Abbey, Helen, *An Examination of the Reed-Frost Theory of Epidemics* , Human Biology, 24:3 (1952:Sept.) p.201

AN EXAMINATION OF THE REED–FROST THEORY OF EPIDEMICS *

BY HELEN ABBEY
The Johns Hopkins University

INTRODUCTION

THE STUDY of the flow of a disease through a population can be approached by making a set of assumptions about the relations among the factors producing the spread of the disease and expressing these assumptions in terms of a mathematical model. The model may then be tested on actual observations of diseases for which the assumptions are thought to be valid.

This is a useful approach since if the model adequately fits the data it lends support to the underlying assumptions of the model. If certain combinations of assumptions give a better fit than others, the simplest assumptions which give a good fit provide working leads for further study of the actual relationships which have been approximated by the model. The estimates of the parameters of the model may be useful in comparing different diseases or the same disease under different environmental conditions.

The present investigation is an application of a specific model (the Reed-Frost model) to observations of certain acute infectious diseases where the assumptions are most likely to be valid. Although there is considerable discussion in the literature of epidemic models, very little testing of the models on actual observations has been done, partly because the necessary data are difficult or impossible to obtain.

Lowell J. Reed and Wade Hampton Frost, in unpublished work used in class lectures at Johns Hopkins University, developed a model which was a modification of one orginally proposed by Soper (1927). Soper

* From the Department of Biostatistics, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md. Department Paper no. 285. This work was done under Navy Contract N6onr-24314.

had postulated a community in which all individuals had equal susceptibility to a disease, equal capacity to transmit it, and the power of passing out of observation when the transmitting period was over. He considered diseases in which the period of infectiousness is short relative to the incubation period. The "law of mass action," which states that the rate of a chemical reaction is proportional to the product of the active masses of the substances, was assumed to apply to the transmission of the disease. Under these conditions, if the time interval is chosen to be the average incubation period of the disease, the number of cases at any time period is proportional to the product of cases and susceptibles in the previous period.

Reed and Frost modified Soper's model to make allowances for the fact that contact between a given susceptible and two or more cases will produce only one new case, an effect which Soper's theory does not consider. Their model is based on the following assumptions:

> The infection is spread directly from infected individuals to others by a certain kind of contact (adequate contact) and in no other way.
>
> Any non-immune individual in the group, after such contact with an infectious person in a given period, will develop the infection and will be infectious to others only within the following time period, after which he is wholly immune.
>
> Each individual has a fixed probability of coming into adequate contact with any other specified individual in the group within one time interval, and this probability is the same for every member of the group.
>
> The individuals are wholly segregated from others outside the group.
>
> These conditions remain constant during the epidemic.

If p is the probability of contact between any two specified individuals in the population in a given interval of time (the period of infectiousness), then $q = 1 - p$ is the probability of their not having contact. Contact, or adequate contact, as used by Reed and Frost, is contact such that, if it occurs between an infectious case and a susceptible, it will produce a new case. The probability of contact, in this sense, depends on the susceptibility or resistance of the host, the infectivity of the parasite, the length of exposure and size of dose necessary to

produce the disease, as well as the environmental conditions necessary for the transfer of the organism.

If C_t is the number of cases produced at time t, then q^{C_t} is the probability that the specified individual will not have contact with any of the C_t cases, and $1 - q^{C_t}$ is the probability that he will have contact with at least one of them. Reed and Frost assumed as Soper did that the infective period is short relative to the incubation period, and that

TABLE 1

Calculation of a theoretical epidemic from the Reed-Frost model

($p = .05$)

TIME PERIOD t	NUMBER OF CASES C_t	NUMBER OF SUSCEPTIBLES S_t	CALCULATION OF C_{t+1} AND S_{t+1}
0	1	100	$C_1 = 100(1 - .95) = 5.00 = 5$
			$S_1 = 100 - 5 = 95$
1	5	95	$C_2 = 95(1 - .95^5) = 21.49 = 21$
			$S_2 = 95 - 21 = 74$
2	21	74	$C_3 = 74(1 - .95^{21}) = 48.80 = 49$
			$S_3 = 74 - 49 = 25$
3	49	25	$C_4 = 25(1 - .95^{49}) = 22.97 = 23$
			$S_4 = 25 - 23 = 2$
4	23	2	$C_5 = 2(1 - .95^{23}) = 1.39 = 1$
			$S_5 = 2 - 1 = 1$
5	1	1	$C_6 = 1(1 - .95^1) = .05 = 0$
			$S_6 = 1 - 0 = 1$
6	0	1	

the time interval is the average length of the incubation period. It follows that if there are S_t susceptible individuals in the population at time t, the expected number of cases produced at the time $t + 1$ is S_t times the probability of contact with at least one case. Or,

$$C_{t+1} = S_t(1 - q^{C_t}). \tag{1}$$

This equation provides a method of stepwise calculation of cases at successive time periods. For example, suppose that the contact rate in a population of 100 susceptibles is .05. If one case is introduced into this population, the epidemic produced by the assumptions of the model will proceed as is shown in Table 1. The first case will produce 5, these five will produce 21, etc., until the epidemic ends with one susceptible remaining in the sixth time period and no new cases being produced.

Soper's equation may be regarded as a first approximation to the Reed-Frost theory. For small numbers of cases relative to the numbers of susceptibles, the two methods are essentially equivalent.

In the models of Soper and of Reed and Frost as presented above, the whole course of the epidemic is determined if the initial conditions are known, and may be calculated step by step as previously indicated. Reed and Frost, however, considered that an epidemic is not uniquely determined by the initial conditions because at each period there are variations due to chance. They used a mechanical device to illustrate the effect of chance variation on the theoretical epidemics.

The concept of chance variation can be introduced into the mathematical model by replacing the equation

$$C_{t+1} = S_t(1 - q^{C_t})$$

by the statement that the probability of C_{t+1} cases in the $(t + 1)$-th interval is

$$P(C_{t+1}) = \frac{S_t!}{C_{t+1}!\, S_{t+1}!}(1 - q^{C_t})^{C_{t+1}}(q^{C_t})^{S_{t+1}} \qquad (2)$$

where S_t and C_t are the observed numbers of susceptibles and cases, respectively, in the t-th interval. Note that this is an ordinary binomial probability, where $1 - q^{C_t}$ is the probability of a susceptible becoming a case in the $(t + 1)$-th interval.

Epidemics can be calculated stepwise from this model, with the aid of random numbers and a table of the cumulative binomial distribution (National Bureau of Standards, 1949). For example, consider an epidemic in a population having a contact rate p, with S_0 susceptibles and C_0 cases in the initial time period. From equation (1), the number of cases *expected* in period one is

$$E(C_1) = S_0(1 - q^{C_0}).$$

The expected number of cases is written $E(C_1)$ to differentiate it from $O(C_1)$, the number actually produced.

The observed number of cases $O(C_1)$ is obtained by drawing it at random from the binomial distribution having a mean $E(C_1)$ and a probability $1 - q^{C_0}$ of becoming a case in period one. This is done by going to the binomial table of partial sums

$$\sum_{s=r}^{n} \binom{n}{s} P^s Q^{n-s}.$$

For a specified value of P and of n, there is a partial sum corresponding to each value of r between $r = 1$ and $r = n - 1$. You choose a partial sum at random, thereby choosing an r at random. In the first time period,

$$n = S_0$$

$$P = 1 - q^{C_0}.$$

Now, draw a random number (of seven digits since the sums are seven-digit numbers). Locate the sum closest to it, for the n and P given above, and read off the corresponding r. This is the observed number of cases $O(C_1)$ in the first time period. The observed number of susceptibles is

$$O(S_1) = S_0 - O(C_1).$$

Proceeding to the next time period, an observed $O(C_2)$ is drawn at random from the binomial distribution having a mean $E(C_1)$ and a probability $1 - q^{O(C_1)}$ of becoming a case; that is, from the sum

$$\sum_{s=r}^{n} \binom{n}{s} P^s Q^{n-s}$$

where

$$n = O(S_1)$$

$$P = 1 - q^{O(C_1)}.$$

The r corresponding to this n and P and to a partial sum chosen at random, is the observed number of cases $O(C_2)$. The observed number of susceptibles is obtained by subtraction:

$$O(S_2) = O(S_1) - O(C_2).$$

The calculations continue in this manner until no new cases are produced. For values beyond the range of the binomial tables, the Poisson or normal distributions may be used as approximations to the binomial.

In the sections which follow, a chi-square test is used to determine whether the theory adequately fits the observed epidemics in human populations. This is a test of whether the variation found to occur in actual epidemics is greater than the binomial variation described above. The test is described in a later section on the application of the model.

SOURCES OF DATA

Although there is a great deal of published data on the reported cases of infectious diseases, most of this is not useful in testing the adequacy of the model because of variable amounts of under-reporting of cases, and of lack of information about the number of susceptibles.

The requirement of a closed population with uniform mixing among its members is most nearly met in institutions or within families. In these groups also, the size of the population is known, and estimates of the number of susceptibles can usually be obtained. Likewise, the diseases on which the model should be tested first are those which come closest to fulfilling the assumptions of the model. Such diseases have the following characteristics:

The period of infectiousness is short relative to the incubation period.

There is uniform susceptibilty to the disease before the attack.

A single (adequate) contact between an infectious case and a susceptible produces the disease.

A single attack of the disease produces lasting immunity.

There are few carriers, or subclinical or missed cases.

The incubation period is constant.

The diseases which approximately satisfy these conditions are the acute infectious diseases of childhood, with measles most nearly fulfilling these conditions, and German measles and chickenpox also being fairly satisfactory. The others, such as scarlet fever, poliomyelitis, diphtheria, mumps and whooping cough not only do not fit these criteria as well, but also are of limited usefulness because they usually produce too few cases in any epidemic period to provide much information on which to test the theory.

The majority of the epidemics used in this study were obtained from the Medical Research Council Special Reports, Numbers 227 and 271, "Epidemics in Schools." These are reports of a study of the incidence of epidemic diseases in Naval and boarding schools in England during the years 1932 to 1939. Table 2 gives some the characteristics of these populations. For a more complete description of the populations, reference may be made to the original reports. The code symbols used to identify the epidemics in the present paper are the same as those

TABLE 2

Characteristics of the population

DESIGNATION	DISEASE	TYPE OF POPULATION	SEX	SIZE OF POPULATION	AGE RANGE*
MB 1932	Chickenpox	English boarding school	M	583	12—18
KB 1933	Measles	"	M	745	13—18
FB 1934	Measles	"	M	515	13—18
QB 1934	Measles and German Measles	"	M	503	13—18
BB 1934	German Measles	"	M	664	13—18 (7.5% under 13)
EB 1934	German Measles	"	M	264	12—18
CB 1934	Chickenpox	"	M	814	10—17
CB 1936	Measles	"	M	830	10—17†
EB 1937	Chickenpox	"	M	273	11—18†
GB 1938	Measles	"	M	245	13—18†
TB 1938	Measles	English Naval school	M	494	13—17†
TB 1939	German Measles	"	M	500	13—17†
RB 1939	Chickenpox	English boarding school	M	657	13—18†
FG 1939	German Measles	"	F	223	11—17†
Aycock	Measles	New England boarding school	M	400	13—18
Mexico	Chickenpox	N. Y. State public school	M and F	548	Grades Kg.—6
Inst. #12 Total	Measles	N. Y. State institution	M and F	258	Under 1—13% 1–4—53% 5–9—29% 10–14— 5%
Inst. #12 Ward C	Measles	N. Y. State institution	M and F	100	‡
Tenements	Measles	7 tenement buildings on a short cul-de-sac	M and F	173§	Under 1—15% 1–4—42% 5–9—43%

* Includes 95% or more of the population, unless otherwise stated.

† Summer term, 1939.

‡ Not available.

§ Population under 10 years of age.

used in the original reports. The reported number of susceptibles is not stated in these two studies, but has been obtained by dividing the total number of cases by the attack rate on susceptibles. The method by which the prior history of disease was obtained could not be ascertained from the reports. It is likely that there is considerable variation in the method among the schools, depending on the routines which were in operation at the beginning of the study.

In addition, single epidemics have been obtained from each of the following sources:

The measles epidemic labelled "Aycock" is taken from "Immunity to Poliomyelitis" by W. Lloyd Aycock, in the *American Journal of Medical Sciences,* September, 1942. It occurred in a New England boys' boarding school in 1934.

The Mexico school chickenpox data are from a study done by Bahlke, Silverman, and Ingraham on ultra-violet light irradiation which is reported in the *American Journal of Public Health,* October, 1949. This was an unirradiated school used as a control. The epidemic occurred in the winter of 1945-1946.

The two epidemics labelled Institution #12, are from "Measles in Institutions for Children, Part 2, Use of Convalescent Serum" by Edward S. Godfrey, Jr., in the *Journal of Preventive Medicine,* January, 1928. In 1923, an epidemic occurred in an institution for normal children in New York State. Ward C, the largest ward which was attacked by the disease, has been separated from the total figures for the institution for comparative purposes. Although convalescent serum was used, the author felt that it had been ineffective in altering the progress of the epidemic or the severity of the disease. No report of the number of susceptibles is given in the article.

The tenement epidemic is reported in the Medical Research Council Special Report No. 120, "An Inquiry into the Relationship between Housing Conditions and the Incidence and Fatality of Measles." The epidemic occurred in the winter of 1925-1926, in seven tenement buildings having a total population of 538, of whom 173 were children under 10 years of age. The buildings were situated on a short cul-de-sac which was used as a playground by the children.

It should be noted from Table 2 that the populations from which the epidemics came differ both in age distribution and in the environmental factors which influence the contact rate. These differences must be considered in making comparisons of the results obtained from fitting the model to the epidemics.

The final set of data (not shown in Table 2) is from "Measles and Scarlet Fever in Providence, R. I. . . ." by Wilson, Bennett, Allen, and Worcester. The figures used are measles cases in Providence, R. I., in 1923, in families of four or more children having one primary case and three susceptibles. They have classified the families into groups which are not mutually exclusive so that some families occur in all the groups and others in only a few or only one. The definitions of the three groups chosen for study here, are given in a later section.

METHOD OF APPLYING THE MODEL

Each epidemic was divided into intervals of the length of the incubation period of the disease in such a way that, in so far as possible, there was a clustering of cases in the middle of each interval. The Reed-Frost model was then applied to the total counts of cases and susceptibles in each interval.

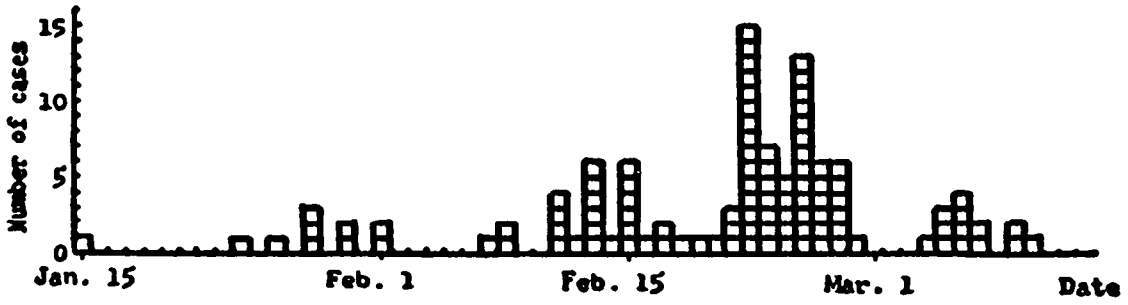

FIG. 1. NUMBER OF CASES OF MEASLES (AYCOCK EPIDEMIC) BY DAY OF OCCURRENCE

For example, Figure 1 shows each case in the Aycock epidemic by day of occurrence. If the incubation period is taken as twelve days and the first case occurs on the fifth day of the interval, there is the following distribution of cases in successive twelve-day periods:

TIME	REPORTED NUMBER OF CASES	REPORTED NUMBER OF SUSCEPTIBLES
0	1	117
1	9	108
2	22	86
3	61	25
4	13	12
5	0	12

An estimate of the contact rate can be obtained from each time period by substituting these cases and susceptibles in the Reed-Frost model and solving the resulting equations for q. Thus from the first interval, if the number of susceptibles at time 0 is 117, we have

$$9 = 117(1 - q_1^1)$$

or $$q_1 = 0.9231$$

and $$p_1 = 0.0769.$$

In the second time period

$$22 = 108(1 - q_2^9)$$

$$q_2^9 = 1 - \frac{22}{108} = 0.7963$$

$$q_2 = 0.9750$$

$$p_2 = 0.0250.$$

Similarly, $p_3 = 0.0546$ and $p_4 = 0.0120$.

The method which is used for estimating a single contact rate for the entire epidemic is the "method of maximum likelihood." This is a standard estimation procedure, and is equivalent to a weighted average of estimates calculated at each point. The maximum likelihood method consists of writing down the "likelihood" of obtaining the observed epidemic:

$$\text{Likelihood} = L = \prod_{t=0}^{n-1} \frac{S_t!}{C_{t+1}!\, S_{t+1}!} (1 - q^{C_t})^{C_{t+1}} (q^{C_t})^{S_{t+1}}, \qquad (3)$$

where S and C are the observed numbers of susceptibles and cases, and q is the parameter to be estimated; and in finding the value of q which makes L a maximum. This value of q is calculated by setting the derivative of $\log L$ with respect to q equal to zero, and solving the resulting equation for q:

$$\sum_{t=0}^{n-1} \frac{C_t C_{t+1} q^{C_t}}{q(1 - q^{C_t})} - \sum_{t=0}^{n-1} \frac{C_t S_{t+1}}{q} = 0.$$

There is no explicit solution, but the equation may be solved by successive approximations.

To test whether the observations fit the mathematical model, the estimated contact rate is substituted in the Reed-Frost equation to determine expected numbers of cases and susceptibles at each time period:

$$E(C_{t+1}) = O(S_t)\,(1 - q^{O(C_t)}),$$

$$E(S_{t+1}) = O(S_t) - E(C_{t+1}).$$

These expected numbers are compared with the observed numbers by a chi-square test of goodness of fit:

$$\chi^2 = \sum_{t=1}^{n-1} \frac{[O(C_t) - E(C_t)]^2}{E(C_t)} + \sum_{t=1}^{n-1} \frac{[O(S_t) - E(S_t)]^2}{E(S_t)}. \qquad (5)$$

The chi-square test is applicable here because the expected numbers in each time period are calculated from the observed numbers in the previous period, and are therefore independent of the previous expected numbers. The test measures whether the variation at each point is of the magnitude of the binomial variation discussed on page 204.

RESULTS OF THE APPLICATION OF THE MODEL TO SCHOOL EPIDEMICS

Tables 3 and 4 and Figure 2 show a comparison of observed and expected numbers of cases. In general the discrepancies are as follows: (1) There are too few expected cases in the early time periods and too many in the later ones. (2) The curve of the expected cases reaches its peak one period later than the curve of the observations. (3) The discrepancies between the observed and expected numbers, as measured by the chi-square test, are very large.

These differences may be due to errors of observation, either because of actual errors in counting susceptibles or cases, or because of a poor choice of the intervals into which the observations are grouped. They may also be due to a failure of the assumptions of the model to represent the actual processes involved in the spread of the disease. These possible causes are examined in some detail in the following sections.

ERRORS IN COUNTING SUSCEPTIBLES

Counting errors are more likely to occur in the susceptibles than in the cases, since the cases were counted as they occurred, while the susceptibles, being defined as persons not having a history of the disease, were counted from events remembered.

The number of susceptibles was therefore assumed to be unknown, and estimates were made both of the initial number of susceptibles and the contact rate, using only the observed number of cases in each time period. These estimates (shown in Tables 3 and 4) were obtained by a

maximum likelihood method similar to that used in estimating the contact rate from observed cases and known numbers of susceptibles.

From these estimates, expected cases were calculated for each time

TABLE 3

Comparison of number of susceptibles reported, with number estimated from the Reed-Frost theory

EPIDEMIC	REPORTED NUMBER OF CASES	INITIAL NUMBER OF SUSCEPTIBLES		RATIO: ESTIMATED TO REPORTED SUSCEPTIBLES	NUMBER OF SUSCEPTIBLES REMAINING AT END OF EPIDEMIC	
		Reported	Estimated		Reported	Estimated
MEASLES						
KB 1933	111	204	112	0.55	93	1
FB 1934	78	109	78	0.72	31	0
QB 1934	103	135	103	0.76	32	0
CB 1936	110	170	110	0.71	60	0
TB 1938	88	123	88	0.72	35	0
GB 1938	62	80	62	0.78	18	0
Aycock	106	118	106	0.90	12	0
Tenements	35	88	41	0.47	53	6
Inst. Total	184	258*	184	≥ 0.71	≤ 74	0
Inst. Ward C	79	100*	79	≥ 0.79	≤ 21	0
GERMAN MEASLES						
BB 1934	258	533	260	0.49	275	2
EB 1934	141	247	141	0.57	106	0
QB 1934	271	427	271	0.63	156	0
TB 1939	72	480	74	0.15	408	2
FG 1939	52	216	52	0.24	164	0
CHICKENPOX						
MB 1932	30	166	30	0.18	136	0
CB 1934	40	236	40	0.17	196	0
EB 1937	34	100	34	0.34	66	0
RB 1939	90	194	90	0.46	104	0
Mexico School	75	188	94	0.50	113	19

* Total population. Susceptibility status not reported.

period. They are shown in Figure 3 in comparison with the observed cases. There is no longer any bias evident in the graphs nor in the runs of signs of observed minus expected cases. The chi-square values have been greatly reduced, and although they are still significantly large, this is of less concern since the bias has been removed. As the contact

TABLE 4

Comparison of goodness of fit of the Reed-Frost model to reported cases and reported susceptibles, and to reported cases and estimated susceptibles

EPIDEMIC	FIT OF MODEL TO REPORTED CASES AND REPORTED SUSCEPTIBLES				FIT OF MODEL TO REPORTED CASES AND ESTIMATED SUSCEPTIBLES			
	χ^2	d.f.	$P(\chi^2)$	Signs*	χ^2	d.f.	$P(\chi^2)$	Signs*
MEASLES								
KB 1933	77.3	5	<0.00001	++−+−	10.7	3	0.01	++−++
FB 1934	139.5	3	<0.00001	++−−	0.0	1	0.99	−+0
QB 1934	59.1	7	<0.00001	−−+++−−−	13.0	5	0.02	−−+++−+
CB 1936	450.0	3	<0.00001	+−−−	16.0	1	0.00006	+−+
TB 1938	30.7	2	<0.00001	+−−	0.0	0		
GB 1938	111.8	3	<0.00001	++−−	16.1	1	0.00006	−++
Aycock	53.1	4	<0.00001	+−+−−	18.4	2	0.0001	−−++
Tenements	13.4	9	0.14	++−+−+−++−	16.0	8	0.04	−+−+−+−++−
Inst. Total					13.3	2	0.001	−+−0
Inst. Ward C					0.1	1	0.70	+−0
GERMAN MEASLES								
BB 1934	189.4	5	<0.00001	+++−−−	9.1	4	0.06	+−−+−−
EB 1934	240.5	4	<0.00001	+++−−−	4.6	2	0.10	−+−+
QB 1934	367.2	2	<0.00001	+−−	0.0	0		
TB 1939	49.1	6	<0.00001	++++−−−	11.7	5	0.04	+−+−+−−
FG 1939	222.7	3	<0.00001	++−−	33.3	1	<0.00001	+−+
CHICKENPOX								
MB 1932	74.2	4	<0.00001	+−−−−	21.6	2	0.00003	+−++
CB 1934	43.8	4	<0.00001	+++−−	5.4	2	0.07	−+−+
EB 1937	88.6	4	<0.00001	+−−+−	41.6	2	<0.00001	+−−+
RB 1939	61.7	5	<0.00001	+++−−−	3.7	3	0.30	++−++
Mexico School	31.0	11	0.001	−+−+++−+−++−	28.4	10	0.002	−+−+++−+−++−

* Sign of observed cases minus expected cases in each generation.

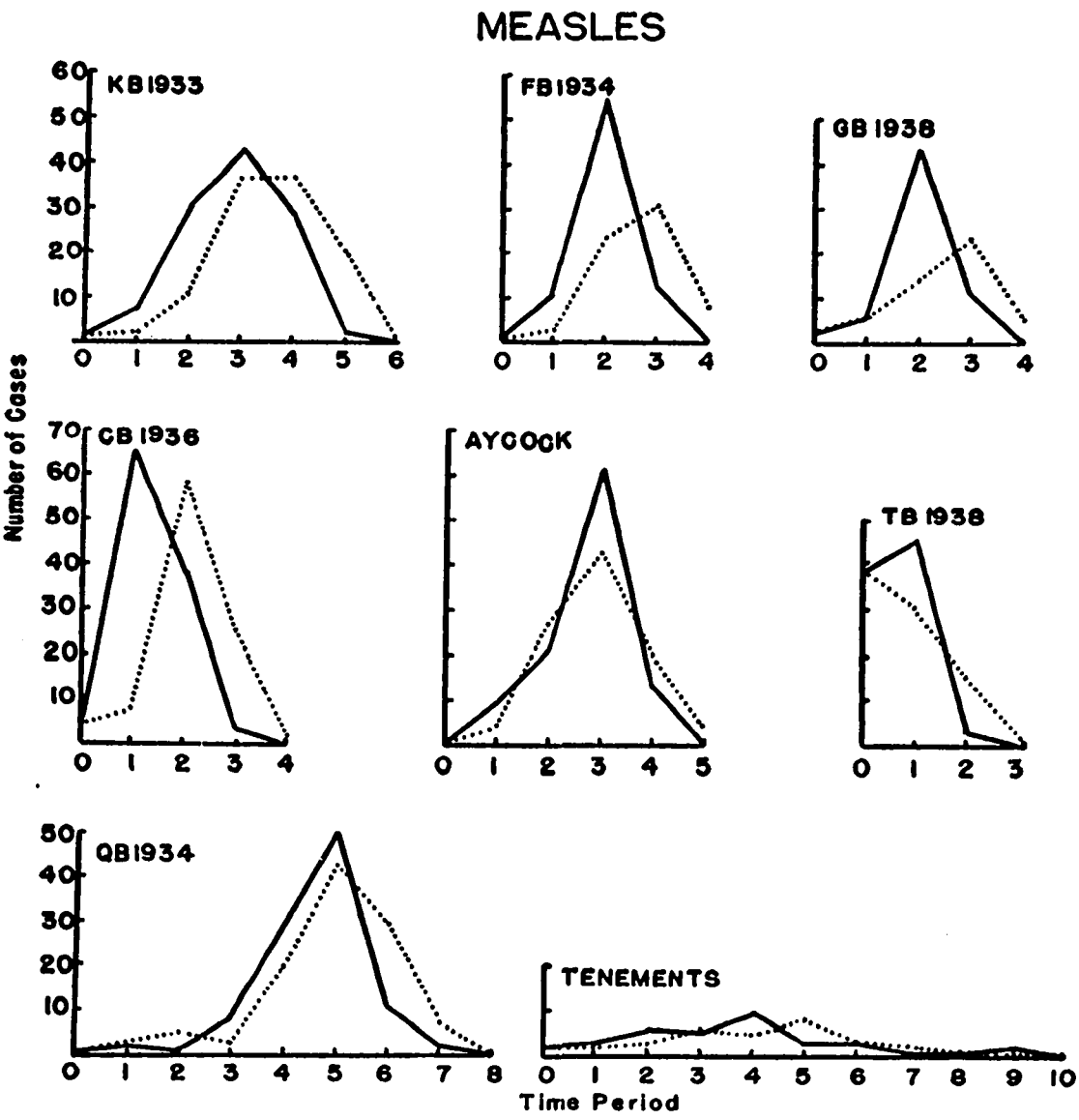

FIG. 2a. COMPARISON OF OBSERVED CASES OF MEASLES WITH CASES CALCULATED FROM REPORTED SUSCEPTIBLES AND ESTIMATED CONTACT RATE

(Solid lines indicate observed cases.)

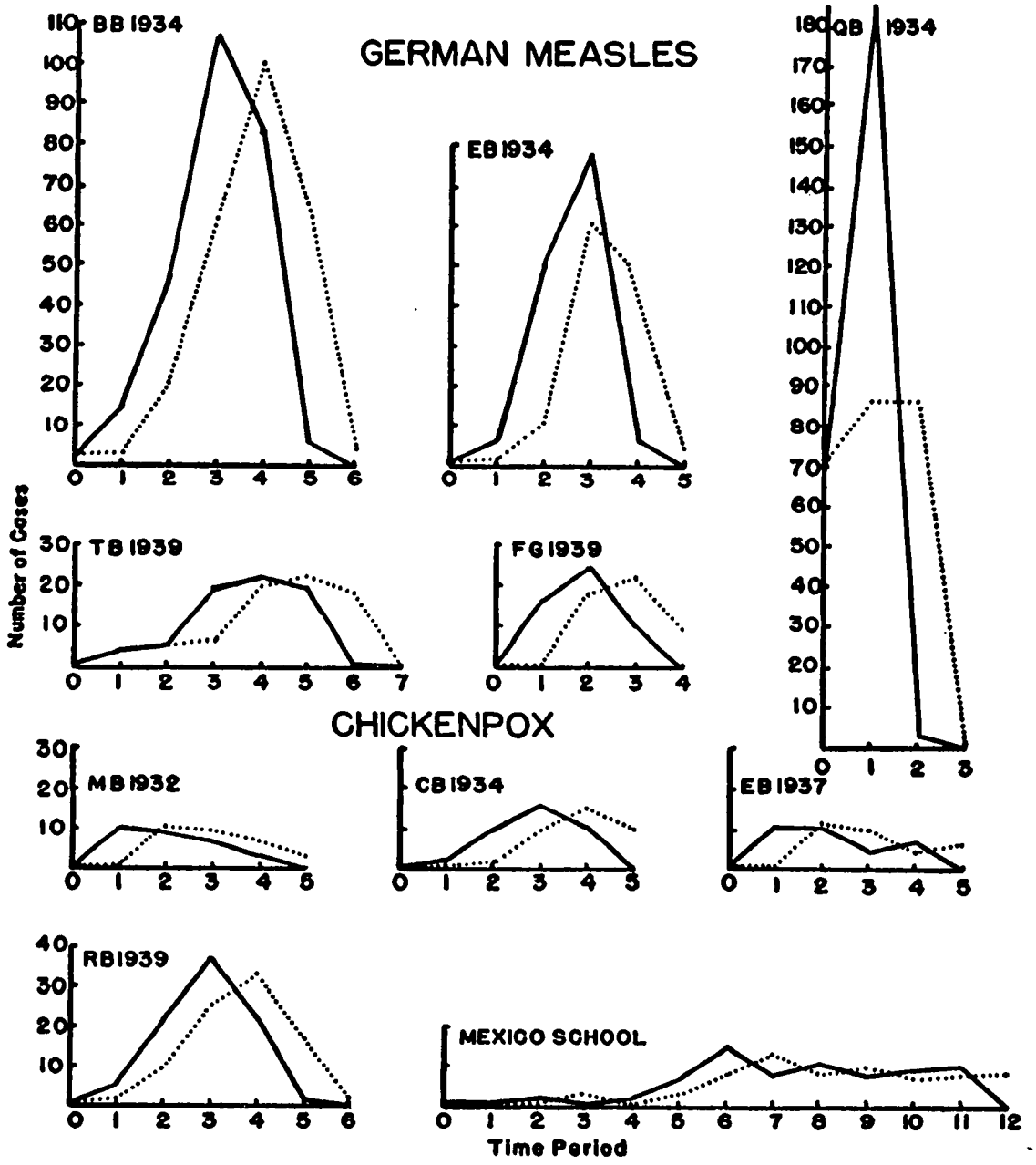

FIG. 2b. COMPARISON OF OBSERVED CASES OF GERMAN MEASLES AND CHICKENPOX WITH CASES CALCULATED FROM REPORTED CASES AND ESTIMATED CONTACT RATE (Solid lines indicate observed cases.)

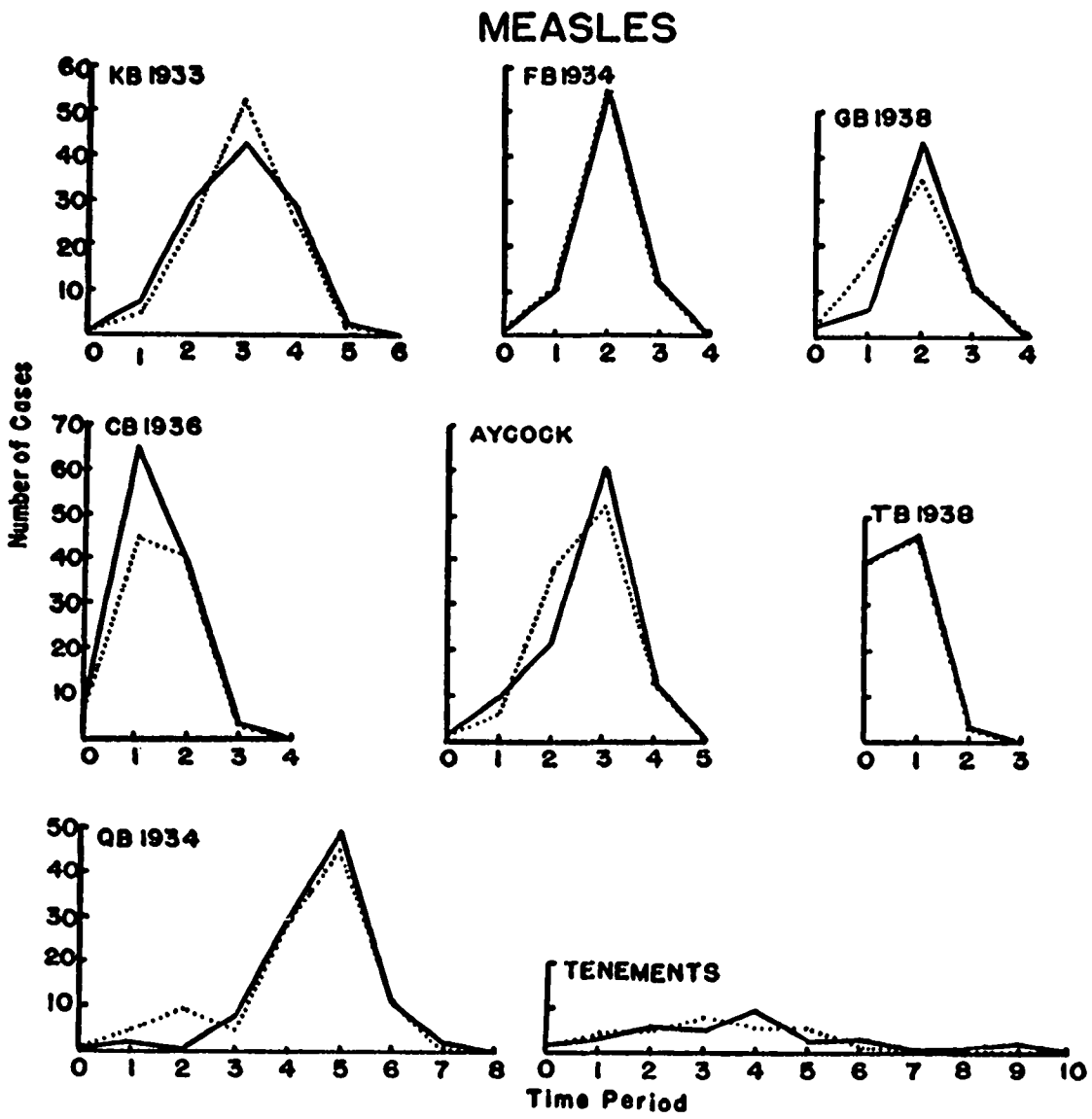

FIG. 3a. COMPARISON OF OBSERVED CASES OF MEASLES WITH CASES CALCULATED FROM ESTIMATED INITIAL SUSCEPTIBLES AND ESTIMATED CONTACT RATE

(Solid lines indicate observed cases.)

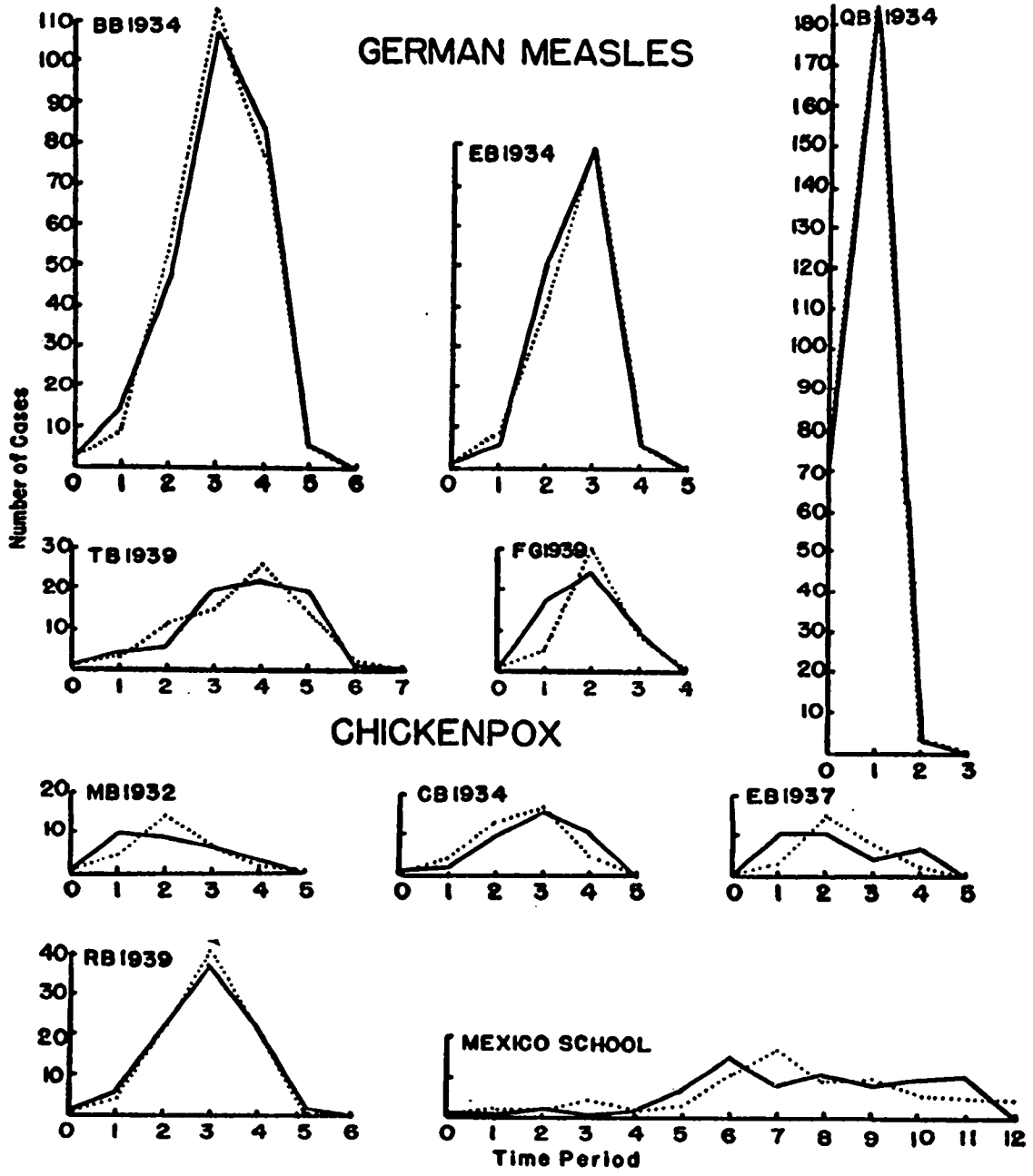

FIG. 3b. COMPARISON OF OBSERVED CASES OF GERMAN MEASLES AND CHICKENPOX WITH CASES CALCULATED FROM ESTIMATED INITIAL SUSCEPTIBLES AND ESTIMATED CONTACT RATE (Solid lines indicate observed cases.)

rate has been defined, it includes all the factors acting to produce the spread of the disease such as the susceptibility or resistance of the host, the length of exposure and size of dose necessary to produce the disease, and the number and kinds of social contacts within the population. The net effect of all of these factors is assumed to be the same for every individual in the population. However, these are human populations made up of individuals who are not precisely alike, but vary in their biological and social characteristics. This variation will contribute to the total variation measured in the epidemic, tending to make it larger than that expected from random effects alone.

A comparison of columns 2 and 4 of Table 3 shows that the two epidemics which were not in institutions, i. e., Tenements and Mexico School, had 14.6 per cent and 20.2 per cent, respectively, of the *estimated* susceptibles remaining at the end of the epidemics. Of the institution epidemics, 14 had no *estimated* susceptibles remaining, one had one, and two had two. This implies that in the institution type of population, there is essentially complete exhaustion of the susceptible population. In these populations, the estimated susceptibles average 70 per cent of the reported susceptibles for the measles epidemics, 42 per cent for German measles, and 28 per cent for chickenpox. Since there is no method for testing the individuals in a population to see if they are actually susceptible to these diseases, we can only look at the estimates to see if they are reasonable in comparison with other available information.

If the school epidemics do, in fact, exhaust the susceptible population, then the reported susceptibles who did not acquire the disease must be assumed not to be susceptible in spite of having no previous history of the disease. Table 5 shows that in the measles epidemics about 7 per cent of the population had no history of the disease at the end of the epidemics. In comparison, a survery of large cities in the United States by Collins (1942) showed that by age 19, 93 per cent of the population report prior history of measles. Since very few persons, presumably less than one per cent, acquire the disease beyond age 19, most of the remaining 7 per cent may be regarded as not susceptible in spite of having no history of the disease. Whether they are not susceptible because of natural immunity or immunity acquired through a forgotten or unrecognized case or by repeated contact with the disease does not matter for the purpose of the present paper.

Similar comparisons for chickenpox and German measles do not show

as good agreement. Collins gives 32 per cent of persons 19 years of age having no history of chickenpox, against 21 per cent of the school populations having no history of the disease at the end of the epidemic; and he gives 69 per cent having no history of German measles against 46 per cent for the school populations. If the overcounting of susceptibles is due to the failure to recognize cases of a disease or to remember its

TABLE 5

Per cent of school populations having no history of the disease at the end of the epidemic

EPIDEMIC	POPULATION	% OF POPULATION HAVING NO HISTORY OF THE DISEASE AT END OF EPIDEMIC
MEASLES		
KB 1933	745	11.3
FB 1934	515	6.6
QB 1934	503	5.4
CB 1936	830	6.6
TB 1936	494	6.5
GB 1938	245	6.9
Aycock	400	3.0
Total	3,732	7.0
GERMAN MEASLES		
BB 1934	664	42.0
EB 1934	264	31.8
QB 1934	503	27.4
TB 1939	500	68.0
FG 1939	223	63.7
Total	2,154	45.6
CHICKENPOX		
MB 1932	583	23.5
CB 1934	814	24.1
EB 1937	273	22.7
RB 1939	657	16.8
Total	2,327	21.4

occurrence, this failure is likely to be of more importance in Collins' data, which is purely historical, than in the school epidemics where, during the period of the study, the cases were recorded as they occurred. This effect would lead to a larger number of reported susceptibles who were actually immune, in the purely historical data. This is in accordance with the differences found between Collins' data and the school epidemics.

CHOICE OF INTERVAL

In order to test the theory, the observations should be grouped into intervals so that the cases in any interval are those which are produced by contact with cases of the previous interval. Since the actual generations of cases were not known in the school epidemics, they were estimated by dividing the observations into intervals of the length of the average incubation period of the disease, centering successive clusters of cases in successive intervals. This method may introduce errors both because the length of the incubation period is not precisely known and because the choice of the center of the interval is a matter of judgment.

For example, in discussing the method of applying the model to the Aycock epidemic of measles, a twelve-day incubation period was used, with the first case occurring on the fifth day of the interval. The model was also fitted to the cases per generation obtained by assuming the first case occurred on the 1st, 2nd, etc., up to the 12th day of the first interval, to determine which of these twelve divisions of the cases into generations fitted the theory best. All possible thirteen and fourteen day intervals were also tried for the Aycock epidemic, and a range of incubation periods and starting points for five of the other epidemics.

In general, the length or position of the interval does not affect the estimate of the initial number of susceptibles. There is variation in the contact rate by interval, but it is less than the variation between epidemics. There is also wide variation in the goodness of fit depending on the interval and endpoints chosen. A study of 12-, 13- and 14-day intervals for measles shows that the range of chi-square values for different endpoints for a given length of incubation period is too large to conclude that one length is substantially better than another. For example, in the Aycock epidemic, the 12-day intervals have chi-square values ranging from 6 to 27, the 13-day intervals from 3 to 48, and the 14-day intervals from 0.7 to 30. There is therefore no evidence from this comparison that any appreciable part of the discrepancy between theory and observations is due to having used a 12- rather than a 13- or 14-day incubation period.

Although the chi-square values vary widely depending on the position of the endpoints of the intervals, the smallest values are not particularly associated with the intervals which would be chosen by inspection of the data as representing generations of the disease. That is, the intervals centering successive clusters of cases in the center of successive intervals

are not consistently better than those centering the clusters at the end of the intervals. Of the intervals which are consistent with the assumed generation process of the disease, the chi-square values are generally large, and if the investigation is limited to these intervals, the divergence of the theory from the observations is not due to the particular choice among them.

VARIATION OF THE CONTACT RATE WITH TIME

If the reported number of susceptibles is used to estimate contact rates at each time period, these rates in general decrease with time (See Fig. 4). Because contact rate as used in this model is affected by the susceptibility of the host, the infectivity of the organism, and the social conditions in the community, a decline in rate may be accounted for by various hypotheses including progressive immunity acquired by repeated contact with the disease, decreasing virulence of the organism, changing environmental factors, or changing relations among hosts. Some of these factors would cause the rate to decline continuously, while others, as for example, isolation procedures, would cause a sharp decline in the rate at the point where these procedures were introduced.

To investigate the hypothesis of a continuous decline, an inspection of the graphs (Fig. 4) suggests that an exponential function may serve as a simple descriptive curve of the rates. Exponential curves were fitted to the contact rates for each epidemic and expected cases calculated from the variable rates so determined. For the group of epidemics as a whole, this procedure was not as effective in improving the fit as estimation of the initial number of susceptibles, although the possibility that there is a decline in the contact rate with time cannot be ruled out.

If the decline is largely due to isolation procedures introduced after the first case is discovered, as the sharp drop in contact rate between the first and second period suggests, a better fit of the model should be obtained by estimating the constant contact rate exclusive of the first interval. When these estimates (and the reported number of susceptibles) are used to calculate expected numbers of cases, the fit of the model, as measured by chi-square, is improved in the three epidemics where the decline in rate between the first and second periods is greatest (Compare Table 4 with Table 6). In all of the other epidemics, however, omitting the first interval is not nearly as effective as estimating the initial number of susceptibles. It therefore appears that the change

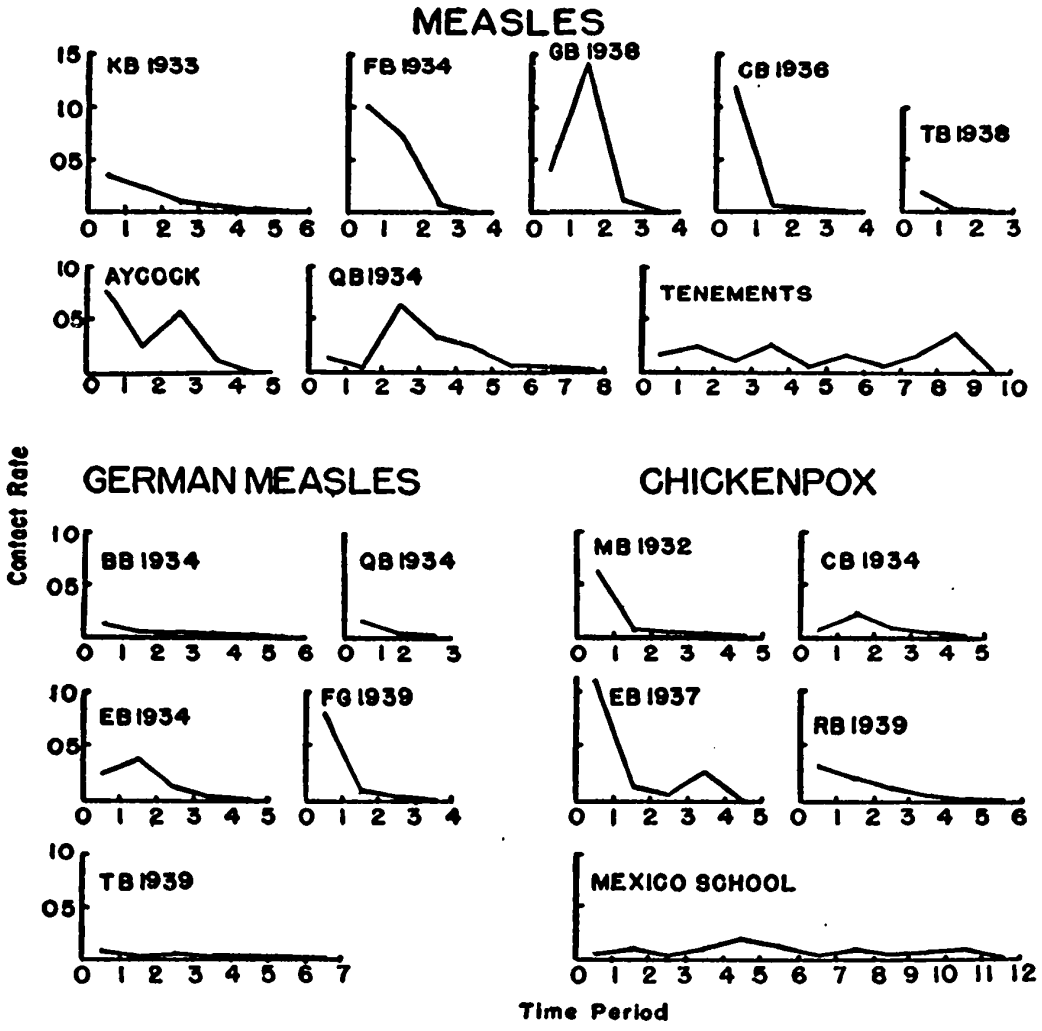

FIG. 4. CONTACT RATES ESTIMATED AT EACH TIME PERIOD FROM REPORTED CASES AND SUSCEPTIBLES

in contact rate in this interval is not in general sufficient to account for the failure of the model to fit the data.

VARIATION IN THE CONTACT AMONG INDIVIDUALS

If the population in which an epidemic occurs is actually a group of smaller populations with some cross-contact among them, this will

TABLE 6

Goodness of fit of the Reed-Frost model to reported cases and reported susceptibles, omitting the first interval

EPIDEMIC	χ^2	d.f.	$P(\chi^2)$	SIGNS*
MEASLES				
KB 1933	62.5	4	<.00001	++−−−
FB 1934	112.3	2	<.00001	+−−
QB 1934	58.9	6	<.00001	−+++−−−
CB 1936	14.0	2	.0009	+−−
GB 1938	111.7	2	<.00001	+−−
Aycock	43.0	3	<.00001	−+−+
Tenements	13.6	8	.09	+−+−+−0+−
GERMAN MEASLES				
BB 1934	148.4	4	<.00001	++−−−
EB 1934	226.8	3	<.00001	++−−
TB 1939	40.0	5	<.00001	0+−−−−
CHICKENPOX				
MB 1932	3.1	3	.38	++−−
CB 1934	40.9	3	<.00001	++−−
EB 1937	13.9	3	.009	+++−
RB 1939	49.9	4	<.00001	++−−−
Mexico School	31.0	10	.0016	+−+++−+−++−

* Sign of observed cases minus expected cases in each generation.

introduce a source of variation not considered in the model. It would be desirable, therefore, to divide the school populations into their probable sub-populations, as for example, school grade or dormitory. However, the information needed to permit such a breakdown is not available.

Some information as to the effects which this variation might produce may be obtained by calculating a theoretical epidemic in a population which is assumed to be made up of sub-populations with some cross-contact among them. Thus, consider a population made up of two sub-populations A and B, with some mixing between them. Let p be the probability that an individual has adequate contact with a member of

his own group, and P be the probability that he has adequate contact with a member of the other group. In time $t + 1$, the probability that an individual in population A escapes the $C_t{}^A$ cases in population A and the $C_t{}^B$ cases in population B is

$$q^{C_t{}^A} \cdot Q^{C_t{}^B}$$

and the probability that he meets at least one case (from either population A or B) is

$$1 - q^{C_t{}^A} \cdot Q^{C_t{}^B}.$$

Then the number of cases produced in population A at time $t + 1$ is

$$C_{t+1}{}^A = S_t{}^A (1 - q^{C_t{}^A} \cdot Q^{C_t{}^B}). \quad (6)$$

Similarly, the number of cases produced in population B at time $t + 1$ is

$$C_{t+1}{}^B = S_t{}^B (1 - q^{C_t{}^B} \cdot Q^{C_t{}^A}). \quad (7)$$

An epidemic starts with the introduction of a case into one of the populations. Stepwise calculations of the epidemics in the two populations proceeds until there are no new cases in either A or B. The two epidemics are then added together to produce an apparent single epidemic.

Epidemics were calculated in this manner for values of the contact rate between populations, P, ranging from zero when the two populations are completely separated, to p, the point where the two become one homogeneous population.

Table 7 shows the results of testing the fit of the Reed-Frost model to the combined epidemics, a single contact rate and an initial number of susceptibles being estimated from the total cases in each time period in the two populations. The deviations of the theory from the combined epidemics are too small to detect the variation in contact rate which had been inserted. When random variation is added to the epidemics at each time period (Table 7), the model still fits the combined epidemics so well that they are indistinguishable from single ones. The order of magnitude of these deviations is much less than that which occurs in the school epidemics.

Because this experimental evidence has been obtained for only a very limited set of conditions, i. e., two populations with selected contact rates within and between them, the results can be no more than suggestive. There is no evidence here that large deviations are due to variation in

TABLE 7

Fit of Reed-Frost model assuming a single homogeneous population, to epidemics calculated from two populations, with selected within and between-population contact rates

SIZE OF EACH POPULATION	TOTAL CASES IN THE TWO POPULATIONS	CONTACT RATE		RATIO OF CONTACT RATES	NO. OF TIME PERIODS	ESTIMATED		FIT OF HYPOTHESIS OF SINGLE POPULATION		
		Within Pop.	Between Pop.			Total Susc.	Single Cont. Rate	χ^2	d.f.	$P(\chi^2)$
Theoretical epidemics										
100	83	.0230	.0000	∞ :1	10	100	.0230	0.00	6	1.00
100	87	.0228	.0002	114:1	11	100	.0230	0.16	7	.99+
100	174	.0226	.0004	56.5:1	15	230	.0074	8.10	10	.62
100	172	.0220	.0010	22:1	13	218	.0090	2.14	10	.99+
100	173	.0214	.0016	13.4:1	12	207	.0104	1.82	9	.99+
100	172	.0198	.0032	6.3:1	11	197	.0120	0.62	8	.99+
100	174	.0115	.0115	1:1	12	200	.0155	0.00	8	1.00
Theoretical epidemics with sampling variation added										
50	90	.04	.004	10:1	9	100	.0246	1.11	6	.98
50	90	.04	.004	10:1	10	108	.0193	7.31	7	.39
50	89	.04	.004	10:1	8	99	.0245	7.14	6	.32
50	83	.04	.004	10:1	9	91	.0278	3.03	6	.81
50	80	.04	.004	10:1	11	93	.0236	4.34	8	.83
50	88	.04	.001	40:1	16	160	.0099	11.78	14	.63
50	84	.04	.001	40:1	13	116	.0149	7.66	10	.65
50	37	.04	.001	40:1	10	41	.0585	4.93	5	.42
50	5	.04	.001	40:1	4	5	.4795	1.46	1	.23

the contact rate within a population. However, to obtain conclusive evidence a more realistic model of the conditions likely to be found in an actual community is needed, for example, one giving each individual several contact rates.

APPLICATION TO FAMILY DATA

The family data presented in this section are from a study by Wilson et al. (1939), of cases of measles in Providence, Rhode Island. He has tabulated the kinds of epidemics which develop in families having one primary case and three susceptibles, classifying the families according to several criteria, three of which have been selected for analysis here. These are, using Wilson's code numbers:

VIII—Four-child families with three susceptibles of all ages under 22, including infants.

IX—Four-child famiiles with three susceptibles under ten years and over 7 months.

XII—Families with four or more children, with three susceptibles of all ages under 22, including infants.

The family is assumed to be the closed population, and the one case and three susceptibles are the initial conditions. There are eight possible epidemics which may occur in these populations:

(1, 0, 0, 0) 1 primary case and no others.
(1, 1, 0, 0) 1 primary case and 1 secondary case.
(1, 1, 1, 0) 1 primary case, 1 secondary case, 1 tertiary case.
(1, 2, 0, 0) 1 primary case, 2 secondary cases.
(1, 1, 1, 1) 1 primary case, 1 secondary case, 1 tertiary case, 1 quaternary case.
(1, 1, 2, 0) 1 primary case, 1 secondary case, 2 tertiary cases.
(1, 2, 1, 0) 1 primary case, 2 secondary cases, 1 tertiary case.
(1, 3, 0, 0) 1 primary case, 3 secondary cases.

Expected numbers of families having each of these kinds of epidemics have been calculated from reported cases and susceptibles, assuming a constant contact rate within the families. Table 8, columns 2, 3, and 4, shows a comparison of observed and expected numbers of cases. The discrepancies are large. When the cases are added together by time

TABLE 8

Observed and expected numbers of families having each possible kind of measles epidemic

Expected numbers calculated from:
Estimated contact rate p;
Estimated contact rate p, and estimated proportion λ, of families with three susceptibles;
Estimated contact rate p_1 in the first period, and p_2 in the later periods.

KIND OF EPIDEMIC	OBSERVED NUMBER OF FAMILIES	EXPECTED NUMBER OF FAMILIES CALCULATED FROM p	χ^2	EXPECTED NUMBER OF FAMILIES CALCULATED FROM p AND λ	χ^2	EXPECTED NUMBER OF FAMILIES CALCULATED FROM p_1 AND p_2	χ^2
VIII							
1,0,0,0	10	7.3	1.0	2.0	31.1	2.3	25.1
1,1,0,0	7	5.1	0.7	2.7	6.6	12.6	2.5
1,1,1,0	3	6.5	1.9	12.6	7.3	5.8	1.3
1,2,0,0	36	8.9	83.0	36.6	0.0	38.0	0.1
1,1,1,1	4	11.3	4.7	1.7	3.0	1.7	3.1
1,1,2,0	3	15.4	10.0	5.0	0.8	1.1	3.3
1,2,1,0	13	57.8	34.7	34.1	13.0	25.7	6.3
1,3,0,0	75	38.7	34.4	56.3	6.2	63.8	1.9
Total	151	151.0	170.4*	151.0	68.0†	151.0	43.6†
IX							
1,0,0,0	4	1.2	6.2	0.6	19.3	0.5	24.5
1,1,0,0	3	0.6	8.5	0.6	9.6	2.4	0.1
1,1,1,0	1	1.0	0.0	3.0	1.3	2.1	0.6
1,2,0,0	8	2.2	15.5	9.9	0.4	11.5	1.1
1,1,1,1	4	3.4	0.1	1.1	7.6	1.6	3.6
1,1,2,0	3	7.3	2.5	3.7	0.1	1.4	1.8
1,2,1,0	10	38.8	21.4	28.0	11.6	24.0	8.2
1,3,0,0	67	45.5	10.1	53.0	3.7	56.6	1.9
Total	100	100.0	64.3*	99.9	53.6†	100.1	41.8†
XII							
1,0,0,0	20	18.8	0.1	7.1	23.4	7.1	23.4
1,1,0,0	22	13.8	4.9	9.0	18.8	18.2	0.8
1,1,1,0	9	15.9	3.0	28.5	13.3	14.1	1.8
1,2,0,0	59	18.8	86.0	61.9	0.1	41.5	7.4
1,1,1,1	8	21.7	8.7	5.0	1.8	8.8	0.1
1,1,2,0	7	25.7	13.6	10.9	1.4	7.2	0.0
1,2,1,0	17	86.4	55.8	58.2	29.2	68.6	38.8
1,3,0,0	107	47.9	72.9	68.5	21.6	83.6	6.6
Total	249	249.0	245.0*	249.1	109.6†	249.1	78.9†

* χ^2 has 6 degrees of freedom.
† χ^2 has 5 degrees of freedom.

period (Table 9, col. 2, 3, 4) it is seen that there is a bias in the sense that too many of the observed epidemics end in the first period. This is similar to the effect observed in the school epidemics.

Wilson, using a somewhat different mathematical model, calculated expected numbers of each kind of epidemic and found discrepancies

TABLE 9

Observed and expected cases of measles in families by time of occurrence

Expected numbers calculated from:
Estimated contact rate p;
Estimated contact rate p, and estimated proportion λ, of families with three susceptibles;
Estimated contact rate p_1 in the first period, and p_2 in the later periods.

TIME PERIOD	OBSERVED NUMBER OF CASES	EXPECTED NUMBER OF CASES CALCULATED FROM p	χ^2	EXPECTED NUMBER OF CASES CALCULATED FROM p AND λ	χ^2	EXPECTED NUMBER OF CASES CALCULATED FROM p_1 AND p_2	χ^2
VIII							
1	340	288	25.8	337	0.1	340	0.0
2	26	106	115.6	58	35.3	35	3.4
3	4	11	8.1	2	2.1	2	2.1
Total	370	405	149.5*	397	37.5†	377	5.5†
IX							
1	248	231	5.4	244	0.4	248	0.0
2	21	58	148.1	39	27.4	30	6.4
3	4	3	0.5	1	9.6	2	2.2
Total	273	292	154.0*	284	37.4†	280	8.6†
XII							
1	519	431	42.5	499	2.4	519	0.0
2	48	175	206.6	113	68.7	106	59.3
3	8	22	10.6	5	1.9	9	0.1
Total	575	628	259.7*	617	73.0†	634	59.4†

* χ^2 has 2 degrees of freedom.
† χ^2 has 1 degree of freedom.

which are essentially the same as those found when using the Reed-Frost model. He does not investigate the possible causes of the discrepancies, but concludes that " with the discrepancy so great between the observations and prediction . . . we may definitely assert that the theory is non-applicable to Providence data for measles." We have investigated some

of the possible sources of variation as was done with the school epidemics to see if there are modifications which improve the fit of the theory to the data.

ERRORS IN COUNTING SUSCEPTIBLES

Although each epidemic of itself contains too little information to provide an estimate of the initial number of susceptibles, they may be combined by assuming that some proportion of the families actually had three susceptibles, the remainder having only two. This proportion may be expressed as an unknown parameter to be estimated jointly with the contact rate. To test whether this assumption improves the fit, the two parameters are used to calculate the expected numbers of families (Table 8) and expected numbers of cases (Table 9). The improvement in fit is substantial, but the discrepancies are still large.

It is noteworthy that the best fit of the model to the observations is obtained when the classification of families limits the susceptibles to the ages where they can be most accurately counted, that is, to children between the ages of six month and ten years (IX); and that the fit is poorest when they can be least accurately counted (XII).

VARIATION IN THE CONTACT RATE WITH TIME

The contact rates, estimated separately for each interval, are as follows:

TIME PERIOD	CONTACT RATE		
	VIII	IX	XII
1	.751	.822	.695
2	.208	.406	.201
3	.571	.800	.471

There is a sharp decline in the rates between the first and second periods, similar to that which was observed in a number of the school epidemics. Although the rates rise again in the last period, this rise may not have much significance because, (1) it is based on small numbers of observations (.800 is based on 5 families), and (2) Wilson, noting that the attack rates in this period increased, stated that in classifying cases as primary, secondary, etc., doubtful cases were classi-

fied with the longer chains rather than the shorter. The effect of this method of classification is to increase the contact rates in the last interval.

Considering, as was done in the school epidemics, the effect of a drop in contact rate after the primary case has been discovered, a joint estimate was made of two contact rates, one in the first period, and the other in the second and third periods. From these estimates, expected numbers of families were calculated (Table 8) and expected numbers of cases (Table 9). Tests of goodness of fit show that the assumption of two contact rates improves the fit of the model over that obtained when a single, constant contact rate is used, and is somewhat more effective than is the assumption of a constant contact rate and an estimated number of susceptibles.

VARIATION IN THE CONTACT RATE AMONG INDIVIDUALS

If the kinds of epidemics are considered separately, it is seen that neither the assumption of overcounting of susceptibles nor a contact rate changing with time is adequate to account for the discrepancies which occur. Because the school data are single epidemics and the family data are composed of 151 (or 100 or 249) epidemics, the latter provide an additional source of variation not previously considered. This is the variation among families in the contact rate. To see whether this variation contributes significantly to the chi-square values of Tables 8 and 9, the intervals have been considered separately. For example, consider only the first time period. It is seen from Table 8 that among the 100 Type IX families, there were 4 who had no cases in the first period, 11 who had one, 18 who had two, and 67 who had three. These data provide an estimate of the assumed constant contact rate which can be used to calculate numbers of families expected from the assumptions of the model. Table 10 shows that the deviations of the observations from the expected numbers are large, and represent an appreciable portion of the total variation in the epidemic. Therefore a possible explanation for much of the discrepancy between the observations and the model may be the variation among families in their contact rates.

The application of the model to epidemics within families has shown three sources of variation in the data which may be of importance in causing discrepancies between the theory and the observations. These are the overcounting of initial susceptibles, the decline in contact rate with time, and the variation in contact rate among families. The first

two of these agree with the findings of the application of the model to school epidemics. The third is an additional source of variation which could not be considered in the school epidemics.

SUMMARY

The Reed-Frost theory of the spread of epidemics has been tested on two series of observations derived from disease and population conditions

TABLE 10

Observed and expected numbers of families having given numbers of cases in the first and second time periods (contact rates estimated separately for each interval)

NUMBER OF CASES	VIII			IX			XII		
	NUMBER OF FAMILIES		χ^2	NUMBER OF FAMILIES		χ^2	NUMBER OF FAMILIES		χ^2
	Observed	Expected		Observed	Expected		Observed	Expected	
				First time period					
0	10	2.3	25.1	4	0.6	21.1	20	7.1	23.6
1	17	21.2	0.8	11	7.8	1.3	46	48.3	0.1
2	49	63.7	3.4	18	36.0	9.0	76	110.1	10.6
3	75	63.8	1.9	67	55.6	2.3	107	83.5	6.6
Total	151	151.0	31.2‡	100	100.0	33.7‡	249	249.0	40.9‡
				*Second time period**					
0†	43	41.4	0.1	11	10.2	0.1	81	77.8	0.1
1	20	23.9	0.6	15	17.0	0.2	34	42.3	1.6
2	3	0.7	0.7	3	1.8	0.8	7	1.9	14.2
Total	66	66.0	1.4§	29	29.0	1.1§	122	122.0	15.9§

* Expected numbers based on observed numbers in first interval.
† Does not include families having no cases in first interval.
‡ χ^2 has 2 degrees of freedom.
§ χ^2 has 1 degree of freedom.

which approximate the assumptions of the theory. The first series is composed of epidemics of measles, chickenpox and German measles in boarding school populations, chiefly, were each school represents a closed universe. The second series consists of epidemics of measles in families each of which is a universe having four (reported) susceptible members.

The model was fitted to the reported cases by generation of the disease and reported number of susceptibles at the beginning of the epidemic. The theory fails to fit either series of observations, the discrepancy

between the theory and the data being consistently in the direction of more observed than theoretical cases in the early generations, and fewer observed than theoretical ones in the later periods.

Removal of this bias, and a substantial improvement in the goodness of fit as measured by chi-square tests, is obtained when the number of susceptibles is assumed to be unknown and estimated from the observed numbers of cases only. The estimated number of susceptibles is in most instances the total number of cases observed, implying that the population is exhausted of susceptibles. Collateral evidence is available to support the hypothesis that the difference between the reported and estimated numbers of susceptibles may be due to the inclusion in the reported numbers, persons who are not in fact susceptible to the disease.

A second hypothesis which also improves the fit of the theory to the observations is that the contact rate declines with time. This is somewhat less effective than the assumption of overcounting of susceptibles, but on the basis of the present analysis, it cannot be entirely ruled out as a source of the discrepancies between the model and the data.

Other sources of variation which were investigated are: variation in fit depending on whether 12, 13, or 14-day incubation periods are used, variation in fit depending on whether or not the epidemic is divided into periods appearing to represent the true generations of the disease, and the amount of error introduced when the population is not homogeneous but composed of sub-populations with some contact among them. The evidence in this paper does not suggest that any of these factors are likely to be important sources of the discrepancies between the theory and the observations.

REFERENCES

AYCOOK, W. L. Immunity to poliomyelitis. Heterologous strains and discrepant neutralization tests: *American Journal of the Medical Sciences*, 204, No. 3, pp. 455-467 (1942).

BAHLKE, A. M., SILVERMAN, H. F., and INGRAHAM, H. S. Effect of ultra-violet irradiation of classrooms on spread of mumps and chickenpox in large rural central schools: *American Journal of Public Health*, 39, No. 10, pp. 1321-1330 (1949).

CHEESEMAN, E. A. Epidemics in schools: An analysis of the data collected during the years 1935 to 1939: *Medical Research Council, Special Report Series*, No. 271. London (1950).

COLLINS, S. D., WHEELER, R. E., and SHANNON, R. D. The occurrence of whooping

cough, chickenpox, mumps, measles, and German measles in 200,000 surveyed families in 28 large cities: Special Study Series No. 1, Division of Public Health Methods, National Institute of Health. Washington (1942).

GODFREY, E. S., JR. Measles in institutions for children: *Journal of Preventive Medicine,* Vol. II, No. 1, pp. 1-33 (1928).

HALLIDAY, J. L. An inquiry into the relationship between housing conditions and the incidence and fatality of measles: *Medical Researoh Counoil, Special Report Series,* No. 120. London (1928).

NATIONAL BUREAU OF STANDARDS. *Tables of the Binomial Probability Distribution.* U. S. Government Printing Office (1949).

SCHOOL EPIDEMICS COMMITTEE. Epidemics in schools. An analysis of the data collected during the first five years of a statistical inquiry: *Medical Research Council, Special Report Series,* No. 227. London (1938).

SOPER, H. E. The interpretation of periodicity in disease prevalence: *Journal of the Royal Statistical Society,* Vol. XCII (New Series), pp. 34-73 (1929).

WILSON, E. B., BENNETT, C., ALLEN, M., and WORCESTER, J. Measles and scarlet fever in Providence, R. I., 1929-1934 with respect to age and size of family. *Proceedings of the American Philosophical Society,* 80, No. 3, pp. 357-476 (1939).

Errata pertaining to the paper *Statistical Analysis of the A–B–O System in Mixed Populations,* by W. L. Stevens, in the February, 1952 issue of *Human Biology.* Page 16, equation (2. 24) : transpose last bracket and exponent 2, so that numerator reads

$$n(i_{PP}i_{QQ} - i_{PQ}^2).$$

Page 23, Table 3, last symbol in second line of symbols, for *wp* read *wP.*