each participant will therefore be 50 c.c. containing 10,000 i.u. (as calculated above).

Storage of test materials and feeding of experimental rats. The two bottles of each solution shall be stored at about 0° C. The test animals shall only be dosed two or three times a week so that the approximate amount of each solution needed for dosing at one time can be chipped out of the stock bottle into a small container without removing it from the cold store. Any residue after a day's dosing should be discarded and not returned to the stock bottle. The solution needed for dosing should be melted by placing the small container in warm water (not above 50° C.).

The solution should be administered from a dropping pipette, previously calibrated, by weighing the drops, to deliver a drop of 20 mg. weight. (A glass pipette, the external diameter of whose tip is about 2 mm., yields a drop of 15–20 mg. weight, when held vertically.) If the same pipette is washed out and dried and used for all solutions, a complete similarity of dosage will be ensured.

As soon as the biological experiment is complete, the two bottles (opened and unopened, if both survive) of each of the three solutions, oil, concentrate and β-carotene, shall be returned *immediately* to Dr Morton for tests on stability. The dates on which the bottles were first opened and on which the feeding tests were ended should be stated. Participants can make their own tests for stability from time to time throughout the

biological experiment, if they wish, but reliable tests for stability on the oil and the concentrate can only be made after saponification, owing to the diluteness of the feeding solutions.

Conditions for the biological experiments. Each of the four experimental groups should include as many rats as possible, between five and ten in each group, litters and sexes being adjusted as evenly as possible. All the comparisons should be simultaneous. The period of feeding the test materials should be 4 or 5 weeks. All rats should receive the same number of drops of coconut oil, whether as solution or as coconut oil only.

II. THE U.S.P. REFERENCE OIL

Vitamin A sub-committee

It has been decided to make a spectrophotometric and biological test on the U.S.P. reference cod-liver oil (oil X), and the methods to be followed are similar to those used in the preceding investigation.

Biological tests will be made to compare oil X with the international β-carotene standard.

Spectroscopic tests will be made by Drs Morton and Edisbury to determine $E^{1\%}_{1\,cm.}$ 325 mμ for oil X at the beginning of the experiment. They will also be made on the diluted solutions actually fed to the experimental animals, at the beginning and end of the biological tests, to ascertain whether the stability of the feeding solution has been maintained throughout. Colorimetric and spectroscopic tests will be made in the same way to test the stability of the standard β-carotene solution used in the biological tests.

It is hoped that the test materials will be ready early in April. Experimental rats should therefore be prepared for this date, as the materials should be tested biologically as soon as possible after preparation.

Diluent. A peroxide-free coconut oil saturated with hydroquinone to produce a 0·05 % solution, as described by Morgan (*Biochem. J.* **28**, 1180) has been used as diluent. A supply will be sent to all participants for equalizing the amount of solvent administered to test animals receiving the smaller doses.

Strength of the feeding solutions. The solution of β-carotene contains 200 i.u. of vitamin A potency per g. and the solution of oil X approximately that amount. Oil X, the official potency of which is 3000 i.u./g. has been diluted 1 in 12·5 with coconut oil.

Dosage. The daily doses of β-carotene standard per rat shall be 2 and 4 i.u., i.e. 10 and 20 mg. of solution. The doses of oil X shall be 10 and 20 mg. of solution daily.

The doses may be varied to suit the differing stocks of rats, but they must all bear the above relation to one another.

Dilution. The dilution of the oil has been carried out by Drs Morton and Edisbury, and of the β-carotene standard by the Department of Biological Standards. The actual solutions to be fed will thus be identical for all the participants, and each will receive two bottles of each solution, one of which is for use and the other to be opened only in emergency, apart from which it is to be retained unopened for reference. Dr Morton will also retain a certain amount of each material, stored at 0° C. as a check.

Bottling of the feeding solutions. The feeding solution of oil X will be bottled in wide-mouthed bottles of 2 oz. capacity, filled with CO_2 of which each participant will receive two. The total amount of material issued to each participant will therefore be about 80 c.c. containing 16,000 (approx.) i.u.

The feeding solution of β-carotene will be an amount of 50 g. (10,000 i.u.) probably in 2 × 20 c.c. + 1 × 10 c.c. bottles as before.

Storage of test materials and feeding of experimental rats. The two bottles of each solution be stored at about 0° C. The test animals shall only be dosed two or three times a week, so that the approximate amount of each solution needed for dosing at one time can be chipped out of the stock bottle into a small container without removing it from the cold store. Any residue after a day's dosing should be discarded and not returned to the stock bottle. Any small loose crumbs of material, particularly discoloured, should be rejected. The feeding solutions may be stored in an atmosphere of inert gas if participants can arrange to refill the container without removing it from the refrigerator. They should report if they have done this or not.

The solution needed for dosing should be melted by placing the small container in warm water (not above 50° C.). It should be administered from a dropping pipette, previously calibrated, by weighing the drops, to deliver a drop of about 20 mg. weight. (A glass pipette, the external diameter of whose tip is about 2 mm. yields a drop of about 15 mg. weight, when held vertically.) If the same pipette is washed out and dried and used for all three solutions, a complete similarity of dosage will be ensured.

As soon as the biological experiment is complete, the two bottles (opened and unopened, if both survive) of each of the two solutions, oil, and β-carotene, shall be returned *immediately* to Dr Morton for tests on stability. The dates on which the bottles were first opened and on which the feeding tests were ended should be stated.

Conditions for the biological experiments. Each of the four experimental groups should include as many rats as possible, between five and ten in each group, litters and sexes being adjusted as evenly as possible. All the comparisons should be simultaneous. The period of feeding the test materials should be 5 weeks. All rats should receive the same number of drops of coconut oil, whether as solution or as coconut oil only. The test period should be reckoned to commence from the first day of dosing.

The experiments should include some negative controls.

III. THE SOLID VITAMIN A 2-NAPHTHOATE

Vitamin A sub-committee

It has been decided to make a spectrophotometric and biological test on the solid vitamin A 2-naphthoate prepared by Dr T. H. Mead of the British Drug Houses. The test will be a co-operative one, similar to the two already carried out.

Biological tests will be made to compare a solution of the vitamin ester with the international β-carotene standard.

Spectroscopic tests will be made by Dr R. A. Morton to determine $E^{1\%}_{1\,cm.}$ 325 mμ for the ester. They will also be made on the feeding solutions at the beginning and end of the biological tests, to ascertain whether the stability of the feeding solutions has been maintained

throughout. Colorimetric and spectroscopic tests will be made in the same way to test the stability of the standard carotene solution used in the biological tests.

Diluent. Arachis oil containing 0·01 % quinol has been used as solvent for both the ester and the carotene. A bottle containing about 20 g. will be supplied to each participant for equalizing the doses. The density of the oil is 0·92.

Strength of the feeding solutions. The solution of β-carotene in arachis oil contains 200 i.u. of vitamin A potency per g. and should not be diluted if this can be avoided. Each participant will receive 50 g. of this solution (10,000 i.u.) from the Department of Biological Standards, National Institute of Medical Research, Hampstead, N.W., to which application should be made for further supply should that first issued not prove enough.

The strength of the solution of vitamin A 2-naphthoate has been calculated so that it may be approximately the same as that of the carotene solution (i.e. 200 i.u./g.). The calculation has been based on the findings of Underhill & Coward (*Biochem. J.* **35**, 599) that the biological value of this ester is 2,225,000 i.u./g. Each participant will receive an ampoule containing 20 g. of this solution sealed in nitrogen from the British Drug Houses, from whom more may be obtained if necessary.

Dosage. The daily doses of carotene standard per rat shall if possible be 2 and 4 i.u., i.e. 10 and 20 mg. of solution. They may however be varied to suit the differing stocks of rats but they *must* all bear the above relation to one another.

Storage of test solutions and feeding of experimental animals. The carotene solution and the ester solution should be stored at 0° C. in the dark. The amount of carotene solution required for feeding each day may be withdrawn after melting the solution.

When the ester solution is required for the test, the entire contents of the ampoule should be transferred to a boiling tube provided with a well-fitting rubber bung. This latter carries inlet and outlet tubes for nitrogen. The inlet reaches to the bottom of the boiling tube and both gas leads are provided with well fitting 'stops'. Such an apparatus is provided with the ampoule of ester solution.

Arachis oil solidifies at 0° and it will, therefore, be necessary to melt the solution completely before taking out doses for feeding. This may be carried out by keeping the closed tube immersed in water at 40° C. up to the level of oil in the tube. When the contents have melted it is very important to remove the stops and pass a stream of nitrogen for a short period to ensure homogeneity and minimize risk of error, which might occur if some of the naphthoate crystallized out in the cold. The walls of the tube near the mouth should be kept free from oil to prevent swelling of the rubber bung. When the required amount of solution for feeding has been withdrawn the boiling tube should be flushed with nitrogen, closed and replaced in the refrigerator as quickly as possible.

The dose may be administered from a dropping pipette, previously calibrated by weighing the drops to deliver a drop of about 20 mg. weight. If the same pipette is washed out and used for both the carotene and the ester solution, a complete similarity of dosage will be ensured. If the doses are administered once or twice a week instead of every day the test solutions will have less opportunity for deterioration.

Stability tests. As soon as the biological experiment is completed or as soon as the first supply of either solution has been used up all except an amount of about 2 g., that amount at least of each solution should be sent at once to Dr R. A. Morton, Chemical Laboratory, The University, Liverpool. In order to be sent through the post the solutions may be transferred to small bottles or tubes, which should be completely filled except for a small bubble of nitrogen to allow for thermal expansion. It is hoped that, if the above precautions are observed, the test solutions will remain stable for the period of the test. It is not, however, absolutely certain that this will be so. The stability of the feeding solutions will be watched throughout, and if they are found to be deteriorating a fresh supply will be issued. Participants are urged to complete the whole experiment as quickly as possible and, if they have the necessary apparatus, to test the stability of their own feeding solution each week. If they find deterioration they should report it to the Secretary of the Vitamin A Sub-committee at once.

Conditions for the biological experiment. Each of the four experimental groups should include as many rats as possible, between ten and twenty in each group, litters and sexes being adjusted as evenly as possible. All the comparisons should be simultaneous. The period of administrating the test materials should be five weeks. All rats should receive the same number of drops of arachis oil whether as solution or as arachis oil only. The test period should be reckoned to begin from the first day of dosing. The experiments should include some negative controls.

No suggestions are made as to the basal diet for the experimental animals but every participant should give this special consideration in order to see if there is any alteration which can be made to ensure that deficiency of vitamin A is the only limiting factor.

REFERENCES

Bartlett, M. S. (1937*a*). *Proc. Roy. Soc.* A, **160**, 268.

Bartlett, M. S. (1937*b*). *J. Roy. Statist. Soc. Suppl.* **4**, 158.

British Pharmacopoeia Commission, General Medical Council (1936). *Report of the Sub-committee on the Accuracy of Biological Assays*, p. 10.

Bliss, C. I. (1935). *Ann. Appl. Biol.* **22**, 325.

Coward, K. H. (1933). *Biochem. J.* **27**, 443.

Fieller, E. C. (1941). *J. Roy. Statist. Soc. Suppl.* **7**, 1.

Fisher, R. A. (1941). *Statistical Methods for Research Workers*, 8th ed. Table IV. London: Oliver and Boyd.

Fisher, R. A. & Yates, F. (1943). *Statistical Tables for Biological, Agricultural and Medical Research*, 2nd ed. Table III. London: Oliver and Boyd.

Hume, E. M. M. (1937). *Nature, Lond.*, **139**, 467.

Hume, E. M. M. (1939). *Nature, Lond.*, **143**, 22.

Hume, E. M. M. (1943). *Nature, Lond.*, **157**, 535.

Irwin, J. O. (1937). *J. Roy. Statist. Soc. Suppl.* **4**, 1.

Yates, F. (1933). *Emp. J. Exp. Agric.* **1**, 129.

(*MS. received for publication* 14. VI. 43.—Ed.)