

A STATISTICAL DISCUSSION OF THE RELATIVE EFFICACY OF DIFFERENT METHODS OF TREATING PNEUMONIA *

RAYMOND PEARL

BALTIMORE

Recently Head¹ contributed an interesting discussion of the results of treating post-influenzal pneumonia by the open as contrasted with the close ward method. The general result apparently was to show that the latter method was greatly superior to the former, as evidenced by the case fatality rates under the two modes of procedure. While the author definitely draws this conclusion and states with confidence that the study of the 1,400 cases he dealt with "points the way to the more successful management of this disease," he also raises a rather obvious point of criticism to the results as they stand, using the following words:

The question at once arises, Is the lowered mortality here shown in favor of the closed ward treatment a real gain in the management of the disease over the open ward method, or are the favorable results only a coincidence, an expression of a lowered mortality arising naturally in the latter part of the epidemic? It has been stated by Abrahams, Hallows and French that the mortality is higher and the disease more severe in the early part of epidemics of influenza complicated by pneumonia.

The problem raised is an interesting one statistically, and one of a sort which continually arises in medical literature. A new method of treatment for a disease is devised and tried in a number of cases with a case fatality rate lower, by a greater or smaller amount, than the case fatality rate which had prevailed under some other mode of treatment. The problem in any such case is this: Is the lowering of the case fatality rate in the new circumstances so great that it cannot reasonably be supposed to have arisen by chance alone? Usually, no adequate evidence on this point is given by the physician, chiefly for the reason that he has no technical knowledge of the proper mathematical tests to apply. The potency of random sampling in bringing about divergences is apparently but little understood outside of professional statistical circles.

As an illustration of the effects of random sampling let us consider a hypothetical case. In any large city, or a state, or indeed, any large population aggregate, the average age at death of persons dying at the same calendar date should be identical for all dates, except for the influence of two factors, viz: (a) chance, or random sampling, and (b) long seasonal waves arising from the fact particularly that rela-

* Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University, No. 6.

1. Head, G. D.: Treatment of Pneumonia, J. A. M. A. **72**:1268 (May 3) 1919.

tively more infants die in hot summer weather than in the colder seasons of the year. In any short period, say ten consecutive days, the second factor (b) would not operate, and we should expect the persons dying on each of these consecutive ten days to show the same average age, except for the fluctuations due to chance alone. How considerable these fluctuations may be is shown in Table 1, which gives the number of deaths and the average age at death of those dying during ten consecutive days in 1916 in Baltimore.²

TABLE 1.—MEAN AGE AT DEATH OF THOSE DYING IN THE STATED DAYS IN BALTIMORE

Date	Number of Deaths	Mean Age at Death in Years
Jan. 13, 1916.....	31	30.16
14.....	40	43.80
15.....	27	40.59
16.....	48	48.21
17.....	32	48.34
18.....	41	51.90
19.....	39	46.82
20.....	31	52.39
21.....	39	51.62
22.....	57	39.40
Total for the year.....	10,668	39.83

Here we have a fluctuation in the average, based on samples of from thirty to fifty individuals, amounting to more than twenty-two years, arising from random sampling alone. Such an illustration emphasizes the fact that before conclusions can safely be drawn from differences between numbers it is necessary to know something about the "probable errors" of those numbers.

The particular theorem in probability which is applicable in problems of the type exemplified in Head's paper referred to above was first set forth by Pearson.³ I have recently discussed this subject further.⁴

Let it be supposed that a first sample of $n = p + q$ be drawn from the population, p denoting the number of times the event dealt with occurs in the n trials, and q the number of times it fails.

Write

$$\bar{p} = \frac{p}{n}, \quad \bar{q} = \frac{q}{n},$$

whence of course

$$\bar{p} + \bar{q} = 1.$$

We then have for the chief constants of the error distribution for a second sample, of magnitude m , drawn from the same population the following values:

2. I am deeply indebted to Dr. John H. Blake, Commissioner of Health, and Dr. W. T. Howard, Assistant Commissioner of Health of Baltimore, for kindly allowing me the use of the records of the Health Department for the study of this and other problems now being investigated in this laboratory.

3. Pearson, K.: *Phil. Mag.*, March, 1907, p. 365.

4. Pearl, R.: *Amer. Nat.* 51:144, 1917.

$$\text{Mean}^s = m\bar{p} + \frac{m}{n+2} (\bar{q} - \bar{p}), \quad (\text{i})$$

$$\text{Mode} = \text{the integral portion of } m\bar{p} + \bar{p} \quad (\text{ii})$$

$$\text{Standard deviation} = \left\{ m \left[\bar{p} + \frac{\bar{q} - \bar{p}}{n+2} \right] \left[\bar{q} - \frac{\bar{q} - \bar{p}}{n+2} \right] \left[1 + \frac{m-1}{n+3} \right] \right\}^{1/2} \quad (\text{iii})$$

These values are entirely general and independent of the values of n , m , p and q . Under certain circumstances, as when n is very large as compared with m , and neither p nor q is very small, (i) and (iii) are obviously capable of being put in much simpler form and still giving a sufficiently close approximation to the true result.

Let us now apply this theorem to the data given by Head.¹ From the figures he gives we may state the problem in these terms: If under the open ward treatment, out of a sample of 966 patients with acute pneumonia 135 died, what would be the probable number to die in a second sample of 435 acute pneumonia patients from the same population, given the same treatment?

We have here, in our mathematical notation

$$\begin{aligned} n &= 966 \\ m &= 435 \\ p &= 135 \\ q &= 831 \\ \bar{p} &= \frac{135}{966} = .1397 \quad \bar{q} = \frac{831}{966} = .8603 \end{aligned}$$

Whence from (i) and (ii) we readily deduce

$$\begin{aligned} \text{Mean deaths expected in second sample} &= 61.12 \pm .28 \\ \text{Modal, or most probable number of deaths in second sample} &= 61 \\ \text{Standard deviation} &= 8.72 \end{aligned}$$

But the actually observed number of deaths in the second sample, under the close ward treatment, was fourteen instead of sixty-one. Hence we may safely conclude that under the close ward treatment significantly fewer persons died than would have been expected to die on the basis of chance, if the same force of mortality had prevailed in the latter period as did in the former.

But can it be assumed that the same force of mortality was impinging, and its results were mitigated only by the new method of treatment? As we have seen, Head himself expresses some doubt on this point, and surely no safe conclusion as to the merits of the treatment can be drawn until this point is settled. In order to settle it we have used data from a comparable outbreak of post-influenzal pneumonia where the same treatment was followed throughout the outbreak. Would there be in such a case a falling off of the case-fatality rate in the latter part of the epidemic corresponding entirely or in some degree to that observed by Head? Through the kindness of

5. From the origin at the lower range end, or $r=0$.

Dr. Eugene L. Opie, to whom I wish to express my indebtedness, I am able to present data of this sort for the Camp Pike epidemic.

Table 2 gives the cases and deaths in the Camp Pike epidemic. The cases have already been published by Opie⁷ *et al.* The deaths by days he kindly furnished me in manuscript.

TABLE 2.—CASES AND DEATHS OF PNEUMONIA AT CAMP PIKE, ARK., FROM SEPT. 1 TO OCT. 31, 1918

Day September	Cases of Pneumonia	Deaths from Pneumonia	Day October	Cases of Pneumonia	Deaths from Pneumonia
1	1	0	1	126	16
2	3	0	2	142	11
3	2	0	3	113	20
4	1	0	4	129	32
5	3	0	5	86	20
6	3	1	6	89	22
7	4	0	7	93	29
8	3	0	8	76	30
9	1	0	9	42	30
10	3	0	10	66	28
11	4	0	11	54	32
12	1	0	12	34	20
13	4	0	13	15	18
14	7	1	14	13	13
15	10	0	15	9	13
16	10	0	16	19	11
17	7	0	17	8	8
18	5	3	18	8	8
19	11	..	19	9	7
20	12	..	20	11	1
21	10	1	21	5	5
22	13	1	22	6	1
23	15	5	23	11	5
24	13	3	24	5	8
25	25	3	25	7	2
26	16	0	26	3	0
27	23	0	27	1	5
28	30	3	28	1	1
29	40	11	29	3	1
30	35	11	30	0	1
			31	0	0
Totals.....				1,499	441

In order to treat these data in a way comparable to the others, it is necessary to divide the epidemic at a corresponding point. In the Camp Wheeler epidemic (Head's data) the change in treatment occurred after 966 out of 1,401 cases had occurred. The same stage of the Camp Pike epidemic was reached $\frac{966}{1,401} \times 1,499 = 1,034$ cases had occurred. The nearest date to this point, as can be determined by summing the "case" columns of Table 2, is October 6. At the end of that day 1,000 cases of pneumonia had occurred in the Camp Pike epidemic. We may then take through October 6 as the first phase of the Camp Pike epidemic, so far as concerns cases, this phase corresponding to the "open ward" period of the Camp Wheeler data. We have next to consider the deaths which belong to these 1,000 cases occurring prior to the end of October 6. In general, it appears from the data in hand that there was, on the average, a lag of about eight

7. Opie, E. L., Freeman, A. W., Blake, F. G., Small, J. C., and Rivers, T. M.: J. A. M. A. **72**:556 (Feb. 22) 1919.

days between case and death curves in the epidemic of post-influenzal pneumonia in camps. Accepting this figure, as in accord to a first degree of approximation at least, with the facts, we have October 14 as the date for dividing the Camp Pike death curve into two portions, which will correspond approximately to the two different treatment moieties of the Camp Wheeler data.

In the same notation as before we have for Camp Pike:

Cases of pneumonia through October 6 = $n = 1,000$

Deaths from pneumonia through October 14 = $p = 364$, whence $q = 636$, and $\bar{p} = .364$, and $\bar{q} = .636$

Cases of pneumonia beginning October 7 and going to end of the epidemic = $m = 499$

The problem then may be stated in this way: If, with no change of treatment, 364 patients died out of a sample of 1,000, what is the probable number of deaths in a second sample of 499 cases? By the same method as before the data give

Mean deaths expected in second sample = $18.77 \pm .40$

Modal, or most probable number of deaths in second sample = 182

Standard deviation = 13.15

Now, the actual number of deaths in the last 499 cases of the Camp Pike epidemic was only seventy-seven instead of the expected 182. But there had been no change in method of treatment. Hence, it is clear that the statement of Abrahams, Hallows and French, quoted from Head at the beginning of this paper, is true. Fewer deaths in proportion to cases occur in the later as compared with the earlier portion of these epidemic outbreaks, quite without change of treatment.

This result obviously tends to cast reasonable doubt on the efficacy of the close ward as compared with the open ward treatment of these epidemic pneumonias. It is necessary, however, to make a further quantitative comparison before drawing any final conclusion. In the second sample (latter part) from the Camp Pike epidemic the actual deaths formed 42 per cent. of the deaths expected on the basis of chance from the results shown in the first sample from the same epidemic ($\frac{77 \times 100}{182} = 42$ per cent.). In the Camp Wheeler epidemic (Head's data) the actual deaths under the close ward treatment formed only 23 per cent. of the deaths expected on the basis of chance from the results shown in the first sample from the same epidemic, under the open ward treatment ($\frac{14 \times 100}{61} = 23$ per cent.).

This result gives the significant comparison. The whole matter may be summarized in this way. While it is true that the case fatality rate tends under a constant form of treatment to be markedly lower in the later portions of epidemic outbreaks of pneumonia, nevertheless Head's data show, when given proper mathematical analysis, that under the close ward treatment only about half (23 per cent. versus

42 per cent.) as many deaths occurred relatively in the latter part of the epidemic as would have occurred under the open ward method of treatment, after making allowance for the normally diminishing case fatality rate of later portions of the epidemic. To that degree, then, Head's conclusions are justified, and obviously constitute an important contribution to the sum of medical knowledge regarding the proper treatment of pneumonia.

COMMENT

The first purpose of this paper is to find out, by appropriate mathematical treatment of the data, the real truth as to the relative merits of two methods of treating epidemic pneumonia. Because of the lack of such analysis the results as originally published by Head, comparing the open ward and close ward treatments, were scientifically inconclusive.

The second purpose of the paper is to illustrate by a concrete example, not only the importance, but indeed the absolute necessity of mathematical tests of the validity of results and conclusions, if medicine is ever to measure up to the standards of scientific logic and accuracy which prevail in some, at least, of the other branches of biologic and physical science. Many, indeed most, of the problems of practical medicine in a broad sense are either essentially statistical problems, or their statistical phase is a vitally important one in reaching correct conclusions. The medical man is thoroughly familiar with the necessity for clinical and laboratory tests. Only in rare instances does he realize the necessity of applying to his results before drawing conclusions from them, whether for purposes of publication or practice, a proper test for the magnitude of the probable errors arising from random sampling. It is a commonplace of medical literature to find the most positive and sweeping conclusions drawn from a sample so meager as to make scientifically sound conclusions of any sort utterly impossible. It is not too much to say that the investigator in the field of medicine should be as familiar with the probable error test as he is with the Wassermann test.

SUMMARY

In this paper it is shown that in deciding as to the merits of different methods of treatment of a disease it is necessary to take account of the probable errors arising from random sampling. In the case of Head's results on the pneumonia epidemic in Camp Wheeler the reduction in the fatality rate following the institution of the close ward treatment is shown to be only in part a result of the treatment, the remainder being the normal reduction in the case-fatality rate in the later portion of epidemics.