THE STATISTICAL TREATMENT OF MEASUREMENTS OF THE CARCINOGENIC PROPERTIES OF TARS (PART I) AND MINERAL OILS (PART II)*

BY J. O. IRWIN, SC.D., D.SC., WITH THE ASSISTANCE OF NANCY GOODMAN, B.A.
Of the Statistical Staff, Medical Research Council

(With 9 Figures in the Text)

CONTENTS

GENERAL INTRODUCTION

A. THE PROBLEM

That contact with tars or mineral oils may produce skin cancers in man has been known for some time. The table on p. 363 gives the cases and deaths from epitheliomatous ulceration reported to the Factory Department of the Home Office between 1923 and 1936.† Among workers exposed to contact with mineral oils, by far the largest proportion of cases and deaths occur among cotton-mule spinners, the scrotum being, as a rule, attacked. Between 1923 and 1936 there were 918 cases of mule spinners' cancer reported, of which 305 were fatal (*Report for* 1936, p. 48). The Factory Department of the Home Office has analysed the time elapsing between commencement of employment and manifestation of the disease in 562 persons, in whom the growth occurred in the scrotum primarily in 554. The time varies from 13 to 70 years, the median being 46 years (*Report for* 1936, p. 49).

It will be seen that, apart from any other consideration, the number of cases and deaths is sufficiently large to justify research into the problem. Much research has in fact been done, the object being to find out what sort of tars or oils are carcinogenic, and to obtain some measure of their carcinogenicity. Here C. C. Twort and J. M. Twort have been pioneers. They have carried out experiments on something like 100,000 mice with this end in view.

B. BIOLOGICAL MEASUREMENT

The measurement of some particular effect of a substance by biological means is a matter of much difficulty. The ultimate stage should be to determine the specific activity per unit weight or volume of the substance, but we have to start by measuring some response which the substance produces. Here, at the outset, it will be necessary to make a choice between the types of animal it is possible to use, between the various responses which might be used

* This work was completed before the war and the paper ready by the early part of 1940; war-time conditions have, however, made publication impossible until now.

† *Annual Reports of the Chief Inspector of Factories and Workshops for the Years* 1923–36. London.

in the animals chosen, and between the possible measures of the response adopted.

When all this has been decided, before specific activity can be measured it is necessary to have a standard substance; a unit of activity can then be defined as the specific activity of a given amount of the standard. Assuming the relation between dosage (or amount given to the animals) and response to have been worked out for the standard, the number of units in a given amount of unknown preparation is then the number of units of the standard which produces the same response as does the given amount of unknown preparation. But further difficulties may arise. The relation between dosage and response (it is the average response of a group of animals that is under consideration) may change with season or other environmental conditions or with the stock of animals used. This can be overcome by always testing the unknown preparation against the standard at the same time and under the same conditions. Further, it is necessary, strictly speaking, if, for example, m units of standard produce the same response as n grams of the unknown preparation that rm units should produce the same response as rn grams of the unknown, otherwise the relative activities are different at different levels and measurement of specific activity becomes impossible. Also the results from different kinds of animals, in which a specific effect can be produced, should agree. There are a number of substances, e.g. vitamins and antitoxins, for which these conditions are approximately fulfilled, but the measurement of the activity of carcinogenic substances has hardly proceeded beyond the first stage —the measurement of the response which the substance produces.

This paper is largely concerned with the statistical analysis of certain of these measures of response in mice. In the later part of the paper, examples are given of their use in examining the relation between the carcinogeneity and certain physical and other characteristics of the oils, specific gravity, refractivity, the fall in refractivity of oil injected into the peritoneal cavity of mice, and place of origin of the oils. The relation to saturation is also considered. The latter part of the paper must be regarded as illustrative rather than exhaustive. For example, viscosity is undoubtedly related to carcinogeneity but it is not considered here. The object is to show the kind of use which can be made of statistical methods in this field. The records of some 60,000 mice have been used, but the labour

Cases of and deaths from epitheliomatous ulceration, 1923–36

	1923		1924		1925		1926		1927		1928		1929	
	C.	D.	C.	D.	C.	D.	C.	D.	C.	D.	C.	D.	C.	D.
Pitch and tar:														
Patent fuel works (pitch)	14	1	11	—	25	3	27	—	20	—	20	—	27	1
Tar distilling (pitch and tar)	14	1	15	2	23	4	18	3	14	1	23	4	36	3
Gas works (pitch and tar)	6	1	1	2	9	2	15	9	14	5	17	13	16	10
Other industries	2	—	4	1	6	—	16	2	10	3	8	2	8	1
Paraffin:														
Shale-oil works	6	—	2	1	4	—	2	—	6	—	2	1	4	—
Mineral oil:														
Cotton-mule spinning	15	1	79	17	78	35	88	20	101	31	101	36	54	24
Other industries	1	—	11	1	15	11	21	15	9	9	4	3	20	11
Total	58	4	123	24	160	55	187	49	174	49	175	59	165	50

	1930		1931		1932		1933		1934		1935		1936	
	C.	D.	C.	D.	C.	D.	C.	D.	C.	D.	C.	D.	C.	D.
Pitch and tar:														
Patent fuel works (pitch)	24	1	29	2	16	1	32	1	25	1	27	2	31	—
Tar distilling (pitch and tar)	41	1	15	1	26	3	39	3	50	2	48	3	42	4
Gas works (pitch and tar)	23	6	14	8	17	13	16	9	22	14	19	7	13	6
Other industries	9	2	18	4	11	2	7	1	5	1	7	3	7	2
Paraffin:														
Shale-oil works	—	—	2	—	1	—	3	—	—	—	3	2	3	—
Mineral oil:														
Cotton-mule spinning	82	21	60	19	57	22	39	23	61	24	62	20	41	12
Other industries	15	5	18	12	3	3	7	3	7	3	5	1	5	3
Total	194	36	156	46	131	44	143	40	170	45	171	38	142	27

C. = cases; D. = deaths.

of the statistical investigation, though considerable, is a small thing compared with the labour of carrying out the experiments and the ingenuity used in devising them.

The data have been very kindly placed at our disposal by Dr C. C. Twort and Mr J. M. Twort. Without their co-operation this investigation could never have been carried out. Our conclusions must not, however, be regarded as committing them in any way.

PART I. EXPERIMENTS WITH TARS

A. INTRODUCTION

C. C. Twort and J. M. Twort have devised a number of measures of carcinogenic potency which they have used extensively in the interpretation of the many experiments they have performed on the carcinogenic effect of tars and mineral oils on mice. A description of these measures is given in several of their papers, notably Twort & Twort (1930, 1931, 1933).

They suggest three measures of potency which they call Methods I, II and III. In order to understand these it is necessary to give a brief description of the nature of their experiments.

As a rule, they utilize 100 mice in each experiment. The mice are painted with the tar or oil in question with a camel-hair brush over an area of the back approximately 5–10 mm. in diameter. Painting is either performed five times a week or twice a week. Experiments in which the former procedure is adopted are called daily, those in which the latter procedure is adopted are called bi-weekly.

For each mouse a record is kept of the date of development of a tumour, of the date when the tumour became malignant and of the date when the mouse died. All tumours are said to start by being benign. In both Methods I and II separate calculations of the potency measures are made based (i) on all tumours and (ii) on malignant tumours only. These are called by the Tworts the benign and malignant potencies.

Method I. In Method I, the percentage of surviving animals, 'which are bearing or have borne tumours', are calculated for each week of the experiment, (i) all tumours and (ii) malignant tumours being considered separately.

The cumulative total of these percentages over the whole experiment is then divided by the corresponding cumulative total taken from a standard curve and the result multiplied by 100. This gives the 'benign potency' when *all* tumours are used, and the 'malignant potency' when only malignant tumours are used. The final potency is taken as half the sum of these two measures.

The standard curve used is given in one of their papers (Twort & Twort, 1933). In it, for total tumours, the percentage of surviving animals with tumours is taken to increase uniformly from zero at the end of 10 weeks, to 20 % at the end of 20 weeks, then uniformly again to 60 % at the end of 30 weeks, and to 100 % at the end of 35 weeks. For malignant tumours the same percentages as in the former case are all taken to occur 10 weeks later.

The disadvantages of Method I are given by the Tworts (1931, p. 205). Briefly they think that it makes an inadequate allowance for variations in the time a tumour-bearing animal remains alive, takes too much account of early tumours which are not of great significance and sometimes disappear, and of very resistant animals. They also think that the interval of 10 weeks between the two standard curves is not always appropriate.

Method II. In Method II, the probability of a tumourless animal getting a tumour in the ensuing week is calculated at the beginning of each week. The cumulative total of these probabilities over the whole experiment is then divided by a standard divisor (based on the cumulative total of a hypothetical set of probabilities), which varies with the duration of the experiment. The result gives the potency, the 'benign' potency when all tumours are used, and the 'malignant' potency when only malignant tumours are used. The final potency by Method II is taken as half the sum of the two. The standard divisors when *all* tumours are used in the calculations are shown in a table (Twort & Twort, 1933, p. 299); when only malignant tumours are used, one-third of the duration of the experiment is deducted before entering the table, i.e. it is assumed that the probability of getting a malignant tumour after x weeks is the same as the probability of getting a benign tumour after $2x/3$ weeks, as far as the standard curve is concerned.

In Method II, when less than five animals survive tumourless at the end of an experiment, the Tworts regard the experiment as having terminated in the week during which one or more of the last five animals to survive tumourless succumbs to a tumour. This they do, they say, 'partially to eliminate the very resistant animals'.

The Tworts frequently use the unweighted mean of the potencies derived from Methods I and II as their final potency.

Method III. From the probabilities of getting a tumour each week, it is possible to calculate the number of animals out of an original 100 who would

get tumours each week, if there were no deaths. From this, we may obtain the number of animals who would have got tumours at the end of a given number of weeks. Hence we may obtain the date when 25% of the original animals would have got tumours.

Using eleven experiments, the Tworts have constructed a graph (1933, p. 314) connecting this date with the mean potency obtained from Methods I and II. Method III consists in calculating this date and then reading off the potency from the graph.

The Tworts have used these methods extensively in endeavouring to interpret their very voluminous data. Many alternatives might have been used, and it has been thought advisable to calculate a number of alternative measures or indices of carcinogenic potency and to correlate these with the Tworts' measures. The measures chosen are as follows:

(1) Total number of tumours.

(2) Dates when (*a*) 25%, (*b*) 50%, (*c*) 75% of surviving animals had tumours.

(3) Dates when (*a*) 25%, (*b*) 50%, (*c*) 75%* of all animals had tumours.

(4) The expectations of tumourless life.

(1), (2) and (3) were considered by the Tworts themselves, but they rejected the first three alternatives in favour of their potency measures. They have stated that in some cases 'absence of tumours was of more significance than presence of a tumour'.

These measures have been calculated (α) for all tumours and (β) for malignant tumours only. (1), (2) and (3) are self-explanatory.† (4) requires some explanation. It is the average time an animal would survive tumourless if we followed a number of animals through the course of an experiment and none of them died. Most of the experiments are of 25 weeks' duration, some are 40 weeks' duration and a few of other durations. The expectation of tumourless life limited to 25 weeks has always been calculated. This is the average period an animal would survive tumourless if none died and survivals of a longer duration than 25 weeks are treated as 25 weeks. In the case of experiments of other than 25 weeks' duration, the expectation of tumourless life for the duration of the experiment has also been calculated.

Table 1 shows a typical working sheet for a bi-weekly tar experiment. Columns (3) and (6) are easily compiled from the original data. Columns (1) and (7) then easily follow by successive additions or subtractions or both. Column (2) is obtained by dividing the entries in column (1*b*) by those in column (1*c*). Column (4) is obtained by deducting from the entry in column (1*a*) half of the entry in column (6*a*), since an animal which dies tumourless is allotted half a week of exposure in the week of its death. Column (5) is obtained by dividing the entries in column (3) by those in column (4) and gives the probability of getting a tumour each week. Subtracting the entries in column (5) from unity gives those in column (8), and starting with an arbitrary base of 10,000 the entries in column (9) are then obtained by successive multiplications with those in column (8).

The 25, 50 and 75% points are then obtained by linear interpolation, and the expectation of tumourless life by summing the entries in column (9), taking only half the entries in the first and last weeks and dividing by 10,000. The last entry in column (7) gives us the total number of tumours.

B. INTERCORRELATIONS OF THE VARIOUS MEASURES OF POTENCY IN 100 DAILY TAR EXPERIMENTS

The durations of the experiments were as follows:

Duration in weeks	Frequency	Duration in weeks	Frequency
23	1	40	22
25	54	45	2
26	2	50	1
30	2	60	2
33	2	70	1
35	5	Total	94*

* Six of the 100 experiments were not used. In four there had been previous application of another substance, and in two the mice had been painted on both rump and neck.

Correlation coefficients between the Tworts' measures of potency and the alternative measures have been worked out separately (i) using all tumours and (ii) using malignant tumours only. These are shown in Tables 2 and 3.

(*a*) *Correlations based on all tumours compared with correlations based on malignant tumours only*

Taking the data based on all experiments together, every correlation based on the malignant-tumour data is larger than the corresponding correlation on the benign-tumour data; but the differences are of no importance except in the case of correlations with the total number of tumours and perhaps expectation of tumourless life.

[The significance of this fact becomes more doubtful when experiments of different durations are examined separately. There is no doubt that

* This rarely occurs. In only two out of 100 daily tar experiments did as many as 75 animals get tumours.

† Definition (2) is not quite unambiguous. It sometimes occurred that the percentage of surviving animals with tumours went up to a maximum and then declined, possibly owing to variations in resistance among the animals. In this case the 75% point might be reached twice. Where this happened the earlier date was taken.

Table 1. C51. *Bi-weekly tar experiment: benign tumours (i.e. all tumours)*

Week	(1) Animals living at beginning of week (a) Tumour-less	(b) With tumours	(c) Total	(2) Percentage of living animals with tumours	(3) Animals getting tumours each week	(4) Animals exposed to risk	(5) q_x	(6) Deaths during week (a) Tumour-less	(b) With tumours	(c) Total	(7) Total tumours to beginning of week	(8) p_x	(9) l_x
1	100	—	100	—	—	—	—	—	—	—	—	—	10,000
2	100	—	100	—	—	—	—	4	—	4	—	—	10,000
3	96	—	96	—	—	—	—	19	—	19	—	—	10,000
4	77	—	77	—	—	—	—	18	—	18	—	—	10,000
5	59	—	59	—	—	—	—	5	—	5	—	—	10,000
6	54	—	54	—	—	—	—	6	—	6	—	—	10,000
7	48	—	48	—	—	—	—	3	—	3	—	—	10,000
8	45	—	45	—	—	—	—	—	—	—	—	—	10,000
9	45	—	45	—	—	—	—	3	—	3	—	—	10,000
10	42	—	42	—	1	41	0·0244	2	—	2	—	0·9756	10,000
11	39	1	40	2·50	6	37·5	0·1600	3	1	4	1	0·8400	9,756
12	30	6	36	16·67	2	30	0·0667	—	—	—	7	0·9333	8,195
13	28	8	36	22·22	1	28	0·0357	—	—	—	9	0·9643	7,648
14	27	9	36	25·00	4	27	0·1481	—	1	1	10	0·8519	7,375
15	23	12	35	34·29	2	23	0·0870	—	1	1	14	0·9130	6,283
16	21	13	34	38·24	5	20·5	0·2439	1	2	3	16	0·6561	5,736
17	15	16	31	51·61	6	14	0·4286	2	—	2	21	0·5714	4,337
18	7	22	29	75·86	—	—	—	—	—	—	27	1·0000	2,478
19	7	22	29	75·86	1	6·5	0·1538	1	—	1	27	0·8462	2,478
20	5	23	28	82·14	1	5	0·2000	—	2	2	28	0·8000	2,097
21	4	22	26	84·62	2	4	0·5000	—	2	2	29	0·5000	1,678
22	2	22	24	91·67	1	2	0·5000	—	2	2	31	0·5000	839
23	1	21	22	95·45	1	1	1·0000	—	—	—	32	0·0000	420
24	—	22	22	100·00	—	—	—	—	4	4	33	—	—
25	—	18	18	100·00	—	—	—	—	1	1	33	—	—
26	—	17	17	100·00	—	—	—	—	—	—	33	—	—
Totals					33			67	16	83			

25 % point, 13·0
50 % point, 16·0
75 % point, 17·0

25 % point, 16·7

Expectation of tumourless life 15·4

Table 2. *Daily tar experiments. Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculations based on all tumours)*

25-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when the following percentage of all mice had tumours: 25 %	50 %	Expectation of tumourless life 25 weeks
Method I	−0·83	−0·83	−0·84	0·33	−0·78	−0·69	−0·87
	(50)	(50)	(48)	(50)	(46)	(40)	(50)
Method II	−0·77	−0·80	−0·81	0·36	−0·73	−0·66	−0·84
	(41)	(41)	(39)	(41)	(37)	(32)	(41)
Mean of I and II	−0·80	−0·83	−0·85	0·35	−0·78	−0·67	−0·87
	(41)	(41)	(39)	(41)	(37)	(32)	(41)
Mean of I and II with III	−0·70	−0·70	−0·74	0·32	−0·68	−0·70	−0·76
	(53)	(53)	(51)	(53)	(49)	(43)	(53)

40-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when 25 % of all mice had tumours: 25 %*	Expectation of tumourless life: 25 weeks	40 weeks
Method I	−0·94	−0·98	−0·94	0·71	−0·92	−0·95	−0·99
	(20)	(20)	(18)	(22)	(17)	(22)	(22)
Method II	−0·73	−0·87	−0·94	0·64	−0·85	−0·88	−0·91
	(17)	(17)	(15)	(19)	(14)	(19)	(19)
Mean of I and II	−0·82	−0·93	−0·96	0·66	−0·89	−0·92	−0·96
	(17)	(17)	(15)	(19)	(14)	(19)	(19)
Mean of I and II with III	−0·83	−0·92	−0·94	0·71	−0·87	−0·74	−0·95
	(20)	(20)	(18)	(22)	(17)	(90)	(22)

All experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when the following percentage of all mice had tumours: 25 %	50 %	Expectation of tumourless life 25 weeks
Method I	−0·74	−0·75	−0·78	0·50	−0·73	−0·75	−0·92
	(85)	(84)	(78)	(88)	(73)	(53)	(87)
Method II	−0·67	−0·70	−0·73	0·49	−0·67	−0·68	−0·78
	(69)	(68)	(63)	(72)	(58)	(40)	(71)
Mean of I and II	−0·71	−0·73	−0·77	0·50	−0·70	−0·71	−0·80
	(69)	(68)	(63)	(72)	(58)	(40)	(71)
Mean of I and II with III	−0·62	−0·64	−0·66	0·46	−0·60	−0·64	−0·74
	(88)	(87)	(81)	(91)	(76)	(56)	(90)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Of the 40-week experiments, in only nine was the 50 % point reached, an insufficient number for correlation purposes.

Table 3. *Daily tar experiments. Correlations between the Tworts' potency measures and a number of alternative measures of potency.* (*Calculations based on malignant tumours*)

25-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when 25 % of all mice had tumours 25 %*	Expectation of tumourless life 25 weeks
Method I	−0·89	−0·72	−0·50	0·57	−0·67	−0·90
	(47)	(40)	(33)	(48)	(36)	(48)
Method II	−0·72	−0·51	−0·53	0·60	−0·49	−0·79
	(40)	(34)	(28)	(41)	(30)	(41)
Mean of I and II	−0·83	−0·64	−0·52	0·59	−0·55	−0·87
	(40)	(34)	(28)	(41)	(30)	(41)
Mean of I and II with III	−0·84	−0·72	−0·71	0·61	−0·47	−0·89
	(52)	(45)	(37)	(53)	(41)	(53)

40-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours 25 %†	Total no. of tumours	Expectation of tumourless life: 25 weeks	40 weeks
Method I	−0·82	0·75	−0·86	−0·94
	(16)	(21)	(21)	(21)
Method II	−0·82	0·86	−0·63	−0·86
	(13)	(18)	(18)	(18)
Mean of I and II	−0·89	0·85	−0·77	−0·95
	(13)	(18)	(18)	(18)
Mean of I and II with III	−0·89	0·95	−0·72	−0·93
	(16)	(21)	(21)	(21)

All experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when 25 % of all mice had tumours 25 %*	Expectation of tumourless life 25 weeks
Method I	−0·75	−0·76	−0·79	0·57	−0·75	−0·95
	(76)	(62)	(48)	(83)	(52)	(83)
Method II	−0·74	−0·75	−0·76	0·63	−0·72	−0·89
	(63)	(54)	(42)	(71)	(41)	(70)
Mean of I and II	−0·76	−0·76	−0·78	0·64	−0·74	−0·93
	(62)	(53)	(41)	(70)	(41)	(70)
Mean of I and II with III	−0·75	−0·76	−0·78	0·63	−0·65	−0·94
	(81)	(67)	(52)	(89)	(57)	(89)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Of the 40-week experiments, in only ten was the date reached when 25 % of animals had tumours. These are insufficient for correlation purposes.

† Of the 40-week experiments, in only thirteen was the date reached when 50 % of surviving animals had tumours, and in only nine the date when 75 % had tumours. These are insufficient for correlation purposes.

correlations with the total number of tumours are higher when based on malignant-tumour data only, indeed, the difference in the 25-week experiments is rather striking. Correlations with the date when 25 % of surviving animals have tumours are also (with two exceptions) greater in the subgroups when based on the malignant-tumour data only. There is, however, no uniform difference shown by the correlations with expectation of tumourless life in the experiments of 25 weeks' duration, while in those of 40 weeks' duration the correlations are actually *lower* when based on malignant-tumour data only. Again, in the experiments of 25 weeks' duration, correlations with the dates when 50 and 75 % of surviving animals have tumours are, with one exception, *smaller* when based on malignant tumour data.]

Correlation coefficients have also been worked out between the Tworts' potency measures based (i) on all tumours and (ii) on malignant tumours only, which they term respectively the 'benign' and 'malignant' potencies. These are shown in Table 4.

Table 4. *Correlation coefficients between the Tworts' 'benign' and 'malignant' potencies*

	25-week experiments	40-week experiments	All experiments
Method I	0·81 (48)	0·78 (21)	0·91 (83)
Method II	0·68 (41)	0·61 (18)	0·83 (71)
Mean of I and II	0·79 (41)	0·74 (18)	0·89 (70)
Mean of I and II with III when used	0·76 (53)	0·83 (21)	0·85 (89)
Expectation of tumourless life	0·77 (54)	0·58 (21)	0·88 (93)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

In the last line the correlations between the expectations of tumourless life on the two bases are given for comparison. There is a substantial measure of agreement between the measures of potency on the two bases.

(*b*) *The results based on all tumours*

Method I compared with Method II

A comparison of Method I with Method II shows that the correlations of other measures with the Tworts' Method I are, with one exception, higher than with their Method II. Individual differences are not as a rule significant, exceptions are correlations with the dates of 25 and 50 % of surviving animals with tumours in the 40-week experiments, with the expectation of tumourless life (limited to 40 weeks) in the 40-week experiments and with the expectation of tumourless life (limited to 25 weeks) for experiments of different durations taken together. The consistency of the effect is sufficient to show its reality; it is not, however, large.

Dates when 25 and 50 % of surviving animals had tumours compared with dates when 25 and 50 % of all animals had tumours

In the 25-week experiments the correlations with the former measures are a little higher than the latter in every case. Individual differences are not significant. In the 40-week experiments there appears to be no difference, apart from the high correlation of 0·98 between the date when 50 % of surviving animals have tumours and the Tworts' Method I. For all experiments taken together there appears to be no difference.

Expectation of tumourless life and total number of tumours compared with the other alternative measures

Expectation of tumourless life appears to correlate slightly more highly with the Tworts' potencies than the measures based on a single date, while the correlations between total number of tumours and the Tworts' potencies are decidedly lower than the others. In the case of the 25-week experiments, they are very low.

Measures based on single dates compared

Of the measures based on single dates there are slightly higher correlations with the later than the earlier dates when these are based on the percentage of surviving animals with tumours; with dates based on percentage of total animals with tumours the reverse seems to be true. The effect is in any case very small, individual differences as a rule being insignificant.

The date when 50 % of all animals have tumours suffers as a measure of potency from the disadvantage that it is not reached in about 40 % of the experiments. For these it cannot therefore serve as a measure to potency at all. This applies also to the corresponding 25 % date in about 16 % of experiments. The date when 75 % of all animals had tumours was only reached in two experiments, each of 25 weeks' duration.

40-week experiments and 25-week experiments compared

The 40-week experiments give somewhat higher correlations than the 25-week experiments.

The important points are:

(1) Total number of tumours gives lower correlations with the Tworts' measures than the other alternatives.

(2) Expectation of tumourless life gives the highest correlations of all alternatives, though apart

from the total number of tumours there is very little to choose between them.

(3) Method I seems slightly better than Method II.

(c) *The results based on malignant tumours only*

Method I compared with Method II

The Tworts' Method I again shows on the whole slightly higher correlations with the alternatives than his Method II. The only significant differences are for the correlations with 25 % of surviving animals with tumours and with expectation of tumourless life for the 25-week experiments and for all experiments together. These are all in favour of Method I.

Dates when 25 % of surviving animals had tumours compared with dates when 25 % of all animals had tumours

The former measure shows a higher correlation with the Tworts' potencies than the latter. This is quite marked in the 25-week experiments but not so noticeable when all experiments are taken together.

Expectation of tumourless life and total number of tumours compared with the other alternative measures

Expectation of tumourless life again correlated more highly with the Tworts' potencies than the other measures. The total number of tumours correlates more highly with the Tworts' potencies for data based on malignant tumours than for data based on all tumours. Indeed, this measure sometimes correlates more highly with them than the date when 75 % of surviving animals had tumours, and almost as highly as the date when 25 % of all animals had tumours. (The value of 0·95 which occurs when Method III is introduced appears to be anomalous.)

Measures based on single dates compared

Of the measures based on single dates, the earlier dates show somewhat higher correlations with the Tworts' potencies than the later dates when these are based on percentage of surviving animals with tumours in the 25-week experiments. There is no difference observable for all periods. The former result is the opposite to that observed when the correlations are based on all tumours.

All measures based on single dates here suffer from the defect that the dates are not reached in an appreciable proportion of experiments. Even the date when 25 % of surviving animals had tumours is only free from this defect in the experiments of 25 weeks' duration.

40-week experiments and 25-week experiments compared

The 40-week experiments again give somewhat higher correlations than the 25-week experiments.

The important points are:

(1) The date when 75 % of surviving animals have tumours gives the lowest correlations with the Tworts' potency measures. 'Total number of tumours' gives the next lowest correlations, but higher ones than when they are based on the data from all tumours.

(2) Expectation of tumourless life gives the highest correlations with the Tworts' potency figures.

(3) Method I is slightly better than Method II.

(d) *Discussion.—Significance of differences between the Tworts' potency measures*

On the whole, the correlations obtained from this body of data are high. We cannot, of course, assert that one of the alternative measures is better than another because it correlates more highly with the Tworts' potency measures. To do this would be to argue in a circle, since the purpose of this inquiry is to examine all the measures critically.

A priori, however, one would anticipate that 'total number of tumours' would be an unsatisfactory measure, since it takes no account whatever of variations in the death-rate of the animals and makes no attempt to eliminate them. For the same reason one would have anticipated that 'expectation of tumourless life' would be a good measure since it eliminates the effect of varying death-rates among the tumourless and uses all the available data, being for the latter reason better than a single date. For this reason, it is gratifying to find that it correlates most highly with the Tworts' measures, giving as a rule a correlation of 0·8 or 0·9, sometimes higher. It is, of course, a somewhat similar function of the data, but one to which one is justified in thinking a good deal of importance may be attached, since it is *a priori* reasonable and has a clear meaning.

We must not, however, assume without examination that because two alternative measures of potency correlate highly they will necessarily agree in the order of potency in which they place two experiments. This will depend not only on the magnitude of the correlation, but on the variability of the potency in the type of experiment dealt with. Suppose x and y are two different measures of potency and that the correlation coefficient between them is r. Then the standard deviation of y when x is known is $\sigma_y \sqrt{(1-r^2)}$, assuming the relation between y and x is linear. Let the true values of x in two experiments with different substances be x_1 and x_2, and the expected values of y which corre-

spond to these be m_1 and m_2. Then the difference between the observed values of y in two experiments will have an expected value $m_2 - m_1$, with a standard error of $\sigma_y \sqrt{[2(1-r^2)]}$. Thus, if $m_2 - m_1$ is less than twice this, or roughly less than $3\sigma_y\sqrt{(1-r^2)}$, it will happen more often than once in forty times that the measure y will place the substances in the opposite order of potency from the measure x. Thus we cannot be sure of the order of carcinogenicity in which the substances are placed by two values of y unless they differ by more than $3\sigma_y\sqrt{(1-r^2)}$.

In presenting this argument, we have first assumed for simplicity that the regression of y on x is linear and that the arrays are of equal scatter. This turns out not to be the case with the measures here dealt with, and in spite of the high correlations we shall find significant departures from linearity and strong indications of unequal dispersion of arrays. In this case the difference between two observed values of y, y_1 and y_2, corresponding to two values of x, x_1 and x_2, will have a standard error of $\sqrt{(\sigma_1^2 + \sigma_2^2)}$, where σ_1 and σ_2 are the two-array standard deviations corresponding to x_1 and x_2. It is only when the difference between two values of y exceeds twice this that we can regard them as differentiating the two substances with regard to potency. Hence it will follow that the potency difference at one level of potency required to differentiate two substances in regard to carcinogenicity will be different from that at another. In what follows we take x to be the expectation of tumourless life and y to be one of the Tworts' potency measures, not because the former is necessarily a better measure than the latter, but because it is a rational alternative.

The object here is to find how much the Tworts' measures should differ in two experiments so as to make it reasonably certain that the verdict should not be reversed when a rational alternative is used.

This seems to be a better line of approach than calculating the sampling errors of the Tworts' potency measures. Such a calculation would tell us whether two such measures differed significantly, but would not throw any light on the question of whether they were measuring the desired attribute. When, however, we know that their verdict is 'almost certainly' in agreement with a rational alternative, we have found the kind of 'significant difference' we require.

We start by giving Table 5, showing the means and estimated standard deviations of the Tworts' potency measures, and of the expectation of tumourless life, which will be useful for reference in the sequel.*

The main points to be noticed in it, in passing, are:

(1) The very much lower potencies in the 40-week experiments than in the 25-week experiments.

Table 5. *Daily tar experiments. Means and estimated standard deviations of potencies and expectations (limited to 25 weeks)*

Duration of experiment		Method I		Method II		Mean of I and II		Mean of I and II with III	
		P.	E.	P.	E.	P.	E.	P.	E.
25 weeks	Benign (Mean)	456	16·4	694	16·7	363	16·7	685	16·3
	(S.D.)	211	2·6	427	2·7	306	2·7	237	2·6
	n*	50	50	41	41	41	41	53	53
	Malignant (Mean)	1084	22·4	911	22·5	990	22·5	1022	22·4
	(S.D.)	537	1·4	444	1·4	478	1·4	505	1·4
	n	48	48	41	41	41	41	53	53
40 weeks	Benign (Mean)	93	23·0	93	23·3	89	23·3	108	23·0
	(S.D.)	53	1·7	86	1·5	67	1·5	82	1·7
	n	22	22	19	19	19	19	22	22
	Malignant (Mean)	82	24·9	107	24·9	90	24·9	116	24·9
	(S.D.)	72	0·2	105	0·2	83	0·2	106	0·2
	n	21	21	18	18	18	18	21	21
All durations	Benign (Mean)	308	19·0	448	19·3	369	19·3	457	18·9
	(S.D.)	241	3·9	440	4·5	331	4·5	468	4·4
	n	87	87	71	71	71	71	90	90
	Malignant (Mean)	675	23·4	585	23·4	625	23·4	668	23·3
	(S.D.)	634	1·6	526	1·6	573	1·6	587	1·6
	n	83	83	70	70	70	70	89	89

* n = number of cases. P. = potency; E. = expectation.

* We use the term 'estimated standard deviation' for $\sqrt{[S(x-\bar{x})^2/(n-1)]}$ which is preferable to $\sqrt{[S(x-\bar{x})^2/n]}$ for small or comparatively small groups. It seems desirable to retain the term 'standard deviation' for the latter expression.

(2) The very much greater relative variability in the Tworts' potency measures than in the expectation of tumourless life. The former is always greater than 34 % and often greater than 50 %, and sometimes approaching 100 %, while the latter is always less than 24 % and sometimes as low as 1 %.

We go on, by way of example, to examine the relation between expectation of tumourless life (calculated from the benign-tumour data), and the Tworts' Method I benign potency for the 25-week experiments. Table 6 gives the means, estimated S.D.'s and coefficients of variation of the potency figures for the given values of the expectation of tumourless life.

The numbers in each group are very small, but the data suffice to show (i) that the regression is not linear, much bigger falls in potency corresponding to a rise of 1 week in the expectation occurring at the beginning than at the end of the series, and (ii) that the S.D. of the potencies and, for these particular data, the coefficient of variation also, declines from a maximum at an expectation of 13 weeks.

The non-linearity of regression is significant (we have in fact $r = -0{\cdot}86$, $\eta = 0{\cdot}96$, $z = 1{\cdot}2$ with 1 % point $= 0{\cdot}55$). With these small numbers the rise from 43·1 to 129·6 is not significant ($z = 1{\cdot}10$, 5 % point $= 2{\cdot}72$), but the fall from 129·6 to 62·7 ($z = 0{\cdot}73$, 5 % point $= 0{\cdot}65$) is significant. However, with such small numbers the results are apt to be somewhat irregular, as an examination of the other cases showed, and from a practical point of view little would be gained by examining the variability of the Tworts' potency figures at more than about three levels of the expectation of tumourless life. For the general statement of the results a wider grouping has therefore been adopted.*

Table 7 shows the results obtained for the Tworts' benign potency measures (based on all tumours), 25-week and 40-week experiments being shown separately. Table 8 shows the same results for all experiments together. Tables 9 and 10 give the results for the Tworts' malignant potency measures.

The use of these tables may be illustrated by one or two examples. Suppose two experiments with 100 animals on different tars gave 'benign potencies' by Method I as 1000 and 600 respectively. These may be taken to come into the first group (Table 8), so that each has a standard error of 113; twice the standard error of the difference is accordingly* $2{\cdot}8(2\sqrt{2})$ times this or 316, and the difference being 400 may be taken as significant. If one result had been 900 and the other 600, one could *not* have asserted a significant difference in potency. (It is unwise to extrapolate backwards for 1000 using columns 3 and 5 of Table 8, because an examination of the data in weekly groupings of 'expectation' shows that the S.D. does not rise greatly above the figures in the 12–15 group.)

Table 6. *Daily tar experiments. Means, estimated standard deviations and coefficients of variation of the potency figures for given values of the expectation of tumourless life*

Method 1, benign potency; 25-week experiments.

Expectation	No. of observations	Mean	Standard deviation	Coefficient of variation
12 and under 13 weeks	2	1086	43·1	4·0
13 „ 14 „	6	700	129·6	18·5
14 „ 15 „	9	560	80·7	14·4
15 „ 16 „	9	463	62·7	13·5
16 „ 17 „	8	392	29·2	7·5
17 „ 18 „	6	338	19·6	5·8
18 „ 19 „	2	290	18·4	6·3
20 „ 21 „	4	210	14·1	6·7
21 „ 22 „	3	173	2·7	1·5
23 „ 24 „	1	122	—	—

If one potency had been 500 and the other 300, a linear interpolation gives us a standard error of 71 for the first potency and 38 for the second; twice the standard error of the difference is

$$2{\cdot}0\sqrt{[(71)^2 + (38)^2]} = 161,$$

and the difference is not significant. Other cases may be treated similarly.

It is to be noted that if the two 'benign' potencies calculated from single experiments both fall in the last group, twice the standard error of their difference is in the neighbourhood of 100, potencies under 100 calculated from single experiments should therefore be regarded as equivalent. The same applies to malignant potencies under about 300. Of course, if a number of experiments, n say, have been carried out on the same tar, the figures in column 5 should be divided by $\sqrt{n}$ to obtain the standard error of the potency figure.

* The average variance of the potency figures was obtained in the weekly groups of expectation, which were combined together; the square root gave the 'average' S.D. for given expectation.

Table 7. *Daily tar experiments. Benign potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
		25-week experiments			
Method I	12–15*	17	671	1162–432	99
	15–18	23	406	555–314	44
	18–24	10	206	303–122	13
Method II	12–15	12	1181	2006–819	242
	15–18	19	617	904–292	153
	18–24	10	256	383– 90	68
Mean of I and II	12–15	12	921	1531–676	141
	15–18	19	512	695–303	88
	18–24	10	231	343–106	38
Mean of I and II with III	12–15	19	1115	2898–676	214
	15–18	24	534	1076–241	156
	18–24	10	231	343–106	38
		40-week experiments			
Method I	20–25	22	93	169– 4	20
Method II	20–25	19	93	276– 2	46
Mean of I and II	20–25	19	89	220– 3	32
Mean of I and II with III	20–25	22	108	274– 3	33

* Strictly 11·95–14·95, and similarly for the other groups except 20–25, which includes 25·0.

Table 8. *Daily tar experiments. Benign potency for given expectation of tumourless life. (All experiments together)*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	12–15*	18	651	1116–315	113
	15–18	24	401	555–290	44
	18–21	14	205	303–135	32
	21–24	16	124	175– 69	22
	24–25	15	38	106– 1	30
Method II	12–15	13	1120	2006–388	274
	15–18	19	617	904–292	153
	18–21	11	289	383–193	56
	21–24	14	134	301– 34	63
	24–25	14	35	161– 2	43
Mean of I and II	12–15	13	877	1531–352	175
	15–18	19	512	695–303	88
	18–21	11	250	343–198	39
	21–24	14	129	238– 53	39
	24–25	14	36	134– 2	36
Mean of I and II with III	12–15	20	1077	2898–352	235
	15–18	25	531	1076–241	153
	18–21	14	255	343–198	38
	21–24	16	134	238– 53	46
	24–25	15	37	134– 2	35

* Strictly 11·95–14·95, and similarly for the other groups except 24–25, which includes 25·0.

Table 9. *Daily tar experiments. Malignant potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
		25-week experiments			
Method I	18–21*	9	1679	2350–293	534
	21–24	30	1139	1617–515	134
	24–25	9	306	477– 17	133
Method II	18–21	7	1263	2161–269	446
	21–24	26	990	1513–153	222
	24–25	8	350	602– 33	174
Mean of I and II	18–21	7	1430	2256–281	522
	21–24	26	1072	1450–919	143
	24–25	8	320	521– 25	147
Mean of I and II with III	18–21	10	1563	2256–281	429
	21–24	34	1054	1564–600	156
	24–25	9	295	521– 25	162
		40-week experiments			
Method I	24–25	21	82	243– 3	53
Method II	24–25	18	107	392– 8	82
Mean of I and II	24–25	18	90	292– 7	60
Mean of I and II with III	24–25	21	116	382– 7	81

* Strictly 17·95–20·95, and similarly for the other groups except 24–25, which included 25·0.

Table 10. *Daily tar experiments. Malignant potency for given expectation of tumourless life. (All experiments together)*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	18–21*	9	1679	2350–293	534
	21–24	32	1088	1617–108	170
	24–25	42	147	477– 1	98
Method II	18–21	7	1263	2161–269	446
	21–24	27	958	1350–130	227
	24–25	36	174	602– 3	119
Mean of I and II	18–21	7	1450	2256–281	522
	21–24	27	1037	1450–119	165
	24–25	36	155	521– 2	101
Mean of I and II with III	18–21	10	1563	2256–281	429
	21–24	36	1015	1564–119	174
	24–25	43	170	521– 2	105

* Strictly 17·95–20·95, and similarly for the other groups except 24–25, which includes 25·0.

A comparison of Table 7 with Table 8 and Table 9 with Table 10 shows as a rule little differences between the S.D.'s in corresponding groups. An exception occurs with the 'malignant' potencies in the 40-week experiments where the S.D.'s are systematically lower than when all experiments are taken together. In dealing with the 40-week experiments, therefore, the figures in Table 9 may be used. In other cases there are no systematic differences, indeed, in only two cases are the corresponding S.D.'s significantly different.

(α) 'Benign' potency, Method I, 25 weeks, groups 18–24, which gives a S.D. of 13 in Table 7 as against 27 for the corresponding group in Table 8 ($z = 0{\cdot}72$; 5% point $= 0{\cdot}67$; 1% point $= 0{\cdot}99$).

(β) 'Malignant' potency, mean of Methods I and II with III, 25 weeks, groups 24–25, which gives 162 against 105 for the corresponding group in Table 10 ($z = 0{\cdot}43$; 5% point $= 0{\cdot}41$; 1% point $= 0{\cdot}5$).

The difference is in opposite directions in the two cases, and (β), at any rate, might be fortuitous. It is therefore suggested that in cases of doubt the larger of the two corresponding S.D.'s be used so as to be on the side of caution.

(e) *The variability of 'potency' in experiments with the same substance*

Twenty experiments were available all of 25 weeks' duration and all performed on a 1% solution in chloroform of tar B 19. The mean and standard deviation of potency for these twenty experiments are given in Table 10A.

Table 10A. *Mean and standard deviation of potency in twenty experiments with the same substance (tar* B 19, 1% $CHCl_3$)

	'Benign' potency		'Malignant' potency	
	Mean	S.D.	Mean	S.D.
Method I	515	182	1316	401
Method II	819	408	1012	318
Mean of I and II	662	284	1152	337

Comparing these results with those of Tables 7 and 9, it will be seen that, taking the values of the mean into consideration, the results are more variable than one would have expected from those tables. This suggests that there may be significant differences of potency, even in different portions of the same tar; or, alternatively, that the variation in animal susceptibility between these twenty experiments was greater than the average.

However that may be, the results suggest that the values of the S.D. in Tables 7–10, which we have decided to use as the 'standard error' of a single experiment, large as they are, are by no means too large.

C. INTERCORRELATIONS OF THE VARIOUS MEASURES OF POTENCY IN 118 BI-WEEKLY TAR EXPERIMENTS

If a daily and a bi-weekly tar experiment are each carried out on the same tar, the Tworts' potency measures would not be the same for the two experiments. The Tworts have given a table for converting the one into the other (Twort & Twort, 1931, p. 210). They say: 'Where actual experiments have not provided us with the necessary figures, we have filled in with the approximate figures which theory demands.' The factor used for converting potencies calculated from bi-weekly to potencies calculated from daily experiments is taken to be from 33 to 25 when the former potencies are between 0·1 and 0·9. It is taken as 23 when the former potency is 1, declines by 2 for each unit increase of potency up to 5, then by 1 for each unit increase of potency up to 10, then by 1 for each 5 units' increase in potency up to 30, by 0·5 for each of 5 units' increase in potency up to 40, then by 0·5 for each 10 units' increase in potency up to 70, while for potencies of 80, 90 and 100 it is taken to be 3·25, 3 and 2·8 respectively.

The potencies of many of the tars are outside this range, and we are not concerned with conversion here in any case. It seems necessary, however, to emphasize that both of the Tworts' potency measures and the alternatives with which they are here compared are measures of response, and they must be different either if a different response is measured or if the treatment to which the response is made is altered in any way.

The durations of the 118 bi-weekly tar experiments were as follows:

Duration in weeks	Frequency	Duration in weeks	Frequency
22	5	38	1
25	30	40	49
28	1	41	1
30	3	45	12
33	1	48	1
34	1	50	4
35	3	57	1
36	1	60	2
37	1	85	1
		Total	118

Four experiments in which there had been previous applications of another substance were omitted from consideration.

Method III was used in seven experiments only (four of 25 weeks', two of 40 weeks' and one of 57 weeks' duration). These experiments, therefore, have not been used in correlating the Tworts' potencies with alternative measures.

In only thirteen experiments did as many as 50% of all animals get tumours; four were of 25 weeks'

duration, seven of 40 weeks and two of other durations. This group was used as a whole to correlate the data when 50 % of all animals got tumour with the Tworts' 'benign potency' measures, but the numbers in the separate groups of 25 weeks' and duration and three of other durations. This 25 % date was therefore used to correlate with the Tworts' 'malignant potency' measures for the group as a whole, but not for the subgroups.

It never happened that as many as 50 % (and

Table 11. *Bi-weekly tar experiments. Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculation based on all tumours)*

25-week experiments

	Date when the following percentage of surviving mice had tumours			Total no. of tumours	Date when all mice had tumours 25 %	Expectation of tumourless life 25 weeks
	25 %	50 %	75 %			
Method I	−0·78 (27)	−0·89 (26)	−0·81 (23)	0·29 (27)	−0·67 (15)	−0·90 (27)
Method II	−0·57 (26)	−0·75 (25)	−0·74 (22)	0·43 (26)	−0·55 (14)	−0·83 (26)
Mean of I and II	−0·64 (26)	−0·80 (25)	−0·77 (22)	0·41 (26)	−0·61 (14)	−0·86 (13)

40-week experiments

	Date when the following percentage of surviving mice had tumours			Total no. of tumours	Date when 25 % of all mice had tumours 25 %	Expectation of tumourless life	
	25 %	50 %	75 %			25 weeks	40 weeks
Method I	−0·83 (47)	−0·84 (45)	−0·86 (34)	0·79 (49)	−0·74 (27)	−0·96 (49)	−0·91 (49)
Method II	−0·69 (45)	−0·73 (43)	−0·77 (33)	0·73 (47)	−0·65 (25)	−0·87 (47)	−0·79 (47)
Mean of I and II	−0·77 (45)	−0·81 (43)	−0·85 (33)	0·80 (47)	−0·73 (25)	−0·95 (47)	−0·87 (47)

All experiments

	Date when the following percentage of surviving mice had tumours			Total no. of tumours	Date when following percentage of all animals had tumours		Expectation of tumourless life
	25 %	50 %	75 %		25 %	50 %	25 weeks
Method I	−0·68 (109)	−0·76 (99)	−0·80 (79)	0·45 (114)	−0·68 (51)	−0·66 (13)	−0·95 (109)
Method II	−0·56 (106)	−0·65 (96)	−0·71 (77)	0·45 (111)	−0·63 (48)	−0·65 (12)	−0·87 (106)
Mean of I and II	−0·61 (106)	−0·70 (96)	−0·75 (77)	0·46 (111)	−0·65 (48)	−0·66 (12)	−0·92 (106)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

40 weeks' duration were too small to be used. In no experiment did 75 % of all animals attain tumours.

In only eighteen experiments did as many as 25 % of all animals get malignant tumours; eight were of 25 weeks' duration, seven of 40 weeks' *a fortiori* as many as 75 %) of all animals got malignant tumours. With these exceptions the analysis proceeded as with the daily tar experiments. Tables 11 and 12 show the results and may be compared with Tables 2 and 3.

(a) *Correlations based on all tumours compared with correlations based on malignant tumour only*

There are no very consistent differences between the two series of correlations, which are of the same order of magnitude. The following points of consistency may, however, be noted.

(2) The date when 25 % of surviving animals have tumours and the total number of tumours both give the *higher* correlation in the malignant data once out of three times for all experiments taken together, every time in the 25-week experiments and twice out of three times in the 40-week experiments.

Table 12. *Bi-weekly tar experiments. Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculation based on malignant tumours)*

25-week experiments

	Date when the following percentage of surviving mice had tumours			Total no. of tumours	Expectation of tumourless life 25 weeks
	25 %	50 %	75 %		
Method I	−0·82 (23)	−0·70 (18)	−0·76 (13)	0·38 (27)	−0·89 (27)
Method II	−0·78 (22)	−0·71 (18)	−0·54 (13)	0·67 (26)	−0·91 (26)
Mean of I and II	−0·84 (22)	−0·76 (18)	−0·74 (13)	0·56 (26)	−0·94 (26)

40-week experiments

	Date when the following percentage of surviving mice had tumours			Total no. of tumours	Expectation of tumourless life	
	25 %	50 %	75 %		25 weeks	40 weeks
Method I	−0·65 (37)	−0·52 (26)	−0·43 (17)	0·71 (49)	−0·89 (49)	−0·73 (49)
Method II	−0·78 (36)	−0·69 (25)	−0·47 (16)	0·94 (47)	−0·67 (47)	−0·90 (47)
Mean of I and II	−0·79 (36)	−0·70 (25)	−0·53 (16)	0·89 (47)	−0·88 (47)	−0·90 (47)

All experiments

	Date when the following percentage of surviving mice had tumours			Total no. of tumours	Date when 25 % of all mice had tumours 25 %	Expectation of tumourless life 25 weeks
	25 %	50 %	75 %			
Method I	−0·51 (84)	−0·55 (62)	−0·61 (43)	0·34 (114)	−0·51 (18)	−0·94 (109)
Method II	−0·71 (82)	−0·74 (61)	−0·82 (42)	0·59 (111)	−0·79 (17)	−0·94 (106)
Mean of I and II	−0·61 (82)	−0·64 (61)	−0·71 (42)	0·44 (111)	−0·60 (17)	−0·96 (106)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

(1) For all experiments taken together the measures based on single dates as well as the total number of tumours correlate *more highly* with Method II when calculated from the malignant-tumour data than when calculated from the benign-tumour data. With Method I and with the mean of I and II the reverse is the case.

(3) The dates when 50 and 75 % of surviving animals have tumours always correlate *less highly* with the Tworts' measures for the malignant-tumour data than for the benign-tumour data in the 25- and 40-week experiments.

The correlation coefficients between the Tworts' potency measures based (i) on all tumours (benign

potencies), and (ii) on malignant tumours only (malignant potencies) are given in Table 13. These

Table 13. *Correlation coefficients between the Tworts' 'benign' and 'malignant' potencies*

	25-week experiments	40-week experiments	All experiments
Method I	0·75 (27)	0·87 (49)	0·83 (114)
Method II	0·71 (26)	0·72 (47)	0·86 (111)
Mean of I and II	0·78 (26)	0·93 (47)	0·91 (111)
Expectation of tumourless life	0·87 (30)	0·80 (49)	0·86 (113)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

again show a substantial agreement between the two measures of potency. The correlations in the 40-week experiments are a little higher than in the case of the daily tars (see Table 4), but otherwise the two series are of the same order of magnitude.

The correlations between the expectation of tumourless life on the two bases are again given for comparison. They are of the same order of magnitude as between the other measures.

(*b*) *The results based on all tumours*

Method I compared with Method II

A comparison of Method I with Method II shows that the correlations of other measures with the Tworts' Method I are with one exception higher than with their Method II, a precisely similar result to that obtained with the daily tars (see Table 2).

Except in the case of expectation of tumourless life, individual differences are not significant. We have, as in the case of the daily tars, a small but consistent effect.

The correlations for the daily tar experiments are as a rule a little higher than for the bi-weekly tar experiments. This is always the case for correlations between dates when percentages of survivors have tumours and both Methods I and II.

Dates when 25 and 50 % of surviving animals had tumours compared with the dates when 25 and 50 % of all animals had tumours

When all experiments are taken together, there are no significant differences between the correlations with the former and latter measures.

When the 25-week and 40-week experiments are considered separately, the former measures correlate more highly with the Tworts' measures than the latter. Individual differences are not significant but the effect is consistent. The same result was found for the 25-week experiments on daily tars (Table 2).

Both the latter measures suffer from the disadvantage that the dates are not attained in a large number of experiments.

Expectation of tumourless life and total number of tumours compared with the other alternative measures

Expectation of tumourless life correlates more highly with the Tworts' potencies than the measures based on a single date, while the correlations between the total number of tumours and the Tworts' potencies are, with the exception of the 40-week experiments, decidedly lower than the others. In the 40-week experiments, the correlations between the total number of tumours and the Tworts' potencies are about as high as the correlations between the measures based on a single date and the Tworts' potencies. Somewhat similar results were found for the daily tars (see Table 2).

Measures based on single dates compared

Of the measures based on single dates there is a slight tendency to higher correlations with the later than the earlier dates when these are based on percentage of surviving animals with tumours; with dates based on percentage of total animals with tumours there is no difference. A somewhat similar result was obtained with the daily tars (see Table 2).

40-week experiments and 25-week experiments compared

The 40-week experiments again give somewhat higher correlations than the 25-week experiments.

The important points are again:

(1) Total number of tumours gives on the whole lower correlations with the Tworts' measures than the other alternatives.

(2) Expectation of tumourless life gives the highest correlations of all the alternatives.

(3) Method I seems slightly better than Method II.

(*c*) *The results based on malignant tumours only*

Method I compared with Method II

Here Method II gives, as a rule, somewhat higher correlations with the alternatives than Method I. The significantly higher values occur for total number of tumours and expectation of tumourless life (limited to 40 weeks) in 40-week experiments; total number of tumours and all dates based on percentages of surviving mice with tumours, in all experiments together. There is, however, one case where Method II gives a significantly *lower* correlation, namely, with expectation of tumourless life limited to 25 weeks (in the 40-week experiments). This is the opposite result from that obtained with the benign-tumour data for the bi-weekly tar experiments and in both cases for the daily tar experiments.

Dates when 25 % of surviving animals had tumours compared with dates when 25 % of all animals had tumours

There are no differences here between the correlations of these respective measures with the Tworts' potencies. The disadvantage of the latter measure has been pointed out already.

Expectation of tumourless life and total number of tumours compared with the other alternative measures

Expectation of tumourless life again correlates more highly with the Tworts' potencies than the other measures. (There is one exception, the correlation of 0·67 with Method II in the 40-week experiments.) Total number of tumours tends to correlate more highly with the Tworts' potencies for data based on malignant tumours than for data based on all tumours. The effect appears clearly in the 25-week and 40-week experiments but is masked on pooling. The correlations are particularly high in the 40-week experiments. This was also the case with the daily tars (see Table 3).

Measures based on single dates compared

Of the measures based on single dates, there is a tendency for the earlier dates to correlate more highly with the Tworts' potency measures than the later dates both in the 25-week and 40-week experiments, but the effect is very small and masked, if not reversed, on pooling. This is the opposite result to that observed with the benign-tumour data. These opposite results were also observed with the daily tars (see Tables 2 and 3). The disadvantage of measures based on single dates has already been pointed out and applies here also.

40-week experiments and 25-week experiments compared

The 40-week experiments give higher correlations between the Tworts' potencies and the total number of tumours than the 25-week experiments, but between the Tworts' potencies and the other alternative measures somewhat lower correlations. The lower result is different from that observed with the benign-tumour data, and in both cases with the daily tars (see Tables 2 and 3).

The important points are:

(1) Total number of tumours as a rule give lower correlations with the Tworts' potency measures than the other alternatives but correlate surprisingly highly with them in the 40-week experiments.

(2) Expectation of tumourless life gives the highest correlations with the Tworts' potency measures.

(3) Contrary to the other cases examined, Method II is here somewhat better than Method I.

(d) Significance of differences between the Tworts' potency measures

The variability of the Tworts' potency measures at given levels of expectation of tumourless life have again been used to obtain tests of significance of the differences between the values they give in different experiments. Table 14 gives the means and standard deviations of the potencies and expectations for the bi-weekly tar experiments.

The absolute figures are, of course, not comparable with Table 5, since they refer to different tars and different frequencies of application, but the two tables exhibit similar features. We notice:

(1) The very much lower potencies in the 40- than the 25-week experiments.

(2) The very much greater variability in the Tworts' potency measures than in the expectation of tumourless life. The former always exceeds 47 %, is sometimes greater than 100 % and occasionally greater than 150 %. The latter is at most 14 % and sometimes as low as 1 %.

It may be noted that it is in the 40-week experiments that the expectation limited to 25 weeks is the least variable particularly when based on the *malignant*-tumour data, where it falls to 1 %. Now it might be urged that the use of an expectation limited to 25 weeks in the 40-week experiments artificially reduced its variability as a measure of potency, and to some extent this must be true. For instance, two 40-week experiments which behaved alike in the first 25 weeks would have the same expectation (limited to 25 weeks), whatever happened subsequently and would be bound to fall in the same potency group, when this measure was used. That this circumstance is not important in practice seems to be shown (i) by the high correlations between this measure and the Tworts' potencies—for data based on *all* tumours they are actually slightly higher than when the expectation is calculated for the full 40 weeks (see Tables 11 and 12 and also 2 and 3 for the daily tar experiments) —and (ii) by the high correlations between the expectations limited to 25 weeks and to 40 weeks, which are respectively 0·92 and 0·76 for data based on all tumours and malignant tumours only. (The corresponding figures for the daily tar experiments were respectively 0·92 and 0·86.) The use throughout of the expectation limited to 25 weeks would therefore seem to be justified.

Tables 15–18 give the mean potencies in the different expectation groups, and the average standard deviations for the given expectation. Table 15 gives the results obtained for the Tworts' benign-tumour data (based on all tumours), 25-week and 40-week experiments being shown separately. Table 16 shows the same results for all experiments together. Tables 17 and 18 give the results for the

Table 14. *Bi-weekly tar experiments. Means and estimated standard deviations of potencies and expectations (limited to 25 weeks)*

		Method I		Method II		Mean of I and II	
		P.	E.	P.	E.	P.	E.
25 weeks	Benign (Mean)	334	18·4	450	18·4	393	18·4
	(S.D.)	158	2·5	377	2·5	265	2·5
	*n**	27	27	26	26	26	26
	Malignant (Mean)	668	23·3	643	23·3	663	23·3
	(S.D.)	577	1·4	494	1·4	517	1·4
	n	27	27	26	26	26	26
40 weeks	Benign (Mean)	100	23·0	104	22·9	101	22·9
	(S.D.)	79	2·0	123	2·0	96	2·0
	n	49	49	47	47	47	47
	Malignant (Mean)	82	24·8	110	24·9	96	24·9
	(S.D.)	130	0·3	102	0·3	106	0·3
	n	49	49	47	47	47	47
All weeks	Benign (Mean)	153	22·0	184	22·0	168	22·0
	(S.D.)	153	3·0	267	3·0	207	3·0
	n	109	109	106	106	106	106
	Malignant (Mean)	231	24·0	245	24·4	240	24·4
	(S.D.)	406	1·0	363	1·0	377	1·0
	n	109	109	106	106	106	106

* *n* = number of cases. P. = potency; E. = expectation.

Table 15. *Bi-weekly tar experiments. Benign potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
		25-week experiments			
Method I	13–16*	6	531	901–406	34
	16–19	10	352	478–290	42
	19–23	11	211	344–123	54
Method II	13–16	6	891	1812–484	197
	16–19	9	464	982–228	199
	19–23	11	197	334–109	50
Mean of I and II	13–16	6	711	1357–445	114
	16–19	9	411	730–260	120
	19–23	11	205	289–140	39
		40-week experiments			
Method I	16–19	2	342	448–235	—†
	19–22	13	167	228–144	22
	22–25	34	60	137– 2	20
Method II	16–19	2	502	662–341	—†
	19–22	12	195	263–148	39
	22–25	33	47	124– 4	25
Mean of I and II	16–19	2	422	555–288	—†
	19–22	12	179	213–151	23
	22–25	33	54	118– 3	20

* Strictly 12·95–15·95, and similarly for the other groups except 22–25, which includes 25·0.

† Each of these two experiments was in a different weekly group of expectation, no estimate of the S.D. was therefore possible.

Table 16. *Bi-weekly tar experiments. Benign potency for given expectation of tumourless life. (All experiments together)*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	13–16*	6	531	901–406	34
	16–19	15	355	478–235	63
	19–22	24	189	344–123	45
	22–25	64	57	171– 0	21
Method II	13–16	6	891	1812–484	197
	16–19	14	488	982–228	172
	19–22	23	197	334–113	55
	22–25	63	44	170– 0	26
Mean of I and II	13–16	6	711	1387–445	114
	16–19	14	423	730–260	104
	19–22	23	192	289–144	36
	22–25	63	51	167– 0	21

* Strictly 12·95–15·95, and similarly for the other groups except 22–25, which includes 25·0.

Table 17. *Bi-weekly tar experiments. Malignant potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
		25-week experiments			
Method I	19–23*	9	1297	2123–347	323
	23–25	18	354	940– 0	208
Method II	19–23	9	1122	1780–638	242
	23–25	17	390	1042– 0	254
Mean of I and II	19–23	9	1210	1952–723	233
	23–25	17	374	978– 0	198
		40-week experiments			
Method I	23–25	49	82	880– 0	42
Method II	23–25	47	110	404– 0	81
Mean of I and II	23–25	47	96	610– 0	58

* Strictly 18·95–22·95; the group 23–25 includes 25·0.

Table 18. *Bi-weekly tar experiments. Malignant potency for given expectation of tumourless life. (All experiments together)*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	19–23*	11	1203	2123–347	355
	23–25	98	122	940– 0	98
Method II	19–23	11	1069	1780–638	206
	23–25	95	150	1042– 0	124
Mean of I and II	19–23	11	1136	1952–696	233
	23–25	95	136	978– 0	92

* Strictly 18·95–22·95; the group 23–25 includes 25·0.

Tworts' malignant-tumour data. These tables may be used in the same way as Tables 7–10 for testing the significance of the differences between the Tworts' potencies for different experiments.

Table 17 shows a decided difference between the 25-week and 40-week experiments in the mean and variability of the Tworts' malignant potencies for the group of expectation 23–25 weeks. Significantly and systematically lower values are given in the 40-week experiments. The same result was obtained reasonable size. Nevertheless, it is of interest to compare the results for the daily and bi-weekly tar experiments. The data in Tables 8 and 10 have accordingly been rearranged with the same grouping as Tables 16 and 18 and the direct comparison is made in Tables 19 and 20.

There is remarkable agreement between the mean potencies in the same expectation groups. This is a very interesting result, as we are dealing with entirely different data in the daily and bi-weekly

Table 19. *Benign potency for given expectation of tumourless life. Daily and bi-weekly experiments compared. (All experiments)*

	Expectation of tumourless life (limited to 25 weeks)	Mean potency		Average S.D. for given expectation			
		Daily	Bi-weekly	Daily		Bi-weekly	
Method I	13–16*	549 (25)†	531 (6)	100	80	34	57
	16–19	348 (18)	355 (15)	29		63	
	19–22	177 (17)	189 (24)	27	27	45	29
	22–25	65 (25)	57 (64)	27		21	
Method II	13–16	926 (19)	891 (6)	236	202	197	179
	16–19	508 (15)	488 (14)	144		172	
	19–22	234 (14)	197 (23)	54	52	66	36
	22–25	56 (22)	44 (63)	51		26	
Mean of I and II	13–16	791 (19)	711 (6)	226	125	114	107
	16–19	442 (15)	425 (14)	81		104	
	19–22	220 (14)	192 (23)	33	36	36	26
	22–25	67 (22)	51 (63)	42		21	

* Strictly 12·95–15·95, and similarly for other groups except 22–25, which includes 25·0.
† Figures in brackets give the number of experiments in each group.

Table 20. *Malignant potency for given expectation of tumourless life. Daily and bi-weekly experiments compared. (All experiments)*

	Expectation of tumourless life (limited to 25 weeks)	Mean potency		Average S.D. for given expectation	
		Daily	Bi-weekly	Daily	Bi-weekly
Method I	19–23*	1308 (33)†	1203 (10)	289	355
	23–25	216 (49)	122 (98)	139	98
Method II	19–23	1066 (29)	1069 (11)	259	206
	23–25	197 (40)	150 (95)	162	124
Mean of I and II	19–23	1170 (29)	1136 (11)	251	233
	23–25	188 (40)	136 (95)	141	92

* Strictly 18·95–22·95, but 23–25 includes 25·0.
† Figures in brackets give the number of experiments in each group.

in the daily tar experiments in the group 24–25 weeks (see Table 9). Otherwise the results in the corresponding groups of the sets of experiments of different durations agree well. Table 17 may be used for tests of significance in the 40-week experiments, otherwise in cases of doubt it is on the side of caution to use the larger of two alternative S.D.'s provided by Tables 15 and 16 or Tables 17 and 18.

It was not found possible to use the same groupings of expectation for the Tables 15–18 as for Tables 7–10 and at the same time obtain groups of tar experiments and a different frequency of application. It shows that the relation between expectation and the Tworts' measures is a stable one.

The average S.D.'s for given expectation are on the whole somewhat greater in the daily tar experiments. Differences in the groups 13–16 are not significant owing to the small numbers. For benign potencies the daily tars show an excess of variability in the first and last groups and a defect in the two middle groups. On taking wider groupings, 13–19 and 19–25, the greater variability

of the daily tar experiments is more clearly brought out.

It must, of course, be remembered that if a daily and a bi-weekly tar experiment are each given the same potency measure or expectation, the tar used in the bi-weekly experiment must be the more potent since the frequency of application is less.

The Tworts frequently use the unweighted mean of the 'benign' and 'malignant' potencies as their final measures of potency, each of these being as a rule the unweighted mean of the two results arrived at by Methods I and II.

In order to have an approximate figure for the standard error of these results, it was therefore necessary to calculate the average S.D. of this measure for given expectation of tumourless life. The expectation used was the 'benign' expectation, i.e. that calculated from the data including all tumours in the manner explained in Table 1 and p. 365. It is sufficient to take experiments of all durations together. The results are given in Table 21, which is placed here for reference.

D. CONCLUSIONS AND SUMMARY

The tar experiments were examined before the oil experiments at the Tworts own request because it was thought that the greater range of potencies which they exhibit would give more opportunity for the intercorrelations of the various alternative measures to become manifest. It is to be noted, however, that the correlations remain high in the 40-week experiments, although these are of lower mean potency and cover a smaller range of potencies than the 25-week experiments.

It seems safe to say that, with the tars at any rate, had the Tworts used any reasonable alternative measure of potency, instead of their own which was, like theirs, a measure of response and a function of time, they would not have been led to very different results. Caution is, however, required in interpreting differences between their potency measures, in different experiments, since the standard errors are large.

It should be emphasized that the Tworts' potency

Table 21. *'Final' potency for given expectation of tumourless life (All durations together)*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
		Daily experiments			
Method I	13–16*	23	953	1600–212	251
	16–19	18	615	879–299	107
	19–22	17	221	472–119	92
	22–25	26	53	160– 0	36
Method II	13–16	19	1038	2033–259	287
	16–19	15	695	903–279	126
	19–22	14	246	452–119	74
	22–25	23	64	215– 0	70
Mean of I and II	13–16	19	1004	1817–235	263
	16–19	15	663	889–289	104
	19–22	14	236	400–143	69
	22–25	23	56	239– 0	52
		Bi-weekly experiments			
Method I	13–16	6	938	1372–410	317
	16–19	15	516	901–210	166
	19–22	24	196	606–106	101
	22–25	64	45	177– 0	23
Method II	13–16	6	1073	1602–561	285
	16–19	14	583	823–280	159
	19–22	23	213	638– 75	106
	22–25	63	51	247– 0	38
Mean of I and II	13–16	6	1006	1487–637	275
	16–19	14	558	860–291	144
	19–22	23	205	622–109	110
	22–25	63	48	199– 0	29

* Strictly 12·95–15·95, and similarly for other groups except 22–25, which includes 25·0.

figures and the alternatives with which they have been compared are all measures of response. When these differ significantly in two experiments, the conclusion that the tar actually given in one experiment (at the level of dosage at which it *was* given) was more carcinogenic than in the other is clearly justified. But the only way in which one could justify a statement that one tar was n times as carcinogenic as another, i.e. contained n times as many carcinogenic units, would be comparing the amounts of the two tars necessary to produce the same response, at some level of dosage below that at which increasing doses no longer lead to increasing response. This comparison cannot be made until the dosage-response relation has been worked out, and this has not so far been done. If it could be done, it would be possible to compare the 'specific carcinogenic activity' of some tar or oil with that of a standard in the same way as is ordinarily done in the biological assay of vitamins and antitoxins.

SUMMARY TO PART I

1. The Tworts' potency measures correlate highly with all the reasonable alternative single measures of potency which have been examined.

2. We are inclined to prefer expectation of tumourless life as a measure because it has a meaning which can be more clearly interpreted. This, however, may be partly a matter of personal preference; in any case it correlates more highly with the Tworts' measures than any of the other alternatives considered.

3. Both in the daily and in the bi-weekly tar experiments, total number of tumours gives on the whole the lowest correlations with the Tworts' measures. In the 40-week experiments these correlations are, however, higher than elsewhere, being often as high as the correlations of the Tworts' measures with measures based on single dates. On all grounds it seems an undesirable measure to use.

4. In the daily tar experiments the date when 75 % of surviving animals have tumours gave the lowest correlations with the Tworts' potency measures when based on malignant tumours only, and seems therefore an undesirable measure to use. In general, it would be possible to use as measures the dates when 25 or 50 % of surviving animals have tumours, or the date when 25 % of *all* animals have tumours, but measures based on single dates appear less satisfactory than the Tworts' measures or than the expectation of tumourless life. With the exception, perhaps, of the date when 25 % of surviving animals have tumours, they all suffer from the further disadvantage that the dates are never reached in an appreciable number of experiments.

5. The variabilities of the Tworts' potency figures in experiments which have the same expectation of tumourless life have been calculated, and these can be used to test the significance of differences between potencies calculated from different experiments. The results suggest that the Tworts' measures are too sensitive for highly potent tars, in the sense that large differences may occur in their numerical values at the upper end of the scale without indicating any significant difference of potency.

6. The variability of the Tworts' 'potencies' has been calculated in twenty experiments with the same substance (tar B 19, 1 % in chloroform). The results are somewhat more variable than the calculation in conclusion 5 would have suggested. Hence the 'standard errors' of the Tworts' 'potencies' which are based on the calculations of conclusion 5, large as they are, are by no means too large.

7. All the alternative measures which have been compared are measures of response. When these differ significantly in two experiments, the conclusion that the tar actually given in one experiment (at the level of dosage at which it was given) was more carcinogenic than the other is clearly justified. But the only way in which one could justify a statement that one tar was so many times as carcinogenic as another would be by comparing the amounts of the two tars necessary to produce the same response at some sufficiently sensitive level of dosage. This comparison cannot be made until the dosage-response relation has been worked out, which is not so far the case.

PART II. EXPERIMENTS WITH OILS

A. INTRODUCTION

The design of the oil experiments was the same as that of the tar experiments dealt with in the first part of this report; the introduction to Part I may therefore be taken as applying to the oil experiments also, as far as the nature of the experiments and the Tworts' methods of calculating potency are concerned.

We have examined 410 oil experiments comprising the records of 41,000 mice. All these were 'daily' experiments. With the exception of eight experiments with saponifiable oils, which gave rise to no tumours, all the oils were mineral oils. Twenty-four other oil experiments gave rise to no tumours at all, and a further 117 to no malignant tumours.

The alternative measures used for comparison with the Tworts' measures were:

(1) Total number of tumours.

(2) Dates when (*a*) 25%, (*b*) 50%, (*c*) 75% of *surviving* animals had tumours.

(3) Dates when 25% of *all* animals had tumours.

(4) The expectation of tumourless life.

The potency of the oils is much less than that of the tars, and the number of tumours obtained is, of course, much smaller. In consequence, the dates (2) were not always reached. For benign tumours they were reached in a sufficient number of experiments to permit of correlation with the Tworts' measures, with the exception of the 75% points in experiments of 50 weeks' duration or greater. For malignant tumours they were not always reached but have been used wherever possible. The date (3) could be used for benign tumours in about 100 experiments but not for malignant tumours at all.

Experiments which yielded no tumours have been included in the intercorrelation of the Tworts' measures with measures (1) and (4).

B. INTERCORRELATIONS OF THE VARIOUS MEASURES OF POTENCY IN 410 DAILY OIL EXPERIMENTS

The durations of the experiments were as follows:

Duration in weeks	Frequency	Duration in weeks	Frequency
18	1	40	65
19	2	43	1
22	1	44	1
23	1	45	69
25	39	47	1
27	2	49	1
28	2	50	38
30	2	52	1
31	2	54	1
32	2	55	1
33	3	56	1
34	4	57	1
35	124	60	33
36	2	70	3
37	2	100	1
38	1		
39	2	Total	410

Correlation coefficients between the Tworts' measures of potency and the alternative measures have been worked out separately, (i) using all tumours, and (ii) using malignant tumours only. This has been done for experiments of respectively 25, 35, 40, 45, 50 and 60 weeks' duration and finally for all experiments together. These are shown in Tables 22–26. Tables 22 and 23 give the results for individual durations when the correlations are based on all tumours, Tables 24 and 25 when they are based on malignant tumours only. Table 26 gives both sets of results for all experiments taken together.

(*a*) *Correlations based on all tumours compared with correlations based on malignant tumours only*

A consideration of the results for all experiments together (Table 26) shows no systematic differences between the magnitudes of the correlation on the two bases. The expectation of tumourless life limited to 25 weeks has been used in the calculations based on all tumours, while for those based on malignant tumours only the expectation limited to 35 weeks has been used. A tumour naturally becomes malignant some time after its first appearance, and to use an expectation limited to 25 weeks in the latter case would have been to neglect the period in which many malignant tumours occur. Using these two measures, the correlations are of the same order of magnitude (0·9), whether all tumours or malignant tumours only are used. Correlations with total number of tumours are, when based on malignant tumours only, higher for Method II and lower for Method I than when based on all tumours. The same applies to the date when 25% of surviving mice had tumours; the result being just significant.

[It is of some interest and importance to consider the experiments of different durations separately.*

We consider first the total number of tumours. In the experiments of 45, 50 and 60 weeks' duration there are no significant differences in the correlations between this measure and the Tworts' potencies calculated on the two bases.

In the 25-week experiments, this measure correlates more highly with the Tworts' Method I and less highly with Method II when based on all tumours than when based on malignant tumours only. In the 35-week experiments, it correlates less highly with Method II when based on all tumours than when based on malignant tumours only, while for Method I there is no significant difference. In the 40-week experiments it correlates more highly with Method I when based on all tumours than when based on malignant tumours only, while for Method II there is no difference. These results are in conformity with those found for all periods together.

We next consider dates when 25 and 50% of surviving mice had tumours. These measures correlate more highly with the Tworts' potencies when based on all tumours than when based on malignant

* The reader will find this section easier to follow if he looks at Tables A–G of Appendix I, p. 417, where the data of Tables 22–26 are rearranged so that each correlation coefficient comes next to the one with which it is compared.

Table 22. *Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculations based on all tumours)*

25-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Expectation of tumourless life 25 weeks
Method I	−0·78 (29)	−0·89 (24)	−0·87 (14)	0·87 (38)	−0·93 (38)
Method II	−0·69 (24)	−0·64 (21)	−0·69 (13)	0·49 (31)	−0·72 (31)
Mean of I and II	−0·83 (24)	−0·81 (21)	−0·80 (13)	0·73 (31)	−0·91 (31)
Mean of I and II with III	−0·72 (30)	−0·78 (25)	−0·84 (15)	0·74 (38)	−0·85 (38)

35-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when 25 % of all mice had tumours* 25 %	Expectation of tumourless life: 25 weeks	35 weeks
Method I	−0·73 (112)	−0·75 (99)	−0·79 (83)	0·63 (124)	−0·59 (63)	−0·90 (124)	−0·86 (124)
Method II	−0·69 (108)	−0·75 (97)	−0·84 (81)	0·61 (120)	−0·58 (63)	−0·89 (120)	−0·84 (120)
Mean of I and II	−0·71 (108)	−0·76 (97)	−0·82 (81)	0·62 (120)	−0·59 (63)	−0·89 (120)	−0·85 (120)
Mean of I and II with III	−0·71 (112)	−0·75 (99)	−0·82 (83)	0·62 (124)	−0·59 (63)	−0·89 (124)	−0·85 (124)

40-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours	Date when 25 % of all mice had tumours* 25 %	Expectation of tumourless life: 25 weeks	35 weeks	40 weeks
Method I	−0·82 (53)	−0·84 (45)	−0·87 (28)	0·87 (65)	−0·79 (15)	−0·96 (65)	−0·97 (65)	−0·97 (65)
Method II	−0·50 (44)	−0·62 (36)	−0·72 (22)	0·76 (56)	−0·68 (12)	−0·74 (56)	−0·76 (56)	−0·72 (56)
Mean of I and II	−0·63 (44)	−0·70 (36)	−0·79 (22)	0·84 (56)	−0·71 (12)	−0·92 (56)	−0·87 (56)	−0·85 (56)
Mean of I and II with III	−0·56 (53)	−0·69 (45)	−0·83 (28)	0·70 (65)	−0·67 (15)	−0·88 (65)	−0·82 (65)	−0·80 (65)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Of the 25-week experiments, in only five was the date reached when 25 % of animals had tumours; the 50 % date was never reached. Of the 35-week experiments, in only nine was the 50 % date reached, and in no case the 75 % date. Of the 40-week experiments, in only two was the 50 % date reached, the 75 % date was never reached. These numbers are insufficient for correlation purposes.

Table 23. *Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculations based on all tumours) (continued)*

45-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours: 25 %	50 %	75 %	Total no. of tumours†	Expectation of tumourless life: 25 weeks	35 weeks	40 weeks	45 weeks
Method I	−0·82	−0·81	−0·48	0·85	−0·86	−0·92	−0·93	−0·91
	(50)	(37)	(19)	(66)	(66)	(66)	(66)	(66)
Method II	−0·68	−0·77	−0·89	0·68	−0·75	−0·76	−0·80	−0·80
	(47)	(34)	(16)	(63)	(63)	(63)	(63)	(63)
Mean of I and II	−0·89	−0·91	−0·80	0·87	−0·88	−0·94	−0·97	−0·95
	(47)	(34)	(16)	(63)	(63)	(63)	(63)	(63)
Mean of I and II with III	−0·75	−0·81	−0·88	0·81	−0·88	−0·93	−0·94	−0·93
	(53)	(40)	(22)	(69)	(69)	(69)	(69)	(69)

50-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours*: 25 %	50 %	Total no. of tumours†	Expectation of tumourless life: 25 weeks	35 weeks	40 weeks	45 weeks	50 weeks
Method I	−0·73	−0·95	0·77	−0·90	−0·90	−0·88	−0·85	−0·82
	(25)	(14)	(35)	(35)	(35)	(35)	(35)	(35)
Method II	−0·72	−0·94	0·81	−0·85	−0·86	−0·95	−0·84	−0·81
	(25)	(13)	(35)	(35)	(35)	(35)	(35)	(35)
Mean of I and II	−0·64	−0·95	0·78	−0·86	−0·87	−0·84	−0·82	−0·79
	(24)	(13)	(34)	(34)	(34)	(34)	(34)	(34)
Mean of I and II with III	−0·57	−0·73	0·78	−0·62	−0·71	−0·72	−0·73	−0·73
	(26)	(15)	(37)	(37)	(37)	(37)	(37)	(37)

60-week experiments

Tworts' measures	Date when the following percentage of surviving mice had tumours*: 25 %	50 %	Total no. of tumours†	Expectation of tumourless life: 25 weeks	35 weeks	40 weeks	45 weeks	50 weeks	60 weeks
Method I	−0·71	−0·90	0·81	−0·93	−0·997	−0·99	−0·97	−0·95	−0·88
	(21)	(13)	(33)	(33)	(33)	(33)	(33)	(33)	(33)
Method II	−0·58	−0·69	0·76	−0·59	−0·67	−0·70	−0·72	−0·75	−0·79
	(18)	(10)	(30)	(30)	(30)	(30)	(30)	(30)	(30)
Mean of I and II	−0·75	−0·97	0·92	−0·93	−0·99	−0·99	−0·99	−0·98	−0·95
	(18)	(10)	(30)	(30)	(30)	(30)	(30)	(30)	(30)
Mean of I and II with III	−0·68	−0·94	0·82	−0·93	−0·97	−0·97	−0·97	−0·95	−0·90
	(21)	(13)	(33)	(33)	(33)	(33)	(33)	(33)	(33)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Of the 50-week experiments, in only eight was the date reached when 75 % of surviving animals had tumours; of the 60-week experiments in only six. These numbers are insufficient for correlation purposes.

† Of the 45-week experiments, in only ten was the date reached when 25 % of all animals had tumours; the corresponding figures for the 50-week and 60-week experiments are respectively two and one. The 50 % date was never reached. These numbers are insufficient for correlation purposes.

Table 24. *Correlations between the Tworts' potency measures and a number of alternative measures of potency (Calculations based on malignant tumours only)*

	25-week experiments		35-week experiments					40-week experiments					
			Date when the following percentage of surviving mice had tumours*			Expectation of tumourless life		Date when the following percentage of surviving mice had tumours*			Expectation of tumourless life		
Tworts' measures	Total no. of tumours*†	Expectation of tumourless life 25 weeks	25 %	50 %	Total no. of tumours†	25 weeks	35 weeks	25 %	50 %	Total no. of tumours†	25 weeks	35 weeks	40 weeks
Method I	0·54	−0·80	−0·66	−0·43	0·65	−0·63	−0·83	−0·74	−0·57	0·68	−0·22	−0·70	−0·89
	(38)	(38)	(62)	(24)	(124)	(124)	(124)	(27)	(12)	(65)	(65)	(65)	(65)
Method II	0·85	−0·78	−0·59	−0·53	0·80	−0·73	−0·85	−0·45	−0·18	0·83	−0·32	−0·70	−0·85
	(31)	(31)	(61)	(24)	(120)	(120)	(120)	(23)	(12)	(56)	(56)	(56)	(56)
Mean of I and II	0·68	−0·87	−0·67	−0·53	0·77	−0·71	−0·89	−0·58	−0·31	0·96	−0·32	−0·76	−0·93
	(31)	(31)	(61)	(24)	(120)	(120)	(120)	(23)	(12)	(56)	(56)	(56)	(56)
Mean of I and II with III	0·67	−0·85	−0·67	−0·53	0·77	−0·72	−0·88	−0·59	−0·31	0·78	−0·37	−0·79	−0·93
	(38)	(38)	(62)	(24)	(124)	(124)	(124)	(27)	(12)	(65)	(65)	(65)	(65)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Of the 25-week experiments, in only six was the date reached when 25 % of surviving animals had malignant tumours; of the 35-week experiments in only six was the 75 % date reached; of the 40-week experiments in only one was it reached. These numbers are insufficient for correlation purposes.

† In none of these experiments was the date reached when 25 % of *all* animals had tumours.

Table 25. *Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculations based on malignant tumours only) (continued)*

45-week experiments

Tworts' measures	Date when 25 % of surviving mice had tumours*	Total no. of tumours†	Expectation of tumourless life		
			35 weeks	40 weeks	45 weeks
Method I	−0·65	0·74	−0·78	−0·79	−0·82
	(13)	(66)	(66)	(66)	(66)
Method II	−0·09	0·78	−0·58	−0·75	−0·78
	(11)	(63)	(63)	(63)	(63)
Mean of I and II	−0·05	0·85	−0·70	−0·85	−0·88
	(11)	(63)	(63)	(63)	(63)
Mean of I and II with III	−0·78	0·81	−0·79	−0·87	−0·89
	(16)	(69)	(69)	(69)	(69)

50-week experiments

Tworts' measures	Date when 25 % of surviving mice had tumours*	Total no. of tumours†	Expectation of tumourless life			
			35 weeks	40 weeks	45 weeks	50 weeks
Method I	0·15	0·88	−0·79	−0·80	−0·82	−0·88
	(8)	(35)	(35)	(35)	(35)	(35)
Method II	0·25	0·77	−0·60	−0·64	−0·67	−0·76
	(8)	(35)	(35)	(35)	(35)	(35)
Mean of I and II	0·23	0·87	−0·72	−0·75	−0·78	−0·86
	(8)	(34)	(34)	(34)	(34)	(34)
Mean of I and II with III	0·23	0·84	−0·48	−0·61	−0·72	−0·83
	(8)	(37)	(37)	(37)	(37)	(37)

60-week experiments

Tworts' measures	Total no. of tumours*†	Expectation of tumourless life				
		35 weeks	40 weeks	45 weeks	50 weeks	60 weeks
Method I	0·69	−0·68	−0·83	−0·91	−0·91	−0·76
	(33)	(33)	(33)	(33)	(33)	(33)
Method II	0·78	−0·84	−0·85	−0·91	−0·91	−0·92
	(30)	(30)	(30)	(30)	(30)	(30)
Mean of I and II	0·77	−0·84	−0·87	−0·94	−0·98	−0·89
	(30)	(30)	(30)	(30)	(30)	(30)
Mean of I and II with III	0·82	−0·68	−0·80	−0·89	−0·93	−0·87
	(33)	(33)	(33)	(33)	(33)	(33)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Of the 45-week experiments, in only eight was the date reached when 50 % of surviving animals had malignant tumours; in only three was the 75 % date reached. Of the 50-week experiments, the 50 % date was reached in only twenty-one, the 75 % date in only one. Of the 60-week experiments, the 25 % date was reached in only five, the 50 % and 75 % dates in only one. These numbers are insufficient for correlation purposes.

† In none of these experiments was the date reached when 25 % of all animals had malignant tumours.

Table 26. *Correlations between the Tworts' potency measures and a number of alternative measures of potency. (Calculations based (a) on all tumours, (b) on malignant tumours only, for all experiments together)*

Tworts' measures	(a) All tumours: Date when the following percentage of surviving mice had tumours 25%	50%	75%	Total no. of tumours	Date when 25% of all mice had tumours*	Expectation of tumourless life 25 weeks	(b) Malignant tumours: Date when the following percentage of surviving mice had tumours† 25%	50%	Total no. of tumours‡	Expectation of tumourless life 35 weeks
Method I	−0·64 (320)	−0·71 (261)	−0·75 (181)	0·64 (402)	−0·60 (103)	−0·91 (398)	−0·57 (129)	−0·63 (53)	0·36 (402)	−0·87 (343)
Method II	−0·54 (294)	−0·64 (238)	−0·70 (166)	0·59 (374)	−0·60 (99)	−0·86 (370)	−0·67 (118)	−0·54 (52)	0·79 (374)	−0·90 (324)
Mean of I and II	−0·60 (293)	−0·69 (238)	−0·74 (166)	0·62 (373)	−0·61 (99)	−0·90 (369)	−0·70 (118)	−0·69 (52)	0·60 (373)	−0·91 (323)
Mean of I and II with III	−0·58 (325)	−0·67 (266)	−0·72 (185)	0·61 (407)	−0·61 (105)	−0·89 (403)	−0·66 (133)	−0·63 (56)	0·59 (407)	−0·90 (348)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* In only fourteen experiments was the date reached when 50% of all mice had tumours, the 75% date was never reached.
† In only sixteen experiments was the date reached when 75% of surviving mice had tumours (malignant).
‡ In none of the experiments was the date reached when 25% of all mice had malignant tumours. These numbers are insufficient for correlation purposes.

tumours only.* The individual differences are significant in the following cases only: for the 25% date, in the 45-week experiments, correlations with the mean of Methods I and II, and in the 50-week experiments, correlations with Methods I and II; for the 50% date in the 25-week experiments correlations with Method I.

Turning now to the expectation of tumourless

---

* This result for Method II requires a little discussion, since for the 25% date it is the opposite from that obtained when all experiments are pooled together. The results are given in the following table, together with the Tworts' mean potencies obtained by Method II and the mean dates, for the four groups.

Correlations between the Tworts' Method II and the dates when 25% of surviving mice had tumours

Duration of experiment	Correlations (a) All tumours	Correlations (b) Malignant tumours	Means (a) 25% date	Means (a) Method II	Means (b) 25% date	Means (b) Method II
35 weeks	−0·69 (108)	−0·59 (61)	18	193	32	113
40 weeks	−0·50 (44)	−0·45 (23)	22	70	38	35
45 weeks	−0·68 (47)	+0·09 (11)	28	25	44	24
50 weeks	−0·72 (25)	+0·25 (8)	39	10	48	12
All durations	−0·54	−0·67	25	133	36	84

(a) gives the higher correlations for each separate period yet (b) gives the higher when all periods are pooled. Why is this? In the first place, it must be remembered that a number of experiments which were not exactly of 35, 40, 45 and 50 weeks' duration are included in 'All durations'. When these are excluded the correlations become (a) −0·62, (b) −0·69, so that the excess of (b), though now smaller and not significant, persists. It must be remembered that when the experiments are pooled, the amount of scatter about the final regression line depends on two factors, (i) the amount of scatter of the four individual groups about their own regression lines, and (ii) the amount of scatter of the means of the individual groups about a regression line fitted to them.

Even if the scatter due to (i) were less in every group for (a) than for (b), yet that due to (ii) might be less for (b) than for (a), and the combined effect might be to make the correlation in the pooled experiments higher in case (b) than in case (a). This is, in fact, the case. The regression of mean potency on mean date is clearly non-linear in both cases but more so in case (a) than in case (b). The point is brought out by a simple analysis of variance and co-variance within and between groups.

life, we find that for expectations limited to the same number of weeks, the correlations with the Tworts' measures are almost always higher when based on all tumours than when based on malignant tumours only. This excess is significant, for Method I, in the 25-week experiments in one case; in every case in the 35-week experiments; in the 40-week experiments in every case for the expectation limited to 25 weeks, in one case (Method I) for the expectation limited to 35 weeks and for that limited to 40 weeks; in the 45-week experiments it is significant in all cases except Method II which shows a difference in the same direction; in the 50-week experiments it is significant only for Method II, when the expectations are limited to 35 and 40 weeks; in the 60-week experiments it is significant, except for Method II which shows a difference in the opposite direction when the expectations are limited to 35, 40 and 45 weeks. There are a few exceptions to this rule when the correlations based on malignant tumours only are higher; the only significant differences in this direction are in the 40-week experiments (expectation limited to 40 weeks, mean of Methods I and II, and mean of Methods I and II with III), and in the 60-week experiments, where for Method II the correlations are uniformly higher when based on malignant tumours only.

The excesses in favour of the correlation based on all tumours are as a rule in the neighbourhood of 0·1 or 0·2. One might have anticipated this excess when expectations limited to the same time in each case are considered. A tumour becomes malignant some time after its first appearance; to curtail the expectation at a given time will exclude consideration of what happened after that time, and this will make the expectation based on malignant data a somewhat less complete index of the result of the experiment. This fact makes the expectation limited to 25 weeks an almost useless measure for the experiments of longer durations when the calculations are based on malignant tumours only; in the 40-week experiments its correlation with the Tworts' measures had dropped to 0·2 or 0·3; it would have been useless to calculate it for experiments of longer duration than this.

We may, however, compare with expectation limited to 25 weeks based on all tumours the expectation limited to 35 weeks based on malignant tumours only.

The correlation of these with the Tworts' measures are shown in Table 27. The results show that the correlations with the Tworts' measures of the expectation limited to 25 weeks based on all tumours are as a rule higher than those of the expectations limited to 35 weeks based on malignant tumours only. The difference is not great and only sometimes significant. That this difference does not appear

		$S(x-\bar{x})^2$	$S(y-\bar{y})^2$	$S(x-\bar{x})(y-\bar{y})$	r
Within experiments of same durations	(*a*)	9,221	4,324,651	−103,785	−0·52
	(*b*)	920	370,296	− 9,737	−0·53
Between experiments of different durations	(*a*)	10,301	1,406,286	−102,708	−0·85
	(*b*)	2,784	163,630	− 20,394	−0·96
Total	(*a*)	19,522	5,730,937	−206,493	−0·62
	(*b*)	3,604	533,926	− 30,131	−0·69

$x = 25\%$ date; y = potency.

The inter-group correlation is greater for (*b*) than for (*a*), while the average intra-group correlations are the same. That the latter correlation is no greater for (*a*) than for (*b*) is due to differences of weighting resulting from the addition of squares and products of deviations from the separate groups. These are as follows:

	$S(x-\bar{x})^2$		$S(y-\bar{y})^2$		$S(x-\bar{x})(y-\bar{y})$	
Weeks	(*a*)	(*b*)	(*a*)	(*b*)	(*a*)	(*b*)
35	3,732	686	3,863,588	322,807	− 82,169	−8,701
40	1,996	148	429,191	40,991	− 14,688	−1,100
45	2,500	7	29;366	6,092	− 5,798	+ 19
50	993	79	2,506	406	− 1,130	+ 45
Total	9,221	920	4,324,651	370,296	−103,785	−9,737

It will be seen that while in case (*b*) the effect of the 35-week experiments predominates in all three terms, in case (*a*) the value of $S(x-\bar{x})^2$ is much increased by the inclusion of the 40-, 45- and 50-week experiments, with a corresponding lowering of the average correlation. The net effect of factors (i) and (ii) is to raise the correlation in case (*b*) in the pooled data, and the paradox is explained. It has been thought worth while to discuss this point at length, since a similar discrepancy occurs when the correlations with the expectation of tumourless life are examined. See pp. 391–2.

when all experiments are taken together (see Table 26) is another instance of the phenomenon discussed in detail in the footnote on p. 390.]

To summarize, an examination of the experiments of different durations separately shows a tendency for the date when 25% of surviving animals have tumours to correlate more highly with the Tworts' measure when based on all tumours than when based on malignant tumours only. A similar tendency is shown by the expectation of tumourless life limited to 25 weeks based on all tumours when it is compared with the expectation of tumourless life limited to 35 weeks based on malignant tumours only. The effect is not very large, and not apparent when the data of all experiments are pooled.

Differences between the magnitude of the correlations based on all tumours and on malignant tumours only are not even as clear-cut for the oils as they are for the tars, where they are also not of great importance (see Part I, p. 365).

Correlation coefficients have been worked out, for the oils just as for the tars, between the Tworts' potency measures based (i) on all tumours, and (ii) on malignant tumours only, which he terms respectively the 'benign' and 'malignant' potencies. These are shown in Table 28. In the last line the correlations between the expectations of tumourless life on the two bases are given for comparison. While there is again a substantial amount of agreement between the measures of potency on the two bases, this is not as high for the oils as for the tars (see Part I, p. 369), and becomes rather low in individual cases. *A priori*, one would expect the 'benign' and 'malignant' potencies to measure rather different things. If this is true, it is more marked in the oils than in the tars.

Table 27. *Correlations with the Tworts' measures of (a) expectation of tumourless life limited to 25 weeks based on all tumours, and (b) expectation of tumourless life limited to 35 weeks based on malignant tumours only*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
35-week experiments	−0·90 (124)	−0·83 (124)	−0·89 (120)	−0·85 (120)	−0·89 (120)	−0·89 (120)	−0·89 (124)	−0·88 (124)
40-week experiments	−0·96 (65)	−0·70 (65)	−0·74 (56)	−0·70 (56)	−0·92 (56)	−0·76 (56)	−0·88 (65)	−0·79 (65)
45-week experiments	−0·86 (66)	−0·78 (66)	−0·75 (63)	−0·58 (63)	−0·88 (63)	−0·70 (63)	−0·88 (69)	−0·79 (69)
50-week experiments	−0·90 (35)	−0·79 (35)	−0·85 (35)	−0·60 (35)	−0·86 (34)	−0·72 (34)	−0·62 (37)	−0·48 (37)
60-week experiments	−0·93 (33)	−0·68 (33)	−0·59 (30)	−0·84 (30)	−0·93 (30)	−0·84 (30)	−0·93 (33)	−0·68 (33)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

Table 28. *Correlation coefficients between the Tworts' 'benign' and 'malignant' potencies*

	25-week experiments	35-week experiments	40-week experiments	45-week experiments	50-week experiments	60-week experiments	All experiments
Method I	0·75 (38)	0·73 (124)	0·64 (65)	0·56 (66)	0·34 (35)	0·75 (33)	0·66 (402)
Method II	0·42 (31)	0·73 (120)	0·46 (56)	0·32 (63)	0·55 (35)	0·80 (30)	0·67 (374)
Mean of I and II	0·76 (31)	0·77 (120)	0·63 (56)	0·48 (63)	0·50 (34)	0·83 (30)	0·75 (373)
Mean of I and II with III	0·74 (38)	0·77 (124)	0·65 (65)	0·72 (69)	0·47 (37)	0·73 (33)	0·74 (407)
Expectation of tumourless life*	0·61 (39)	0·76 (124)	0·76 (65)	0·66 (69)	0·30 (38)	0·74 (33)	0·77 (349)

Numbers in brackets are the number of pairs of observations on which the corresponding correlations are based.

* Expectation limited to 25 weeks for experiments of 25 weeks' duration; in all other periods, expectation limited to 35 weeks has been used.

(b) *The results based on all tumours*

Method I compared with Method II

For all experiments together, the correlations of

the dates when 25, 50 and 75 % of surviving animals have tumours with the Tworts' Method I are slightly higher than with their Method II. The same is true of the expectation of tumourless life. There is no difference for the total number of tumours. These results are confirmed by the separate consideration of the experiments of different durations. With no significant exception, the dates when 25, 50 and 75 % of surviving animals have tumours correlate more highly with the Tworts' Method I than with Method II, and the same is true of expectation of tumourless life. Total number of tumours also show a slight excess in favour of Method I, but this is significant in only one case (45-week experiments). A similar result was noted with the tar experiments (Part I, p. 378).

Date when 25 % of surviving animals had tumours compared with date when 25 % of all animals had tumours

The latter measure is not as important for oils as for tars, as it is seldom reached. However, a comparison is possible for the experiments of 35 and 40 weeks' duration and for all experiments together. The 35-week experiments show in each case a small difference in favour of the former date, but individual differences are not significant. Neither the 40-week experiments nor 'all experiments together' show any systematic or significant differences. A somewhat similar result was obtained for the tars (Part I, p. 379).

Expectation of tumourless life and total number of tumours compared with the other alternative measures

Considering 'all experiments together', expectation of tumourless life correlates more highly with the Tworts' measures than measures based on a single date. Total number of tumours correlates with them to a slightly lower extent than measures based on a single date. As regards expectation, this result is completely confirmed by the separate consideration of the experiments of different durations. The results for 'total number of tumours' are more irregular, and the separate results of the experiments of different durations do not suggest any systematic difference between the magnitude of their correlations with the Tworts' measures and the magnitude of the correlations with the latter of the measures based on single dates. Total number of tumours correlates more highly with the Tworts' measures for oils than it does for tars (Part I, pp. 367, 376), while, on the whole, the reverse is true for measures based on single dates. Thus the disadvantage which was noted in the case of tar for 'total number of tumours' disappears in the case of the oils, and it must not be forgotten that measures based on single dates always suffer from the disadvantage that the dates may not be reached in particular experiments.

Measures based on single dates compared

As in the case of tars there are somewhat higher correlations with the later than the earlier dates when these are based on the percentage of surviving animals with tumours. Only the 25 % date, however, can be considered of much importance for oils, the later dates being reached too seldom.

Experiments of different durations compared

It is not possible to detect any systematic differences between the magnitudes of corresponding correlations in sets of experiments of different durations. There is no sign that the association between the Tworts' measures and alternative measures is related to the duration of the experiments.

The important points are:

(1) Expectation of tumourless life, as with the tar experiments, gives the highest correlations of all the alternatives.

(2) While the total number of tumours correlates more highly with the Tworts' measures for oils than for tars, measures based on single dates correlate somewhat less highly, so that for oils there is very little difference between them in this respect. Measures based on single dates suffer from the disadvantage that they may not be reached in particular experiments, and only the date when 25 % of surviving animals have tumours can be considered of any importance for oils.

(3) As with the tars, Method I seems slightly better than Method II.

(*c*) *The results based on malignant tumours only*

Method I compared with Method II

For all experiments together there are no significant differences between the correlations of the dates when 25, 50 and 75 % of surviving animals have tumours with the Tworts' Method I and with their Method II. The same is true for the expectation of tumourless life. Total number of tumours, however, correlates more highly with Method II than with Method I. These results are confirmed by the separate consideration of the experiments of different durations. The correlation of total number of tumours with Method I is decidedly lower for all experiments together than for any of the separate durations.

Expectation of tumourless life and total number of tumours compared with the other alternative measures

Considering 'all experiments together' expectation of tumourless life limited to 35 weeks corre-

lates more highly with the Tworts' measures than with the measures based on a single date. The correlations between total number of tumours and the Tworts' measures are irregular (the value of 0·36 for Method I has already been considered, see p. 385), but on the whole they are of the same magnitude as the correlations with them of measures based on a single date. On the whole a separate consideration of the experiments of different durations confirms these results, but certain qualifications have to be made. The expectation of tumourless life limited to the duration of the experiment, in all cases, correlated more highly with the Tworts' measures than any of the other alternatives. But it has already been pointed out that when malignant-tumour data only are considered, the expectations of tumourless life limited to a fixed period, say 25 or 35 weeks, tends to become a less satisfactory measure as the duration of the experiment is increased and its correlation with the Tworts' measures to fall. This is very noticeable with the expectation limited to 25 weeks, but that limited to 35 weeks remains satisfactory. However, its correlation with the Tworts' measures starts to drop in the 40-week experiments, and from this point onwards its correlation with the Tworts' measures is no higher than that of total number of tumours, when the experiments of different durations are considered separately. For data based on malignant tumours only, the date when 25 % of surviving animals had tumours is too seldom reached to be of importance.

Measures based on single dates compared

In view of what has just been said, this comparison is of no importance for malignant-tumour data, but the very few experiments which are available for the comparison (35- and 40-week experiments) show an advantage in favour of the 25 % date.

Experiments of different durations compared

Apart from the tendency, already mentioned, of expectation limited to a fixed period to correlate less highly with the Tworts' measures as the duration of the experiment is increased, there are no systematic differences between corresponding correlations in sets of experiments of different durations. The correlation between total number of tumours and Method I does, indeed, rise as the duration of the experiment is increased from 25 to 50 weeks, but as it drops again at 60 weeks no stress can be laid on this. Measures based on single dates become useless in experiments of more than 35 weeks' duration.

The important points are:

(1) Expectation of tumourless life, as with the tar experiments, gives the highest correlations of all the alternatives. This is true without qualification for expectation limited to the duration of the experiment, but for expectation limited to 35 weeks there is a small drop in the correlation for experiments of 40 weeks' duration or more. For these it gives no higher correlation than total number of tumours.

(2) The measures based on single dates are practically useless, being reached in too few experiments. Correlations with total number of tumours are reasonably high but somewhat irregular.

(3) The total number of tumours correlates more highly with Method II than with Method I.

(*d*) *Correlations in tar experiments and in oil experiments compared*

The correlations between the Tworts' measures and the alternative measures of potency considered are, on the whole, of the same order of magnitude for the oil experiments as for the tar experiments.

One of the best ways of making the comparison is to study the 25- and 40-week experiments which occur both for the tars and oils (see Part I, pp. 365–70). Comparing first the results based on all tumours, the correlations between the Tworts' measures and the dates when 25, 50 and 75 % of surviving animals have tumours are lower in the oil experiments than in the tar experiments of 40 weeks' duration; there is no systematic difference in the 25-week experiments. Correlations between the Tworts' measures and the total number of tumours are higher in the oil experiments than in the tar experiments, the difference being much greater in the 25-week than in the 40-week experiments. Going on to compare the results based on malignant tumours only, we find that the correlations between the Tworts' measures and the date when 25 % of surviving animals have tumours are again lower in the oil experiments than in the tar experiments of 40 weeks' duration; no comparison is possible for 25-week experiments. Correlations between the Tworts' measures and total number of tumours are higher in the oil experiments than in the tar experiments of 25 weeks' duration; in those of 40 weeks' duration there is no systematic difference. The difference between oils and tars in the magnitude of this correlation is not as great for the data based on malignant tumours only as for the data based on all tumours.

The general agreement between the oil and tar results is very satisfactory, since the oils are less potent than the tars, and also the range of potency in them is much less. The chief differences between the two sets of results are:

(1) For data based on all tumours, whereas in the tar experiments total number of tumours correlates less highly with the Tworts' measures than measures based on single dates, this is not so in the oil experi-

ments, there being very little difference between them.

(2) The measures based on single dates are of even less use for oils than for tars, the dates being reached in too few experiments.*

There are a number of smaller differences between the oil and tar results, for instance, in the relative behaviour of Methods I and II, but these are not important.

(e) *Significance of differences between the Tworts' potency measures*

The remarks made on pp. 370, 371 of Part I may be taken as applying to oils as well as to tars, with the qualification that 'total number of tumours' is a rather better measure for oils than for tars.

As with the tars, we start this section by giving (in Tables 29–31) the means and standard deviations of potencies and expectations. We add also the same constants for 'total number of tumours'. The latter measure, though somewhat crude, gives a concrete idea of the magnitude of the phenomenon it is desired to measure. The main points to be noticed are:

(1) Potency, however measured, declines with increasing duration of experiment after 35 weeks. This is doubtless because experiments which yielded very few tumours initially were continued longer in order to get some measure of what was obviously a low carcinogenic potency. There is no appreciable difference in potency between the 25- and 35-week experiments, although the latter yielded more tumours. The tumourless animals must have lived longer in the latter case, the number of tumours being higher because animals were exposed to risk longer. This emphasizes the drawbacks of 'total number of tumours' as a measure.

(2) The relative variability of the Tworts' measures is much greater than that of the expectation of tumourless life. For the Tworts' benign potencies the coefficient of variation is in the neighbourhood of 100 %, and even greater for their malignant potencies. For the expectation of tumourless life based on all tumours, it varies from 4 or 5 % to over 20 %, for the expectation based on malignant tumours only from less than 1 % to about 4 %. The figure of 4 % was attained for the expectation limited to 35 weeks in the 35-week experiments; for the expectation limited to 60 weeks in the 60-week experiments the coefficient of variation was 1 %. The corresponding means were 33·7 and 59·1 weeks, very close to the maxima possible, 35 and 60 weeks. This was bound to be the case, since the total number of malignant tumours in an experiment with 100 animals is so small. The average for all experiments was 3; about one-third of the experiments had none and about one-sixth one; the maximum was 18. Although the Tworts' 'malignant potencies' show much greater variation, we shall show that two of these measures have to differ rather widely before the difference can be regarded as significant.

Our next step, as with the tars, is to calculate the means and standard deviations of the Tworts' potencies for a given value of the expectation of tumourless life, in order to provide a basis for assessing the experimental error of their results. In the first place, since it is important to compare the daily oil and daily tar results, we consider only the daily oil experiments of 25 and of 40 weeks' duration. The means and standard deviations of the Tworts' potencies for given expectation have been calculated for these experiments and are presented (in Tables 32 and 33) exactly as in Tables 7 and 9. For the benign potencies there is a very fair agreement between oils and tars in the mean potencies for given expectation, but the standard deviations, though irregular, are somewhat greater in the 25-week oil experiments. The malignant potencies for given expectation and their variabilities are lower in the oil experiments and the difference cannot be ascribed to chance. Nevertheless, we are justified in saying that the difference is not large, seeing that in two *single* tar experiments a difference of 300 in malignant potency would not be significant.

Before putting the means and standard deviations of the Tworts' potencies for given expectation into a tabular form in which they can be used, it has been necessary to consider two points: (i) To what period should the expectation of tumourless life which we are to use as a basis be limited? (ii) Should all experiments be grouped together or should they be subdivided in some way? Of the experiments of less than 35 weeks' duration most lasted 25 weeks; for these, therefore, there is no alternative but to use an expectation limited to 25 weeks. However, it has already been noted that an expectation limited to 25 weeks becomes an almost useless measure for the experiments of longer durations (p. 391); for experiments of 35 weeks' duration or more, therefore, an expectation limited to 35 weeks has been used. That this choice is justified seems to be shown by Table 34 of correlations between expectation of tumourless life limited to 35 weeks and expectation limited to the duration of the experiment. Naturally the correlation falls somewhat as the duration of the experiment in-

* At first sight it might appear that this difficulty could be overcome by taking the date when, say, 5 or 10 % of surviving animals had tumours. On further consideration, it does not seem likely that this would provide a good measure. There are only 100 animals in each experiment; the result would depend on the behaviour of very few animals and probably be subject to a large sampling error.

Table 29. *Means and estimated standard deviations of potencies and expectations*

Duration of experiment			Potency: Method I	Method II	Mean of I and II	Mean of I and II with III	Expectation limited to (weeks): 25	35	40	45	50	60	Total no. of tumours
25 weeks	Benign	(Mean)	156	186	178	154	21·9	—	—	—	—	—	8·7
		(S.D.)	149	207	166	159	2·6	—	—	—	—	—	9·8
		*n**	38	31	31	38	38	—	—	—	—	—	38
	Malig-nant	(Mean)	62	37	50	47	24·8	—	—	—	—	—	0·8
		(S.D.)	155	64	107	99	0·3	—	—	—	—	—	1·1
		n	38	31	31	38	28	—	—	—	—	—	38
35 weeks	Benign	(Mean)	160	177	170	167	21·0	24·9	—	—	—	—	25·6
		(S.D.)	136	187	163	162	3·0	5·9	—	—	—	—	16·5
		n	124	120	120	124	124	124	—	—	—	—	124
	Malig-nant	(Mean)	46	72	59	58	24·8	33·7	—	—	—	—	4·3
		(S.D.)	49	71	57	57	0·3	1·5	—	—	—	—	4·5
		n	124	120	120	124	124	124	—	—	—	—	124
40 weeks	Benign	(Mean)	81	57	64	79	22·7	28·8	31·6	—	—	—	16·8
		(S.D.)	61	92	73	98	2·0	4·5	6·0	—	—	—	12·0
		n	65	56	56	65	65	65	65	—	—	—	65
	Malig-nant	(Mean)	14	24	19	20	24·97	34·7	39·3	—	—	—	1·9
		(S.D.)	16	34	24	24	0·07	0·5	1·0	—	—	—	2·6
		n	65	56	56	65	65	65	65	—	—	—	65

* n = number of experiments.

Table 30. *Means and estimated standard deviations of potencies and expectations*

Duration of experiment			Potency: Method I	Method II	Mean of I and II	Mean of I and II with III	Expectation limited to (weeks): 25	35	40	45	50	60	Total no. of tumours
45 weeks	Benign	(Mean)	41	20	29	35	23·9	31·3	34·5	37·4	—	—	14·9
		(S.D.)	34	23	25	33	1·1	3·1	4·2	5·6	—	—	9·7
		*n**	66	63	63	69	69	69	69	69	—	—	69
	Malig-nant	(Mean)	7	11	8	12	—	34·9	39·7	44·3	—	—	2·1
		(S.D.)	7	14	9	18	—	0·2	0·5	0·8	—	—	2·6
		n	66	63	63	69	—	66	66	66	—	—	69
50 weeks	Benign	(Mean)	20	8	13	13	24·5	33·3	37·3	40·9	44·2	—	9·6
		(S.D.)	23	9	14	14	0·6	1·8	2·7	3·7	4·8	—	7·7
		n	35	35	34	37	37	37	37	37	37	—	37
	Malig-nant	(Mean)	3	4	3	4	—	34·9	39·8	44·6	49·3	—	1·5
		(S.D.)	4	6	5	5	—	0·3	0·5	0·7	1·0	—	2·0
		n	35	35	34	37	—	37	37	37	37	—	37
60 weeks	Benign	(Mean)	10	3	6	6	24·8	34·2	38·7	43·1	47·3	54·9	7·3
		(S.D.)	16	5	10	10	0·4	1·3	1·9	2·6	3·8	5·3	7·1
		n	33	30	30	33	33	33	33	33	33	33	33
	Malig-nant	(Mean)	1	2	1	2	—	34·98	39·9	44·9	49·8	59·1	1·5
		(S.D.)	2	4	3	4	—	0·03	0·2	0·3	0·5	1·5	2·4
		n	33	30	30	33	—	33	33	33	33	33	33

* n = number of experiments.

Table 31. *Means and estimated standard deviations of potencies and expectations. (All experiments together)*

	Method I	Method II	Mean of I and II	Mean of I and II with III	Expectation limited to 25 weeks	Expectation limited to 35 weeks†	Total no. of tumours
Benign (Mean)	109	111	109	107	22·5	29·0	16·8
(S.D.)	123	167	143	141	2·7	5·6	14·1
n*	402	374	373	407	403	348	407
Malignant (Mean)	31	39	35	35	—	34·4	2·9
(S.D.)	61	54	52	52	—	1·1	3·3
n	402	374	373	407	—	348	407

* n = number of experiments. † Excluding experiments of less than 35 weeks' duration.

Table 32. *Daily oil experiments. Benign potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency		Observed range	Average S.D. for given expectation	
	25-week experiments						
Method I	15 and under 18*	3	514		624–386	104	
	18 ,, 21	9	254	170	393– 47	75	53
	21 ,, 24	14	115		201– 20	36	
	24 and 25	12	40		114– 0	40	
Method II	15 and under 18	3	543		747–395	64	
	18 ,, 21	8	207	221	323–128	42	187
	21 ,, 24	10	233		667– 34	242	
	24 and 25	10	16		55– 0	15	
Mean of I and II	15 and under 18	3	529		686–436	20	
	18 ,, 21	8	244	204	358–154	38	86
	21 ,, 24	10	173		408– 58	109	
	24 and 25	10	24		84– 0	25	
Mean of I and II with III	15 and under 18	3	529		686–436	20	
	18 ,, 21	10	196	174	358– 7	92	94
	21 ,, 24	13	158		408– 58	95	
	24 and 25	12	22		84– 0	24	
	40-week experiments						
Method I	15 and under 18	2	248		313–183	—	
	18 ,, 21	8	160		213–119	25	
	21 ,, 24	31	95		175– 31	19	
	24 and 25	24	18		65– 0	16	
Method II	15 and under 18	1	653		—	—	
	18 ,, 21	4	137		167– 88	44	
	21 ,, 24	28	61		110– 16	25	
	24 and 25	23	8		28– 0	7	
Mean of I and II	15 and under 18	1	483		—	—	
	18 ,, 21	4	148		178–108	38	
	21 ,, 24	28	77		142– 24	20	
	24 and 25	23	12		36– 0	10	
Mean of I and II with III	15 and under 18	2	454		483–425	—	
	18 ,, 21	8	178		491– 46	49	
	21 ,, 24	31	76		142– 24	21	
	24 and 25	24	14		67– 0	14	

Strictly 14·95–17·95, and similarly for the other groups except the last which includes 25·0.

creases, but the advantage of summarizing our results in as few groups as possible seems to outweigh any disadvantage arising from this circumstance.

The final results are shown in Tables 35–40, which give the means and standard deviations of (i) 'benign', (ii) malignant and (iii) 'final' potencies for given expectation of tumourless life. The expectations used are 'benign' expectations (i.e. based on *all* tumours) in cases (i) and (iii) and 'malignant' expectations (i.e. based on malignant tumours only) in case (ii). The groupings used here are the same as those in Tables 15–21. Table 35 may be compared with Table 19 and Table 36 with Table 20.

The results for the oils are in quite good agreement with those for the tars, as far as *benign potency* is concerned, but the means and variabilities of the *malignant potencies* for given expectation are smaller in the oils. Table 37 may be compared with Table 21. The comparison shows that the means and variabilities of *final potency* for given expectation are quite different in the oils from the tars. Tables 38–40 cannot be directly compared with the corresponding tables for the tars, as they are based on an expectation limited to 35 weeks, whereas for the tars an expectation limited to 25 weeks was used. Nevertheless, the lower expectation groups of Table 38 do not suggest anything contrary to the results of Table 19. A relatively small number of animals will have remained tumourless after 25 weeks in these groups, so that the expectation would not have been much altered had it been limited to 25 weeks. Table 40 confirms the conclusion drawn from Table 37 that the means and variabilities of *final potency* for given expectation are quite different in the oils from the tars.

It is not difficult to see the reason for this last fact. Ninety-four daily tar experiments yielded in all 4442 tumours, of which 2595 became malignant, or 58·4 %. 409 daily oil experiments yielded 6741 tumours of which 997 became malignant, or 14·8 %. Thus of a tar and an oil which have the same *benign potency* the latter will, *on the average*, have a lower *malignant potency*, and the final potency, which is half the sum of the two, will be lower. Further, the 'final potency' for a given 'benign expectation' will also be lower. Thus final potency may be a misleading measure in a group of experiments where there is much variation in the proportion of tumours which become malignant. The difficulty is inherent in the attempt to find a single measure of all aspects of a complex phenomenon.

Nevertheless, one practical question to which it is useful to have an answer is the following. Given a measure of potency calculated by the Tworts' method of, say, 300, for a single experiment, what is its experimental error?

The complete answer is that it depends on the substance experimented with, the duration of the experiment, the frequency of application, and by which method the potency was calculated. Yet a careful analysis of the results of this section shows that the standard deviation of potency at a given potency level, though by no means constant when these factors are varied, is of an order of magnitude which may be roughly estimated independently of

Table 33. *Daily oil experiments. Malignant potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
	25-week experiments				
Method I	23 up to 25*	38	62	860–0	62
Method II	23 up to 25	31	37	222–0	37
Mean of I and II	23 up to 25	31	50	486–0	47
Mean of I and II with III	23 up to 25	38	47	486–0	51
	40-week experiments				
Method I	24 and 25	65	12	75–0	17
Method II	24 and 25	56	22	174–0	33
Mean of I and II	24 and 25	56	17	124–0	24
Mean of I and II with III	24 and 25	65	18	124–0	24

* Strictly 22·95 up to and including 25·0.

Table 34. *Correlations between expectations of tumourless life limited* (i) *to* 35 *weeks, and* (ii) *to the duration of the experiment*

Duration of experiment	(*a*) All tumours	(*b*) Malignant tumours only
40 weeks	0·99	0·92
45 weeks	0·97	0·91
50 weeks	0·94	0·78
60 weeks	0·90	0·60

Table 35. *Daily oil experiments of less than 35 weeks' duration. Benign potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	10 and under 13*	1	916	—	—
	13 ,, 16	3	443	624–491	210
	16 ,, 19	8	390	533–320	64
	19 ,, 22	17	210	324– 47	53
	22 ,, 25	26	73	163– 0	39
Method II	10 and under 13	1	1203	—	—
	13 ,, 16	3	581	830–165	362
	16 ,, 19	8	415	605–272	71
	19 ,, 22	15	211	667–106	141
	22 ,, 25	19	65	209– 0	33
Mean of I and II	10 and under 13	1	1060	—	—
	13 ,, 16	3	512	686–189	280
	16 ,, 19	8	402	485–296	30
	19 ,, 22	15	216	408–138	70
	22 ,, 25	19	76	344– 0	57

* Strictly 9·95–12·95, and similarly for the other groups except the last, which includes 25·0.

Table 36. *Daily oil experiments of less than 35 weeks' duration. Malignant potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	23 up to 25*	55	64	860–0	64
Method II	23 ,, 25	46	47	222–0	46
Mean of I and II	23 ,, 25	46	54	486–0	53

* Strictly 22·95 up to and including 25·0.

Table 37. *Daily oil experiments of less than 35 weeks' duration. Final potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 25 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	10 and under 13*	1	555	—	—
	13 ,, 16	3	373	742–133	324
	16 ,, 19	8	274	440–223	78
	19 ,, 22	17	125	259– 23	48
	22 ,, 25	26	45	135– 0	34
Method II	10 and under 13	1	679	—	—
	13 ,, 16	3	316	429–103	184
	16 ,, 19	8	276	354–215	23
	19 ,, 22	15	124	333– 59	77
	22 ,, 25	19	50	333– 0	64
Mean of I and II	10 and under 13	1	617	—	—
	13 ,, 16	3	344	585–118	234
	16 ,, 19	8	275	362–219	30
	19 ,, 22	15	126	205– 74	44
	22 ,, 25	19	41	172– 0	30

* Strictly 9·95–12·95, and similarly for the other groups except the last, which includes 25·0.

Table 38. *Daily oil experiments of 35 weeks' duration or more. Benign potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 35 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	13 and under 16*	6	574	788–463	93
	16 ,, 19	19	300	494–183	92
	19 ,, 22	27	223	347–152	45
	22 ,, 25	29	147	243–114	22
	25 ,, 28	44	105	153– 53	17
	28 ,, 31	52	68	140– 6	18
	31 ,, 34	83	29	63– 2	10
	34 ,, 35	83	4	68– 0	7
Method II	13 and under 16	6	743	1152–485	246
	16 ,, 19	18	401	653–208	124
	19 ,, 22	26	236	367–111	62
	22 ,, 25	26	126	189– 66	29
	25 ,, 28	38	76	165– 7	25
	28 ,, 31	48	37	83– 8	19
	31 ,, 34	79	14	39– 2	9
	34 ,, 35	83	2	22– 0	3
Mean of I and II	13 and under 16	6	659	970–474	169
	16 ,, 19	18	354	562–216	80
	19 ,, 22	26	250	332–136	41
	22 ,, 25	26	136	192– 96	20
	25 ,, 28	38	90	146– 30	18
	28 ,, 31	48	51	92– 29	10
	31 ,, 34	79	21	50– 2	7
	34 ,, 35	82	3	45– 0	5

* Strictly 12·95–15·95, and similarly for the other groups except the last, which includes 35·0.

Table 39. *Daily oil experiments of 35 weeks' duration or more. Malignant potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 35 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	25 and under 28*	2	210	242–179	—
	28 ,, 31	6	152	196–120	20
	31 ,, 34	58	66	221– 10	38
	34 ,, 35	277	7	60– 0	8
Method II	25 and under 28	2	291	349–233	—
	28 ,, 31	6	211	324– 74	94
	31 ,, 34	56	102	258– 0	47
	34 ,, 35	260	12	101– 0	16
Mean of I and II	25 and under 28	2	251	296–206	—
	28 ,, 31	6	181	229– 97	50
	31 ,, 34	56	85	204– 5	35
	34 ,, 35	259	9	70– 0	11

* Strictly 24·95–27·95, and similarly for the other groups except the last, which includes 35·0.

that variation and which declines as the level of potency declines. Table 41 has been constructed to illustrate this. A rough interpolation has been made in the preceding tables, and the standard deviation of potency for given expectation of tumourless life at potency levels of 400, 300, 200, 100 and 50 has been estimated. The data only justify a very approximate calculation as the results are too irregular for anything else, the irregularity being particularly marked in Tables 35 and 37. Table 41 tells us, for instance, that a 'benign potency' of 400 calculated from a single experiment has a standard error of something between 60 and 140, while one of 50 has a standard error of something between 10 and 50. This is not very exact information, though by a knowledge of the type of experiment the estimate can be improved. But the information suffices to show that the experimental errors of the Tworts' potencies are large, and therefore that considerable caution should be exercised in drawing conclusions from them.

The standard error of the difference between two potencies may be calculated from the standard errors of the single potencies by the usual rule, and a difference which exceeds twice this may be regarded as significant.*

C. THE RELATION OF POTENCY TO PHYSICAL CHARACTERISTICS AND PLACE OF ORIGIN OF THE OIL

The Tworts' have laid great stress on the importance of physical characteristics and place of origin of the oil in indicating its carcinogenic potency. Up to 1935 they were laying great stress on specific gravity and refractivity; more lately they have been inclined to stress the prognostic value of the fall in refractive index of the oil when it has been allowed to remain in the peritoneal cavity of mice for a certain period of time.

The analysis in this section is confined to 'straight' oils. By 'straight' oils the Tworts appear to mean ordinary commercial mineral oils as received by them, which did not subsequently receive any further chemical treatment or undergo any blending.

(a) *Potency and physical characteristics*

There were 143 different straight oils for which some physical data were available. The potencies of 128 of these oils were determined by a single experiment each, in the remaining oils more than

* The careful *statistical* reader may ask:

'When potencies are low and the expectation of tumourless life consequently near its limiting value, will not the distribution of potencies for a given value of the expectation be far from normal and the usual rule of taking differences exceeding twice the standard error as significant be invalidated?'

The answer is as follows: When potencies are very low, the distribution of potencies for given expectation is very skew and leptokurtic (i.e. having a higher top and longer tail than the normal curve). The skewness, however, does not affect the symmetry of the frequency distribution of the difference between two potencies at the same potency level, which will be symmetrical whether the original distribution is skew or not. The effect of the leptokurtosis depends on the exact form of the distribution, which we do not know. The value of β_2 (the Pearsonian measure of leptokurtosis) for the distribution of the difference between two potencies at the same potency level is given by

$$B_2 = 3 + \tfrac{1}{2}(\beta_2 - 3),$$

where β_2 is the value for the original distribution. We give below the values of β_1, β_2 and B_2 for certain distributions of potency, at given levels, in experiments of 35 weeks' duration or more.

The only values of B_2 which differ significantly from the normal value of 3·0 are 5·0, 11·4 and 24·3. If we make the not unplausible assumption that the distribution of the difference between two potencies can be satisfactorily graduated by a Pearsonian curve (Type VII), it is easy to calculate the frequency between $\pm 2\sigma$ and $\pm 3\sigma$. We find:

	Percentage frequency between	
	$\pm 2\sigma$	$\pm 3\sigma$
∞	95·3	98·7
24·3	95·2	98·7
11·4	95·1	98·8
5·0	95·0	99·1
3·0 (normal value)	95·4	99·7

The effect of the leptokurtosis is quite negligible; the effect of errors of estimation in the standard error will, of course, be much greater.

	Expectation limited to 35 weeks	β_1	β_2	B_2	Standard error of B_2 (normal values)
'Benign potencies'	25–28	0·0	2·8	2·9	0·37
	28–31	2·4	6·9	5·0	0·34
	31–34	0·7	3·4	3·2	0·27
	34–35	34·3	45·6	24·3	0·27
'Final potencies'	25–28	0·0	2·8	2·9	0·37
	28–31	0·6	2·8	2·9	0·34
	31–34	0·5	3·0	3·0	0·27
	34–35	13·7	19·8	11·4	0·27

one experiment each had been performed as follows:

No. of experiments	No. of oils
2	9
3	3
4	2
6	1

But the experiments on the same oil were not as a rule of the same duration, and since some doubt was felt about the propriety of averaging results from

(2) '*Malignant*' *potency* (calculated as in (1)).

(3) '*Final*' *potency* (mean of 'benign' and 'malignant' potencies).

(4) '*Benign*' *expectation limited to* 25 *weeks* (i.e. expectation limited to 25 weeks calculated from *all* tumours).

(5) '*Malignant*' *expectation limited to* 35 *weeks* (i.e. expectation limited to 35 weeks calculated from *malignant* tumours only).

(1), (2) and (3) are the Tworts' own measures. When dealing with (5), experiments of less than

Table 40. *Daily oil experiments of* 35 *weeks' duration or more. Final potency for given expectation of tumourless life*

	Expectation of tumourless life (limited to 35 weeks)	No. of experiments	Mean potency	Observed range	Average S.D. for given expectation
Method I	13 and under 16*	6	365	483–306	68
	16 ,, 19	19	193	315–103	55
	19 ,, 22	27	142	228– 81	33
	22 ,, 25	29	94	135– 62	23
	25 ,, 28	44	61	91– 27	12
	28 ,, 31	52	39	65– 9	11
	31 ,, 34	83	16	34– 2	6
	34 ,, 35	83	2	34– 0	4
Method II	13 and under 16	6	494	693–345	135
	16 ,, 19	18	261	383–172	74
	19 ,, 22	26	165	263– 95	46
	22 ,, 25	26	96	170– 39	34
	25 ,, 28	38	56	106– 5	19
	28 ,, 31	48	27	72– 4	13
	31 ,, 34	79	11	31– 2	7
	34 ,, 35	83	1	15– 0	3
Mean of I and II	13 and under 16	6	430	588–325	101
	16 ,, 19	18	230	338–164	44
	19 ,, 22	26	154	225– 88	36
	22 ,, 25	26	96	152– 53	23
	25 ,, 28	38	59	89– 16	14
	28 ,, 31	48	32	57– 14	9
	31 ,, 34	79	13	36– 3	6
	34 ,, 35	83	2	22– 0	3

* Strictly 12·95–15·95, and similarly for the other groups except the last, which includes 35·0.

experiments of different durations, all except two of these oils, on one of which two experiments of 45 weeks' and on the other two of 40 weeks' duration had been performed, were omitted.* For these two oils the results of the two experiments were averaged. We thus had 130 oils; in 118 of these the specific gravity and refractivity were both available.

The measures of potency used were as follows:

(1) '*Benign*' *potency* (mean of Methods I and II, or Method III where the Tworts used that instead).

* The effect of including these omitted oils is discussed in Appendix II.

35 weeks' duration were excluded. This left ninety-seven experiments for which specific gravity and refractivity were available.

The physical measures were as follows:

Specific gravity. Specific gravity at 60° F.

Refractivity. Twort and Lyth used the formula $R=(n-1)/d$, where n = refractive index at 20° C. (i.e. 68° F.), d = density at 60° F.

Refractive index is largely influenced by density, but in refractivity the effect of density is eliminated. The effect of using density at 60° F., instead of density at 68° F., will be small and for practical

Table 41. *Average standard deviation of potency at certain potency levels*

	Potency level	Oil experiments, 35 weeks' duration and over Method of calculating potency			Oil experiments, under 35 weeks' duration Method of calculating potency			Daily tar experiments Method of calculating potency			Bi-weekly tar experiments Method of calculating potency			Highest and lowest
		I	II	Mean of I and II	I	II	Mean of I and II	I	II	Mean of I and II	I	II	Mean of I and II	
'Benign potencies'	400	90	120	90	90	100	100	60	110	70	60	140	100	60–140
	300	90	90	60	60	100	100	30	80	50	30	100	70	30–100
	200	40	50	30	50	100	100	30	50	40	30	60	40	30–100
	100	20	30	20	40	50	50	30	50	40	30	40	30	20– 50
	50	10	20	10	40	50	50	30	50	30	30	30	20	10– 50
'Malignant potencies'	400	—	—	—	—	—	—	160	180	170	140	150	130	130–180
	300	—	—	—	—	—	—	150	170	150	120	140	120	120–170
	200	—	—	—	—	—	—	140	160	140	100	130	100	100–160
	100	40	50	40	—	—	—	—	—	—	—	—	—	40– 50
	50	30	50	20	50	50	50	—	—	—	—	—	—	20– 50
'Final potencies'	400	—	110	90	—	—	—	100	90	80	140	130	130	80–140
	300	60	80	60	140	100	100	90	80	70	120	120	120	60–140
	200	50	60	40	60	100	100	80	70	70	100	100	110	40–110
	100	20	40	30	40	50	50	50	70	60	50	60	50	20– 70
	50	10	10	10	30	50	50	40	60	50	20	40	30	10– 60

purposes constant for all oils. Ordinary light (electric) was used with an Abbé refractometer.

The means, standard deviations and coefficients of variation of the measures used, for this series of oils, are given in Table 42.

Table 42. *Means, standard deviations and coefficients of variation of the biological and physical measures*

	Mean	Standard deviation	Coefficient of variation %
Benign potency	62·8	99·5	158
Malignant potency	13·7	21·3	155
Final potency	38·1	56·0	147
Benign expectation (25 weeks)	23·6	1·9	8·1
Malignant expectation (35 weeks)	34·7	0·5	1·4
Specific gravity	0·8994	0·0194	2·2
Refractivity	0·5538	0·0032	0·6

The Tworts' potencies are very variable measures, the other measures have much less variability.

In the first place, correlation coefficients were worked out between the biological measures of carcinogenic potency and specific gravity and refractivity. The results are given in Table 43.

Table 43. *Correlation coefficients between biological and physical measures*

	Specific gravity	Refractivity
Benign potency	+0·026	+0·59
Malignant potency	−0·016	+0·62
Final potency	+0·009	+0·60
Benign expectation (25 weeks)	−0·094	−0·69
Malignant expectation (35 weeks)	+0·040	−0·58

The correlations with specific gravity are quite insignificant, but refractivity possesses very considerable prognostic value. Since, however, the Tworts have laid considerable stress on specific gravity and refractivity as indices, the multiple regression equations between the five biological measures and specific gravity and refractivity were worked out. The multiple correlation coefficients were almost identical with the correlations with refractivity only (see Table 44), showing that if refractivity is used, specific gravity adds nothing further to the accuracy of the prediction. It was, however, clear that the relation between the biological measures and refractivity was not linear; using a parabolic regression line much improved the fit. It was found, nevertheless, that a parabola did not give a very good fit to the mean potencies for given refractivities at the lower end of the range (since it rose again slightly for decreasing refractivities). This was remedied by using a cubic curve which gave a further small improvement. These curves are shown in Figs. 1–4, together with their equations; correlations between the observed and fitted values are given in Table 44. The curves should be regarded simply as good graduations of the observed facts over the range of observation. They would not necessarily hold good outside this range.

The conclusion to be drawn from this section is that refractivity has a very considerable value as an index of carcinogenic potency, but we gain nothing by using specific gravity in addition. Refractivity already takes specific gravity into account.

Table 44. *Relation between biological measures of potency, specific gravity and refractivity. Correlation coefficients between observed and fitted values*

	Specific gravity and refractivity (linear regression equation)	Refractivity only	
		Parabola	Cubic
Benign potency	0·59	0·68	0·68
Malignant potency	0·62	0·81	0·83
Final potency	0·61	0·73	0·73
Benign expectation (25 weeks)	0·69	0·73	—
Malignant expectation (35 weeks)	0·58	0·72	—

(*b*) *Potency and refractive-index fall*

The fall in the refractive index of an oil which has been allowed to remain in the peritoneal cavities of mice for a certain period of time and is then recovered can hardly be called a purely physical characteristic of the oil. It is partly biological and partly physical. *Refractive-index fall* was in each case determined after the oil had been allowed to remain *one week* in the bodies of mice. Sometimes one mouse was used for the test, occasionally four, more often five, and once ten. Sometimes two tests were performed for each oil, in which case a weighted mean of the two results was used.

Of the 130 oils mentioned above (p. 402), the refractive index fall was available in 123; when experiments of less than 35 weeks' duration were excluded (as was necessary in dealing with malig-

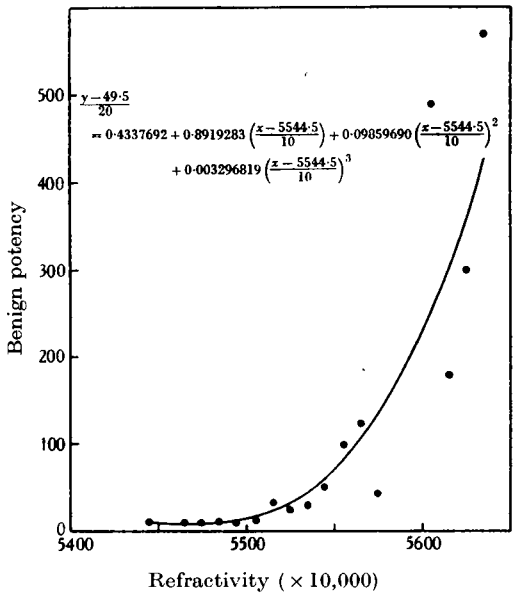

Fig. 1. Average benign potency for given refractivity.

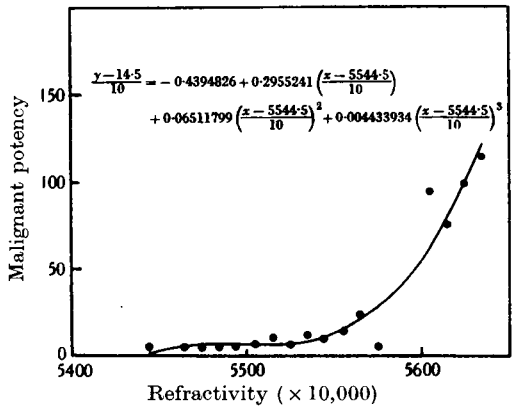

Fig. 2. Average malignant potency for given refractivity.

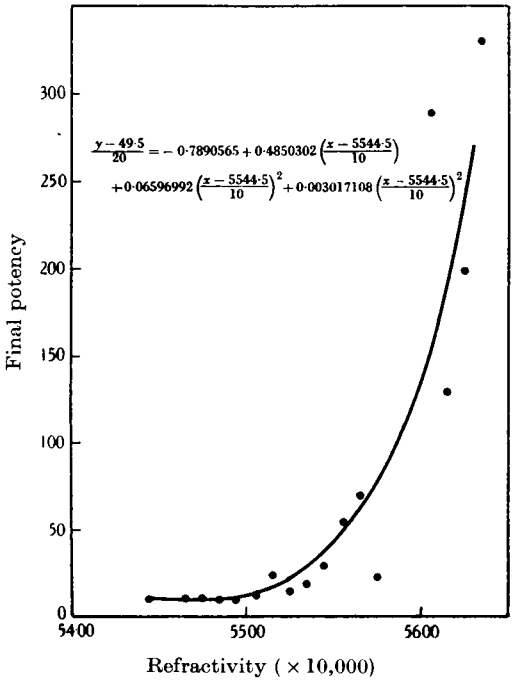

Fig. 3. Average final potency for given refractivity.

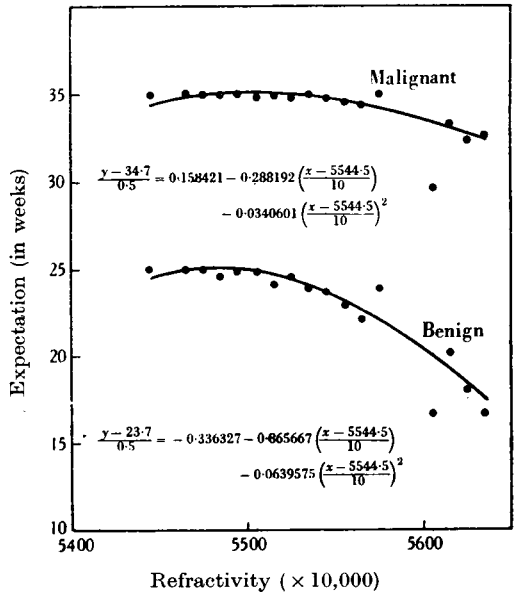

Fig. 4. Average expectation of tumourless life for given refractivity*

* Limited to 25 weeks for benign expectation, 35 weeks for malignant.

nant expectation), 101 oils could still be used. The mean refractive-index fall for the 123 oils was 0·0066 with a standard deviation of 0·0034, so that it is a rather variable measure. The correlation coefficients with the five measures of potency are as follows:

Benign potency	+0·33
Malignant potency	+0·18
Final potency	+0·29
Benign expectation	−0·47
Malignant expectation	−0·25

These associations are rather low; with the malignant potency and expectation they are little beyond the 5% level of significance. The average measures for given refractive-index falls are shown in Figs. 5–9, together with the regression straight lines. For

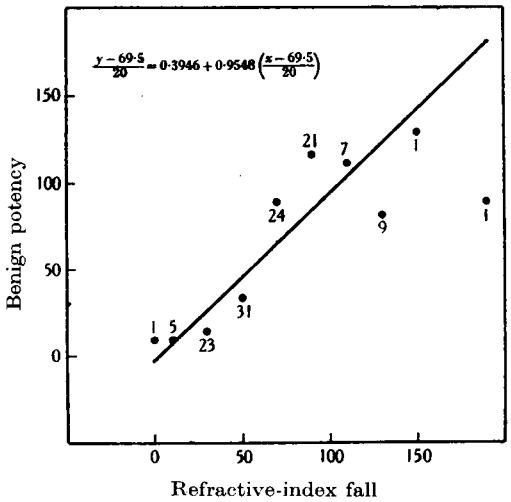

Fig. 5. Average benign potency for given refractive-index fall. Note: numbers against points are the numbers of observations upon which the corresponding points are based.

benign expectation, the regression straight line gives a fair fit, in the other cases it may be regarded as giving the general trend, but clearly where the association is so low it cannot be expected to fit closely, nor would any appreciable improvement result from the employment of curved regression lines. From the evidence presented by these oils, it would seem that refractive-index fall is not such a good index of carcinogenic potency as is refractivity.

(c) *Potency and place of origin*

We have in all ninety-six 'straight oils' classified by place of origin. The classification is that given by the Tworts themselves. Table 45 gives for each of these oils the final potency and the benign expectation of tumourless life limited to 25 weeks. (It also gives for each oil the refractivities and refractive-index falls which will be used in sections (*d*) and (*e*).) These two measures of potency will be sufficient for our purpose, the other three measures, were they used, would tell us much the same story

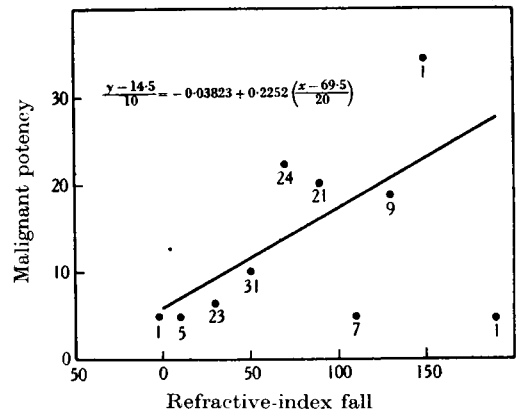

Fig. 6. Average malignant potency for given refractive-index fall. Note: numbers against points are the numbers of observations upon which the corresponding points are based.

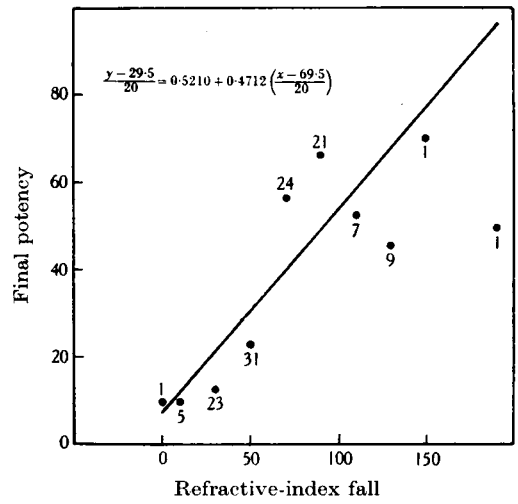

Fig. 7. Average final potency for given refractive-index fall. Note: numbers against points are the numbers of observations upon which the corresponding points are based.

owing to the high intercorrelations between these five measures.

Table 46 gives the mean 'final' potencies and 'benign' expectations for the oils from each place of origin, together with their standard errors.

There is no doubt that in these oils there are significant differences in potency according to place

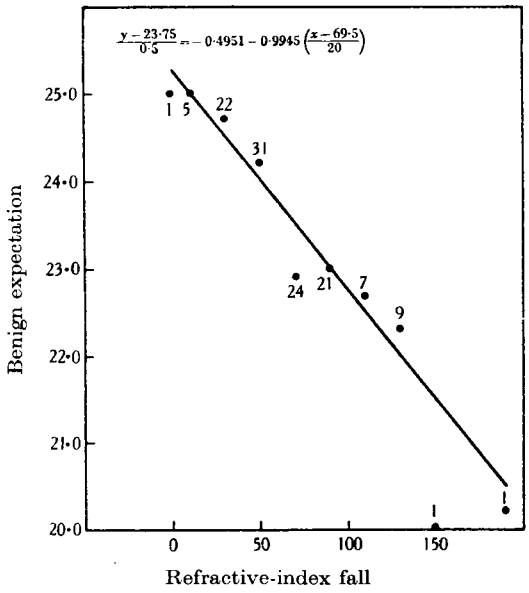

Fig. 8. Average benign expectation for given refractive-index fall. Note: numbers against points are the numbers of observations on which the points are based.

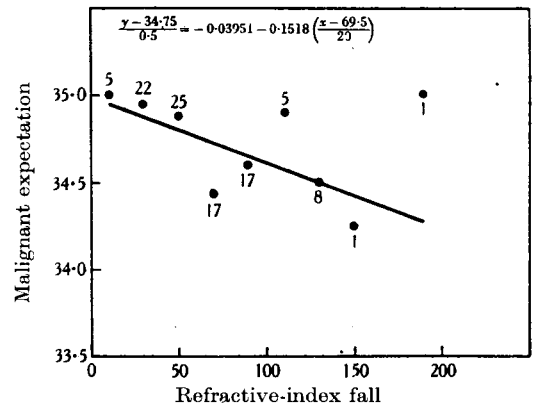

Fig. 9. Average malignant expectation for given refractive-index fall. Note: numbers against points are the numbers of observations upon which the points are based.

Table 45. *Biological and physical measures of individual oils*

Shale:										
Final potency	124	128	167	179	216	228	293	330	338	
Benign expectation	20·2	20·3	18·0	18·6	17·4	17·5	16·7	15·4	16·6	
Refractivity	5615	5615	5627	5613	5625	5621	5604	—	5633	
Refractive-index fall	97	74	74	89	63	71·7	—	—	90	
Borneo:										
Final potency	13	17								
Benign expectation	24·4	23·8								
Refractivity	5537	5537								
Refractive-index fall	—	—								
Persian:										
Final potency	8	11	12	42	61	74	78	82	278	
Benign expectation	24·5	24·6	24·4	22·3	22·4	22·8	21·6	20·2	18·6	
Refractivity	5572	5574	5559	5564	5566	5560	5569	5565	5563	
Refractive-index fall	96	110	34	80·5	73	75·5	129	138·2	92·5	
Russian:										
Final potency	0	0	0	0	0·2	2	4	5·5	6	30
Benign expectation	25·0	25·0	25·0	25·0	25·0	25·0	25·0	25·0	24·8	24·2
Refractivity	5447	5469	5470	5506	5506	5505	5502	5502	5503	5500
Refractive-index fall	0	4	3	53·5	20·6	41	26	49·5	58·8	49·5
Persian-Russian:										
Final potency	204									
Benign expectation	21·2									
Refractivity	5553									
Refractive-index fall	85									
Rumanian:										
Final potency	53									
Benign expectation	22·1									
Refractivity	5573									
Refractive-index fall	194									

Table 45. *Biological and physical measures of individual oils (continued)*

Mid-Continental:										
Final potency	10	10·5	31	34						
Benign expectation	23·5	24·6	23·4	20·7						
Refractivity	5561	5561	5555	5566						
Refractive-index fall	138	86·5	1205	122						
Venezuelan:										
Final potency	6	50	56	79	95	117				
Benign expectation	24·3	23·9	23·2	21·8	20·9	21·3				
Refractivity	5552	5554	5559	5555	5555	5563				
Refractive-index fall	60	63	134	1048	55·5	139				
Mexican:										
Final potency	2·5	3	12	16	18	18	53	62		
Benign expectation	24·8	25·0	24·2	24·6	22·7	35·85	21·1	21·5		
Refractivity	5549	5600	5546	5545	5595	5552	5544	5549		
Refractive-index fall	63·8	95	58	89·2	1225	90	1005	97·8		
Southern Mexican:										
Final potency	2·5	3·6								
Benign expectation	25·0	24·9								
Refractivity	5529	5524								
Refractive-index fall	60	79								
Mexican-Venezuelan:										
Final potency	26	33	37	172						
Benign expectation	22·4	24·1	22·4	22·5						
Refractivity	5554	5560	5566	5554						
Refractive-index fall	105	84·2	1013	1064						
Pennsylvanian:										
Final potency	1	1·5	1·5	1·5	2	3	5	10	22	30
Benign expectation	25·0	24·9	24·9	24·9	24·6	24·8	24·2	24·8	24·6	22·9
Refractivity	5544	5543	5548	5555	5546	5546	—	5552	5542	5544
Refractive-index fall	27	24	27	45	36·8	42·5	48	42	55	39
Final potency	37	44	61	73						
Benign expectation	23·1	23·2	22·8	22·3						
Refractivity	5546	5546	5545	5549						
Refractive-index fall	48	20	51·7	59						
Californian:										
Final potency	0	2·5	10	12	50	64	77			
Benign expectation	25·0	24·7	24·5	24·3	22·0	22·6	20·0			
Refractivity	5473	5485	5539	5540	5537	5543	5541			
Refractive-index fall	49	48·5	57	94	64·8	—	142			
Texan:										
Final potency	+*	1	1	2	2	4	4	4	4	8
Benign expectation	25·0	24·9	25·0	24·8	24·9	24·4	24·7	24·7	24·8	24·5
Refractivity	5507	5523	5507	5516	5507	5522	5509	5494	5514	5525
Refractive-index fall	36	40	34	49	98	57	87	81	55	90
Final potency	8	12	25	29	30	42				
Benign expectation	24·1	24·2	24·5	23·6	22·6	23·2				
Refractivity	5514	5517	5546	5512	5508	5516				
Refractive-index fall	74	74	58·2	64	82	61				
Peruvian:										
Final potency	2	2	2·5							
Benign expectation	24·9	25·0	25·0							
Refractivity	5481	5475	5479							
Refractive-index fall	32	81·8	51							

Refractivity figures and refractive-index falls multiplied by 10,000.

* + means a trace (counted as 0 in calculation).

Table 46. *Mean 'final potencies' and 'benign expectations' by place of origin*

Origin	Final potency Mean	Final potency S.E.	Benign expectation Mean	Benign expectation S.E.	No. of observations
Shale	222·5	27·3	17·9	0·5	9
Borneo	15·0	2·0	24·1	0.3	2
Persian	71·8	27·6	22·4	0·7	9
Russian	4·8	2·9	24·9	0·1	10
Persian-Russian	204·0	—	21·2	—	1
Roumanian	33·0	—	22·1	—	1
Mid-Continental	21·4	6·5	23·1	0·8	4
Venezuelan	67·2	15·9	22·6	0·6	6
Mexican	23·1	7·9	23·5	0·5	8
South Mexican	3·1	0·6	25·0	0·05	2
Mexican-Venezuelan	67·0	35·1	22·9	0·4	4
Pennsylvanian	20·9	6·5	24·1	0·3	14
Californian	30·8	12·1	23·3	0·7	7
Texan	11·0	3·3	24·4	0·2	16
Peruvian	2·2	0·2	25·0	0·03	3

of origin. The mean potency of the shale oils is significantly higher than that of the oils from any other source of origin, except possibly for the solitary anomalous Persian-Russian oil. This is shown not only by comparing the difference between pairs with their standard errors, but by the almost complete absence of overlap (see Table 45) between the results of the individual shale-oil experiments and the other experiments.* The Russian, Texan, South Mexican and Peruvian oils form a group of low carcinogenic potency, but the South Mexican and Peruvian oils are so few in number that we cannot feel any certainty about them.

The results suggest arranging the oils in four groups as in Table 47. However, the differences between the Persian oil, at the top of the second group, and the Pennsylvanian and Borneo oils, at the bottom of the third group, are of doubtful significance, so that we cannot attach much importance to the difference between the oils in the second and third groups. We can, however, be reasonably sure that the potency of the shale oils is high and of the Texan and Russian oils low.

Table 47. *Rank of oils according to two different potency measures*

	Final potency	Benign expectation
Shale	1	1
Persian-Russian	2	2
Persian	3	4
Venezuelan	4	5
Mexican-Venezuelan	5	6
Roumanian	6	3
Californian	7	8
Mexican	8	9
Mid-Continental	9	7
Pennsylvanian	10	10·5
Borneo	11	10·5
Texan	12	12
Russian	13	13
South Mexican	14	14·5
Peruvian	15	14·5

Of course, we do not know in the least whether these oils can be regarded as a random sample of the oils from the countries from which they came (the Tworts, too, have been careful to emphasize the need for caution in this respect), but we shall give further reasons later for believing the difference in carcinogenic potency between shale oils and Russian and Texan oils to be genuine.

(*d*) *Potency, place of origin and refractivity*

It is of interest to inquire whether refractivity is of any value in distinguishing between the carcino-

* The statistical reader may care to note that the analysis of variance within and between places is as follows:

Final potency

	Sum of squares	Degrees of freedom	Mean square
Between places	380608	14	27186
Within places	152062	81	1877
Total	532670	95	—

$$z = \tfrac{1}{2}\log_e\left(\frac{27186}{1877}\right) = 1{\cdot}34. \qquad 1\,\%\text{ point} = 0{\cdot}4.$$

Benign expectation

	Sum of squares	Degrees of freedom	Mean square
Between places	350·39	14	25·03
Within places	132·00	81	1·63
Total	482·39	95	—

$$z = \tfrac{1}{2}\log_e\left(\frac{25{\cdot}03}{1{\cdot}63}\right) = 1{\cdot}37. \qquad 1\,\%\text{ point} = 0{\cdot}4.$$

This again shows that there are significant differences in potency between oils from different places. A theoretical point that may be urged against the use of this test here is that there are clearly significant differences in variance from place to place, so that our estimate of variance within places is not, as it should be, an estimate of *one* variance but an estimate of the average of a number of variances. For this reason, the standard errors in Table 46 have been calculated from the individual estimates of variance and not the pooled estimate. There can be no doubt about the existence of significant *place* differences; for three different lines of approach lead to that conclusion, the standard errors of Table 46, the analysis of variance and the absence of overlap between individual experiments in different places.

genic potency of oils from the same place of origin, or whether it is only the usually wider differences between oils of different origin that it can pick out.

A statistical technique particularly adapted to answering this question is known as 'analysis of covariance'. If y be one of our potency measures, x refractivity, the sum of the squares of the deviations of observations from the mean for each of these measures, which may be denoted by $S(y-\bar{y})^2$ and to $S(x-\bar{x})^2$, can be split up into two portions due, respectively, to differences *between* different places of origin and to differences *within* the same place.* The same may be done with the product sum of deviations $S(x-\bar{x})(y-\bar{y})$. We thus have the following scheme:

Between places	$S_1(x^2)$	$S_1(y^2)$	$S_1(xy)$
Within places	$S_2(x^2)$	$S_2(y^2)$	$S_2(xy)$
Total	$S(x-\bar{x})^2$	$S(y-\bar{y})^2$	$S(x-\bar{x})(y-\bar{y})$

The total correlation coefficient between x and y is, of course, given by

$$r=\frac{S(x-\bar{x})(y-\bar{y})}{\sqrt{[S(x-\bar{x})^2\times S(y-\bar{y})^2]}}.$$

The correlation coefficient *between* places is given by

$$r_1=\frac{S_1(xy)}{\sqrt{[S_1(x^2)\times S_1(y^2)]}},$$

and that *within* by

$$r_2=\frac{S_2(xy)}{\sqrt{[S_2(x^2)\times S_2(y^2)]}}.$$

Applying this technique to our refractivity and two measures of potency (final potency and benign expectation), we obtain the following results for the ninety-four oils for which both refractivity and the potency measures are available:

The correlation within places is completely insignificant, while that between places is high. The former conclusion is confirmed by working out the correlation coefficients between the potency measures and refractivity for each type of oil separately; they are shown in Table 48. Not one of them is significant and the signs fluctuate; the only consistency is that the two measures agree in the sign of the association between potency and refractivity indicated, a high expectation, of course, denoting a low potency. With such extremely small numbers, little attention can be paid to the magnitude of the coefficients (with two observations, for instance, the coefficient *must* either be $+1$ or -1); but had there been a real correlation between potency and refractivity in oils from the same place, the signs of

Table 48. *Correlation coefficients between potency measures and refractivity*

Origin	'Final' potency	'Benign' expectation	No. of observations
Shale	+0·231	−0·285	8
Borneo	—	—	2
Persian	−0·300	+0·302	9
Russian	+0·275	−0·200	10
Persian-Russian	—	—	1
Roumanian	—	—	1
Mid-Continental	+0·031	−0·619	4
Venezuelan	+0·716	−0·498	6
Mexican	−0·374	+0·203	8
South Mexican	−1·0	+1·0	2
Mexican-Venezuelan	−0·467	+0·146	4
Pennsylvanian	−0·089	+0·122	13
Californian	+0·575	−0·582	7
Texan	+0·250	+0·000	16
Peruvian	+0·189	−0·754	3

Final potency (y) and refractivity (x)

	$S(x-\bar{x})^2$	$S(y-\bar{y})^2$	$S(x-\bar{x})(y-\bar{y})$	r
Between places	113,522	311,954	153,074	0·812
Within places	16,328	138,802	3,412	0·072
Total	129,850	450,756	156,486	0·647

Benign expectation (y) and refractivity (x)

	$S(x-\bar{x})^2$	$S(y-\bar{y})^2$	$S(x-\bar{x})(y-\bar{y})$	r
Between places	113,522	295,305	−5,127·97	−0·886
Within places	16,328	125,197	− 175·99	−0·123
Total	129,850	420,502	−5,303·96	−0·718

* This possibility is a consequence of the identities

$$S(x-\bar{x})^2=S(x-\bar{x}_p)^2+\Sigma n_p(\bar{x}_p-\bar{x})^2,$$

$$S(y-\bar{y})^2=S(y-\bar{y}_p)^2+\Sigma n_p(\bar{y}_p-\bar{y})^2,$$

$$S(x-\bar{x})(y-\bar{y})=S(x-\bar{x}_p)(y-\bar{y}_p)+\Sigma n_p(\bar{x}_p-\bar{x})(\bar{y}_p-\bar{y}),$$

where $\bar{x}_p$ denotes the mean of a particular place, n_p the number of oils from that place, $S(x-\bar{x}_p)^2$ the sum of the squares of the deviations of all the x observations from the mean of their own place, $\Sigma n_p(\bar{x}_p-\bar{x})^2$ the sum of the squares of the deviations of the place means from the general mean, with similar meanings for the (y^2) and (xy) sums. Above we have, for short, denoted $S(x-\bar{x}_p)^2$ by $S_2(x^2)$, $\Sigma n_p(\bar{x}_p-\bar{x})^2$ by $S_1(x^2)$, with similar meanings for the other symbols. The correlation coefficient 'within places' is analogous to the ordinary partial correlation coefficient. If the correlation between x and y for a constant value of z is the same at all levels of z, the partial correlation coefficient estimates this correlation. If not, it estimates an average of the correlations between x and y at different levels of z. We cannot calculate a partial correlation between potency and refractivity holding place constant because place is not numerically measurable. The 'within place' correlation calculated from the analysis of covariance serves the same purpose.

the coefficients would have been consistent from place to place.

This conclusion is of some importance, because it means that while refractivity will be of considerable importance in comparing the carcinogenic potency of widely differing oils, it could not discriminate between two rather similar oils, for instance, two shale oils.

(e) *Potency, place of origin and refractive-index fall*

The same analysis may be repeated for refractive-index fall; ninety-one oils were available. We have the following results:

Final potency (y) and refractive-index fall (x)

	$S(x-\bar{x})^2$	$S(y-\bar{y})^2$	$S(x-\bar{x})(y-\bar{y})$	r
Between places	68,001	258,578	46,959	0·354
Within places	41,360	129,739	16,975	0·232
Total	109,361	388,317	63,934	0·310

Benign expectation (y) and refractive-index fall (x)

	$S(x-\bar{x})^2$	$S(y-\bar{y})^2$	$S(x-\bar{x})(y-\bar{y})$	r
Between places	68,001	254·11	−1985·1	−0·478
Within places	41,360	122·02	− 872·3	−0·388
Total	109,361	376·13	−2857·4	−0·446

The correlation between potency and refractive-index fall in oils from different places is lower than that between potency and refractivity. But the correlation between potency and refractive-index fall in oils from the same place though low is significant, whereas there was no correlation between potency and refractivity in oils from the same place. This conclusion is confirmed by the correlation coefficients between potency and refractive-index fall for each place separately, shown in Table 49. As mentioned above, the magnitudes of these coefficients for such small groups have little meaning, but the signs are almost entirely consistent. With one exception they are all positive for the correlations with final potency, and with two exceptions all negative for the correlations with benign expectation. The probabilities of this occurring by chance, if there were no association, are $13/4096 = 0{\cdot}0032$ and $79/4096 = 0{\cdot}0193$ respectively.

Table 49. *Correlation coefficients between potency measures and refractive-index fall*

Origin	'Final' potency	'Benign' expectation	No. of observations
Shale	+0·004	+0·286	7
Persian	+0·164	−0·401	9
Russian	+0·447	−0·399	10
Persian-Russian	—	—	1
Roumanian	—	—	1
Mid-Continental	+0·226	−0·449	4
Venezuelan	+0·465	−0·279	6
Mexican	+0·391	−0·534	8
South Mexican	+1·0	−1·0	2
Mexican-Venezuelan	+0·448	−0·968	4
Pennsylvanian	+0·450	−0·386	14
Californian	+0·791	−0·806	6
Texan	+0·151	−0·287	16
Peruvian	−0·136	+0·790	3

This is again an interesting conclusion, because it tells us that refractive-index fall is of some, though severely limited, value in distinguishing between oils that are rather closely alike, while refractivity is of no use at all. Refractivity, on the other hand, is of considerably more use than refractive-index fall for distinguishing more widely different oils. We have used the term 'severely limited' because the correlation with refractive-index fall in oils from the same place accounts for only about 10% of the variance in carcinogenic potency.

(f) *Relation between the biological and physical measures. Summary*

Of the total variance in potency of the oils examined the correlation with refractive-index fall accounts for about 15%, of the variance of the place means about 20%. Refractivity accounts for about 50% of the total variance in potency and for about 70% of the variance of the place means. While there is no correlation, in oils from the same place, between potency and refractivity, refractive-index fall here accounts for about 10% of the variance in potency.

Refractivity correlates at least as highly with 'final potency' as do any of the alternative measures discussed in Section B, p. 385 (see Table 26), with the Tworts' potencies, with the exception only of expectation of tumourless life. It must therefore be regarded as a good indicator, the disadvantage that it cannot distinguish between rather similar oils being shared to a great extent by *all* measures, biological and physical, that have been discussed.

D. SATURATION, REFRACTIVITY, CARCINOGENIC POTENCY AND PLACE OF ORIGIN*

The Tworts have expressed the view that it is the unsaturated hydrocarbon constituents (and probably unsaturated ring structures) that are responsible for the carcinogenicity of mineral oils (see, for instance, Twort & Twort, 1931, p. 224). The relations of saturation to refractivity, therefore, become of considerable importance.

* We must acknowledge our indebtedness, in the preparation of this section, to Mr E. A. G. Shrimpton, whose knowledge of chemistry and patience in discussion has been of the greatest help.

(a) General relation of refractivity to saturation

Unsaturated hydrocarbons are those which have one or more double bonds in their structural formulae while saturated hydrocarbons have not. The consequence of this is that the unsaturated hydrocarbons will take into combination certain other substances (in particular the halogens) and that saturated hydrocarbons will not. One way of measuring the degree of unsaturation of an oil is, therefore, to determine its iodine value which is the amount of iodine per gram of oil which can be taken up in combination under certain specified conditions. Permanganate values, which have the same sort of definition, may also be determined. The Tworts have determined iodine and permanganate values for some of their oils, and to these reference will again be made later; here we are concerned with the relations between saturation and refractivity. It is stated in text-books on organic chemistry (see, for instance, von Richter, 1929, p. 52) that the refractivity of a liquid hydrocarbon may be calculated from the atomic refractivities of its elements together with a certain allowance for any double bonds in its structural formula. These atomic refractivities are constants which depend on the atom in question (in this case either carbon or hydrogen), on the wave-length of the light used and on the particular definition used for calculating refractivity.* For the definition $r=(n-1)/d$ which the Tworts have used and the sodium D line, the formula for the refractivity of a hydrocarbon with u carbon atoms, v hydrogen atoms and w double bonds is

$$r=\frac{4{\cdot}71u+1{\cdot}47v+2{\cdot}64w}{12u+v},$$

where $4{\cdot}71$ = atomic refractivity of carbon, $1{\cdot}47$ = atomic refractivity of hydrogen, $2{\cdot}64$ = allowance for a double bond. The atomic weight of carbon is 12 and of hydrogen 1, so that the denominator is the molecular weight of the compound. It follows from this that the effect of unsaturation is to increase refractivity, other things being equal; also since it is possible to show that for fixed u and w the value of r given by the formula increases as v increases, of two compounds with the same number of carbon atoms and double bonds, the one with the greater number of hydrogen atoms will have the higher refractivity.

In order to gain a better appreciation of the sort of differences in refractivity that might arise from these two circumstances Table 50 was constructed. The olefines and naphthenes listed in the *Handbook of the Petroleum Industry* (vol. **1**, 1922, p. 472) were tabulated and the theoretical values of the refractivities were calculated. Where the observed refractive indices and specific gravities were given, the actual refractivities were calculated from these data. The olefines are *all unsaturated with one double bond* and are chain hydrocarbons; the naphthenes are ring structures and are saturated. Olefines and naphthenes were selected because (see the *Petroleum Technologists' Pocket-Book*, 1923, p. 208) crude shale oils are stated to consist largely of olefines and crude Baku oils, which form the greater part of the Russian production, of naphthenes.* Unfortunately, the composition and constitution of lubricating oils, with which we are mainly concerned, may be quite different from that of the crudes from which they are derived; these values, therefore, have no direct relevance to the oils examined by the Tworts. The table is given purely for illustrative purposes; the observed and theoretical values are, it appears, not in conspicuously close agreement, but it does illustrate the tendency, in a comparison of two hydrocarbons with the same number of carbon atoms, one unsaturated and the other saturated, for the unsaturated to have the higher refractivity. It also shows that other things being equal reducing the proportion of hydrogen to carbon atoms also reduces refractivity.

Further, the table gives an idea of the order to magnitude of refractivities that are likely to arise (the greatest value is 0·5792, the lowest 0·5264); in the Tworts' oils whose physical characteristics we have examined the highest value was 0·5621 (a shale) and the lowest 0·5447 (a Russian), but again it must be emphasized that the Tworts' oils will be mixtures, mainly of lubricating fractions of complex structure. The refractivity of a mixture will be intermediate between the refractivities of the hydrocarbons which comprise it.

(b) Refractivity and saturation in lubricating oils

The next step, therefore, was to try and find out something about the constitution of mineral lubricating oils. Rather unexpectedly we found that comparatively little is known about their precise structure. For instance, Nash & Bowen in their work on the principles and practice of lubrication

* There are apparently two definitions to choose from, Gladstone's formula $r=(n-1)/d$ (where r = refractivity, n = refractive index, d = density) and Lorentz's formula $r=(n^2-1)/(n^2+2)d$. The Tworts have used the former. The Tworts did not in fact use the sodium D line, but ordinary white (electric) light. The effect of this difference on the refractivity is negligible.

* The terms olefines and naphthenes are apparently not always used in quite the same sense by chemists. One chemist told us that ring structures with olefine side chains are sometimes referred to as olefines. Ordinary naphthenes (for instance, those listed in the *Handbook of the Petroleum Industry*) are saturated, yet there is a reference below, quoted from Nash & Bowen (1929), to 'unsaturateds' which are possibly polynuclear naphthenes.

(1929) say (p. 177): 'The petroleum lubricant hydrocarbons are considered to be saturated and "unsaturated" compounds, probably of the naphthenic or polynuclear naphthenic type, and the existence of unstable bridge structures is also possible', and again (p. 171) 'the "unsaturateds", which are apparently not olefines or acetylenes, are render the correlation of the known hydrocarbons with the fraction that can be isolated from petroleum, difficult.'*

This being so, any direct numerical treatment of the relation between saturation and refractivity in lubricating oils is clearly impossible; nevertheless, it seems likely that of two mineral lubricating oils,

Table 50. *Theoretical and observed refractivities (olefine and naphthene series)*

No. of carbon atoms	Olefine series C_nH_{2n}		Naphthene series C_nH_{2n}		C_nH_{2n-2}		C_nH_{2n-4}		C_nH_{2n-6}	
n	Theoretical value*	Observed	Theoretical value*	Observed†	Theoretical value*	Observed	Theoretical value*	Observed	Theoretical value*	Observed
6	0·5778	0·5792	0·5464	0·5472	—	—	—	—	—	—
6	—	—	0·5464	0·5483	—	—	—	—	—	—
6	—	—	0·5464	0·5464	—	—	—	—	—	—
7	—	—	0·5464	0·5474	—	—	—	—	—	—
7	—	—	0·5464	0·5539	—	—	—	—	—	—
7	—	—	0·5464	0·5485	—	—	—	—	—	—
7	—	—	0·5464	0·5509	—	—	—	—	—	—
8	0·5700	—	0·5464	0·5501	—	—	—	—	—	—
8	—	—	0·5464	0·5488	—	—	—	—	—	—
8	—	—	0·5464	0·5525	—	—	—	—	—	—
8	—	—	0·5464	0·5512	—	—	—	—	—	—
8	—	—	0·5464	0·5519	—	—	—	—	—	—
8	—	—	0·5464	0·5465	—	—	—	—	—	—
9	0·5674	—	0·5464	0·5454	0·3315	0·5470	—	—	—	—
9	—	—	0·5464	0·5510	—	—	—	—	—	—
9	—	—	0·5464	0·5519	—	—	—	—	—	—
9	—	—	0·5464	0·5515	—	—	—	—	—	—
9	—	—	0·5464	0·5513	—	—	—	—	—	—
9	—	—	0·5464	0·5515	—	—	—	—	—	—
9	—	—	0·5464	0·5513	—	—	—	—	—	—
10	0·5653	0·5686	0·5464	0·5515	0·5330	0·5491	—	—	—	—
10	—	—	0·5464	0·5539	—	—	—	—	—	—
12	0·5621	—	—	—	0·5353	0·5480	—	—	—	—
13	0·5609	0·5574	—	—	0·5361	0·5442	—	—	—	—
14	0·5599	0·5485	—	—	—	—	—	—	—	—
15	0·5590	0·5498	—	—	0·5375	0·5481	—	—	—	—
16	0·5582	0·5473	—	—	0·5381	0·5303	0·5296	0·5465	—	—
17	0·5575	—	0·5464	0·5495	0·5386	0·5363	0·5306	0·5367	—	—
17	—	—	0·5464	0·5488	—	—	—	—	—	—
18	0·5569	—	—	—	—	—	0·5315	0·5437	—	—
19	—	—	0·5464	0·5504	—	—	0·5323	0·5466	—	—
20	0·5558	—	—	—	—	—	—	—	—	—
21	—	—	—	—	0·5401	0·5373	0·5336	0·5443	—	—
22	0·5550	—	0·5464	0·5493	0·5404	0·5454	—	—	—	—
23	0·5546	—	0·5464	0·5511	—	—	0·5348	0·5434	—	—
24	0·5543	—	0·5464	0·5526	0·5409	0·5467	0·5353	0·5427	—	—
25	—	—	—	—	—	—	0·5357	0·5445	—	—
26	0·5537	—	—	0·5512	—	—	—	—	—	—
27	0·5534	—	—	—	0·5415	0·5444	—	—	0·5264	0·5454
30	0·5527	—	—	—	—	—	—	—	—	—

* The theoretical values are as follows:

Olefine series: C_nH_{2n} $[0{\cdot}5464+(0{\cdot}1886)/n]$.
Naphthene series: C_nH_{2n} $[0{\cdot}5464]$.
C_nH_{2n-2} $[0{\cdot}5464-(1{\cdot}8471)/(14n-2)]$.
C_nH_{2n-4} $[0{\cdot}5464-(3{\cdot}6943)/(14n-2)]$.
C_nH_{2n-6} $[0{\cdot}5464-(7{\cdot}3886)/(14n-8)]$.

† Where several values are given with the same number of atoms, they correspond to different structural formulae.

possibly polynuclear naphthenes with unstable groupings or "bridged structures"....Our knowledge of the higher members of this series is deficient, and as already pointed out, although petroleum lubricants correspond to the formulae of this series from ultimate analysis, yet there are many differences in physical and chemical properties that the one with the higher refractivity will be the more unsaturated for the following reasons.

* A more recent article on 'Scottish Shale Oil' by G. H. Smith, A. G. Grant and S. Allen (*Petroleum Science*, **4**, 3099, 1938) also says: 'As regards the individual hydrocarbons present in shale oil, little is known of the heavier fractions.'

In the first place crude oils are separated into fractions of varying character by a process of distillation. The petroleum spirit which has the lowest boiling-point is separated off first, then the kerosene or illuminating oils, then fuel oils and finally lubricating oils which have the highest boiling-point, somewhere about 300° C. at atmospheric pressure. Now the boiling-point of a hydrocarbon increases with the number of carbon atoms present; in consequence there will be a lower limit, probably about 15, to the number of carbon atoms present in a lubricating oil. For instance, the numbers of carbon atoms in the formulae for five different lubricating oils listed by Nash & Bowen (1929, p. 175) are 20, 27, 30, 23, 25. The relevance of this will be seen shortly.

Now consider two lubricating oils which have equal proportions of constituents with given numbers of carbon and hydrogen atoms; clearly the one which contains the greater proportion of unsaturated constituents will have the higher refractivity. But this might conceivably be upset if the proportions of constituents with given numbers of carbon and hydrogen atoms were not the same in the two oils. The result is not very likely to be upset by differences in the proportion of constituents with given numbers of carbon atoms, because the effect of varying the numbers of carbon atoms above 15 is not very great. It might, however, be upset by differences in the proportion of constituents with given numbers of hydrogen atoms, as Table 51 illustrates. Saturated hydrocarbons with the formulae C_nH_{2n}, C_nH_{2n-2}, C_nH_{2n-4}, C_nH_{2n-8} are known to occur in lubricating oils (see Nash & Bowen, 1929, p. 178). We may compare their refractivities (theoretical values) with one another and with the refractivities they would have if they contained one double bond, again purely for illustrative purposes.* We have taken $n = 15$, 20, 25 and 30. We see that for hydrocarbons with the same formula, the unsaturated have the higher refractivities, the variation in the number of carbon atoms has relatively little effect, but variation in the number of hydrogen atoms associated with a given number of carbon atoms might lead to unsaturateds with no higher refractivities than the saturateds. The above argument is, therefore, not sufficiently strong. It probably places too much reliance on 'other things being equal'.

A more direct if less theoretical argument is to examine the iodine values which are themselves measures of degree of unsaturation. In eleven oils examined by the Tworts for which both refractivity and iodine values were available, a correlation of 0·81 between these two characteristics was obtained.

Thus, from whichever point of view the problem is considered, we do seem to come to the conclusion that the more unsaturated oils will, at any rate on the average, have the higher refractivities.

(c) *Saturation and refractivity in Russian and shale oils*

According to the *Petroleum Technologists' Pocket-Book* crude shale oil consists largely of olefines and Russian oil of naphthenes. Olefines are unsaturated and naphthenes saturated hydrocarbons. If, therefore, it could be shown that the unsaturateds present in crude shales lead to unsaturateds in the lubricating oils derived from them, while the Russian lubricating oils derived from saturated crudes were themselves saturated, we should have reason to believe that the results were not particular properties of the samples examined by them but were true in general. For one could then expect shale lubricating oils to have high refractivities and Russian lubricating oils to have low refractivities, and in consequence the former to be on the average relatively highly and the latter only slightly carcinogenic.

Saturated crudes, it appears,* give rise only to saturated lubricants, but it does not follow that unsaturated crudes give rise to unsaturated lubricants. Subsequent to distillation lubricating oils are refined by acid treatment, and drastic acid treatment may remove not only unsaturated chain structures but also unsaturated ring structures. Even mild acid treatment may remove unsaturated chain structures such as the olefines. The lubricating oils used in the spinning industry (spindle oils) are not very highly refined,† but the treatment might well remove any olefines. So any unsaturated structures that remain are likely to be ring structures. But here again we are impeded in our inquiry because the constitution of the higher boiling fractions of shale oils is largely unknown; we do not know whether they contain unsaturated ring structures. The only evidence is the high refractivities of the shale lubricating oils examined by the Tworts. So we come back to our starting point. Can the shale oils examined by the Tworts be regarded as a random sample of shale oils in general? We do not know; but this much may be said. While the stated origin of some of the foreign oils may be doubtful, and we cannot know how representative they are of the fields from which they come, the shale oils are all Scottish. There is very little reason

* Hydrocarbons with the formula C_nH_{2n} do occur with one double bond; they are ordinary olefines. Whether hydrocarbons with the formulae used for illustration and one double bond have been actually noted to occur, we do not know.

* From information given to us by C. C. Twort and by other chemists.

† C. C. Twort tells us that treatment with about 2% of sulphuric acid is usual.

to believe that they would have been a consciously biased selection, especially in a direction which is highly unfavourable to the producers.

To sum up, there is reason to believe that Russian lubricating oils will, as a rule, be saturated, have low refractivities and be relatively non-carcinogenic; there is some reason but not quite such a good reason to believe that shale lubricating oils will contain unsaturated constituents, have high refractivities and be relatively carcinogenic. This is certainly true of the Russian and shale lubricating oils examined by the Tworts. What evidence there is points to the presence of unsaturated ring structures rather than unsaturated chain structures as responsible, and supports the Tworts' view that the (carcinogenically) active constituents are of the nature of unsaturated ring-structured hydrocarbons. We must, however, guard against regarding unsaturation as the whole explanation; it may be necessary but can hardly be a sufficient condition.

Table 51. *Comparison of the refractivities of certain hydrocarbons (theoretical values of refractivities)*

	No. of carbon atoms (n)	C_nH_{2n}	C_nH_{2n-2}	C_nH_{2n-4}	C_nH_{2n-8}
Saturated	15	0·5464	0·5375	0·5285	0·5098
One double bond	15	0·5590	0·5502	0·5413	0·5227
Saturated	20	0·5464	0·5398	0·5330	0·5192
One double bond	20	0·5558	0·5493	0·5426	0·5289
Saturated	25	0·5464	0·5411	0·5357	0·5248
One double bond	25	0·5539	0·5487	0·5434	0·5325
Saturated	30	0·5464	0·5420	0·5375	0·5285
One double bond	30	0·5527	0·5483	0·5439	0·5439

For example, one of the Tworts' oils was a spirit derived from shale with a refractivity of 0·5619, and, therefore, very probably unsaturated, but with no carcinogenic potency at all.

E. SUMMARY TO PART II

1. Section A of this report is introductory; Section B deals with the correlations between the Tworts' measures of carcinogenic potency and certain alternative measures, and their use in testing the significance of differences between the potencies of different oils. Section C deals with the relation of potency to physical characteristics and place of origin of the oil. Section D deals with the relations between the saturation, refractivity, carcinogenic potency and place of origin of oils.

2. The correlations between the Tworts' measures and the alternative measures of potency considered are, on the whole, of the same order of magnitude for the oil experiments as for the tar experiments. This is a very satisfactory result, since the oils are less potent than the tars and also the range of potency in them is much less. The chief differences between the two sets of results are:

(i) For date based on all tumours, whereas in the tar experiments total number of tumours correlate less highly with the Tworts' measures than measures based on single dates, this is not so in the oil experiments, there being very little difference between them.

(ii) The measures based on single dates are of even less use for oils than for tars, the dates being reached in too few experiments.

3. Expectation of tumourless life, as with the tar experiments, gives the highest correlations with the Tworts' measures of all the alternatives considered. This is true without qualification for expectation limited to the duration of the experiment. When only malignant-tumour data are taken into account, the expectation limited to 25 weeks becomes useless for experiments of more than 40 weeks' duration. In the expectation limited to 35 weeks there is a small drop in the correlation for experiments of 40 weeks' duration or more. For these it gives no higher correlation than total number of tumours.

4. For data based on *all tumours*, total number of tumours correlates more highly with the Tworts' measures for oils than for tars. For data based on malignant tumours only there is little difference.

5. Differences between the magnitude of the correlations based on all tumours and on malignant tumours only are not even as clear-cut for the oils as they are for the tars, where they are also not of great importance.

6. Correlations between the Tworts' measures based (i) on all tumours and (ii) on malignant tumours only show a substantial amount of agreement between the two measures. This agreement is, however, not so good for oils as for tars, the correlations becoming rather low in individual cases. One would expect 'benign' and 'malignant' potencies to measure rather different things; if this is true, it is more marked in the oils than in the tars.

7. For oils as for tars, the variabilities of the Tworts' potency figures in experiments which have the same expectation of tumourless life have been calculated and may be used to test the significance of differences between potencies calculated from

different experiments. When data of *all tumours* are used, the means and variabilities of the Tworts' potencies at given levels of the expectation of tumourless life show good agreement in tars and oils. When, however, data for *malignant tumours only* are used, the means and the variabilities of the Tworts' potencies at given levels of the expectation of tumourless life are lower in oils than in tars. If the Tworts' 'final potencies' which are the mean of their 'benign' and 'malignant' potencies are used, the means and variabilities of these measures are found to be less for oils than for tars at the same level of expectation of tumourless life (calculated from all tumours). The reason is that on the average 58% of tumours become malignant in the tar experiments, while in the oil experiments only 15% did so. Thus final potency may be a misleading measure in a group of experiments where there is much variation in the proportion of tumours which become malignant. This difficulty is inherent in the attempt to find a single measure for all aspects of a complex phenomenon.

8. The experimental error of a 'potency figure' calculated by the Tworts' methods depends on the amount of animal variation, the substance experimented with, the duration of the experiment, the frequency of application and on the method by which the potency is calculated. Nevertheless, the standard deviation of potency at a given potency level, though by no means constant when these factors are varied, is of an order of magnitude which may be roughly estimated independently of that variation and which declines as the level of potency declines. The experimental errors of the Tworts' potencies are large, and considerable caution has to be exercised in drawing conclusions from them.

9. The relation of the Tworts' 'potencies' and of expectation of tumourless life to certain physical characteristics and to place of origin in about 100 'straight' oils has been examined. The physical characteristics in question were specific gravity, refractivity and the fall in refractive index of an oil which has been allowed to remain one week in the bodies of mice. No significant relation was found with specific gravity. Refractivity has a very considerable value as an index of carcinogenic potency, but, in these data, we gain nothing by using specific gravity in addition. Refractive-index fall has some value but not as much as refractivity. The carcinogenic potency of the shale oils examined is high and that of the Russian and Texan oils low. Other oils occupy an intermediate position and their differences are of doubtful significance. In these oils refractivity has no value for discriminating oils from the same place, but refractive-index fall has a small but definite value.

10. Of the total variance in potency of the oils examined the correlation with refractive-index fall accounts for about 15%, of the variance of the means of oils from difference places about 20%. Refractivity accounts for about 50% of the total variance in potency and for about 70% of the variance of the place means. While there is no correlation in oils from the same place, between potency and refractivity, refractive-index fall accounts for about 10% of the variance in potency.

11. Other things being equal unsaturated hydrocarbons have higher refractivities than saturated hydrocarbons and might be expected, on this account, to be on the average more carcinogenic. But not enough is known of the chemical constitution of lubricating oils to tell whether other things are equal or to give a general numerical treatment of the problem. There is reason to believe that Russian lubricating oils will, as a rule, be saturated, have low refractivities and be relatively non-carcinogenic; there is some reason, but not perhaps quite such good reason, to believe that shale lubricating oils will contain unsaturated constituents, have high refractivities and be relatively carcinogenic. *This is certainly true of the Russian and shale oils examined by the Tworts.* What evidence there is points to the presence of unsaturated ring structures rather than unsaturated chain structures as responsible, and supports the Tworts' view that the carcinogenically active constituents are of the nature of unsaturated benzenoid (i.e. ring-structured) hydrocarbons.

REFERENCES

Day, D. T. (1922). *Handbook of the Petroleum Industry*, **1**.

Nash, A. W. & Bowen, A. R. (1929). *The Principles and Practice of Lubrication.*

Redwood, B. & Eastlake, A. W. (1923). *Petroleum Technologists' Pocket-Book.*

Richter, V. von (1929). *Organic Chemistry*, **1**. (Translated and revised by E. Spielmann.)

Smith, G. R., Grant, G. G. & Allen, S. (1938). Scottish Shale Oil. *Petroleum Science*, **4**, 3099.

Twort, C. C. & Twort, J. M. (1930). *J. Hyg., Camb.*, **29**, 373–9.

Twort, C. C. & Twort, J. M. (1931). *J. Industr. Hyg.* **13**, 204–26.

Twort, C. C. & Twort, J. M. (1933). *Amer. J. Cancer*, **17**, 293–320.

(*MS. received for publication* 7. viii. 45.—Ed.)

APPENDIX I (TO PART II, SECTION B)

Intercorrelations between different measures of potency arranged according to the duration of the experiments

DAILY OIL EXPERIMENTS

A. *Experiments of 25 weeks' duration*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:								
25 %	−0·78 (29)	— (5)	−0·69 (24)	— (4)	−0·83 (24)	— (4)	−0·72 (30)	— (6)
50 %	−0·89 (24)	— (2)	−0·64 (21)	— (1)	−0·81 (21)	— (1)	−0·78 (25)	— (2)
75 %	−0·87 (14)	— (1)	−0·69 (13)	— (1)	−0·80 (13)	— (1)	−0·84 (15)	— (1)
Total no. of tumours*	0·87 (38)	0·54 (38)	0·49 (31)	0·85 (31)	0·73 (31)	0·68 (31)	0·74 (38)	0·67 (38)
Expectation of tumourless life	−0·93 (38)	−0·80 (39)	−0·72 (31)	−0·78 (31)	−0·91 (31)	−0·87 (31)	−0·85 (38)	−0·85 (38)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Benign tumours: In only five cases was the date when 25 % of all animals had tumours reached. The 50 % point was never reached. Malignant tumours: The 25 % point was never reached.

B. *Experiments of 35 weeks' duration*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:*								
25 %	−0·73 (112)	−0·63 (62)	−0·69 (108)	−0·59 (61)	−0·71 (108)	−0·67 (61)	−0·71 (112)	−0·67 (62)
50 %	−0·75 (99)	−0·43 (24)	−0·75 (97)	−0·53 (24)	−0·76 (97)	−0·53 (24)	−0·75 (99)	−0·53 (24)
75 %	−0·79 (83)	—	−0·84 (81)	—	−0·82 (81)	—	−0·82 (82)	—
Total no. of tumours	0·63 (124)	0·65 (124)	0·61 (120)	0·80 (120)	0·62 (120)	0·77 (120)	0·62 (124)	0·77 (124)
Date when 25 % of all animals had tumours†	−0·59 (63)	—	−0·58 (63)	—	−0·59 (63)	—	−0·59 (63)	—
Expectation of tumourless life (in weeks):								
25	−0·90 (124)	−0·63 (124)	−0·89 (120)	−0·73 (120)	−0·89 (120)	−0·71 (120)	−0·89 (124)	−0·72 (124)
35	−0·86 (124)	−0·83 (124)	−0·84 (120)	−0·85 (120)	−0·85 (120)	−0·89 (120)	−0·85 (124)	−0·88 (124)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Malignant tumours: The 75 % point was reached in six cases.

† Benign tumours: In only nine cases was the 50 % point reached and in no case the 75 % point. Malignant tumours: The 25 % point was never reached.

C. *Experiments of 40 weeks' duration*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:*								
25 %	−0·82 (53)	−0·74 (27)	−0·50 (44)	−0·45 (23)	−0·63 (44)	−0·58 (23)	−0·56 (53)	−0·59 (27)
50 %	−0·84 (45)	−0·57 (12)	−0·62 (36)	−0·18 (12)	−0·70 (36)	−0·31 (12)	−0·69 (45)	−0·31 (12)
75 %	−0·87 (28)	—	−0·72 (22)	—	−0·79 (22)	—	−0·83 (28)	—
Total no. of tumours	0·87 (65)	0·68 (65)	0·75 (56)	0·83 (56)	0·84 (56)	0·96 (56)	0·70 (65)	0·78 (65)
Date when 25 % of all animals had tumours†	−0·79 (15)	—	−0·68 (12)	—	−0·71 (12)	—	−0·67 (15)	—
Expectation of tumourless life (in weeks):								
25	−0·96 (65)	−0·22 (65)	−0·74 (56)	−0·32 (56)	−0·92 (56)	−0·32 (56)	−0·88 (65)	−0·37 (65)
35	−0·97 (65)	−0·70 (65)	−0·76 (56)	−0·70 (56)	−0·87 (56)	−0·76 (56)	−0·82 (65)	−0·79 (65)
40	−0·97 (65)	−0·89 (65)	−0·72 (56)	−0·85 (56)	−0·85 (56)	−0·93 (56)	−0·80 (65)	−0·93 (65)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Malignant tumours: The 75 % point was reached in only one experiment.

† Benign tumours: In only two cases was the 50 % point reached. The 75 % point was never reached. Malignant tumours: The 25 % point was never reached.

D. *Experiments of* 45 *weeks' duration*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:*								
25%	−0·82 (50)	−0·65 (13)	−0·68 (47)	0·09 (11)	−0·89 (47)	−0·05 (11)	−0·75 (53)	−0·78 (16)
50%	−0·81 (37)	—	−0·77 (34)	—	−0·91 (34)	—	−0·81 (40)	—
75%	−0·48 (19)	—	−0·89 (16)	—	−0·80 (16)	—	−0·88 (22)	—
Total no. of tumours†	0·85 (66)	0·74 (66)	0·68 (63)	0·78 (63)	0·87 (63)	0·85 (63)	0·81 (69)	0·81 (69)
Expectation of tumourless life (in weeks):								
25‡	−0·86 (66)	—	−0·75 (63)	—	−0·88 (63)	—	−0·88 (69)	—
35	−0·92 (66)	−0·78 (66)	−0·76 (63)	−0·58 (63)	−0·94 (63)	−0·70 (63)	−0·93 (69)	−0·79 (69)
40	−0·93 (66)	−0·79 (66)	−0·80 (63)	−0·75 (63)	−0·97 (63)	−0·85 (63)	−0·94 (69)	−0·87 (69)
45	−0·91 (66)	−0·82 (66)	−0·80 (63)	−0·78 (63)	−0·95 (63)	−0·88 (63)	−0·93 (69)	−0·89 (69)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Malignant tumours: The 50% point was reached in only five cases (Methods I, II and mean of I and II), and in only eight cases (mean of Methods I and II with III). The 75% point was reached in two and three cases respectively.

† Benign tumours: The 25% point was reached in eight cases (Method I), seven cases (Method II and mean of I and II) and ten cases (mean of Methods I and II with III). The 50% point was never reached. Malignant tumours: The 25% point was never reached.

‡ For malignant tumours, the expectation limited to 25 weeks was not calculated.

E. *Experiments of* 50 *weeks' duration*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:*								
25%	−0·73 (25)	0·15 (8)	−0·72 (25)	0·25 (8)	−0·64 (24)	0·23 (8)	−0·57 (26)	0·23 (8)
50%	−0·95 (14)	—	−0·94 (13)	—	−0·95 (13)	—	−0·73 (15)	—
Total no. of tumours†	0·77 (35)	0·88 (35)	0·81 (35)	0·77 (35)	0·78 (34)	0·87 (34)	0·78 (37)	0·84 (37)
Expectation of tumourless life (in weeks):								
25‡	−0·90 (35)	—	−0·85 (35)	—	−0·86 (34)	—	−0·62 (37)	—
35	−0·90 (35)	−0·79 (35)	−0·86 (35)	−0·60 (35)	−0·87 (34)	−0·72 (34)	−0·71 (37)	−0·48 (37)
40	−0·88 (35)	−0·80 (35)	−0·95 (35)	−0·64 (35)	−0·84 (34)	−0·75 (34)	−0·72 (37)	−0·61 (37)
45	−0·85 (35)	−0·82 (35)	−0·84 (35)	−0·67 (35)	−0·82 (34)	−0·78 (34)	−0·73 (37)	−0·72 (37)
50	−0·82 (35)	−0·88 (35)	−0·81 (35)	−0·76 (35)	−0·79 (34)	−0·86 (34)	−0·73 (37)	−0·83 (37)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Benign tumours: In only eight cases was the 75% point reached: Malignant tumours: The 50% point was reached in only two cases, the 75% point in one only.

† Benign tumours: In only two cases was the 25% point reached, and in no case was the 50% point reached. Malignant tumours: The 25% point was not reached.

‡ The expectation of tumourless life limited to 25 weeks was not calculated for malignant tumours.

F. *Experiments of* 60 *weeks' duration*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:*								
25%	−0·71 (21)	—	−0·58 (18)	—	−0·75 (18)	—	−0·68 (21)	—
50%	−0·90 (13)	—	−0·69 (10)	—	−0·97 (10)	—	−0·94 (13)	—
Total no. of tumours†	0·81 (33)	0·69 (33)	0·76 (30)	0·78 (30)	0·92 (30)	0·77 (30)	0·82 (33)	0·82 (33)
Expectation of tumourless life (in weeks):								
25‡	−0·93 (33)	—	−0·59 (30)	—	−0·93 (30)	—	−0·93 (33)	—
35	−1·00 (33)	−0·68 (33)	−0·67 (30)	−0·84 (30)	−0·99 (30)	−0·84 (30)	−0·97 (33)	−0·68 (33)
40	−0·99 (33)	−0·83 (33)	−0·70 (30)	−0·85 (30)	−0·99 (30)	−0·87 (30)	−0·97 (33)	−0·80 (33)
45	−0·97 (33)	−0·91 (33)	−0·72 (30)	−0·91 (30)	−0·99 (30)	−0·94 (30)	−0·97 (33)	−0·89 (33)
50	−0·95 (33)	−0·91 (33)	−0·75 (30)	−0·97 (30)	−0·98 (30)	−0·98 (30)	−0·95 (33)	−0·93 (33)
60	−0·88 (33)	−0·76 (33)	−0·79 (30)	−0·92 (30)	−0·95 (30)	−0·89 (30)	−0·90 (33)	−0·87 (33)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Benign tumours: In only six cases was the 75% point reached. Malignant tumours: The 25% point was reached in five cases, the 50 and 75% points in one only.

† Benign tumours: In only one case was the 25% point reached, and in no case the 50% point. Malignant tumours: 25% point was not reached.

‡ The expectation of tumourless life limited to 25 weeks was not calculated for malignant tumours.

G. *Experiments of all durations*

	Method I		Method II		Mean of I and II		Mean of I and II with III	
	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only	All tumours	Malignant tumours only
Date when following percentage of surviving mice had tumours:*								
25 %	−0·64 (320)	−0·57 (129)	−0·54 (294)	−0·67 (118)	−0·60 (293)	−0·70 (118)	−0·58 (325)	−0·66 (133)
50 %	−0·71 (261)	−0·63 (53)	−0·64 (238)	−0·54 (52)	−0·69 (238)	−0·69 (52)	−0·67 (266)	−0·63 (56)
75 %	−0·75 (181)	—	−0·70 (166)	—	−0·74 (166)	—	−0·72 (185)	—
Total no. of tumours	0·64 (402)	0·36 (402)	0·59 (374)	0·79 (374)	0·62 (373)	0·60 (373)	0·61 (407)	0·59 (407)
Date when 25 % of all animals had tumours†	−0·60 (103)	—	−0·60 (99)	—	−0·61 (99)	—	−0·61 (105)	—
Expectation of tumour-less life (in weeks):								
25	−0·91 (398)	—	−0·86 (370)	—	−0·90 (369)	—	−0·89 (403)	—
35	—	−0·87 (343)	—	−0·90 (324)	—	−0·91 (323)	—	−0·90 (348)

Numbers in brackets are the number of observations upon which the corresponding correlations are based.

* Malignant tumours: The 75 % point was reached in sixteen cases.

† Benign tumours: The 50 % point was reached in fourteen cases, the 75 % point was never reached. Malignant tumours: The 25 % point was never reached.

APPENDIX II (TO PART II, SECTION C)

The effect of including twelve oils whose omission is mentioned on p. 402 is here considered. Of the fifteen oils mentioned on p. 402, on which several experiments had been performed, there were twelve on which two experiments each of the same duration had been performed. All the twelve were omitted. Two out of the three were included in the analysis of potency, specific gravity and refractivity; the third could not be because its specific gravity was doubtful. The twelve were the only omissions, apart from omissions due to one or more of the characteristics used not being given, or being marked by Dr Twort as of doubtful accuracy.

Among the twelve oils were three Borneo oils of high specific gravity and considerably higher potency than the remaining two Borneo oils which *were* included. Accordingly, it seems desirable to re-examine the relation of carcinogenic potency to specific gravity and place of origin when the twelve oils are included. We then have 130 oils available. For the twelve oils now added, the potencies from different experiments on the same oil have been averaged. The correlation coefficient between final potency and specific gravity is now 0·041 as against 0·009 with the twelve oils omitted. Both values are quite negligible. The correlation ratio (of final potency on specific gravity) is 0·273, while the average value to be expected by chance if there were no correlation is 0·278.

If the twelve oils previously omitted are now included in the study of the relation between potency and place of origin Table 46 is modified as shown in Table A. There each of the twelve oils has been allotted the average 'final potency' and 'benign expectation' of the different experiments performed on it.* Comparison with the original Table 46 shows that the alterations are trivial except for the Borneo oils. Table 47 is modified as shown in Table B.

Table A. *Mean 'final potencies' and 'benign expectations' by place of origin*

Origin	Final potency Mean	Final potency S.E.	Benign expectation Mean	Benign expectation S.E.	No. of oils
Shale	227·6	22·4	17·8	0·4	11
Borneo	109·7	77·2	21·9	1·3	5
Persian	71·8	27·6	22·4	0·7	9
Russian	10·8	5·7	24·8	0·1	12
Persian-Russian	204·0	—	21·2	—	1
Persian-Mexican	37·0	—	22·6	—	1
Roumanian	64·8	11·8	22·3	0·2	2
Mid-Continental	19·2	5·5	23·4	0·7	5
Venezuelan	70·6	13·9	22·2	0·6	7
Mexican	23·1	7·9	23·5	0·5	8
South Mexican	3·1	0·7	25·0	0·05	2
Mexican-Venezuelan	67·0	35·1	22·9	0·4	4
Pennsylvanian	25·3	6·6	23·9	0·3	16
Californian	30·8	12·1	23·3	0·7	7
Texan	11·0	3·3	24·4	0·2	16
Peruvian	2·2	0·2	25·0	0·03	3

Now let us consider the Borneo oils. The average result for the Borneo oils has the largest standard error of any place of origin in the table. In 'final

* The results have also been calculated, counting each experiment as a separate oil, the results do not differ in any material particular from those given.

Table B. *Rank of oils according to two different potency measures*

	Final potency	Benign expectation
Shale	1	1
Persian-Russian	2	2
Borneo	3	3
Persian	4	6
Venezuelan	5	4
Mexican-Venezuelan	6	8
Roumanian	7	5
Persian-Mexican	8	7
Californian	9	9
Pennsylvanian	10	12
Mexican	11	11
Mid-Continental	12	10
Texan	13	13
Russian	14	14
South Mexican	15	15½
Peruvian	16	15½

potency' it does not differ significantly from any other single place of origin in the table. In 'benign expectation' it differs significantly from the shale and Russian and perhaps from the South Mexican and Peruvian oils. In fact, it has been transferred by including the previously omitted oils from the bottom of the two middle doubtful groups to the top. The average 'benign expectation' has been shortened by 2·2 weeks. The individual values for the five Borneo oils are for 'final potency' 13, 17, 40·4, 61·5 and 416·5 and for 'benign expectation' 24·4, 23·8, 22·8, 21·8 and 16·8. The large range of potency should be noted, in fact these five oils differ by almost as much as five oils selected at random without regard to place of origin. This is the common-sense reason why classification of potency by place of origin in the middle groups is so uncertain. The specific gravities of these five oils are 0·940, 0·950, 0·949, 0·943 and 0·939 respectively. All these specific gravities are high, but the first two oils including the one with the highest specific gravity are certainly not particularly carcinogenic, while the one with the highest carcinogenic potency has the lowest specific gravity.