THE FREQUENCY OF CANCER DEATHS IN THE SAME HOUSE AND IN NEIGHBOURING HOUSES

By PERCY STOCKS

Medical Statistical Officer, General Register Office, London

CONTENTS

I. Introduction

The problem as to the frequency with which multiple cases of, or deaths from, a disease may be expected to occur in the same house has been attacked by Troup and Maynard (1911–12), Karl Pearson (1911–12, 1911–12 *a*, 1913), McKendrick (1911–12), Greenwood and Yule (1920), Greenwood (1910, 1931) and others. Troup and Maynard obtained a general expression for the frequency of occurrence, and the standard deviation of the frequency, of s cases in a house when n cases of a disease occur in a town with m houses, assuming each house to contain the same number of persons, all equally liable to contract it. Karl Pearson (1911–12) considered the problem as analogous to throwing balls at random into equally accessible pigeon-holes, and showed that since the chance that any one case will fall into any one house is $\frac{1}{m}$, and that it will not is $\frac{m-1}{m}$, the numbers of houses having 0, 1, 2, 3, ... cases should be given by the successive terms of the binomial $m\left\{\frac{m-1}{m}+\frac{1}{m}\right\}^n$. In a later paper (1913) he showed that the correspondence between the observed and expected frequencies could be measured by applying the ordinary χ^2 test to the frequencies, omitting the houses having no cases.

In applying this formula to Dr Webb's data for 377 cases of cancer occurring at Madeley between 1837 and 1910, he found (1911–12 *a*) that houses with three or four cases occurred more frequently than would be expected, but considered that without further and more ample data from other districts it would be unwise to draw conclusions from this.

Prof. Greenwood (1910, 1931) has discussed other possible formulae applicable to problems of this kind. Lumière and Vigne (1932, 1933), investigating 6703 cancer deaths in Lyons during 20 years, found that out of 23,258 houses 18,231 had no cancer death in the period, 3769 had 1, 953 had 2, 228 had 3, 53 had 4, 16 had 5, 5 had 6, 2 had 7 and 1 had 8 deaths. Comparing with this the frequency distribution of 6703 births the corresponding frequencies were 18,248, 3736, 982, 212, 51, 18, 6, 3, 2, and for 6703 deaths from all causes they were 18,197, 3952, 913, 210, 46, 13, 4, 2, 1. From the similarity of these distributions with one another and with the Paris frequencies of cancer deaths obtained by Besson (1923), they concluded that there was no statistical evidence for the existence of cancer houses in Lyons, a conclusion later criticised by Chaton (1933).

Some of the alleged evidence for "cancer houses" is found on examination to be merely grotesque owing to the definition, or lack of definition, of what constitutes a house. Thus Léon Imbert (1933), investigating the houses wherein 1023 cases of cancer occurred in Marseilles during 1929–32, found 8 houses having 2 cases and 2 houses having 3 cases each. But one of the latter was *la caserne des douanes* with 4000 population, whilst the other sheltered 150 people, and not one of the houses recording more than one case in the four years had less than 15 persons living in it.

It is evident of course that even if it were conclusively proved that multiple cancer cases in a house do occur more frequently than would be expected on the basis of a random distribution in houses *equally* populated, this might arise from several causes and need not indicate either infection, the influence of a common environment, or a tendency to run in families. The principal difficulty is that cancer is highly selective as regards age, and the assumption that the populations of houses are equally liable to cancer is far from being true. A family consisting of a young married couple and several children is at no appreciable risk to cancer within a short period of years, whereas a household consisting mainly of older people may contain several individuals for whom the risk is large. Moreover, since cancer occurs most frequently in advanced age, when one person has died of cancer the remaining persons in the house must tend as a rule to have a mean age less than before, and the risk of a second case occurring soon after is therefore affected by the first case. On the other hand, the high correlation between the ages of husband and wife must tend to increase the probability of two cases occurring in a house. Greenwood and Yule (1920) examined the effect of some departures of this kind from the random distribution assumed by Troup, Maynard and Pearson, in order to apply them to the study of the frequency of multiple accidents.

The difficulties of making allowance for the age factor alone are almost insuperable in the cancer problem, and for this reason an attempt has been made in the present paper to eliminate the effects of this and other factors, without evaluating them, by the device of (1) comparing the frequencies of pairs of cancer deaths in (*a*) the same house and (*b*) in pairs of neighbouring houses at different intervals apart in the same street, and (2) repeating the process for pairs of persons aged 55–75 resident in a sample of the same streets at the census of 1931. The data employed are the records of every cancer death occurring in the city of Bristol during the six years 1922–7 and in the city of Worcester during the ten years 1921–30, comprising a total of some 3500 deaths. For the Bristol records I am greatly indebted to Dr R. H. Parry, Medical Officer of Health, for lending me a card index of the deaths from cancer, each street having a separate card on which was entered the year, number of house and classification according to organ affected, of each cancer death. For the Worcester data I am indebted to the Medical Officer of Health, Dr A. J. B. Griffin, and his staff, for entering each separate cancer death upon a card, giving in addition to the above information the age, sex, occupation and other information on the death certificate, and to Miss M. N. Karn, M.A., of the Galton Laboratory, University College, London, for analysing and tabulating these deaths.

II. Frequency of multiple cancer deaths in the same house in relation to that expected on a random basis

In Bristol 2895 cancer deaths occurred in the six years 1922–7, and at the 1921 census there were 72,472 occupied dwellings. On a random basis of distribution, using the formula of Karl Pearson and denoting the number of houses by m and the number of deaths by n, we should expect the frequencies $f_0, f_1, f_2, f_3, \ldots$ of houses having 0, 1, 2, 3, ... deaths to be given by the successive terms of the expansion $m\left\{\frac{1}{m}+\frac{m-1}{m}\right\}^n$, and hence it follows that

$$\frac{f_1}{f_0}=\frac{n}{m-1},\quad \frac{f_2}{f_1}=\frac{n-1}{2(m-1)},\quad \frac{f_3}{f_1}=\frac{(n-1)(n-2)}{6(m-1)^2},\quad \frac{f_4}{f_1}=\frac{(n-1)(n-2)(n-3)}{24(m-1)^3}.$$

Thus $f_1=0{\cdot}039947f_0$; $f_2=0{\cdot}019967f_1$; $f_3=0{\cdot}0002657f_1$; $f_4=0{\cdot}00000265f_1$; and since $f_1+2f_2+3f_3+4f_4+\ldots=2895$, the expected frequencies are $f_1=2781{\cdot}675$, $f_2=55{\cdot}542$, $f_3=0{\cdot}739$, $f_4=0{\cdot}007$. The observed numbers of houses having 1, 2, 3, 4 deaths were respectively 2757, 64, 2, 1, and comparing these with the expected frequencies, combining the last three, since f_3 and f_4 are very small, $\chi^2=2{\cdot}257$, giving $P=0{\cdot}14$ for a single degree of freedom. Such a poor fit might arise through chance, the frequencies of multiple cases not being in excess of expectation to a significant degree.

In Worcester 694 cancer deaths occurred in the decade 1921–30, and the number of occupied dwellings in 1921 was 11,555. Thus $f_1=0{\cdot}060066f_0$; $f_2=0{\cdot}029990f_1$; $f_3=0{\cdot}0001497f_1$; $f_4=0{\cdot}00000119f_1$ and the expected frequencies are $f_1=654{\cdot}450$, $f_2=19{\cdot}627$, $f_3=0{\cdot}098$, $f_4=0{\cdot}001$. The observed numbers were 658, 18, 0, 0, from which $\chi^2=0{\cdot}253$. The Worcester data therefore show no significant deviation either from what might be expected from a purely random distribution of deaths.

III. FREQUENCY OF PAIRS OF CANCER DEATHS IN RELATION TO THE INTERVAL BETWEEN THEM

In Bristol the cancer deaths were distributed over the 6 years as follows:

1922	1923	1924	1925	1926	1927	Total
442	491	481	478	477	526	2895

Assuming an equal distribution over the six years, the probability that a death selected at random would occur in the same year as some other death also selected at random would be 1/6; the probability that the two deaths would occur in consecutive years would be 10/36, for if the first death occurred in 1922 or 1927 the chance of this would be 1/6, and if the first death occurred in any of the other years it would be 2/6; similarly the probability that the deaths would occur in years separated by one would be 8/36, in years separated by two 6/36, in years separated by three 4/36, and in years separated by four 2/36. Hence the frequencies of pairs of deaths, whether in the same house or in other specified pairs of houses, should, if the distribution were random, be in the following ratio:

Same year	Consecutive years	Next but 1	Next but 2	Next but 3	Next but 4
3	5	4	3	2	1

The observed frequencies for pairs of Bristol cancer deaths, and the frequencies expected on this basis of distribution of the totals, were as shown in Table I.

Table I

Cancer deaths, Bristol 1922–7	Same year	Consecutive years	Next year but 1	Next year but 2	Next year but 3	Next year but 4	Total pairs
Pairs of deaths occurring in the same street and in:							
Same house	20	23	12	12	9	—	76
Houses of consecutive numbers	9	26	17	17	3	2	74
Houses of alternate numbers	18	28	21	18	17	3	105
Houses whose numbers differed by 3	18	18	12	9	12	—	69
Houses whose numbers differed by 4	7	25	24	11	7	2	76
Houses whose numbers differed by 5	11	22	12	12	6	8	71
Total frequency of above pairs	83	142	98	79	54	15	471
Expected distribution of total	78·5	130·8	104·7	78·5	52·3	26·2	471
Houses whose numbers differed by more than 5	544	864	702	494	336	130	3070
Expected distribution of total	512	853	682	512	341	170	3070
Pairs of deaths occurring in different streets and in:							
Houses whose numbers differed by 5 or less	65	100	81	—	—	—	246
Expected distribution of total	62	102	82	—	—	—	246
Houses whose numbers differed by more than 5	327	591	435	—	—	—	1353
Expected distribution of total	338	564	451	—	—	—	1353

Houses of consecutive numbers are not usually next-door houses. In English towns there are two systems of numbering houses in streets, namely (1) "Consecutive" numbering, by which the numbers run 1, 2, 3, ..., x, along one side of the street and then $x+1$, $x+2$, $x+3$, ..., h, in the *reverse* direction along the other side, and (2) "Alternate" numbering, by which the odd numbers 1, 3, 5, ... run along one side of the street and the even numbers 2, 4, 6, ... run in the same direction along the other side. The method of alternate numbering is much the more common, and hence houses of alternate numbers (*i.e.* next number but 1) are more likely to be next-door houses than are houses of consecutive numbers. In Table II and the lower part of Table III the houses are paired not by their numbers alone but after taking into account the system of numbering used in the street in question. This is more fully explained in Section IV, with regard to the analysis used in Tables VI and VII.

For all pairs of cancer deaths in houses of the same street whose numbers differed by not more than 5, $\chi^2 = 6{\cdot}491$ and $P = 0{\cdot}26$, giving a reasonable fit with expectation. Dealing only with neighbouring houses on the same side of the street, the observed and expected distributions of the total pairs were as shown in Table II. For deaths in the same house the distributions according

Table II

Cancer deaths, Bristol 1922–7	Same year	Con-secutive years	Next year but 1	Next year but 2	Next year but 3	Next year but 4	Total pairs
Pairs of cancer deaths occurring in:							
Same house	20	23	12	12	9	—	76
Expected distribution of total	12·7	21·1	16·9	12·7	8·4	4·2	76
Next door houses	17	37	15	17	14	4	104
Next door but 1	9	21	26	12	10	2	80
Next door or next but 1	26	58	41	29	24	6	184
Expected distribution of total	30·7	51·1	40·9	30·7	20·4	10·2	184

to time interval give $\chi^2 = 10{\cdot}098$, $P = 0{\cdot}07$, the significance of the difference being in doubt. More of these pairs of cancer deaths occurred in the same year than would be expected, namely 26·32 per cent. with a probable error of 3·41, instead of 16·71 per cent., the excess being of doubtful significance. Whilst the expected pairs of deaths in (a) the same year, (b) next year or next but 1, (c) next year but 2, 3 or 4, would be in the ratio $a : b : c = 10{\cdot}5 : 31{\cdot}5 : 21$, the observed pairs were in the ratio 20 : 35 : 21.

For deaths in neighbouring houses, that is to say next door or next door but 1, $\chi^2 = 4{\cdot}110$, $P = 0{\cdot}40$, and no such tendency is observed. When houses are paired in streets taken at random, as shown in the lower part of Table I, there is also reasonably good agreement between the observed and expected frequencies, this being included merely as a check. For houses of the same street whose numbers differed by more than 5, shown in the middle of Table I, $\chi^2 = 12{\cdot}845$, $P = 0{\cdot}025$, and there is a tendency for pairs of deaths to occur slightly more frequently in years close together than would be expected. To test whether the expected frequencies would be appreciably altered if the exact totals of deaths in each year had been used instead of assuming them equal, the expected frequencies were recalculated on this basis and found to be 513, 851, 681, 512, 343, 170, which agree closely with the more approximate estimates.

In Worcester the cancer deaths were distributed over the ten years as follows:

1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	Total
72	60	54	58	75	54	67	87	97	70	694

Assuming an equal distribution throughout the ten years, the probabilities that a death selected at random would occur in the same year, next year, next but 1, ... to some other death selected at random would be in the ratio 5 : 9 : 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1, and the expected distribution of frequencies of any number of pairs of deaths on a random basis can be calculated. The observed frequencies of pairs of cancer deaths in houses differing in number by 0, 1, 2, ... and occurring in years separated by intervals of 0, 1, 2, ... years are set out in Table III.

Table III

Cancer deaths in Worcester, 1921–30	Occurring in years differing by x, where x is										Total pairs
	0	1	2	3	4	5	6	7	8	9	
irs of cancer deaths in the same street and in:											
me house	2	5	4	2	3	—	1	1	—	—	18
:ouses of consecutive numbers	5	5	2	3	8	3	5	3	1	4	39
:ouses of alternate numbers	5	4	4	1	5	4	—	2	—	—	25
:ouses whose numbers differed by 3	2	4	4	1	2	3	1	2	2	—	21
,, ,, ,, 4	3	7	3	11	3	2	2	3	1	—	35
,, ,, ,, 5	4	3	2	2	—	1	—	—	—	—	12
,, ,, ,, 6	4	3	5	3	2	4	1	1	3	—	26
,, ,, ,, 7	—	2	3	1	—	—	2	—	—	—	8
,, ,, ,, 8	1	4	3	4	1	1	3	—	—	—	17
Total	26	37	30	28	24	18	15	12	7	4	201
Expected distribution of total	20·1	36·2	32·2	28·1	24·1	20·1	16·1	12·1	8·0	4·0	201
rs of cancer deaths in neighbouring houses on same side of street:											
ext door	5	8	5	2	2	3	4	1	—	1	31
ext door but 1	4	7	3	6	5	3	2	4	1	—	35
ext door but 2	4	3	5	3	3	5	1	1	4	—	29
ext door but 3	1	2	4	6	2	1	3	—	1	—	20
Total	14	20	17	17	12	12	10	6	6	1	115
Expected distribution of total	11·5	20·7	18·4	16·1	13·8	11·5	9·2	6·9	4·6	2·3	115

Combining all the 201 pairs, the agreement between the observed and expected distributions according to time interval is good except for pairs in the same year, which show an apparent excess not in itself significant. For the whole series $\chi^2 = 2{\cdot}321$, $P = 0{\cdot}98$; for neighbouring houses on the same side of the street 115 pairs give $\chi^2 = 2{\cdot}328$, $P = 0{\cdot}98$.

Dealing with the Bristol and Worcester data, there is therefore no conclusive evidence that when cancer deaths occur in pairs in the same or neighbouring houses there is any tendency for them to be separated by a short rather than a long interval.

IV. FREQUENCY OF PAIRS OF CANCER DEATHS IN THE SAME AND IN NEIGHBOURING HOUSES

The comparison of the frequency of pairs of deaths occurring in the same house with the frequency of pairs consisting of one death occurring in one house and the other in a neighbouring house is complicated by the fact

https://doi.org/10.1017/S0022172400018969 Published online by Cambridge University Press

that streets are not of infinite length. Owing to the small number of houses in many streets, end houses form an appreciable fraction of the total, and the expected frequency of pairs is modified for end houses.

Let there be n persons at risk in each house, and let the numbers of streets having only 1, 2, 3, ..., m houses be $s_1, s_2, s_3, \ldots, s_m$. In a street with h houses there are hn persons at risk, and the number of possible pairs in the same house in the street $= \frac{1}{2}hn\,(n-1)$; the number of possible pairs in consecutively numbered houses $= (h-1)\,n^2$, in houses differing in number by two $= (h-2)\,n^2$, and in houses differing in number by $d = (h-d)\,n^2$.

On the basis of a random distribution of cancer deaths in all houses of the town, therefore:

No. of possible pairs in same house of same street

$$= \tfrac{1}{2}n\,(n-1)\,[s_1 + 2s_2 + 3s_3 + \ldots + ms_m]; \qquad \ldots\ldots(1)$$

No. of possible pairs in consecutive houses by number in same street

$$= n^2\,[s_2 + 2s_3 + 3s_4 + \ldots + (m-1)\,s_m]; \qquad \ldots\ldots(2)$$

No. of possible pairs in houses differing in number by 2 in same street

$$= n^2\,[s_3 + 2s_4 + 3s_5 + \ldots + (m-2)\,s_m]. \qquad \ldots\ldots(3)$$

and so on.

Dividing up all the cancer deaths according to the number of the house in which they occurred regardless of streets, let the totals occurring in No. 1, No. 2, No. 3, ... be $N_1, N_2, N_3, \ldots$. Then if these totals be large and the distribution random,

$$k\,(N_h - N_{h+1}) = s_h = \text{no. of streets having just } h \text{ houses}^1,$$

where k is a constant, approximately equal to the ratio of total houses to total cancer deaths. Hence, since $N_{m+1} = 0$,

$$[s_1 + s_2 + \ldots + s_m] = kN_1$$
$$[s_2 + s_3 + \ldots + s_m] = kN_2, \text{ and so on.}$$

By summation it follows that

$$s_1 + 2s_2 + 3s_3 + \ldots + ms_m = k\,(N_1 + N_2 + \ldots + N_m) = kS_1^m\,(N), \qquad \ldots\ldots(4)$$

$$s_2 + 2s_3 + 3s_4 + \ldots + (m-1)\,s_m = k\,(N_2 + N_3 + \ldots + N_m) = kS_2^m\,(N), \qquad \ldots\ldots(5)$$

$$s_3 + 2s_4 + 3s_3 + \ldots + (m-2)\,s_m = k\,(N_3 + N_4 + \ldots + N_m) = kS_3^m\,(N), \qquad \ldots\ldots(6)$$

and so on, the values $S_x^m\,(N)$ being easily found by summing the frequency distributions according to successive values of N from the end N_m to the points $N_1, N_2, \ldots, N_x$.

Now let the frequency of pairs of deaths in houses of *consecutive* numbers in the same street expected from a random distribution be $M = n^2kS_2^m\,(N)$

[1] kN_h = no. of houses numbered h, and kN_{h+1} = no. of houses with the number $h+1$, so the difference = no. of streets having just h houses. For streets with very few houses this would be complicated by alternate numbering, but the effect on this calculation is of no importance.

according to equations (2) and (5) above, then[1] the expected frequency in houses of the same street whose numbers differ by two $= M.S_3^m(N)/S_2^m(N)$, the expected frequency in houses of the same street whose numbers differ by $p = M.S_{p+1}^m(N)/S_2^m(N)$, and the expected frequency in the same house is given by $\frac{M}{2}.\frac{n-1}{n} S_1^m(N)/S_2^m(N)$.

The values of N, $S(N)$, and the ratios $S_x^m(N)/S_2^m(N)$ are shown for Bristol and Worcester in Table IV.

The actual frequencies of pairs of cancer deaths in consecutively numbered houses, numbers differing by 2, etc., of the same street are given in Table V. The expected frequencies are calculated by equating

$$M\left\{1+\frac{S_3^m(N)}{S_2^m(N)}+\frac{S_4^m(N)}{S_2^m(N)}+\ldots\right\}$$

to the total pairs from consecutive numbers onwards, thus finding M and substituting in the expressions above. Thus in the first series of Bristol pairs (deaths in years separated by not more than 1), omitting those in the same house, there were 390 pairs in houses whose numbers differed by 1 to 8 and

$$M[1+0{\cdot}971+0{\cdot}941+0{\cdot}907+0{\cdot}882+0{\cdot}852+0{\cdot}830+0{\cdot}804]=7{\cdot}187M=390,$$

whence $M = 54{\cdot}265$ and the expected frequencies are obtained by multiplying by the successive factors in the bracket[2].

Table IV

	Bristol, 1922–7			Worcester, 1921–30		
No. of house x or $x-$	No. having a cancer death N_x	Sum of preceding column from bottom $S_x^m(N)$	Ratio $\frac{S_x^m(N)}{S_2^m(N)}$	No. having a cancer death N_x	Sum of preceding column from bottom $S_x^m(N)$	Ratio $\frac{S_x^m(N)}{S_2^m(N)}$
1	82	2674	1·032	28	640	1·046
2	74	2592	1·000	21	612	1·000
3	81	2518	0·971	14	591	0·966
4	87	2437	0·941	21	577	0·943
5	63	2350	0·907	22	556	0·908
6	79	2287	0·882	15	534	0·873
7	56	2208	0·852	28	519	0·848
8	69	2152	0·830	14	491	0·802
9	46	2083	0·804	14	477	0·779
10	60	2037	0·786	16	463	0·757
11–20	556	1977	0·763	121	447	0·730
21–30	339	1421	—	66	326	0·533
31–40	238	1082	—	61	260	—
41–50	186	844	—	46	199	—
51–100	351	658	—	101	153	—
101–200	204	307	—	37	52	—
201–300	49	103	—	9	15	—
301–400	24	54	—	6	6	—
401–500	13	30	—	—	—	—
501–600	13	17	—	—	—	—
601–700	2	4	—	—	—	—
701–800	1	2	—	—	—	—
801–900	1	1	—	—	—	—
	2674	—	—	640	—	—

[1] From equations (2), (3), (5), (6).

[2] Data were not extracted for houses whose numbers differed by 9 or 11, see note to Table V.

https://doi.org/10.1017/S0022172400018969 Published online by Cambridge University Press

It is noticeable that for deaths in years separated by not more than 1 the actual exceed the expected frequencies for pairs of houses whose numbers differ by 2, 4, 6 and 10, and that the reverse is true for those differing by odd numbers, this being clearly seen when the totals for the even and odd intervals up to 8 are compared. Applying the goodness of fit test to these even and odd totals, $\chi^2 = 5{\cdot}57$ with one degree of freedom, giving $P = 0{\cdot}02$, which seems to indicate a significant difference. If the test is applied to the eight frequencies making up the total, however, $\chi^2 = 10{\cdot}935$ and $P = 0{\cdot}14$, so we cannot be sure that these differences are not fortuitous. Small alternating differences might also arise from the fact that in a certain proportion of streets having alternate numbering the whole or some part of one side of the street is not occupied by dwelling houses, and therefore the possible pairs of houses differing by an even number will be in excess, and those differing by an odd number in corresponding defect; in other words the total expected frequencies may not form a smooth descending series in reality, but show rather higher values for the even than the odd intervals.

Table V

Bristol

	Frequencies of pairs of cancer deaths in houses of same street													
		Numbering of the houses differing by												
												1 to 8		
	Same house	1	2	3	4	5	6	7	8	10	12	Odd	Even	Total
Same year	20	9	18	18	7	11	12	9	9	9	7	47	46	93
Next year or next but 1	35	43	49	30	49	34	39	21	32	39	29	128	169	297
Total	55	52	67	48	56	45	51	30	41	48	36	175	215	390
Expected*	22·4	54·3	52·7	51·1	49·2	47·9	46·2	45·0	43·6	42	42	198·3	191·7	390
Next year but 2, 3, or 4	21	22	38	21	20	26	26	15	22	20	26	84	106	190
Expected*	10·9	26·4	25·7	24·9	24·0	23·3	22·5	21·9	21·3	21	20	96·5	93·5	190
All intervals to 3 years	67	69	85	57	67	57	64	35	51	55	50	218	267	485
Expected*	27·9	67·5	65·5	63·5	61·2	59·5	57·5	56·0	54·3	53	52	246·5	238·5	485
All intervals to 5 years	76	74	105	69	76	71	77	45	63	68	62	259	321	580
Expected*	33·3	80·7	78·4	76·0	73·2	71·2	68·7	66·9	64·9	63	62	294·8	285·2	580

* Expected distribution of the total pairs in houses whose numbering differed by 1 to 8. The frequencies for intervals of 9 and 11 were not ascertained, so the expected distribution is based on the continuous series as far as it goes.

For the second series, deaths separated by 2–4 years, $7{\cdot}187M = 190$, and $M = 26{\cdot}437$, giving the expected frequencies shown. Houses differing by even numbers again show an excess in the total observed frequency over that expected, but this is entirely due to houses with numbers differing by 2. Comparing the odd with the even totals $\chi^2 = 3{\cdot}29$ for one degree of freedom, giving $P = 0{\cdot}07$, or applying the test to the eight separate groups the value of χ^2 is $10{\cdot}952$, giving $P = 0{\cdot}14$, so the distributions are not significantly different though the bias is in the same direction as for the first series.

Combining the two series it would seem that the excess over expectation in the frequency of pairs of cancer deaths in houses (of the same street) differing in number by 2 is statistically significant. There were 105 such pairs, the expected number being only 78. As shown below seven-eighths of the houses in a large sample were numbered alternately along the same side of the street, so that houses differing in number by 2 were almost always *next-door* houses. On the other hand, houses differing in number by 7 show an almost equally remarkable defect in frequency, and this is difficult of explanation.

Leaving this for a moment to examine the frequency of pairs of deaths *in the same house*, from census data in 1921 the mean number of persons per dwelling in Bristol was 5, and in 1931 it was 4·5. Hence, making no allowance for the younger not being at any appreciable risk to cancer, n, the mean number at risk in a house, must have a value not greater than 5. Thus the maximal expected frequency would be $\frac{M}{2}.\frac{n-1}{n}.\rho=\frac{2M}{5}.\rho$, where $\rho=\frac{S_1^m(N)}{S_2^m(N)}$ is the correction factor for limited lengths of the streets, shown in Table IV to be 1·032 for Bristol and 1·046 for Worcester. The greatest frequency to be expected from a random distribution is therefore 0·413M for Bristol. In the first series $M=54{\cdot}265$, and the expected number is 22·4, the observed number being 55. In the second series $M=26{\cdot}437$, and the expected number is 10·9, the observed number being 21. Combining all intervals between deaths, 76 *pairs of cancer deaths in the same house were observed, whereas the maximal expected frequency on this random basis of distribution was* 33.

Returning to the data for neighbouring houses, in view of the curious result from Table V, it was ascertained for each street in which a cancer death occurred whether the numbering of the houses was consecutive or alternate, that is to say whether numbering proceeded up one side of the street and down the other or whether odd numbers were on one side and even numbers on the other. It was thus possible to find the observed numbers of pairs of deaths in next-door houses, next door but 1, next door but 2, next door but 3 and next door but 4, which are given for Bristol in Table VI for those pairs of deaths which occurred in years differing by not more than three.

In order to calculate the expected distribution of the total frequency in the new grouping, next door, next door but 1, etc., it is only necessary to weight the expected frequencies, calculated as in Table V, by the proportions of houses in streets having alternate and consecutive numbering respectively. In a random sample of streets there were 14,794 houses numbered alternately and 2602 numbered consecutively, a ratio of 7 : 1. Thus the expected frequency for next-door pairs would be, from Table IV,

$$\tfrac{1}{8}\{7M.S_3^m(N)/S_2^m(N)+M.S_2^m(N)/S_2^m(N)\}=M\{(7\times 0{\cdot}971)+1\}=0{\cdot}975M,$$

since one-eighth of them would, according to expectation, be numbered with consecutive numbers and seven-eighths with numbers differing by 2. The fraction 0·975 is of course merely the appropriate correction for the fact that streets are of limited length. For houses next door but 1, seven-eighths would have numbers differing by 4, and one-eighth would have numbers differing by 2, thus giving $\frac{M}{8}\{(7\times 0{\cdot}907)+0{\cdot}971\}=0{\cdot}915M$. The resulting correction factors are given at the foot of Table VI, and the expected frequencies are found by multiplying them by the value of M in Table V for all intervals up to three years inclusive, viz. $7{\cdot}187M=485$, or $M=67{\cdot}483$.

On this basis there was an excess of observed frequency of pairs of cancer deaths in houses next door to one another amounting to 21, in houses next door but 1 amounting to 6, but no appreciable excess for houses further apart. Thus the ratios of observed to expected frequency of pairs of cancer deaths in the same house, next-door houses, next door but 1, etc., are about 2·4, 1·3, 1·1 and then become unity or less, but whether or not the values 1·3 and 1·1 are

Table VI

Bristol

	Frequency of pairs of cancer deaths in							
		Two houses on same side of street and						
Numbering of houses	Same house	Next door	Next door but 1	Next door but 2	Next door but 3	Next door but 4	Total	Other pairs of houses in Table V, mostly on opposite sides of street
Same house number	67	—	—	—	—	—	—	—
Next number	—	23	—	—	—	—	23	46
Next number but 1	—	64	21	—	—	—	85	—
,, ,, 2	—	—	—	14	—	—	14	43
,, ,, 3	—	—	47	—	20	—	67	—
,, ,, 4	—	—	—	—	—	15	15	42
,, ,, 5	—	—	—	43	—	—	43	21
,, ,, 6	—	—	—	—	—	—	—	35
,, ,, 7	—	—	—	—	29	—	29	22
,, ,, 9	—	—	—	—	—	40	40	15
,, ,, 11	—	—	—	—	—	—	—	50
In same year	20	17	9	14	9	10	59	—
Consecutive years	23	37	21	18	14	29	119	—
Next years but 1	12	15	26	16	14	8	79	—
Next years but 2	12	18	12	9	12	8	59	—
Total	67	87	68	57	49	55	316	274
Correction factors	—	0·975	0·915	0·863	0·846	0·798	—	—
Expected distribution of totals	27·9	65·8	61·7	58·2	57·1	53·9	296·7	293·3

significantly above unity remains in doubt.

The *Worcester* data, when analysed in the same manner as in Table VI, give the frequencies of pairs of cancer deaths in neighbouring houses shown in the lower part of Table III. The totals, correction factors and expected frequencies of pairs are given in Table VII, the last being based on the total of all pairs (excluding the same house) for which the difference in house numbering was less than 9.

Thus from Table III there were 183 pairs, and the sum of the correction factors being, from Table IV,

$$[1+0{\cdot}966+0{\cdot}943+0{\cdot}908+0{\cdot}873+0{\cdot}848+0{\cdot}802+0{\cdot}779]=7{\cdot}119,$$

it follows that $M = 183/7{\cdot}119 = 25{\cdot}706$. Hence the expected frequency in next-door houses in Table VII is $0{\cdot}972 \times 25{\cdot}706 = 25$, and so on. The result shows an excess over expectation in the observed frequency of pairs in next-door houses and next door but 1, as for Bristol, and also for next but 2, and a corresponding defect for houses not so close together. For the five groups comprising the total of 183 pairs, $\chi^2 = 15{\cdot}128$, $P = 0{\cdot}004$, indicating a significant excess in neighbouring houses. In the same house the frequency of pairs of deaths is also excessive, as in Bristol, the observed number being 18 and the expected number 11.

Proceeding, therefore, on the assumption that the *population at risk to die of cancer is uniformly distributed* throughout the houses of the average street, the Bristol and Worcester data alike indicate a significant tendency for cancer deaths to be repeated in the same house more frequently than would be expected by comparison with coincidences in pairs of houses some distance apart.

Table VII

Worcester

	Frequency of pairs of cancer deaths in							
		Two houses on same side of street and						
Interval	Same house	Next door	Next door but 1	Next door but 2	Next door but 3	Total of last 4 columns	Other pairs of houses at a distance	Total excluding same house
0–2 years	11	18	14	12	7	51	—	—
3–5 ,,	5	7	14	11	9	41	—	—
6–9 ,,	2	6	7	6	4	23	—	—
0–9 ,,	18	31	35	29	20	115	68	183
Correction factors	—	0·972	0·919	0·865	0·802	—	—	—
Expected distribution of pairs	10·8	25·0	23·6	22·2	20·6	91·4	91·6	183

Neighbouring houses on the same side of the street also seem to show an excessive frequency of occurrence of pairs of cancer deaths over expectation on this basis. The validity of the assumption on which the expected frequencies are based will be examined in the next section.

V. DISTRIBUTION OF PERSONS OF "CANCER AGE" IN HOUSES

Most studies of the distribution of cancer deaths or cases in houses have tacitly assumed a random distribution of the persons at risk. It may be, however, that old people tend to live in groups of houses close together and younger people in other groups. This is certainly true of districts in a town, or of streets as a whole, but there seemed no *a priori* reason to suppose that such selection would be important with regard to groups of houses within a given street. Larger houses tend to be grouped together in a road, and this may also affect the expectation of coincidences in adjacent houses to a significant degree.

In order to test to what extent persons of the ages specially liable to provide cancer deaths tend to group themselves in the same or adjacent houses, a

https://doi.org/10.1017/S0022172400018969 Published online by Cambridge University Press

sample was taken by alphabetical selection from those streets in which more than one cancer death occurred, and the 1931 census records of the selected streets in Bristol were examined and the number of persons aged 55–75 living in each house ascertained. This age group was used as perhaps the simplest relative index of the cancer risks in the houses concerned. These people at risk were then paired with one another in precisely the same way as the cancer deaths in Table VI, giving the possible number of pairs which could be made of persons aged 55–75 living in the same house, in next-door houses, and so on. The actual frequencies thus obtained, which are of course a function of the arbitrary size of the sample, though the ratios of one to another are independent of it, are shown at the top of Table VIII.

Reducing these frequencies to the same totals as in Table VI, the comparison between the observed numbers of pairs of cancer deaths and the corresponding relative frequencies of pairs of people "at risk" is also shown

Table VIII

Bristol

	Frequency of pairs of cancer deaths and of persons living aged 55–75						
	Same house (1)	Next door (2)	Next door but 1 (3)	Next door but 2 (4)	Next door but 3 (5)	Next door but 4 (6)	Total (2) to (6)
Actual frequency of possible pairs of persons aged 55–75	1179	1463	1256	1181	775	915	5590
Frequencies reduced to comparable totals:							
Cancer deaths, observed	67	87	68	57	49	55	316
Cancer deaths, expected	28	66	62	58	57	54	297
Persons living aged 55–75	67	83	71	67	44	52	316

in Table VIII, with the corresponding distribution of expected deaths on the assumption of random distribution and equal populations at risk (from Table VI). It is now seen that the anomalous features of the Bristol distribution of cancer deaths when compared with expectation, namely the high frequencies for the same house and next-door houses and the low frequency for next door but 3, are repeated in the distribution of living persons aged 55–75.

The distributions of pairs of deaths and of persons living at these ages, arrived at from independent data, show in fact a remarkable agreement; thus dividing into only three groups:

	Same house	Next door or next but one	Next but 2, 3 or 4
Cancer deaths	67	155	161
Persons living, 55–75	67	154	163
Expected on random basis	28	128	169

It only remained to ascertain the reasons for the departure from a random distribution of the population of cancer ages in houses, and a detailed examination of the records revealed the following. In one street providing a large number of cancer deaths, four of the deaths were registered in the six years

https://doi.org/10.1017/S0022172400018969 Published online by Cambridge University Press

1922–7 as occurring in a single house, two in 1923 and two in 1927. The sixty-seven houses in this street included nine common lodging houses registered for 204 lodgers, and the census showed that in six houses the persons aged 55–75 numbered 15, 15, 14, 9, 9, 8 respectively. It is clear that, with a moving population averaging 15 persons of this age in a house the occurrence of four cancer deaths in six years is not remarkable. Another street providing an unusual number of cancer deaths was found to have only sixteen houses occupied in 1931, but three of these had 13, 11 and 9 persons aged 55–75 respectively. The presence of several common lodging houses in a street produces large departures from any theoretical frequencies of expectation of the coincidence of cancer deaths in the same or neighbouring houses, and the segregation of elderly people together in these and other circumstances is evidently sufficient to discount the usefulness of such comparisons with mathematical expectations on a random basis. It is this basis of randomness in the distribution of the population at risk which has been the underlying assumption in almost all the work relating to "cancer houses."

The close agreement between the frequencies of pairs of cancer deaths and of pairs of persons living at the ages specially liable to cancer, when distributed according to the proximity of the houses, shows that the occurrence of a second cancer death in the same or a neighbouring house took place on the whole neither more nor less often than was to be anticipated from the number of persons exposed to risk in the houses concerned. This absence of evidence of any general tendency for multiple cancer deaths to occur in a house, when checked by a control series, confirms the conclusion reached in a recent paper (Stocks and Karn, 1933, p. 246) from a study of histories furnished by patients and controls. In that study, according to 222 women cancer patients 11 previous cases of cancer were known to have occurred in the houses in which they live, the corresponding number for 207 women controls of the same ages being also 11.

VI. DISTRIBUTION OF CANCER DEATHS ACCORDING TO ORGANS AFFECTED

In the Bristol records, collected in the Public Health Department, every cancer death during 1922–7 was classified into one of the following groups according to the site of the primary growth:

A. Buccal cavity;
B. Pharynx, oesophagus, stomach, liver and annexa;
C. Peritoneum, intestines, rectum and colon;
D. Female genital organs;
E. Breast;
F. Skin;
G. Other or unspecified organs.

In the Worcester records the causes of death were entered on the cards as stated on the death certificates, and the classification into groups was carried

https://doi.org/10.1017/S0022172400018969 Published online by Cambridge University Press

out in the Galton Laboratory according to the above scheme. The distributions of the total deaths in each year were as shown in Table IX; under each actual frequency is entered the number expected if the distribution was proportionately the same in every year. Applying the χ^2 test for goodness of fit to the distributions, the following values of P, the probability of a worse fit to the expected numbers being obtained by chance than that observed, are found for Bristol and Worcester:

	A	B	C	D	E	F
Bristol	0·094	0·384	0·470	0·684	0·794	0·583
Worcester	0·914	0·204	0·812	0·840	0·856	0·997

There is, therefore, no evidence that cancer tended to affect a particular part of the body in one year more than another.

Table IX

Cancer deaths, registered and expected, according to site of primary growth

Site of primary growth	Bristol						
	1922	1923	1924	1925	1926	1927	Total
A	14	35	30	31	21	30	161
	24·58	27·30	26·75	26·58	26·53	29·25	
B	135	138	135	123	158	158	847
	129·32	143·66	140·74	139·85	139·56	153·89	
C	77	114	110	103	105	115	624
	95·27	105·83	103·67	103·03	102·81	113·37	
D	56	62	67	61	48	65	359
	54·81	60·89	59·65	59·28	59·15	65·23	
E	59	62	57	63	62	55	358
	54·66	60·84	59·60	59·23	59·61	65·05	
F	3	5	7	3	2	5	25
	3·82	4·24	4·16	4·13	4·12	4·54	
G	98	75	75	94	81	98	521
	79·55	88·37	86·57	86·03	85·85	94·66	
Total	442	491	481	478	477	526	2895

Cancer deaths, registered and expected, according to site of primary growth

Site of primary growth	Worcester										
	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	Total
A	3	3	3	2	4	1	6	4	4	3	33
	3·42	2·85	2·57	2·76	3·57	2·57	3·19	4·14	4·61	3·33	
B	21	16	10	10	17	19	26	22	30	14	185
	19·19	15·99	14·40	15·46	19·99	14·40	17·86	23·19	25·86	18·66	
C	16	16	17	21	21	16	9	27	23	25	191
	19·82	16·51	14·86	15·96	20·64	14·86	18·44	23·94	26·70	19·26	
D	7	7	10	9	8	9	6	13	10	8	87
	9·03	7·52	6·77	7·27	9·40	6·77	8·40	10·91	12·16	8·78	
E	7	3	6	8	10	3	8	9	10	9	73
	7·57	6·31	5·68	6·10	7·89	5·68	7·05	9·15	10·20	7·36	
F	3	—	—	1	1	1	1	1	3	—	11
	1·14	0·95	0·86	0·92	1·19	0·86	1·06	1·38	1·54	1·11	
G	15	15	8	7	14	5	11	11	17	11	114
	11·83	9·86	8·87	9·53	12·32	8·87	11·00	14·29	15·93	11·50	
Total	72	60	54	58	75	54	67	87	97	70	694

In Table X the contingency according to site of the cancer is shown for 290 pairs of deaths occurring in the same year and in the same street in Bristol, taking only those streets where not more than two cancer deaths occurred in the given year. The expected frequencies are the numbers which would result according to a uniform distribution of the totals in Table IX, and are shown below the observed frequencies. The total number of pairs in which the same site was affected in both, that is to say the diagonal total, was 59, which agreed closely with the expected total 57·497. For the whole table $\chi^2 = 21{\cdot}445$, $P = 0{\cdot}718$, so there was no evidence of any tendency for cancer to affect the same part of the body in deaths occurring in the same street in the same year.

In 337 pairs of deaths in the same street in successive years, taking only years having a single death each, the diagonal total of coincidences between sites affected was 60, the expected total being 66·817. In 235 pairs in the same street in alternate years the total coincidences numbered 56, the expected total being 46·594. Thus, combining the 862 pairs of deaths in the same street with intervals of not more than two years, the expected coincidences in regard to site of the cancer were 181, and the observed number 175.

Table X. *Contingency table for cancer deaths by site, in same year and same street in Bristol*

	A	B	C	D	E	F	G	Total
A	2 0·896	8 9·436	11 7·954	4 4·002	5 3·990	— 0·278	6 5·806	36
B	—	31 24·824	40 36·574	21 21·042	14 20·984	1 1·468	31 30·538	138
C	—	—	13 13·473	21 15·504	14 15·456	1 1·078	14 22·498	63
D	—	—	—	3 4·457	12 8·898	2 0·620	14 12·946	31
E	—	—	—	—	5 4·434	— 0·620	12 12·910	17
F	—	—	—	—	—	— 0·020	— 0·900	—
G	—	—	—	—	—	—	5 9·393	5
Total	2	39	64	49	50	4	82	290

In Table XI the contingency for all pairs of cancer deaths in Worcester which occurred in the same street at intervals not exceeding two years is shown, in this instance not confining the selection to streets in which only two deaths occurred at such intervals, but including every possible pair in streets where more than two occurred. The total coincidences to be expected between deaths from cancer of the same site would be, from the totals in Table IX,

$$481 \times \frac{\{33^2 + 185^2 + 191^2 + 87^2 + 73^2 + 11^2 + 114^2\}}{(694)^2} = 97{\cdot}68,$$

the observed number being 90. No evidence that cancer occurring in a given street tended to affect any particular organ is therefore found from the Bristol or Worcester data.

Comparing the site distributions in the two cities, the totals in the last column of each portion of Table IX give $\chi^2 = 16{\cdot}051$ for the seven groups, and

Table XI. *Contingency table for cancer deaths by site, in same street and within two years, in Worcester*

A	B	C	D	E	E	F	G	Total
A	1	7	10	10	4	1	7	40
B	—	37	85	35	33	8	46	244
C	—	—	28	31	18	2	35	114
D	—	—	—	8	13	3	29	53
E	—	—	—	—	2	2	10	14
F	—	—	—	—	—	—	2	2
G	—	—	—	—	—	—	14	14
Total	1	44	123	84	70	16	143	481

hence $P=0{\cdot}013$, so there is a significant difference between the relative frequencies with which different parts of the body were affected in the two cities. The mean annual cancer death rates in 1922–7 were 1·307 in Bristol and 1·337 in Worcester, per 1000 population, and the percentage distributions of the cancer deaths were as follows:

	A	B	C	D	E	F	G	Total
Bristol 1922–7	5·56	29·26	21·56	12·40	12·37	0·86	17·99	100·00
Worcester 1921–30	4·75	26·66	27·52	12·54	10·52	1·58	16·43	100·00

The chief difference is a relative excess of intestinal cancer in Worcester, with a corresponding deficiency in the other groups.

Although no significant variation of the site distribution of deaths from year to year was found either in Bristol or in Worcester, nevertheless when these distributions during 1922–7 are compared with the expected frequencies according to Table IX, there was a tendency for cancer of the digestive tract as a whole (A, B, C) and of unspecified organs (G, chiefly kidney, bladder or bone cancers) to be relatively in excess or defect in the two cities in the same years. Thus if an excess over expectation is denoted by + and a defect by −, then putting the symbol for Bristol first and that for Worcester last, the arrangement in Table XII is found.

Table XII

	1922	1923	1924	1925	1926	1927
A	− +	+ +	+ −	+ +	− −	+ +
B	+ −	− −	− −	− −	+ +	+ +
C	− −	+ +	+ +	− +	+ +	+ −
D	+ −	+ +	+ +	+ −	− +	− −
E	+ −	+ +	− +	+ +	+ −	− +
F	− −	+ −	+ +	− −	− +	+ −
G	+ +	− −	− −	+ +	− −	+ 0

It is noticeable that for groups D, E, F, which denote cancers of the breast, skin and genital organs, there is no sign of any correlation between the deaths in the same year in the two cities, the sign being different 10 times and the same 8 times. For groups A, B, C, cancers of the digestive tract, the signs were alike 13 times out of 18, and for group G, cancer of kidney, bladder, bones, etc., the signs were the same 5 times out of 5. It would be interesting to test this on more extended data for other towns before attempting to draw any conclusions from it.

VII. CONCLUSIONS

1. This investigation of cancer deaths in Bristol and Worcester shows that cancer deaths tend to occur in pairs more frequently in the same or adjacent houses than would be expected if the population at risk was distributed uniformly in the houses and the deaths took place at random in the population. If the assumption of a uniform distribution of the people liable to die of cancer were valid, this result would be suggestive of some kind of "infectious" origin, but an exactly similar study of the distribution of persons aged 55–75 living in houses in Bristol, according to the census, shows that when such persons are paired together in the same manner as the cancer deaths, precisely the same result is reached. To postulate any theory of infection to "explain" the curious distribution of cancer deaths in houses is therefore redundant, since it can be sufficiently explained by the tendency to segregation in the same or adjacent houses of people of those ages at which cancer death is of most frequent occurrence.

2. When two cancer deaths occurred in the same street in the same or adjacent years, there was no tendency for the cancer to be located in the same part of the body rather than in different parts.

3. Significant differences were found between the distributions of cancer deaths according to site in Bristol and Worcester.

REFERENCES

BESSON, L. (1923). Le cancer et les maisons. *Ann. d'Hyg. publ. industr. et sociale*, **1**, 23.

CHATON (1933). Existe-t-il des maisons à cancer? *Bull. Acad. Méd.* **109**, 182.

GREENWOOD, M. (1910). Statistical analyses of data respecting epidemics of plague. *J. Hygiene*, **10**, 416–42.

—— (1931). On the statistical measure of infectiousness. *Ibid.* **31**, 336–51.

GREENWOOD, M., and YULE, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. Roy. Statist. Soc.* **83**, 255–79.

IMBERT, LÉON (1933). A propos des maisons à cancer. *Bull. Acad. Méd.* **110**, 787.

LUMIÈRE, A. and VIGNE, P. (1932). Existe-t-il des maisons à cancer? *Ibid.* **108**, 1288.

—— (1933). Statistiques et maisons à cancer. *Ibid.* **109**, 280.

MCKENDRICK, A. G. (1911–12). A simplified method of calculating frequencies of occurrence from a large number of unequal probabilities. *Biometrika*, **8**, 413–19.

PEARSON, KARL (1911–12). On the appearance of multiple cases of disease in the same house. *Ibid.* **8**, 404–12.

—— (1911–12*a*). On cancer houses, from the data of the late T. Law Webb, M.D. *Ibid.* **8**, 430–4.

—— (1913). Multiple cases of disease in the same house. *Ibid.* **9**, 28–33.

STOCKS, P. and KARN, M. N. (1933). A co-operative study of the habits, home life, dietary and family histories of 450 cancer patients and of an equal number of control patients. *Ann. of Eugenics*, **5**, 237–80.

TROUP, J. McD. and MAYNARD, G. D. (1911–12). Note on the extent to which the distribution of cases of disease in houses is determined by the laws of chance. *Biometrika*, **8**, 396–403.

(*MS. received for publication* 5. XII. 1934.—Ed.)