[1] Zwicky, F., *Proc. Nat. Acad. Sci.*, **28**, 317–320 (1942).
[2] Zwicky, F., *Ibid.*, **28**, 150 (1942); *Astrophys. Jour.*, **95**, 555 (1942).
[3] Emden, R., *Gaskugeln*, Teubner, Leipzig, 1907.
[4] Hubble, E., *Astrophys. Jour.*, **71**, 231 (1930).
[5] Zwicky, F., *Th. von Kármán Anniversary Volume*, May, 1941, p. 137.

THE EPIDEMIC CURVE

BY EDWIN B. WILSON AND MARY H. BURKE

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated July 15, 1942

In 1929 Soper[1] developed a theory of the epidemic curve based on tracing the rise and fall of the disease by "generations" of successive groups of infectious cases. If in the ith generation there are C_i infectious cases and S_i susceptibles and if there be a constant number A of susceptibles per infectious generation coming into the population, his equations to determine the number of cases and the number of susceptibles in the next generation were[2]

$$\frac{C_{i+1}}{C_i} = \frac{S_i}{m}, \tag{1a}$$

and

$$S_{i+1} = S_i - C_{i+1} + A, \tag{1b}$$

where m is the number of susceptibles necessary for one old case to produce just one new case. In the stationary endemic condition $C_{i+1} = C_i$, $S_i = m$, $S_{i+1} = S_i$ and $C_{i+1} = A$. The length of the generation was taken to be the "incubation period." Thus if a fortnight be taken as the incubation period for measles and if there are 150 children per fortnight coming into the population through birth and growing up (with due allowance for deaths and emigration and immigration) measles could maintain itself in a stationary endemic condition with 150 cases per fortnight, the quantity m would have to be determined by enumerating the susceptibles in the population under such conditions, and might well be of the order of 4 to 5 years' worth of recruits A, which at 150 per fortnight would be 15,600 to 19,500. As measles does not occur in a steady endemic condition but in sharp epidemics, it is necessary according to Soper's equations that m should be variable or that there should be an accumulation of susceptibles to a number considerably greater than m before the epidemic, with an alternative deficit of susceptibles considerably below m after the epidemic.

Already in February, 1928, Dr. Wade H. Frost when delivering the Cutter Lectures in Preventive Medicine at the Harvard Medical School had presented a similar theory based on a somewhat different line of thought.[3] He, too, considered that for the diseases to which his theory would be applicable one could think in terms of generations of infectious persons C_i who being in contact with susceptibles S_i for a certain period would infect some of them who would then become the new generation of infectious persons at an average time later by the "incubation period"; but he recognized that in the intermixture of susceptibles and infectious some of the susceptibles might have multiple contacts with infectious persons and yet could develop at most only one infection apiece. If S_i were the number of susceptibles, $p = 1/S_i$ would be the chance that any particular contact would fall upon any particular susceptible and $q = 1 - 1/S_i$ would be the chance that he would escape. If then there were k_i contacts made between infectious and susceptibles the chance that a susceptible would escape them all would be q^{k_i} and the chance that he would have at least one contact would be $1 - q^{k_i}$. The number of infected would therefore be $S_i(1 - q^{k_i})$. Frost assumed that the number of contacts k_i would be proportional to the numbers of infectious and of susceptibles jointly, or $k_i = rC_iS_i$. Thus his equations corresponding to Soper's 1*a* and 1*b* are

$$C_{i+1} = S_i\left[1 - \left(1 - \frac{1}{S_i}\right)^{rC_iS_i}\right], \tag{2a}$$

and

$$S_{i+1} = S_i - C_{i+1} + A, \tag{2b}$$

except that Frost did not allow for the recruitment of susceptibles at the rate A per incubation period for the reason that he was satisfied at the time to give a theory of the curve of an epidemic so sharp that the number of recruits during the epidemic would not materially influence the course of the epidemic.

It is clear that any such theory as that proposed by Soper or by Frost cannot be expected to explain in quantitative detail the course of any epidemic; any precise theoretical discussion of the epidemic curve must be highly mathematical and difficult and hypothetical. The greatest immediate value of the development of the theory and of attempts at its application to concrete instances must be upon the qualitative side in indicating the sorts of things which may happen under various idealized conditions. It is noteworthy that Soper's paper did accomplish instructive results and that Frost's discussion in his lectures here, but principally at the Johns Hopkins School of Hygiene and Public Health with his students in successive years, has likewise been deemed of great value. It is noteworthy also that although the two theories seem to be different as exemplified in equations (1) and (2) they are in fact pretty much alike. If the number of infectious C_i and the contact rate r are small enough so that only the first two terms of the expansion of $(1 - 1/S_i)^{rC_iS_i}$ need be taken, (2a) becomes

$$C_{i+1} = S_i[1 - (1 - rC_i)] = rC_iS_i$$

and $r = 1/m$ makes (2a) then the same as (1a) even though Soper apparently did not think of $1/m$ as a contact rate.[4]

As the formula (2a) is not easy to compute because of the high powers of numbers very near to 1 to which it leads, it was suggested to Dr. Frost that the "law of small numbers" could be used to modify the formulae without material change in the results so long as the number of susceptibles did not decline too far. Thus $(1 - 1/S)^{rCS} = e^{-rC}$ and

$$C_{i+1} = S_i[1 - e^{-rC_i}], \qquad (3a)$$

and

$$S_{i+1} = S_i - C_{i+1} + A. \qquad (3b)$$

Table 1 gives the calculation of an epidemic where a single infectious case $C_0 = 1$ is introduced into a population of $S_0 = 2000$ susceptibles under the hypothesis that the rate of effective contact of infectious and susceptibles is $r = 0.001$ (each infectious person averages 2 contacts with susceptibles), where the recruitment A is neglected and where the results of formulae (1), (2) and (3) are compared, keeping calculations to the nearest integer.

TABLE 1

COURSE OF HYPOTHETICAL EPIDEMIC, $C_0 = 1$, $S_0 = 2000$, $r = 0.001$

GENERATION	$m = 1000$ FORMULAE (1)	$r = 0.001$ FORMULAE (2)	$r = 0.001$ FORMULAE (3)
1	2	2	2
2	4	4	4
3	8	8	8
4	16	16	16
5	32	32	32
6	62	61	61
7	116	111	111
8	204	186	186
9	317	267	268
10	393	308	309
11	332	265	267
12	171	173	173
13	59	90	89
14	17	41	40
15	5	18	17
16	1	8	7
17		3	3
18		1	1
Total infected	1739	1594	1594
Residual susceptibles	261	406	406

It is clear that the results of calculations by (2) and (3) are essentially identical, and that the elimination of the double contacts makes the epidemic longer, more symmetrical and lower at the peak, and leaves more susceptibles untouched at the end.[5]

If the formulae (3) are used a very neat result may be had for the relation between S_0/m, the ratio of initial susceptibles to the number $m = 1/r$, and S_E/m, the ratio of the number of residual susceptibles to the same number. Indeed

$$S_1 = S_0 e^{-rC_0}, \quad S_2 = S_1 e^{-rC_1}, \quad \ldots \quad S_{k+1} = S_k e^{-rC_k} \qquad (4)$$

Multiplying these together and cancelling $S_1, \ldots, S_k$,

$$S_{k+1} = S_0 e^{-r(C_0 + C_1 + \ldots + C_k)} \qquad (5)$$

In any epidemic to which the theory applies the initial number of cases C_0 introduced into the population of susceptibles would be few compared with the number around the peak, and the terminal is 0 after the epidemic has passed.[6] Hence if S_E be the number of susceptibles left, the total cases $C_0 + C_1 + \ldots + C_k$ must be essentially $S_0 - S_E$, and one has the result

$$S_E = S_0 e^{-r(S_0 - S_E)} \text{ or } \frac{rS_E}{rS_0} = e^{-rS_0(1 - S_E/S_0)} \qquad (6)$$

where $rS_0 = S_0/m$ and $rS_E = S_E/m$. If $F = S_0/m$ and $f = S_E/m$,

$$f/F = e^{-F(1 - f/F)} \qquad (7)$$

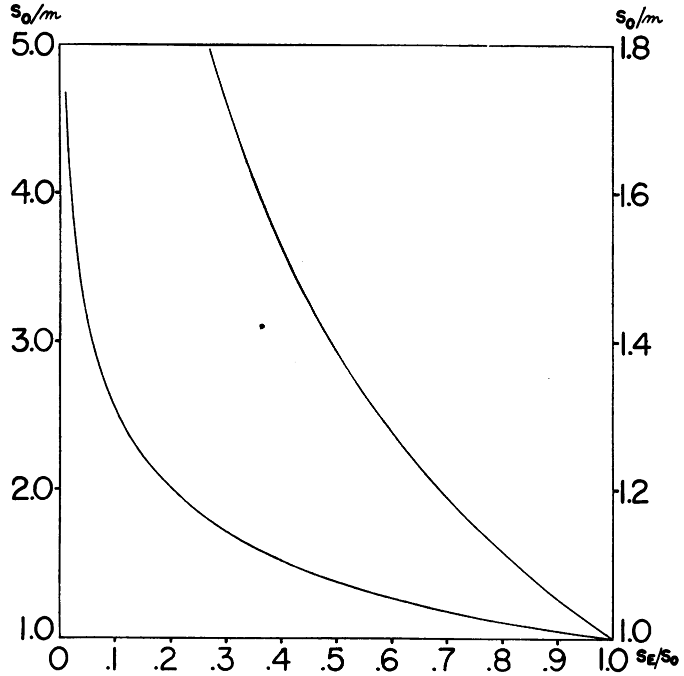
This relation (7) between f and F cannot be solved for either, but a table of corresponding values of F and f may be computed and tabulated as in table 2. The results of the table are given in the figure.

TABLE 2

RELATION BETWEEN THE RATIOS F AND f

$\frac{f}{F} = \frac{S_E}{S_0}$	$F = \frac{S_0}{m}$	$f = \frac{S_E}{m}$	$\frac{f}{F} = \frac{S_E}{S_0}$	$F = \frac{S_0}{m}$	$= \frac{S_E}{m}$
0.01	4.652	0.04652	0.32	1.676	0.5362
0.02	3.992	0.07984	0.35	1.615	0.5653
0.03	3.615	0.1084	0.40	1.527	0.6109
0.04	3.353	0.1341	0.45	1.452	0.6533
0.05	3.153	0.1577	0.50	1.386	0.6932
0.07	2.859	0.2002	0.55	1.329	0.7307
0.10	2.558	0.2558	0.60	1.277	0.7663
0.12	2.409	0.2891	0.65	1.231	0.8000
0.15	2.232	0.3348	0.70	1.189	0.8322
0.17	2.135	0.3629	0.75	1.151	0.8630
0.20	2.012	0.4024	0.80	1.116	0.8926
0.22	1.941	0.4271	0.85	1.083	0.9210
0.25	1.848	0.4621	0.90	1.054	0.9482
0.27	1.794	0.4843	0.95	1.026	0.9746
0.30	1.720	0.5160	1.00	1.000	1.0000

It is clear that if the number of susceptibles at the start were 3.35 m, only 4% of the original susceptibles would remain untouched by the epidemic, 96% would contract the disease. In Panum's Faroe Islands epidemic of measles something like 96% of the population of the villages



Plot of relation between the fraction S_E/S_0 of residual susceptibles to initial susceptibles and the multiple S_0/m which initial susceptibles are of $m = 1/r$(scale on the left), with enlargement for a part of the range (scale on the right).

entered by the disease was actually attacked by it.[7] It is further clear from the table that if S_0/m were only about 2, the fraction of the susceptibles which would escape would be 20%. For Hedrich's analysis of measles epidemics in Baltimore,[8] the table would not be strictly applicable because there was recruitment of the population which for a fairly long-drawn out epidemic of some 8 months might not be negligible. For his epidemic of 1930–1931, the susceptibles at the beginning were 78,968 but they rose to 81,449 during the initial stages where recruitment exceeded cases; from then they fell to 52,111 before the end of the epidemic, rising to 54,408 at its end. The ratio of minimum to maximum is 0.64 whereas that of end to beginning is 0.69. If we enter the table with the value $S_E/S_0 = 0.64$ we find $S_0/m = 1.240$ which would give $m = 65{,}300$ on the base $S_0 = 81{,}000$;

if we enter with $S_E/S_0 = 0.69$ we find $S_0/m = 1.199$ which would give $m = 66{,}000$ on the basis of $S_0 = 79{,}000$. These estimates of m are nearly alike. There were about 13,000 children coming into the susceptible group each year, which means that the number of susceptibles $m = 1/r$ would correspond to about 5 years of births, which may not be an unreasonable result in view of the certainty that both the theory and the calculations can at best be regarded as only approximately representing real conditions.

[1] Soper, H. E., "The Interpretation of Periodicity in Disease Prevalence," *J. Roy. Statist. Soc. London*, **92**, 34–61 (1929).

[2] We shall not restate in detail the conditions under which the theory might be considered as approximately true, nor at this time enter upon a discussion of the question of periodicity. The rigorous equations for a theory involving the basic conceptions and restrictions would seem to be these: First, there are at any time a number of infectious persons $I(t)$ and a number of susceptibles $S(t)$. Second, the rate of loss of susceptibles is $-dS/dt$ and should be set equal to the rate at which susceptibles become infected, namely, $C(t)$, less the rate of recruitment of new susceptibles $A(t)$. Third, the rate $C(t)$ is taken to be proportional to the product of $I(t)$ and $S(t)$. Fourth, the newly infected persons $C(t)dt$ become infectious after a latent time τ and remain infectious for a time σ. Hence

$$\frac{-dS}{dt} = C(t) - A(t), \quad C(t) = r(t)I(t)S(t), \quad I(t) = \int_{t-\tau-\sigma}^{t-\tau} C(t)dt. \qquad (A)$$

The factor r and the rate of recruitment A are generally taken as constant, though Soper shows that probably r or his m which is the reciprocal of r has a seasonal variation. Here the symbols C and A are rates, instead of being numbers of individuals as in (1). The variables C and I may be eliminated to get

$$A - \frac{dS}{dt} = rS(t)\int_{t-\tau-\sigma}^{t-\tau}\left(A - \frac{dS}{dt}\right) = rS(t)\ [A\sigma - S(t-\tau) + S(t-\tau-\sigma)] \qquad (B)$$

which is a differential-difference equation for S.

[3] Dr. Frost's lectures were on Feb. 2–3, 1928, and the dates of my letters to him were Feb. 9, 23 with replies from him dated Feb. 14, Mar. 21. It was in my second letter that I suggested the use of the law of small numbers in the way mentioned below. I strongly urged Dr. Frost to publish his theory of the epidemic curve, but he thought it too slight a contribution.—E. B. W.

[4] If we take Frost's equations and express the condition for a steady state we have $C_i = C_{i+1} = A$ and $A = S[1 - (1 - 1/S)^{rAS}]$. This equation cannot be solved strictly for the relationship $rS = 1$ with $S = m$ for the steady state, the relation between r, A, S for the steady state is more complicated. If we make S constant in (B) we have $1 = rS\sigma$ so that it is $r\sigma$ which takes the place of r in Frost's theory or of $1/m$ in Soper's; this is quite to be expected because the effective rate of generation of new cases must be the product of a contact rate r by a time σ available for making contacts. While for illustrative purposes to show what sorts of things may happen we may try different values of r or $1/m$ in the Frost or Soper theories, and different values of C_0 and S_0, and of A if we wish to admit recruitment, it must be remembered that in efforts to interpret concrete epidemics by the theory, the quantities r, C_0, S_0, A have to be determined from the data and cannot be expected to be determinable except

within rather wide limits. As a matter of fact the contact rate r must be highly variable within any community because contact within the home, within the school, and within the community at large must be at very different rates, so that at best r or m must be a sort of over-all community average of such very different rates.

[5] One of the limitations of Soper's set-up is that if m is sufficiently small, the calculation runs into the impossible situation where C_{i+1} becomes greater than the remaining S_i; for example, with $C_0 = 1$, $S_0 = 2000$, and $m = 500$, the cases in successive generations are 4, 16, 63, 243, 814 (by which time $S = 860$), and the next value of C by (1) comes out at around 1400. Frost's method involving elimination of double contacts seems not to suffer from this defect; calculating by the law of small numbers as in (3) we find with $r = 1/500$ for this case the successive values of C as 4, 16, 62, 224, 611, 764, 250, 27, 2, leaving 40 susceptibles still untouched. The calculations have been carried to tenths and then rounded off at the end in tabulating cases in successive generations. If we go through the detailed calculation with the Frost formulae (2) we find the same integral values of C as with formulae (3). When calculating according to Soper's formula (1*a*) we may be doing him an injustice; after giving that formula he shifted over to the formula

$$\frac{C_{i+1/2}}{C_{i-1/2}} = \frac{S_i}{m}, \text{ i.e., } \frac{\text{no. cases next interval}}{\text{no. cases last interval}} = \frac{\text{no. susceptibles at present}}{m}$$

"since the change in S_i is usually small in the unit interval." This shift is advantageous for the analytical developments upon which he is entering and surely makes no change which would not be within the tolerances of approximations in the theory as applied in concrete cases. It is impossible to determine from his paper whether he based his numerical calculations upon the original form (1*a*) or upon this modified form.

[6] Under the conditions, the epidemic has to die out, as it would not have to if there were recruitment; if one would add the initial cases C_0 which were introduced into the susceptible population S_0 to S_0 itself and use $S_0 + C_0$ in place of S_0 in (6) it would not be necessary to disregard the small number C_0 as mentioned in the text.

[7] Panum, P. L., *Observations Made During the Epidemic of Measles on the Faroe Islands in the Year 1846*, Delta Omega Society, 1940, distributed by the American Public Health Association, New York, N. Y.

[8] Hedrich, A. W., "Monthly Estimate of the Child Population Susceptible to Measles, 1900–1931, Baltimore, Md.," *Amer. J. Hygiene*, **17**, 613–636 (1933).