VOLUME XXIX DECEMBER, 1929 No. 3

THE PERIODICITY OF INFLUENZA.

BY MAJOR GREENWOOD, F.R.S.

TEN years ago, the late Dr John Brownlee published in *The Lancet* a preliminary note giving the reasons which led him to conclude, provisionally, that epidemics of influenza tended to occur at intervals of approximately 33 weeks. In the following year two other short notes were published in the same journal by Mr Spear and by Dr Stallybrass. Dr Stallybrass, from an analysis of Liverpool data (not very fully set out in his paper), confirmed Brownlee's conclusion, but Mr Spear, using the London data, criticised Brownlee's inferences and methods, pointing out—*inter alia*—that an interval between successive epidemics of the order of 52 weeks agreed better with the observations than an interval of 33 weeks or a multiple thereof. In an appendix to the Report on the Pandemic of Influenza 1918–19 issued by the Ministry of Health, Dr Brownlee reaffirmed his original contentions, published some further collateral evidence and replied to Mr Spear's methodological criticism. That was, I think, the end of Dr Brownlee's connection with the subject; to his death he retained, I think, a half-serious faith in the 33 weeks' rule but did not regard the matter as of sufficient epidemiological interest to justify further investigation.

The subject, however, was not allowed to be forgotten. In the first place, an editorial writer in *The Times* warmly praised Brownlee's discovery and sometimes reproved those who had not taken it very seriously. In the second place, Mr Spear, who had criticised the texture of Brownlee's prophetic mantle, himself became a prophet. On, at least, one occasion, Mr Spear prophesied so accurately that his work caught the eye of the journalists while, because his prophetic formula seemed to give particular prominence to the time between the equinoxes, his general plan was, at least *prima facie*, more agreeable than Brownlee's to those who remembered the traditional centring of epidemics about the spring and autumnal equinoxes.

Briefly, the methods of the two investigators were these. Brownlee observed that from 1889 to 1896 epidemics of influenza in London did really occur at intervals of some 33 weeks, but that if this interval brought the date of emergence into the late summer or early autumn the expected epidemic did not appear. He noted, of course, the very flagrant violation of this rule afforded by the great autumn epidemic of 1918 and that, after about 1903, it ceased to describe the London data adequately, but he still held it to be typical or ideal.

Mr Spear, working more pragmatically, claimed—and rightly claimed—that he could describe the London experience better by the following rule. To find the interval from the maximum of one epidemic to that of the next, subtract the number of the week of the year of the first maximum from 34·5, double the difference and the product gives the expected interval in weeks. Thus, in 1890 the week of maximum mortality in London was the third, twice 31·5 is 63 and so we shall expect the next maximum 63 weeks later (actually it was 69 weeks on). Mr Spear applied the plan to the years from 1890 to 1926 and (see Table I) his results were much better than those of Brownlee's rule.

Table I. *Influenza in London.*

(From *Report of London County Medical Officer of Health for* 1925, p. 36.)

Method of forecast	Forecast too early by: 14 weeks or more	Forecast too early by: 5–13 weeks	Forecast correct to within 4 weeks	Forecast too late by: 5–13 weeks	Forecast too late by: 14 weeks or more
I. (Mr Spear's rule)	—	6	20	6	1
II. (Dr Brownlee's rule)	2	6	8	9	8

Since (1) neither Dr Brownlee nor Mr Spear claimed to be able to predict whether an epidemic would be large or small, whether it would kill its hundreds or its thousands, (2) both prophets were completely gravelled by the 1918–19 sequence, (3) even Mr Spear's rule—although decidedly better than Dr Brownlee's—sometimes gave very poor approximations to the "truth," a public health administrator, anxious for a timely warning as to the demand likely to be made upon hospital accommodation, will not be much interested in adjudicating between the prophetic merits of Dr Brownlee and Mr Spear. Perhaps he may recall Dr Johnson's difficulty in determining the precedence of Derrick and Smart[1]. But neither Dr Brownlee nor Mr Spear was setting up as a racing tipster; both investigators were seeking an epidemiological truth which, they hoped, might—more or less distorted—emerge in a numerical sequence or rule and that is an object worth a little more study.

I must, at the outset, admit that I am by no means sure that I *have* grasped the philosophy of Dr Brownlee's undertaking or that Mr Spear really claimed more than a purely pragmatic sanction for his method. I think, however, that Dr Brownlee was flying at higher game than Mr Spear; that he thought of epidemic influenza, *sub specie aeternitatis*, as an unbroken series of events of the following kind. One measures points of time along a horizontal axis, and, starting at any convenient origin, finds that the maxima of successive epidemics occur at the points $y_0, y_1, y_2, \dots y_x$, where y_x denotes the time interval from the origin at which the $(x+1)$th epidemic of the series occurred. Then Brownlee conceived that, for all values of x,

$$y_{x+1} - y_x = \Delta y_x$$

should be substantially constant and that Δy_x was (for London) 33 weeks. He conceived that this constant was some function of the life-cycle of the causal organism, but what he meant by *that* is just as mysterious, to me, as Sir William Hamer's speculations anent mutating ultra-microscopic viruses

[1] See Boswell, *sub anno* 1783.

https://doi.org/10.1017/S0022172400009955 Published online by Cambridge University Press

or as Sydenham's occult and inexplicable changes *in ipsis terrae visceribus.* Dr Brownlee chose this as an ideal series, but, noting that actually epidemics hardly ever occurred in the late summer and autumn, he supposed that if the 33 weeks' interval brought the time-point into an unfavourable season of the year, the epidemic would be "missed." Hence the intervals actually to be expected in real data would not always be 33 weeks but sometimes, indeed more often, 66 weeks. A succession of two or more 66 weeks' intervals might be found, more than two successions of 33 weeks could not be observed. Such was Brownlee's hypothesis and, as Mr Spear has shown, it does not satisfactorily describe the observed facts. Brownlee was aware that it did not describe London experience after 1896 and it is not clear why he attributed typical importance to the series of years which it did describe.

I do not know that Mr Spear thought of a succession of epidemics in this way, but it is not without interest to note that the translation of his method into this language would also yield as an ideal a succession differing from Dr Brownlee's only in that instead of one constant interval there would be two, one succeeding the other regularly. In symbols, we may write

$$y_{x+1} - y_x = [k_1 - y_x + (x-1)\,k_2]\,k_3,$$

where the y_x's have the meaning assigned above and the k's are constants, viz. $k_1 = 34{\cdot}5$, $k_2 = 52$ or 53, $k_3 = 2$. Solving for y_x, we have

$$y_x = \frac{l}{l-1}\left[A + B(x-1) - \frac{Bl}{l-1}\right] + \frac{C}{l^x},$$

where $$\frac{1}{l} = 1 - k_3, \quad A = k_1 k_3, \quad B = k_2 k_3,$$

and C is the constant introduced by summation. In other words

$$y_{x+1} - y_x = \Delta y_x = \frac{Bl}{l-1} + \frac{C}{l^x}\left(\frac{1-l}{l}\right) \qquad \ldots\ldots(1).$$

With Mr Spear's numerical values, $l = -1{\cdot}0$. Hence the intervals between epidemics are alternately $\frac{B}{2} - 2C$ when x is even, and $\frac{B}{2} + 2C$ when x is odd.

This translation into terms of an ideal series of epidemics is interesting in showing a certain resemblance between the methods but it has no other value, and I turn to the real as distinct from the ideal similarity of the two methods. That real similarity is due to the observed facts (of London experience) that the later the emergence of the epidemic in the first half (nearly always the first *quarter*) of the year, the shorter, on the average, the time we have to wait before a new epidemic emerges. The difference between Dr Brownlee and Mr Spear is that Dr Brownlee will not permit epidemics to occur at shorter intervals than 33 weeks, while it is theoretically possible by Mr Spear's rule that the maximum of one epidemic should be only a week later than that of its predecessor (when the first epidemic has its maximum in the 34th week of the year). Since, in 1918, two great epidemics succeeded one another within a much shorter interval than 33 weeks, pragmatic honours would seem to fall to Mr Spear. But the rule secures this triumph at the rather serious cost of demanding that while an epidemic in the 34th week requires another pestilence a week later, an epidemic in the 35th week will keep us free for 103 weeks. Of course this might be amended by making the rule apply not to the *next following* but to the *nearest* 35th week of a year, but we should still have the difficulty of an impossibly short interval when an epidemic happens close to the 35th week of the year.

If we scrutinise the basis of Mr Spear's rule a little more closely, we find it involves a very bold hypothesis. Suppose we have two variables x_1 and x_2 and linear functions of them of the form $k_1 - x_1 + x_2$ and $k_2 - k_3x_1$, where the k's are constants. Then these functions are not independent even if x_1 and x_2 are independent one of another. In fact the coefficient of correlation between the two expressions is

$$\frac{\sigma_1 - \sigma_2 r_{12}}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 r_{12}}} \qquad \text{......(2)},$$

where σ_1 and σ_2 are the standard deviations of x_1 and x_2 and r_{12} the coefficient of correlation, while the coefficient of linear regression of $k_1 - x_1 + x_2$ upon $k_2 - k_3x_1$ is

$$\frac{\sigma_1 - \sigma_2 r_{12}}{k_3\sigma_1} \qquad \text{......(3)}.$$

Now if x_1 is the ordinal number of the week of the year (the measure of the first week of the year is 1, of the second week 2, etc.) in which the maximum of an epidemic occurs, and x_2 the similarly measured maximum of the next year, we have approximately

$$\sigma_1 = \sigma_2 \text{ and } \bar{m}_1 = \bar{m}_2;$$

if $$r_{12} = 0, \ (2) = \frac{1}{\sqrt{2}} \text{ and } (3) = \frac{1}{k_3}.$$

In our special case k_1 is 52 (or 53), and k_3 is -1, so calling $k_1 - x_1 + x_2$ y and referring to the means we reach

$$y - 52 = (1 - r_{12})(\bar{m}_1 - x_1) \qquad \text{......(4)}.$$

This is the linear regression equation of interval between epidemics upon number of week of occurrence of first epidemic of a pair. If $r_{12} = 0$, it becomes

$$y = 52 + \bar{m}_1 - x_1.$$

Now in London experience $\bar{m}_1$ is about 8. So that for London the arithmetical relation would be $y = 60 - x$ (dropping subscripts).

Similarly written Mr Spear's rule is $y = 69 - 2x$, which is just what (4) would give us if we put $r_{12} = -1{\cdot}0$ and $\bar{m}_1 = 8{\cdot}5$. In other words, the Spear rule assumes perfect negative correlation between the ordinal numbers of the weeks containing successive maxima. It will be seen, of course, from (2) that the correlation of length of interval with ordinal number of first maximum is (changing the sign of (2) to replace $-x$ by x) $-0{\cdot}707$ when r_{12} is zero and $-1{\cdot}0$ if r_{12} is $-1{\cdot}0$. In fact, the empirical value for the London data used (see below) is $-0{\cdot}9 \pm 0{\cdot}01$ (for these data, the mean interval is 51·91 weeks and the mean value of x 8·33[1]), suggesting that the successive x's are not independent, as indeed direct calculation shows. One finds that

$$r_{12} = -0{\cdot}51 \pm 0{\cdot}09.$$

But the empirical correlation is very far indeed from perfect. Hence we may ask ourselves which of the three courses open to us is likely to give the least unsatisfactory result. (1) To assume no correlation whatever between the successive ordinal numbers. (2) To assume, with Mr Spear, perfect correlation. (3) To assume an intermediate degree of correlation.

To meet the third eventuality, I used the following artifice. The week of a maximum was measured from the approximate mid-point of the year, *e.g.* the ordinal number 10, say, was replaced by $26 - 9{\cdot}5$, the difference multiplied by 180/52 and treated as an angle, the cosine of which was used as the measure of position in the year. That is, the annual range was from $-\pi/2$ to $+\pi/2$ and the "independent" variable x' forced to lie between 0 and + 1. The regression equation of interval upon this x' was $y = 66{\cdot}287 - 32{\cdot}489x'$. No real difference is effected if we write $y = 66 - 33x'$ and so, "through great varieties of untried being," return to Brownlee's constant twice over! That romantic satisfaction is, however, the only advantage of torturing the variable; the equation $y = 65{\cdot}4 - 1{\cdot}58x$, x being measured as above, gives us essentially the same arithmetical values. It is, indeed, altogether unnecessary to think or write in terms of *interval* at all for most practical purposes. Mr Spear's rule might quite as well be written $x_2 = 17 - x_1$, where x_1 is the ordinal number of one year and x_2 of the next, and independence might have been tested by equating x_2 to a constant, viz. the mean ordinal number, provided we are permitted negative values for our x's. Still, for form's sake only, I have tested intervals. I have used the three "rules,"

$$y = 69 - 2x,\quad y = 60 - x,\quad y = 66 - 33x',$$

the rules of perfect, zero and moderate correlation, and have tested them (1) on data upon which Mr Spear's rule and the rule of moderate correlation were based, viz. London statistics, (2) upon independent data. If the surmise I formed, viz. that the Spear rule was something of an arithmetical mare's nest, were correct, this is what we should expect:—That for the London data, the rules of perfect and moderate correlation would agree slightly better with the observations than the rule of zero correlation, but not greatly better, while for other data all would be bad but the rule of zero correlation rather less bad than the others. That is exactly what *has* happened. In Table II we have London experience (omitting the pandemic years, recalcitrant to both Brownlee's and Spear's handling). If one accepts, as Brownlee was disposed to do, a prediction within 4 weeks of the true date as a success,

[1] In calculation I have measured the 52nd week of a year as having for ordinal number -1, the 51st as having for ordinal number -2, etc. If the distribution were uniform, this must lead to an absurdity—which I have, rather clumsily, avoided in the third artifice of the text. The whole undertaking is, in the nature of things, so rough that the point is of very little importance.

Table II. *Influenza in London.*

Year	Week of maximum	Interval to next maximum	Interval calculated by Spear's rule	Interval calculated by $y=60-x$	Interval calculated by $y=66-33x'$
1890	3	69	63	57	61
1891	20	35	29	40	36
1892	3	63	63	57	61
1893	14	36	41	46	42
1893	50	64	75	63	61
1895	10	47	49	50	48
1896*	5	64	59	55	57
1897	16	41	37	44	40
1898	5	58	59	55	57
1899	11	43	47	49	46
1900	2	60	65	58	63
1901	10	50	49	50	48
1902*	8	48	53	52	52
1903	3	59	63	57	61
1904	10	49	49	50	48
1905	7	61	55	53	53
1906	16	37	37	44	40
1907	1	60	67	59	65
1908*	9	56	51	51	50
1909	12	46	45	48	45
1910	6	54	57	54	55
1911	8	52	53	52	52
1912	8	56	53	52	52
1913*	12	46	45	48	45
1914	5	55	59	55	57
1915	8	61	53	52	52
1916	17	35	35	43	38
1920	12	53	45	48	45
1921	13	41	43	47	43
1922	2	64	65	58	63
1923	14	45	41	46	42
1924*	7	52	55	53	53
1925	6	60	57	54	55
1926	14	42	41	46	42
1927	4	49	61	56	59
1928	1	59	67	59	65
		Mean deviations	3·9	4·1	3·4

* 53-week year.

Mr Spear scores 22 bull's-eyes, the rule of moderate correlation 25, and the rule of zero correlation only 2 fewer hits than that of Mr Spear. Even in London, a prophet who resolutely asserts that the week of a maximum will always be the same week—or says that influenza has an annual period—will do inappreciably worse than Mr Spear. Turning now to other cities, I chose Manchester, Birmingham, Liverpool, Leeds, Sheffield, Edinburgh and Glasgow and went through their records for 1921–9. I have disregarded any maximum determined by fewer than 10 deaths and any interval exceeding 75 weeks. The results are shown in Table III.

In one city only, Manchester, the application of the rule gave no trouble; one could fix a maximum for each year and so had 8 observed intervals to test. Mr Spear's rule scores a bull's-eye 4 times, the zero rule also 4 times and the transformed rule 5 times. If we take account of the deviations, *i.e.* apply an arithmetical criterion, the average error of Mr Spear's rule is 4·9 weeks, of the zero rule 3·5 weeks and of the transformed rule 3·75 weeks.

https://doi.org/10.1017/S0022172400009955 Published online by Cambridge University Press

Table III.

Year	Week of maximum	No. of deaths registered in week of maximum	Observed interval to next maximum	Interval calculated by Spear's rule	Interval calculated by $y=60-x$	Interval calculated by $y=66-33x'$
			MANCHESTER			
1921	19	16	36	31	41	36
1922	3	52	64	63	57	61
1923	15	32	50	39	45	41
1924*	13	44	45	43	47	43
1925	5	23	56	59	55	57
1926	9	17	51	51	51	50
1927	8	88	45	53	52	52
1928	1	12	58	67	59	65
1929	7	119	—	—	—	—
			BIRMINGHAM			
1921	None					
1922	4	57	63	61	56	59
1923	15	24	47	39	45	41
1924*	10	41	55	49	50	48
1925	12†	34	55	45	48	45
1926	15	21	45	39	45	41
1927	8	8	—	—	—	—
1928	None					
1929	10	266	—	—	—	—
			LIVERPOOL			
1921	17	10	42	35	43	38
1922	7	54	—	—	—	—
1923	None					
1924*	9	21	53	51	51	50
1925	9	26	59	51	51	50
1926	16	17	44	37	44	40
1927	8	47	—	—	—	—
1928	None					
1929	6	82	—	—	—	—
			LEEDS			
1921	50	35	78	75	62	61
1923	24	13	38	21	36	34
1924*	10	72	50·5	49	50	48
1925	7·5‡	13	—	—	—	—
1926	None					
1927	10	18	—	—	—	—
1928	None					
1929	9	139	—	—	—	—
			SHEFFIELD			
1921	10§	10	40	49	50	48
1921	50	40	73	75	62	61
1922	None					
1923	19	46	47	31	41	36
1924*	14	15	—	—	—	—
1925	None					
1926	3	15	57	63	57	61
1927	8	44	—	—	—	—
1928	None					
1929	10	114	—	—	—	—

* 53-week year.
† 34 deaths registered in 11th, 12th and 13th weeks.
‡ 13 deaths registered in the 7th and also in the 8th week.
§ A very dubious attribution, 10 deaths registered this week, 6 in the 11th, 10 again in the 12th week.

https://doi.org/10.1017/S0022172400009955 Published online by Cambridge University Press

Table III—*continued.*

Year	Week of maximum	No. of deaths registered in week of maximum	Observed interval to next maximum	Interval calculated by Spear's rule	Interval calculated by $y=60-x$	Interval calculated by $y=66-33x'$
			EDINBURGH			
1921	1	14	55	67	59	65
1922	4	116	—	—	—	—
1923	None					
1924*	13	13	36	43	47	43
1924*	49	11	70	77	64	59
1926	15	10	46	39	45	41
1927	9	27	—	—	—	—
1928	None					
1929	7	68	—	—	—	—
			GLASGOW			
1921	9	21	46	51	51	50
1922	3	202	—	—	—	—
1923	None					
1924*	12	52	42	45	48	45
1925	1	12	48	67	59	65
1925	[49	11	17	77	64	59]
1926	14	81	51	41	46	42
1927	13	11	—	—	—	—
1928	None					
1929	4	236	—	—	—	—

Mean deviations of Table III (omitting Glasgow, 1925–6): 6·3, 4·4, 5·8.

* 53-week year.

Birmingham gives only 5 pairs. Here Mr Spear only hits the bull's-eye once and has a mean error of 6·4 weeks. Random chance scores 2 bull's-eyes and has a mean error of 4·2 weeks, the transformed rule scores 2 bull's-eyes and has a mean error of 6·2 weeks. Liverpool gives only 4 pairs. The scores are: Mr Spear 1 bull's-eye, mean error 6·0 weeks; random succession 3 bull's-eyes, mean error 2·75 weeks; transformed rule 3 bull's-eyes, mean error 5·0 weeks. So far, honours are with randomness. It scores 9 bull's-eyes out of a possible 17 and has a mean error of 3·5 weeks. Mr Spear scores 6 bull's-eyes and has a mean error of 5·6 weeks. The transformed rule makes 10 bull's-eyes but has a mean error of 4·8 weeks. The other cities provide even more dubious material. For what they are worth, however, the results are these. Of 13 "possibles," Mr Spear scores 2 bull's-eyes and has a mean error of 12·5 weeks; chance scores 4 bull's-eyes and has a mean error of 9·5 weeks; the transformed rule has 3 bull's-eyes with a mean error of 11·0 weeks.

As another test I have examined the notification data of the City of Copenhagen since 1920 (Table IV). Here there is no difficulty in assigning the week of maximum *recorded* incidence. Even in 1920, a year of very low incidence, the 50th week with 145 cases differs fairly substantially from any other at all adjacent week, while in each other year the maximum is quite sharp. Thus in 1922, the week of maximum had 8896 notifications, the previous and successive weeks 5845 and 7955 respectively. Treating the data as before, we see, as before, that Spear's rule gives no better results than the rule of independence.

Table IV. *Notifications of influenza in Copenhagen.*

Year	Week of maximum	No. of cases notified in week of maximum	Observed interval to next maximum	Interval calculated by Spear's rule	Interval calculated by $y = 60 - x$	Interval calculated by $y = 66 - 33x'$
1920	50	145	57	75	63	61
1922	3	8896	60	63	57	61
1923	11	977	51	47	49	46
1924*	10	2886	60	49	50	48
1925	17	846	53	35	43	38
1926	18	696	37	33	42	37
1927	3	6725	69	63	57	61
1928	20	1336	36	29	40	36
1929	4	9749	—	—	—	—

* 53-week year.

It is not, I think, necessary to pursue the subject further. I hope to have made clear what assumptions, what extravagantly improbable assumptions, the correctness of Mr Spear's rule must involve, and I have certainly shown that when applied to the rough data of ordinary life it gives no information of either practical value or epidemiological interest. Arithmetical devices of this class are, I believe, quite nugatory.

REFERENCES.

BROWNLEE, J. (1919). *Lancet*, ii, 856.

—— (1919). *Report on the Pandemic of Influenza*, 1918–19 (Stationery Office), p. 575.

SPEAR, J. (1920). *Lancet*, i, 589.

—— (1925). *London County Council, Annual Report*, 1925, Vol. 3 (Public Health), p. 34.

STALLYBRASS, C. O. (1920). *Lancet*, i. 372.

(*MS. received for publication* 30. v. 1929.—Ed.)