VOLUME XXXVI JUNE, 1936 No. 2

NASAL DIPHTHERIA CARRIERS

BY G. DE M. RUDOLF, M.R.C.P., D.P.H., D.P.M., *Bristol*
AND W. R. ASHBY, M.A., M.D., D.P.M., *Leavesden, Watford*

(With 3 Figures in the Text)

THIS paper is a study of 71 adult male carriers of morphological *B. diphtheriae* occurring in an institution. No cases of clinical diphtheria developed, although tests demonstrated that many of the cultures were virulent to guinea-pigs. Others were avirulent.

Nasal swabs were collected at irregular intervals and were continued until three successive negative specimens had been obtained from each patient dealt with in this article.

When carriers other than nasal, and certain cases which do not conform to statistical requirements are omitted, a series of 62 nasal carriers remains for statistical investigation, those omitted being cases in which the first swab was not positive or in which the series did not terminate with three negative specimens.

All these patients were isolated and treated with normal saline gargles and the sniffing of saline up the nose three times a day.

The following investigation is based upon the hypothesis of a constant probability of obtaining a positive swab under these conditions of treatment.

The statistical methods used have been placed in Appendices I–V preceding the Summary.

EXAMINATION OF RESULTS

(1) *Homogeneity of the series*

First we should test the series to see whether every person behaved in much the same way, and responded in the same way to treatment.

To do this, the "length of run", *i.e.* the number of swabs made on each person until three successive negative results were obtained, was recorded for each person. The 62 lengths of run form a distribution.

To test for homogeneity we may compare the actual variance of this distribution with the variance calculated from the (infinite) distribution which we would get by pure chance. A lack of homogeneity will make the actual variance greater than the expected variance.

It may be shown (see Appendix II) that if the chance of getting a negative result is q, then the variance of the distribution of length of run will be

$$V = (1-q)\,(1+3q+6q^2+3q^3+q^4)/q^6.$$

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

In our series, 386 swabs were taken, of which 243 were negative. So q may be estimated as $243/386 = 0{\cdot}6295$. Substituting this value of q in the equation above, we find $V = 36{\cdot}73$. Calculating the variance of our actual results we find it to be 17·78. This is less than the estimated variance, so there is no evidence that the actual variance *exceeds* the estimated variance.

The series, therefore, appears to be reasonably homogeneous.

(2) *Effect of treatment*

To determine whether the length of run of swabs was reduced by the effect of treatment, it is necessary to find the expected mean length of run ($\bar{X}$).

First we find that the actual mean length of run ($\bar{x}$) is 6·226 swabs. We now have to calculate the expected length of run. It may be shown that if q is the chance of getting a negative result, then

$$\bar{X} = \frac{1+q+q^2}{q^3} \qquad \text{.....(A)}$$

(see Appendix I). We have to estimate q from the data. This may be done quite straightforwardly by finding

$$q = \frac{\text{no. of negative swabs}}{\text{total number of swabs}}$$

$$= \frac{243}{386}$$

$$= 0{\cdot}6295.$$

From this, by substituting in equation (A), we find that the expected mean length of run = 8·120 swabs, *i.e.* our actual average length of run (6·226) is about two swabs shorter than one would expect by pure chance. Testing the significance of this difference, we find (see Appendix III) that $P = 0{\cdot}05$, meaning that there is one chance in twenty of getting a deviation as large as this by pure chance. This degree (one in twenty) is usually taken as the dividing line for significance, so it appears likely that there has been some effect from the treatment, but the data are not quite decisive on the point.

Fig. 1 shows a comparison of the expected numbers for each length of run (thin line) against the actual numbers (heavy line). The calculated expected numbers are given in Appendix IV.

(3) *Duration of the carrier period*

The lengths of the periods until three successive negative swabs were obtained were known in many instances. In others, however, an interval of 1 or 2 days occurred between the last positive swab and the first negative, as swabs were not always taken every day. In consequence, there are known lengths of positive periods (from the first to the last positive swabs inclusive), and possible positive periods (from the first positive swab to the day previous to the first negative swab, inclusive). Table I shows that of 19 known durations,

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

10 were of 1 day's duration and all were of less than 9 days. Of 43 possible durations 31 were under 9 days.

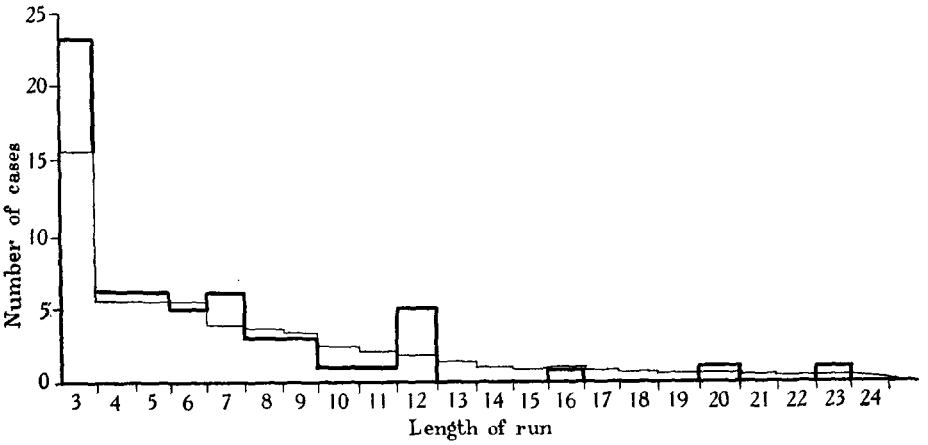

Fig. 1.

Table I. *Duration of carrier period*

Periods in days	No. of cases with known periods	No. of cases with possible periods
1	10	0
2	1	3
3	1	8
4	2	3
5	0	7
6	2	4
7	2	4
8	1	2
9	0	0
10	0	1
11	0	0
12	0	0
13	0	0
14	0	4
15	0	3
16	0	0
17	0	0
18	0	0
19	0	1
20	0	0
21	0	1
36	0	1
51	0	1
	19	43

(4) *Results of examination on successive days*

Table II shows that when swabs are collected daily from patients, a continuous series of positives is rarely found, the swabs being usually mixed, *i.e.* negative and positive.

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

Table II. *Collection of daily swabs*

No. of successive daily swabs	No. of cases	No. giving all positives	No. giving positives and negatives
3	22	2	20
4	14	3	11
5	17	2	15
6	3	2	1

(5) *Negative swabs preceding positive*

In 22 cases two successive negative swabs followed by a positive were obtained. Two negatives are clearly insufficient for pronouncing that the carrier condition has ceased.

The probability of getting n successive negative results from a carrier can be found by proceeding as follows:

The number of tests necessary to achieve any degree of certainty clearly depends on the value of one test, *i.e.* if we have a very reliable test which always gives a positive result when tested on a known positive, then theoretically one negative would be sufficient, while if a test gave only a few positives, then obviously one would require a very large number of negatives before one would accept a case as really negative.

If P = percentage of carriers shown up as positives by n tests, and if p = percentage of positive results given by the test when applied to known positive cases, then

$$P = 100\left\{1-\left(1-\frac{p}{100}\right)^n\right\}.$$

(see Appendix V). p is a measure of the efficiency of the test and is 100 per cent. for a perfect test and is 0 per cent. for a completely useless one. Usually p is known from experience or by actual testing.

We may restate this in another form which may be more convenient. If we have a method which gives p per cent. positives when tested on known carriers, and if we want to make sure of recognising at least P per cent. of the carriers, then we should make n tests, where

$$n = \frac{\log\left(\frac{100}{100-P}\right)}{\log\left(\frac{100}{100-p}\right)}.$$

(This will sometimes give a fractional answer, so we take the next higher integer.) For instance, if p is 37·05 per cent. and we want to be sure of recognising at least 99 per cent. of the carriers, then

$$\begin{aligned} n &= \frac{\log\left(\frac{100}{100-99}\right)}{\log\left(\frac{100}{100-37\cdot05}\right)} \\ &= \frac{2}{0\cdot2010} \\ &= 9\cdot95, \end{aligned}$$

so 10 swabs are necessary.

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

A short table (Fig. 2) is given showing n for given values of P and p. Looking down the first column, one notices that for fairly high certainty, if p is only moderate, say 30 per cent., one needs many more tests than are usually considered necessary.

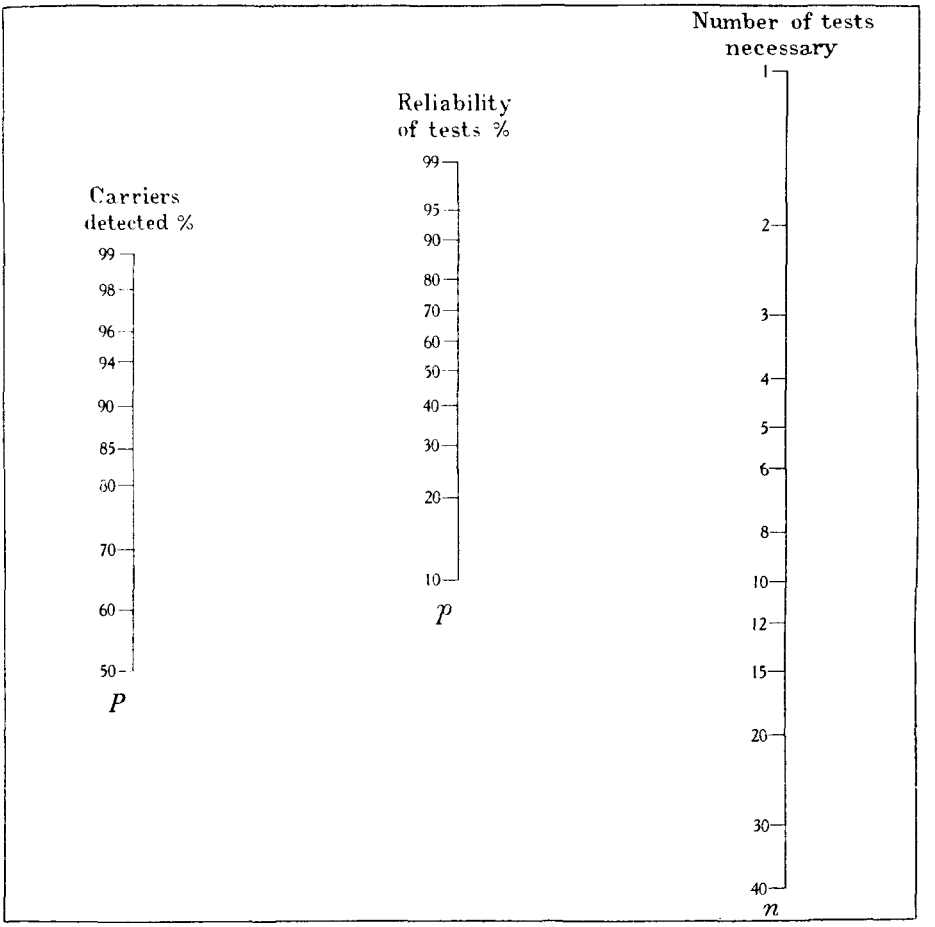

Fig. 2.

Of 160 tests in 33 known carriers, 104 tests were positive and 56 were negative. The swab test was therefore reliable in 65 per cent. (Table IV). From Table III we should therefore require five swabs per person in order to make certain of recognising 99 per cent. of the carriers in a population receiving treatment with normal saline. We eliminated the first positive result in each case, as this defines the carrier, and all swabs from and including the last positive preceding the first of three final negatives. The first positive and the last positive showed that the patients were carriers, and were eliminated as these specimens could not in any circumstances be negative and were thus biased (see Table IV).

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

Table III. *Number of tests required to determine varying percentages (P per cent.) of all cases*

	P %				
p %	99	95	90	80	50
100	1	1	1	1	1
90	2	2	1	1	1
80	3	2	2	1	1
70	4	3	2	2	1
60	6	4	3	2	1
50	7	5	4	3	1
40	9	6	5	4	2
30	13	9	7	5	2
20	21	14	11	8	4
10	44	29	22	16	7
0	∞	∞	∞	∞	∞

Our thanks are due to Dr I. Walker Hall at whose laboratory at the Department of Preventive Medicine, University of Bristol, the swabs were examined and virulence tests performed.

Table IV. *Results of tests in known carriers, i.e. between the first and last positive swabs* (65 *per cent. positive of* 33 *cases*)

Case No.	Positive	Negative	Total
1	5	0	5
2	2	2	4
3	4	4	8
4	7	1	8
5	5	3	8
8	8	4	12
10	1	1	2
12	3	5	8
13	1	0	1
14	2	2	4
16	12	4	16
18	1	0	1
20	2	0	2
22	1	0	1
24	1	0	1
29	1	0	1
30	1	0	1
32	13	6	19
34	1	2	3
38	3	2	5
39	2	3	5
41	2	0	2
45	2	1	3
46	1	2	3
47	1	2	3
49	4	2	6
50	1	2	3
51	1	2	3
53	6	2	8
54	4	0	4
56	2	0	2
65	4	3	7
67	0	1	1
	104	56	160

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

APPENDIX I

If p is the chance that any swab should come up positive, and $q\ (=1-p)$ the chance that it comes up negative, and if a person stops after three consecutive negative swabs, it is required to find the average number of swabs before stopping, assuming p to be constant for all times, for all persons and for all swabs.

If N_n is the proportion of people still left in after the nth swab, we may find N_n in terms of lower values of n thus: Draw up a table showing the possibilities and proportions. After the nth swab, let there be

a people whose record terminates ... +
b ,, ,, ,, ... + −
c ,, ,, ,, ... + − −
d ,, ,, ,, ... + − − −

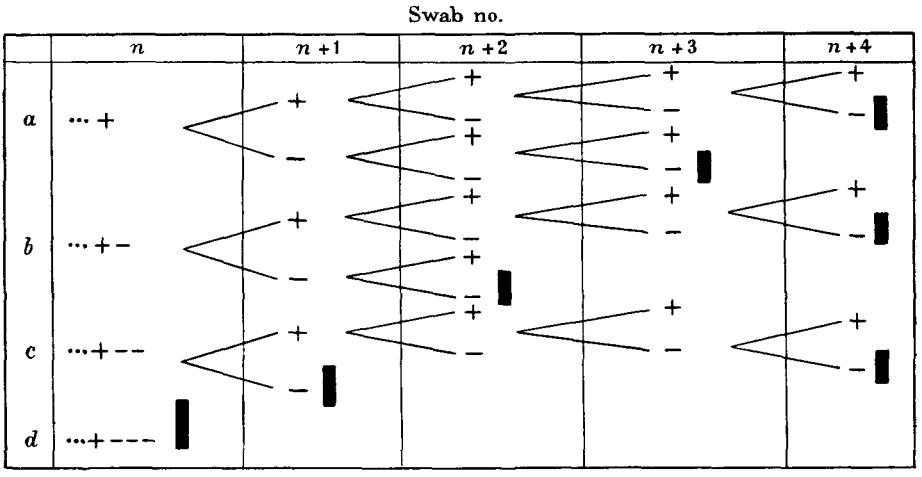

Fig. 3.

These latter (d) will drop out after this swab (having three consecutive negatives), so

$$N_n = a + b + c \qquad \text{.....(1).}$$

Taking four more swabs, we obtain the table of events recorded in Fig. 3.

A heavy line is drawn to indicate that the class has dropped out with three consecutive negatives. Unessential possibilities have been left out for simplicity.

From the table we see that

$$\begin{aligned} N_{n+4} &= N_{n+3} - (pq^3 a + pq^3 b + pq^3 c) \\ &= N_{n+3} - pq^3\,(a+b+c) \\ &= N_{n+3} - pq^3 N_n, \end{aligned}$$

or, rewriting,

$$N_n = N_{n-1} - pq^3 N_{n-4} \qquad \text{.....(2).}$$

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

This gives N in terms of lower values of n.

We now put D_n = proportion of people who drop out after the nth swab. We have $D_n = N_{n-1} - N_n$.

Also
$$N_{n-1} = N_{n-2} - pq^3 N_{n-5}$$
$$N_n = N_{n-1} - pq^3 N_{n-4}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

subtracting
$$D_n = D_{n-1} - pq^3 D_{n-4} \qquad \ldots\ldots(3).$$

Further, by drawing up a table similar to Table I we soon find that
$$D_1 = 0$$
$$D_2 = 0$$
$$D_3 = q^3$$
$$D_4 = pq^3 \qquad \ldots\ldots(4).$$

With these first four values and equation (3) we can now find D_n to any value of n. The actual values of D_n in terms of p and q are not worth giving, as they rapidly become very complicated.

We can now calculate the average run, $= \bar{X}$. We have
$$\bar{X} = \frac{3D_3 + 4D_4 + 5D_5 + 6D_6 + \ldots}{D_3 + D_4 + D_5 + D_6 + \ldots}.$$

Since everyone must drop out ultimately,
$$D_3 + D_4 + D_5 + D_6 + \ldots = 1,$$

so
$$\bar{X} = 3D_3 + 4D_4 + 5D_5 + 6D_6 + \ldots \qquad \ldots\ldots(5).$$

To sum this series, write down successive terms as follows, using equations (3) and (4):
$$D_4 = pq^3$$
$$2D_5 = 2D_4 - 2pq^3 D_1$$
$$3D_6 = 3D_5 - 3pq^3 D_2$$
$$4D_7 = 4D_6 - 4pq^3 D_3$$
$$5D_8 = 5D_7 - 5pq^3 D_4$$
$$6D_9 = 6D_8 - 6pq^3 D_5$$
$$\text{etc.}$$

Add up both sides and strike out common factors of D_n. We get
$$\begin{aligned} 0 &= pq^3 \\ &\quad + D_4 + D_5 + D_6 + D_7 + \ldots \\ &\quad - pq^3 (4D_3 + 5D_4 + 6D_5 + \ldots) \\ &= pq^3 \\ &\quad + 1 - D_3 \\ &\quad - pq^3 (D_3 + D_4 + D_5 + \ldots + 3D_3 + 4D_4 + 5D_5 + \ldots) \\ &= pq^3 + 1 - q^3 - pq^3 (1 + \bar{X}), \end{aligned}$$

whence
$$\bar{X} = \frac{1 + q + q^2}{q^3}.$$

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

More generally, if instead of stopping after three consecutive negatives, we stop after n consecutive negatives, then

$$\bar{X} = \frac{1-q^n}{(1-q)\,q^n}.$$

APPENDIX II

The variance of the distribution is given by the ordinary method as

$$V = 3^2D_3 + 4^2D_4 + 5^2D_5 + 6^2D_6 + \ldots - \bar{X}^2.$$

Setting up the series as in the method for $\bar{x}$, but using multipliers 3^2, 4^2, etc., and adding up, we get an expression with the V series on the right, after a little coaxing, and solving for V and subtracting $\bar{X}^2$, we find

$$V = \frac{(1-q)\,(1+3q+6q^2+3q^3+q^4)}{q^6}.$$

More generally, if we stop after n consecutive negatives,

$$V = \frac{1}{(1-q)^2q^{2n}}\{1-(2n+1)\,q^n+(2n+1)\,q^{n+1}-q^{2n+1}\}.$$

APPENDIX III

The variance of $\bar{x}$ is given at once by $V/62$. To find the variance of $\bar{X}$ is more difficult. We have $\bar{X} = \frac{1+q+q^2}{q^3}$ where $q = \frac{243}{386}$. The variance of 243 is given by the binomial distribution as $386pq = 90{\cdot}02$, so S.D. $(243) = 9{\cdot}488$. To determine the variance of $\bar{X}$ is not easy since it is not a linear function of q, but we may find a reasonable approximation by calculating the 5 per cent. points for 243, converting to $\bar{X}$ and so finding approximately the 5 per cent. points for $\bar{X}$, obtaining four times the S.D. of $\bar{X}$. Thus:

The 5 per cent. points for 243 are 262·0 and 224·0. Assuming these values we find corresponding values of $\bar{X}$ to be 6·843 and 9·807; so

$$\text{S.D.}\,(\bar{X}) = \frac{9{\cdot}807 - 6{\cdot}843}{4} = 0{\cdot}7410,$$

whence the variance of $\bar{X}$ is 0·5491.

We now have

$$\begin{aligned} V\,(\bar{x}-\bar{X}) &= V\,(\bar{x}) + V\,(\bar{X}), \\ &= 0{\cdot}2868 + 0{\cdot}5491 \\ &= 0{\cdot}8360, \end{aligned}$$

so S.D. $(\bar{x}-\bar{X}) = 0{\cdot}914$. Consequently

$$\frac{\bar{x}-\bar{X}}{\text{S.D.}\,(\bar{x}-\bar{X})} = 2{\cdot}07,$$

giving $P = 0{\cdot}05$.

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

Appendix IV

Length of run (swabs)	Number expected (persons)	Number obtained (persons)	Length of run (swabs)	Number expected (persons)	Number obtained (persons)
3	15·47	23	27	0·17	0
4	5·73	6	28	0·14	0
5	5·73	6	29	0·12	0
6	5·73	5	30	0·10	0
7	4·30	6	31	0·09	0
8	3·77	3	32	0·07	0
9	3·24	3	33	0·06	0
10	2·71	1	34	0·05	0
11	2·31	1	35	0·05	0
12	1·97	5	36	0·04	0
13	1·67	0	37	0·03	0
14	1·42	0	38	0·02	0
15	1·20	0	39	0·01	0
16	1·02	1	40	0·01	0
17	0·87	0	41	0·01	0
18	0·74	0	42	0·01	0
19	0·62	0	43	0·01	0
20	0·53	1	44	0·01	0
21	0·45	0	45	0·01	0
22	0·38	0	46	0·01	0
23	0·32	1	47	0·01	0
24	0·28	0	48	0·01	0
25	0·23	0	49–	0·00	0
26	0·20	0		61·97	62

Appendix V

We have $P = 100$ minus the percentage missed. The percentage missed is, by definition, those patients who have returned all n results as negative. As $\frac{p}{100}$ is the chance of getting a positive result, $1-\frac{p}{100}$ is the chance of getting a negative. The chance that all n results shall be negative, purely by chance, is $\left(1-\frac{p}{100}\right)^n$, so the percentage missed is $100\left(1-\frac{p}{100}\right)^n$, whence

$$P = 100\left\{1-\left(1-\frac{p}{100}\right)^n\right\}.$$

Summary

1. A series of 62 nasal carriers of *B. diphtheriae*, virulent and avirulent, has been investigated.

2. Treatment by means of normal saline gargles and sniffing normal saline up both nostrils appeared to reduce the number of positive swabs by two, but the data are not quite decisive on this point.

3. With the above treatment the majority of the patients were carriers for periods of less than 9 days.

4. Two successive negative swabs were followed by positive results in 22 cases.

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press

5. In order to recognise 99 per cent. of the carriers in a population receiving treatment with normal saline, five swabs are required from each person examined.

6. Equations are given by which the expected length of run may be calculated for other conditions.

(*MS. received for publication* 15. II. 1936.—Ed.)

https://doi.org/10.1017/S0022172400043539 Published online by Cambridge University Press