

Biometrika Trust

On "Cancer Houses," from the Data of the Late Th. Law Webb, M.D.

Author(s): Karl Pearson

Source: *Biometrika*, Vol. 8, No. 3/4 (Jan., 1912), pp. 430-435

Published by: [Biometrika Trust](#)

Stable URL: <http://www.jstor.org/stable/2331593>

Accessed: 17/06/2014 12:28

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Biometrika Trust is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

<http://www.jstor.org>

represent the general Jewish population. We must actually measure the "Jewishness" by analysing it into its component factors and measuring them by accepted and specially devised anthropometric methods—such as the nasal index, the cephalic index, the profile angle, etc., together with any number of new facial measurements which it would be quite easy to devise, and also pigmentation determinations. Where, as in man, we cannot be certain of the gametic constitution of an individual, we can only guess at it by an inquiry as to purity of the ancestry and the somatic characters both in Gentile and Jew. This, especially in regard to non-Jewish looking Jews, has been wholly omitted by Dr Salaman in each individual case. From such knowledge we can, by well-known biometric methods, calculate the probable character of the offspring. The result of such an investigation may be to confirm Mendelian theory—as to that I make no assertion. But we stand now, and until this fact has penetrated into the Mendelian consciousness we shall continue to stand, in the position indicated by Weldon ten years ago—"the fundamental mistake which vitiates all work based on Mendel's method is the neglect of ancestry and the attempt to regard the whole effect upon the offspring produced by a particular parent as due to the existence in the parents of particular structural characters; while the contradictory results obtained...show clearly enough that not only the offspring themselves but their race, that is, their ancestry, must be taken into account before the result of pairing them can be predicted."

IV. On "Cancer Houses," from the Data of the late Th. Law Webb, M.D.

By KARL PEARSON, F.R.S.

THE data dealt with in the present paper contain a record of the house distribution of 377 cases of cancer occurring between 1837 and 1910 in the Madeley registration subdistrict. They include both cases of sarcoma and carcinoma. It has not been possible to separate the two, for in many cases the entry is merely cancer. The data include the exact position of the house, its water supply; the age, occupation, and date of death of subject, and the organ affected by the growth. The data consist, besides the record of cases, of large scale maps of the district upon which every inhabited house is marked in with blue wash and every house in which a case of cancer followed by death has occurred in the given period with red wash. A red cross marks existing cases of cancer in the living. The collection was formed by Th. Law Webb, Esq., M.D.

The deaths from cancer are as follows :

1831—1840: [20 ?]*	1841—1850: 14	1851—1860: 20
1861—1870: 48	1871—1880: 41	1881—1890: 77
1891—1900: 81	1901—1910: 88.	

The number of cancer deaths has thus increased very considerably, but the population has increased† and its average age has no doubt very considerably increased, so that it is not possible to say whether this increase of cancer deaths marks an actual increase of cancer.

* Based on four years of the decade only.

† Not very markedly, and lately it has fallen again.

There appear to be very few large houses in the district, most being by their descriptions of the cottage type. Certainly those in which multiple cases of cancer have occurred are often quite small. It is not very easy to determine from the maps the exact number of inhabited houses, but it appears to be about 2865*. Twelve cases of cancer appeared to be living in the district at the date of the return; these are not included in our number of cases. Taking 377 death cases in all and supposing these to be distributed at random among the 2865 houses we should anticipate according to the formulæ of my article (*Biometrika*, Vol. VIII. p. 410).

330·6 houses with one case,	in round numbers	331
21·7 houses with two cases,	„ „	22
·95 houses with three cases,	„ „	1
·03 houses with four cases,	„ „	0

A total is thus obtained of 354 houses with 377 cases and there will be 2511 houses with no cases.

The actually recorded numbers appear to be as follows: 2523 houses with no cases, 315 houses with one case; 20 houses with two cases; six houses with three cases; and one house with four cases. The identification of the houses is not always quite clear, but I think I have erred, if at all, on the side of reducing the multiple houses, e.g. I have supposed nine cases in which the house was not known to have occurred in non-multiple houses. On the other hand, when the sufferer lived in a house up to death, going perhaps to a general hospital or the work-house just before death, I have reckoned the cancer as developed in that house.

If we now substitute in the general formula

$$\chi^2 = S \left\{ s^2 \frac{(p_s - \bar{p}_s)^2}{\bar{p}_s} \right\},$$

we find

$$\chi^2 = \frac{(16)^2}{331} + \frac{4 \times (2)^2}{22} + \frac{9 \times (5)^2}{1} + \frac{16 \times (.97)^2}{.03}$$

$$= \text{about } 727.$$

The probability P is therefore infinitesimal, being widely outside any existing tables for χ^2 and P . We may therefore say that if these numbers be correct the distribution cannot possibly be a random one. Six houses with three cases each and one with four are wholly beyond the bounds of the possible, assuming cancer to be distributed at random among the houses.

I have made a second estimate of this improbability. I have assumed that Dr Law Webb's district actually coincides with the Madeley subregistration district and that a count of houses on the Ordnance Map is likely to be defective †.

The 1841 Census gives the number of houses as 1802 and the population as 8732. The 1851 Census gives 2006 houses and a population of 9848; the 1861 Census provides 2154 houses and a population of 10,733. In 1871 we have 2291 houses and a population of 10,535. In 1881 the numbers are 2359 houses and 10,026 persons. In 1891 we have 2228 houses and 8825 persons. In 1901, 9129 persons and 2196 houses. But it is not quite certain that the boundaries remained absolutely the same. In 1911 the population was 8859 and the *inhabited* houses 2037. I take 2000 as an average number of houses, and if 9000 to 10,000 be the average population then 4·5 to 5·0 are the average number of inhabitants per house.

* Taking about eight cancer deaths a year for the last three decades, this suggests an *average* recognised duration of the disease of about eighteen months.

† The estimate of houses must be very elastic, many houses in the period have come into being, large numbers have ceased to be: the 2865 is a maximum limit.

For

$n=377$

$m=2000$, we find :

$$\left. \begin{array}{l} p_1=312.4 \\ p_2=29.4 \\ p_3=1.84 \\ p_4=0.086 \end{array} \right\} \text{giving a total number of cases} = 377.06,$$

$$\begin{aligned} \chi^2 &= \frac{(2.6)^2}{312.4} + \frac{4(9.4)^2}{20} + \frac{9(4.16)^2}{1.84} + \frac{16(.914)^2}{.086} \\ &= .01 + 17.67 + 84.65 + 155.42 \\ &= 257.75. \end{aligned}$$

P lies therefore outside any table, i.e. the probability that such a distribution could arise from random sampling is only one in many many millions. It seems therefore clear that the reduction of our number of houses by 800 makes no substantial difference in the improbability of our result. The houses with three cases are quite sufficient in themselves—even if we neglected the four-case house—to make the distribution indefinitely improbable.

We have next to consider the sources of this improbability.

The possibilities are :

(a) that these “cancer houses” are larger and contain more inhabitants than the others, thus they would be more likely to have multiple cases.

(b) that constitutions liable to cancer are hereditary and so the “cancer house” marks merely the presence of a “cancer family.”

(c) that certain houses have been inhabited by persons following the same occupation, and that “cancer houses” are those inhabited by persons with a bad occupational mortality for cancer.

(d) that some houses by their environment, or by the presence of some organism render their occupiers more liable to cancer.

I propose to examine the seven instances of three or more cases from the problem of these possibilities.

(A.) No. 1. G—V—. The first case that occurred in this house was that of a station-master, aged 28, cancer of rectum. He died in 1855. The second case was that of a schoolmistress, who died, aged 64, in 1889 from cancer of the uterus. The third case was that of the servant to a nonconformist minister and his wife who were living next door. This servant attended the second case, that of the schoolmistress, in her last illness, and died in 1890, aged 57, of malignant ulceration of the intestines with perforation. The fourth case lived to her 78th year in this house in association with the third, but removed before her death elsewhere. She died at 88 years in 1899 of cancer of the liver.

(B.) No. 2. G—V—. This is the *next* house to A above. The first recorded case in this house was that of the nonconformist minister referred to in A. He died about 1870, aged about 60, of cancer of the stomach. The second case was that of his wife who died in 1880, aged 71, of cancer of the rectum. The third case was the wife of the man who followed the widow of the nonconformist minister. She died of cancer of the breast in 1881, barely a year after the minister's wife.

We have thus in 26 years seven cases associated with two small houses, the servant to the occupants of one living in the other. It is clear that neither the size of these houses, the blood relationship of their occupants, nor any similarity of their occupations will account for these multiple cancer cases.

(C.) S— R—. In this house the servant died in 1862, aged 69, of cancer of the breast. The second case was that of a bricklayer who died in 1885, aged 73, of carcinoma of first part of sigmoid flexure of colon. The third case was that of a woman, whose husband had left her, who died here aged 55, in 1906, of secondary carcinoma.

The adjacent house to this was also a multiple cancer house, being credited with two cases, one of carcinoma of the stomach and one of cancer of the womb.

(D.) C— C—. The first case is that of a man aged 56, who died in 1869 of cancer of the breast. His brother had also died, but not in this house, of cancer of the rectum. On his death he was succeeded in his official position by his clerk, and the latter also retained the same housekeeper. She died in 1885, aged 57, of cancer of the stomach, and the clerk just mentioned died in the same house in 1894, aged 52, of epithelioma of the tongue.

(E.) I— B—, F— M—. The first case in this house was that of a pattern maker who died of cancer of rectum in 1875, aged 53. His father had died of cancer before him. The second case was that of his widow who died in 1885, aged 69, of cancer of the rectum. The third case was that of their daughter in 1888, aged 39, after three years' suffering from cancer of the left breast.

This case may show an hereditary constitutional tendency, but this does not account for *both* husband and wife dying of it. That must either be looked upon as fortuitous, or else it is evidence in favour of (*d*).

(F.) G— B—, T— O—. The first case was that of a woman occupant who died in 1877, aged 52, of carcinoma of the large intestine. The second case was that of a man of superior social class who died in 1895, aged 73, of carcinoma of the stomach. The third case was that of the mother-in-law of the second case, who lived in the house with her daughter; she died in 1902, aged 84, of carcinoma of the liver*.

(G.) M— W—, B— B—. There is a group of five cases here, all in very close proximity. It is not clear how far they are in separate parts of the same original building, but, I think, we may assume three to have been. The first case is one of cancer of the breast in a woman aged 44, in 1862; the second of malignant disease of the stomach in the wife of a nail-maker, she died aged 77 in 1894, and the last one in a woman, who had suffered from epithelioma of hip in 1887, but died of bronchitis. The other two cases were those of a woman who died in 1897, aged 62, of malignant disease of the intestines, and of a blacking-maker, who died, aged 52, of cancer not further defined, in 1865. Undoubtedly this group (G.) is less definite than the others, because the information is less easy to interpret. It did not seem desirable to omit it, for we have clearly a considerable number of cases within a very circumscribed area, but the exact division of the tenements is not clear. If it be omitted, it will not substantially alter the extreme improbability of the general distribution.

From the above specification of the individual "cancer houses," it seems improbable that occupation or heredity constitution had anything to do with the multiple cases; in only one instance do we find parents and offspring; and sex and occupations are most varied. Nor do we believe that the size of house can be concerned in the matter, as it is quite clear that we are dealing with small tenements. The cancer maps of the district show roughly a fairly uniform distribution of the cancer cases, in a district which is largely industrial. But the fact remains that the frequency of multiple case-houses is exceedingly improbable on the basis of a random

* A case of cancer of the breast occurred in the next house in 1876, a woman aged 74. Dr Law Webb writes of the three-case house: "Water, *pump*, often flooded with surface water, when stormy weather; soon dry in summer, drains in close proximity."

distribution of cases. Dr Law Webb's data provide sufficient evidence to justify a demand for a thorough investigation of the subject, such as is not feasible in the case of the individual medical man. They do not finally demonstrate that cancer is more frequent in one house than a second, but they do justify a complete inquiry into the possibility that "cancer-houses" are not wholly a myth, in other words, that immediate environment is in the long run a factor of the frequency of cancer. What is needed is a record of the houses in which cancer has occurred, say for the last 50 or 60 years in (i) a practically fully developed urban district, (ii) a completely agricultural district, and (iii) an industrial area such as occurs frequently in Lancashire or Yorkshire, etc., with relatively small factories, mines or works spread out over a rural district. The examination of the certificates of death of such districts, the careful preparation of "spot maps," and the record of occupations and relationships ought to be a perfectly straight-forward matter, and if it be carried out,—whether it justifies the inference to be drawn from the present data, or does not,—I think Dr Law Webb must be considered as a pioneer in the inquiry.

Experiment. It occurred to me on reading this paper through after completion, that the improbability of multiple cases as measured by the P derived from χ^2 , however much it may appeal to the statistician, might not sufficiently impress the medical mind as demonstrating the non-random character of this cancer distribution. Above all a medical man thinking of cancer as a common disease might fail to appreciate, on reading the individual details of the multiple cases, their extreme improbability taken as a whole.

Accordingly, at the suggestion of my colleague, Dr David Heron, I arranged for the drawing at random of 377 cases out of 2000 possibilities. What we want is something equivalent to drawing 377 times at random a ball out of 2000 balls numbered 1 to 2000 in a bag, each ball being replaced after drawing.

The experiment was arranged in the following manner. A series of numbers of four figures having no exact square root, cube root or reciprocal was taken, and the figures in the seventh decimal place of the first and second and in the tenth decimal place of the third were written down. These formed the last three figures of the numbers. To obtain the first figure in the number, in one case the last figure of the cube, i.e. the twelfth was taken, and in three other cases, the tenth figure of the cube. The whole series of 377 numbers were thus taken directly from tables of cubes, square roots, cube roots and reciprocals (Barlow's). Out of the numbers thus obtained, those beginning with 0, 1, 2, 3, 4 were reckoned as belonging to the first thousand, and those beginning with 5, 6, 7, 8 and 9 to the second thousand; 0000 counted however as 2000. Thus: $3004 = 0004 = 4$, but $5004 = 1004$. We thus had equal chances for every number from 1 to 2000, provided there be no bias in taking numbers consecutively* out of such tables.

The results were as follows :

	Theory	1st Experiment	2nd Experiment	3rd Experiment	4th Experiment
p_1 ...	313	315	321	333	321
p_2 ...	29	31	28	22	28
p_3 ...	2	0	0	0	0
p_4 ...	0	0	0	0	0

The absence of triplets led me to suppose some bias in the tables tending in favour of more uniform distribution than a mere random drawing provides. Accordingly a fifth experiment

* Subject to the omission of numbers with perfect square or cube roots etc., as stated above.

was made in the more lengthy manner of drawing 1508 times a card out of a pack containing 20 cards, numbered in duplicate 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The card was returned, and the pack shuffled and the 1508 draws recorded. Each successive four gave a number which was treated exactly as in the previous four experiments. The results were :

	Theory	Experiment	Cancer
$p_1 \dots$	313	316	315
$p_2 \dots$	29	26	20
$p_3 \dots$	2	3	6
$p_4 \dots$	0	0	1
χ^2	—	approx. 6	approx. 258

It will be seen that the χ^2 of experiment is essentially within the bounds of probability ($P=2$, say), while the cancer-data give a value of χ^2 which is 43 times as large. For the previous set of experiments the χ^2 for the third is about 25, and for the other three about 18—larger values than occur in the shuffling experiment—but of a quite different order to the cancer value. Many arguments—all houses not being of same age, some houses pulled down, and so forth—might be used to account for the multiple cancer cases, but I think these data certainly justify a fuller inquiry into the whole question. They provide some evidence, of more value than mere impression, that the hypothesis of “cancer” houses is worthy of a fuller consideration.

V. Hybridisation of Canaries.

Note on the Communication by C. L. W. NOORDUYN to the Members of the Genetics Congress held in Paris from September 18—23, 1911.

By A. RUDOLF GALLOWAY, M.B., C.M., M.A.

At the Genetics Congress held at Paris in September, 1911, Mr C. L. W. Noorduyn, of Groningen, quoting *Biometrika*, Vol. VII., Nos. 1 and 2, gave from my paper on “Canary breeding” the “Table of Canary Hybrids bred since 1891, arranged to show Plumage Colour.”

This table indicates the rarity of lightly variegated and “clear” hybrids, and Mr Noorduyn regretted that I did not specify the “five or six hybrids” out of the total of 526 that were not bred from ♂ wild bird × ♀ canary, but from the reverse mating of ♂ canary × ♀ wild bird.

He believed the proportion of lightly variegated hybrids to be greater from ♂ canary × ♀ wild bird than from the reciprocal mating.

In view of the prominence so kindly given to this table by Mr Noorduyn, it is necessary for me to say that the information desired is really included in the paper, but refers to five or six exceptional *matings*, not to five or six individual birds as stated by Mr Noorduyn.