less largely due to the laying-up of a reserve store in the body against emergencies which may increase the rate of destruction of the vitamin, the relations of intake to bodily storage are being studied quantitatively.

* Aided by grants from The Nutrition Foundation, Inc.

1 Sherman, H. C., and Campbell, H. L., *Jour. Biol. Chem.*, 60, 5 (1924).

2 Sherman, H. C., and Campbell, H. L., *Proc. Nat. Acad. Sci.*, 14, 852 (1928); *Jour. Nutrition*, 2, 415 (1930).

3 Sherman, H. C., and Campbell, H. L., *Jour. Nutrition*, 14, 609 (1937).

4 Sherman, H. C., Campbell, H. L., and Lanford, C. S., *Proc. Nat. Acad. Sci.*, 25, 16 (1939).

THE LAW OF MASS ACTION IN EPIDEMIOLOGY, II

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated February 19, 1945

If one takes the equation for an epidemic when there are new susceptibles (recruits) coming into the population in the form[1]

$$\frac{dS}{dt} - A = \left(\frac{S}{m}\right)^p \left(\left.\frac{dS}{dt}\right|_{t-\tau} - A\right) \tag{1}$$

and introduces $u = \log (C/A)$, $C = Ae^u$, measuring case rates relative to the recruit rate A instead of relative to m, one finds as the equation for u

$$\left.\frac{du}{dt}\right|_t = \frac{pA}{m} e^{\frac{u(t-\tau)-u(t)}{p}} (1 - e^{u(t)}) + \left.\frac{du}{dt}\right|_{t-\tau}. \tag{2}$$

If t be advanced to $t + \tau/2$, one has

$$\left.\frac{du}{dt}\right|_{t+\tau/2} - \left.\frac{du}{dt}\right|_{t-\tau/2} = \frac{pA}{m} e^{\frac{u(t-\tau/2)-u(t+\tau/2)}{p}} (1 - e^{u(t+\tau/2)}). \tag{3}$$

The first approximation to this equation, neglecting first and higher derivatives upon the right, is

$$\frac{d^2u}{dt^2} = \frac{pA}{m\tau}(1 - e^u) \tag{4}$$

and corresponds to the approximation for the case $p = 1$ made by Soper in his discussion of periodicity. Equation (3) is indeed identical with his, except that $pA/m\tau$ replaces $A\ /\ m\ \tau$. The period of an infinitesimal oscillation is therefore

$$P = 2\pi\sqrt{\frac{m\tau}{pA}}. \tag{5}$$

The oscillations in epidemics of measles are not small. We have shown that, for approximations of this order in the absence of recruits ($A = 0$), we have

$$\frac{m}{p} = \frac{(\text{total cases})^2}{8(\text{peak cases})}, \tag{6}$$

where peak cases means the number of cases during one incubation period at the peak rate, and that if we consider higher approximations the value of m/p is not modified by more than perhaps 2 per cent unless the epidemic is very sharp (i.e., unless peak cases exceed one-sixth the total cases). We propose here to consider the modification in this relationship (6) that is due to the steady accession of susceptibles.

If we integrate (4) and determine the constant of integration so that $u = u_0$ when $du/dt = 0$, i.e., at the peak, we have

$$\frac{du}{dt} = \pm\sqrt{\frac{2pA}{m\tau}}\sqrt{u - e^u - u_0 + e^{u_0}}. \tag{7}$$

If u_1 be the negative root of $u - e^u - u_0 + e^{u_0}$, the half period will be obtained by integrating dt from $u = u_0$ to $u = u_1$ and the whole period will be found by doubling this result to be

$$P = \sqrt{\frac{m\tau}{2pA}}\left[2\int_{u_1}^{u_0}\frac{du}{\sqrt{u - e^u - u_0 + e^{u_0}}}\right] = \sqrt{\frac{m\tau}{pA}}\, f \tag{8}$$

where

$$f = \sqrt{2}\int_{u_1}^{u_0}\frac{du}{\sqrt{u - e^u - u_0 + e^{u_0}}} \tag{9}$$

and can be computed for different values of u_0 or $e^{u_0} = C_0/A$.

We now turn to finding the modification of (6) due to accession of recruits. If P be the period, the number of recruits during that time is PA and this must in turn be equal to the total cases, for in the hypothetical case under consideration everything must return to the same condition after one period. Hence

$$\text{Total cases} = PA = \sqrt{\frac{mA\tau}{p}}\, f. \tag{10}$$

As peak cases are $Ae^{u_0}\tau$, we have in lieu of (6)

$$\frac{m}{p} = \frac{e^{u_0}}{f^2} \frac{(\text{total cases})^2}{\text{peak cases}}. \tag{11}$$

In actual calculation we may obtain P from the total cases as P = total cases/A rather than from (8) and we may obtain m/p as

$$\frac{m}{p} = \frac{(\text{total cases})^2}{A\tau f^2}. \tag{12}$$

Here $A\tau$ are the recruits during an incubation period; to find f one must have a table of values in terms of u_0 or of $e^{u_0} = C_0/A$ computed from (9), viz.,

e^{u_0} = 1	2	4	6	10	15	20	30
f = 2π	6.4	7.1	8.0	9.7	11.5	13.1	15.3

The crucial test of any theory comes with the comparison of the theoretical results with the observed facts. If we had reliable values of m, of total cases, of recruits and of the peak case rate we could now determine the value of p appropriate to the particular relationship (12) derived from the theory. Fortunately, Hedrich[2] published a careful study of measles in Baltimore in which he has estimated the actual number of cases during each month from January, 1900, to December, 1931, and the number of intacts at the beginning of each month. By definition his intacts are children under 15 years who have not had measles. This should be close to the number of susceptibles S. In so far as the fundamental equation (1) is true, we should expect to find m as the value of S at the time when the case rate C was equal to the value one incubation period earlier. There are two such occasions each year around the time when the case rate is maximum and around the time when it is minimum. At minimum, cases are few and no great accuracy can be assigned to an estimate of the time when $C(t) = C(t - \tau)$; at maximum the cases are numerous in epidemic years but in non-epidemic years may be few and irregular. We have made the best estimates we can of the value of m at the peaks and in the troughs and find that the average value for the peaks is 67,000 and for the troughs 66,000. The average date of the peak (which varies from November to June) is estimated as around April 22 and the average date of the trough (which varies from August to November) is estimated as around September 22. The value m = 66,500 represents about $5^1/_2$ years of the estimates of recruits given by Hedrich.

If we take six of the most clear-cut epidemics, i.e., those that rise from low values of the case rate and die away to low values within a single epidemic year from September to August we find

Year	02–03	04–05	12–13	25–26	27–28	30–31
Cases	27,194	16,717	23,069	31,683	28,657	34,978
A (yr.)	11,100	11,000	11,700	12,000	11,500	10,500
P (yr.)	2.4	1.5	2.0	2.6	2.5	3.3
e^{u_0}	10.0	5.7	7.8	11.4	8.8	12.4
m	58,632	59,817	61,146	70,629	65,562	64,054
p	3.4	6.1	4.3	3.7	3.2	2.6

These values of p are certainly not in the neighborhood of 1, and furthermore they show a great variation from epidemic to epidemic.

If we take other clear-cut epidemics of measles from a variety of places, we have no published estimate of m to use and no estimate of the true number of cases of measles. However, we may for some places find a record of measles for a long period of years over which both the child population and the number of cases of measles seem to show little or no trend, and on the reasonable assumption that from 90 to 100 per cent of all persons have measles before the age of 15 we may estimate the fraction φ of cases that are reported. We may also assume[3] that the value of m is 5.5 times the average annual population A under 15, i.e., $m = 5.5A$. Under these assumptions we have[4]

$$P = \frac{\text{total cases}}{A\varphi}, \qquad p = 0.23\left(\frac{A\varphi f}{\text{total cases}}\right), \qquad e^{u_0} = \frac{\text{peak cases}}{\varphi A/24}.$$

The results are given in table 1.

The value of P represents the number of years of recruits which are used up in the epidemic and must vary inversely with the estimate φ of the fraction of cases reported. As this estimate has been made by comparing the average cases reported with the recruits, it has been assumed that the reporting was equally good in all years. Such a period as that of 6.5 found for Minneapolis is not readily reconciled with the history of measles in that city before and after the great epidemic, namely,

30–31	31–32	32–33	33–34	34–35	35–36	36–37	37–38	38–39
1776	232	8148	276	18,022	3372	90	2677	4791

With the estimate of 6776 for annual recruits the total of 39,384 in 9 years would give a ratio of 65 per cent for reporting instead of the 41 per cent obtained from a longer run of years. Probably 65 per cent is high because the years 29–30 and 39–40 were very low. Measles is a very variable disease and any estimate based on a limited number of years must be subject to considerable error; furthermore there is no assurance that the fraction of reporting is the same from year to year, it may be higher in the years of large epidemics than in years relatively free from measles, or inversely. As may be seen from the formulae used in the calculation the value of P varies inversely with the fraction of reporting and that of p

TABLE 1

EPIDEMICS OF MEASLES WITH TOTAL CASES REPORTED, AND ESTIMATES OF RECRUITS PER ANNUM, AN ESTIMATE OF PEAK CASES, THE RATIO OF PEAK TO TOTAL CASES, AN ESTIMATE OF THE FRACTION REPORTED, THE VALUE OF e^{u_0} AND OF THE PERIOD P AND EXPONENT p

PLACE	YEAR	SEPT.	OCT.	NOV.	DEC.	JAN.	FEB.	MARCH	APRIL	MAY	JUNE	JULY	AUG.	TOTAL	A	φ	PEAK	PEAK/TOTAL [a]	e^{u_0}	P	p
Albany, N. Y.	32–33	17	26	51	292	656	632	404	97	104	43	1	2	2,325	1713	0.45	351	0.15	10.9	3.0	2.5
Albany, N. Y.	35–36	1	5	26	111	282	701	547	252	149	108	54	0	2,236	1713	0.45	379	0.17	11.8	2.9	2.9
Albany, N. Y.	38–39	7	4	6	5	14	31	460	841	1049	309	24	7	2,757	1713	0.45	545	0.20	17.0	3.6	2.6
Berkeley, Calif.	30–31	8	16	2	4	18	88	362	661	372	95	15	2	1,643	1006	0.66	344	0.21	12.4	2.5	4.2
Berkeley, Calif.	33–34	6	2	2	18	66	104	153	344	334	90	7	8	1,134	1006	0.66	186	0.16	6.7	1.7	5.4
Berkeley, Calif.	38–39	2	5	14	64	377	865	1,260	372	92	13	6	3	3,073	1006	0.66	649	0.21	23.4	4.6	2.1
Binghamton, N. Y.	34–35	0	0	2	7	8	8	13	206	448	456	107	14	1,269	1161	0.49	246	0.19	10.4	2.2	4.4
Binghamton, N. Y.	37–38	3	0	1	4	4	15	122	269	296	141	25	0	880	1161	0.49	158	0.18	6.7	1.5	6.6
Binghamton, N. Y.	41–42	4	1	5	36	188	372	501	198	38	8	4	1	1,356	1161	0.49	256	0.19	10.8	2.4	4.0
Milwaukee, Wis.	31–32	11	9	11	26	99	587	1,892	4453	5544	1666	183	16	14,497	8927	0.62	2822	0.19	12.2	2.6	3.7
Milwaukee, Wis.	37–38	51	52	145	474	2534	8894	13,837	1595	221	75	31	13	27,922	8927	0.62	7266	0.26	31.5	5.0	2.2
Minneapolis, Minn.	34–35	19	74	234	1293	3828	6402	4,404	1347	310	67	38	6	18,022	6776	0.41	3449	0.19	29.8	6.5	1.3
New Haven, Conn.	30–31	1	5	39	42	108	350	1,428	1432	733	294	22	0	4,454	2587	0.51	767	0.17	13.9	3.4	2.5
New Haven, Conn.	34–35	2	2	7	43	104	301	925	2700	1428	326	20	4	5,862	2587	0.51	1420	0.24	25.8	4.4	2.4
New Haven, Conn.	38–39	3	11	3	12	49	170	479	1162	1480	698	101	20	4,188	2587	0.51	775	0.19	14.1	3.2	2.9
Niagara Falls, N. Y.	35–36	5	70	235	515	597	291	157	49	4	8	6	1	1,938	1352	0.41	307	0.16	13.3	3.5	2.2
Niagara Falls, N. Y.	37–38	2	0	4	3	15	126	329	398	251	73	35	5	1,241	1352	0.41	205	0.17	8.9	2.2	3.9
Niagara Falls, N. Y.	40–41	0	2	2	3	4	6	11	110	676	395	62	7	1,278	1352	0.41	362	0.28	15.7	2.3	5.9
Providence, R. I.	34–35	8	5	1	7	13	57	343	1351	1953	1279	241	17	5,275	4031	0.55	1003	0.19	10.9	2.4	4.0
Providence, R. I.	36–37	0	0	9	77	422	811	1,182	711	472	129	31	4	3,848	4031	0.55	589	0.15	6.4	1.7	5.1
Syracuse, N. Y.	31–32	0	1	1	3	83	728	2,323	2058	1171	587	191	4	7,150	3106	0.52	1156	0.16	17.2	4.4	1.7
Syracuse, N. Y.	38–39	4	0	10	3	89	286	405	697	1224	497	436	7	3,658	3106	0.52	640	0.17	9.5	2.3	4.0
Syracuse, N. Y.	41–42	5	0	0	2	2	25	107	324	1094	2255	610	74	4,498	3106	0.52	1191	0.26	17.7	2.8	4.5

[a] We have stated that ordinarily Peak/Total $<1/6$ and that in that case our approximate formulae should be good to within 2 per cent. For many of the epidemics here listed the ratio is greater than $1/6$ and consequently the approximation will not be so good; we desired, however, to have a wide range of epidemics, regardless of the accuracy of the formulae, particularly as the difficulties in estimating the necessary quantities are such as to introduce a liability to a considerable error in the values of P and p. It should be observed that the values of p for these 23 epidemics are not correlated appreciably with the ratio Peak/Total.

varies directly with the square of that fraction multiplied by f. Thus should we use 65 per cent in place of 41 per cent, we should get $P = 4.1$ and $p = 2.2$ in place of $P = 6.5$ and $p = 1.3$.

As a result of the difficulty of making reliable estimates of the factors of reporting the values of P and p which are entered in table 1 cannot be regarded as individually well determined; but the conclusion from all of them and from the results obtained for Baltimore seems inescapable that: In so far as the relationship $C = (S/m)^p C_{-1}$ may be valid for the course of epidemics of measles, the value of p which is appropriate to the relationship between peak cases and total cases cannot be considered to be unity, i.e., the simple law of mass action which has been so widely used is not in accord with the facts, and it is doubtful whether any fixed value can be brought into satisfactory accord with the course of epidemics in different years and in different places.[5]

[1] The notation will be essentially that of earlier papers, see these PROCEEDINGS, **31**, 24–34 (1945).

[2] Hedrich, A. W., "Monthly Estimates of the Child Population 'Susceptible' to Measles 1900–1931, Baltimore, Md.," *Amer. Jour. Hyg.*, **17**, 613–636 (1933). Such estimates might appear easy to make. If one starts with any level of susceptibles at the beginning of any month, adds the number of recruits and subtracts the number of cases, one obtains the number of susceptibles at the beginning of the next month. The difficulty of scaling up the number of cases reported to the true number of cases is however so serious that the process is not easy to carry out over a long period of time without getting unreasonably high or unreasonably low values of the susceptibles at some times. Hedrich's work seems to be carefully and critically done and we shall base our calculations upon his figures.

[3] The age distribution of reported cases of measles for New York State (exclusive of New York City) and for Massachusetts for the years specified is

NEW YORK STATE (EXCLUSIVE OF N. Y. CITY)

	1932		1933		1934		1935		1936		
AGE	CASES	%	CASES	%	CASES	%	CASES	%	CASES	%	MEAN, %
Under 1	927	1.87	703	1.99	577	2.01	710	1.78	537	1.94	1.92
1	2,015	4.06	1,463	4.15	1,098	3.82	1,339	3.36	1,168	4.21	3.92
2	2,714	5.47	1,879	5.33	1,435	5.00	1,768	4.43	1,422	5.13	5.07
3	3,182	6.41	2,150	6.10	1,658	5.77	2,184	5.47	1,669	6.02	5.95
4	3,634	7.32	2,516	7.13	1,970	6.86	2,442	6.12	1,910	6.89	6.86
5	5,401	10.88	3,806	10.79	3,279	11.41	3,825	9.59	3,171	11.43	10.82
6	7,093	14.29	5,220	14.80	4,434	15.44	5,154	12.92	4,407	15.89	14.67
7	6,263	12.62	4,342	12.31	3,817	13.29	4,865	12.19	3,875	13.97	12.88
8	5,118	10.31	3,427	9.72	2,916	10.15	3,673	9.20	2,903	10.47	9.97
9	3,245	6.54	2,290	6.49	1,990	6.93	2,468	6.18	1,777	6.41	6.51
10–14	6,759	13.62	4,945	14.02	4,007	13.95	6,796	17.03	3,468	12.50	14.22
15–19	1,765	3.56	1,323	3.75	833	2.90	2,340	5.86	765	2.76	3.77
20+	1,505	3.03	1,204	3.41	712	2.48	2,340	5.86	666	2.40	3.44
	49,620	99.98	35,268	99.99	28,726	100.01	39,904	99.99	27,738	100.02	100.00

The variation from year to year is notable. The large percentage of older cases in 1935 is especially noteworthy; the percentages at all ages under 10 are below the average of the percentages and are above the averages at all ages over 10. Contrariwise in 1936 the percentages are above the averages up to age 9 and below them after that age.

MASSACHUSETTS

	1936		1937		1938		1939		1940		
AGE	CASES	%	CASES	%	CASES	%	CASES	%	CASES	%	MEAN, %
Under 1	527	1.95	389	1.91	334	3.27	422	1.62	619	2.92	2.33
1	1,181	4.38	896	4.40	571	5.60	1,062	4.08	1,071	5.06	4.70
2	1,498	5.55	1,293	6.35	651	6.38	1,500	5.76	1,467	6.93	6.19
3	1,864	6.91	1,426	7.00	781	7.66	1,841	7.07	1,924	9.08	7.54
4	2,431	9.01	1,792	8.80	1,025	10.05	2,079	7.98	1,828	8.63	8.89
5	3,403	12.61	2,260	11.10	1,339	13.13	2,954	11.34	2,448	11.56	11.95
6	4,977	18.44	3,749	18.41	1,950	19.12	4,437	17.03	3,500	16.53	17.91
7	4,089	15.15	3,068	15.06	1,406	13.78	3,818	14.66	3,075	14.52	14.63
8	2,445	9.06	2,156	10.59	794	7.78	2,813	10.80	1,983	9.36	9.52
9	1,249	4.63	1,058	5.19	371	3.64	1,707	6.55	1,185	5.60	5.12
10–14	2,470	9.15	1,677	8.23	609	5.97	2,592	9.95	1,378	6.51	7.96
15–19	443	1.64	322	1.58	212	2.08	475	1.82	371	1.75	1.77
20+	410	1.52	280	1.37	157	1.54	347	1.33	330	1.56	1.46
	26,987	100.00	20,366	99.99	10,200	100.00	26,047	99.99	21,179	100.01	99.97

Again the variation from year to year is large. The percentages in 1939 are below the averages for all ages under 7. In comparison with the mean percentages of New York State the Massachusetts averages are higher under 8 and lower over 8; the cumulated percentages under 8 are Massachusetts 74.2, New York 62.1. It is clear that no percentage distribution can be assigned that is valid in both States in all years. How much differential there is in the factors of reporting by age in either State is unknown. The effect of the different age distributions in the two States could be eliminated but would make no really substantial modification.

If one bases an actuarial calculation upon the mean percentages one finds that the average number of immunes in the population is more than $5^1/_2$ years of recruits in Massachusetts and still more in New York State. On the other hand the figures given by Selwyn Collins (Public Health Reports, April 5, 1929) obtained from a large number of surveys in which was tabulated by age the percentage of children who had had measles, indicate much higher attack rates for children at early ages than those found here in either State. His fitted curve

$$y = 89\ (1 - e^{0.00586-0.04868x-0.02899x^2})$$

for the percentage who have had measles by age x, while representing well the observations has the obvious defect which inheres in all such series, namely, that the asymptotic percentage is too far below 100 to be representative of the true situation with respect to measles. There is no telling how the percentages should be scaled up to represent the true situation but the figures obtained from the curve give, respectively, 10.5, 12.7, 13.3, 12.3, 10.5, at ages 1, 2, 3, 4, 5 in place of 4.7, 6.2, 7.5, 8.9, 12.0 in Massachusetts. Conversion of the Massachusetts figures to rates would modify the percentages in a minor way. Clearly any estimate of the average number of susceptibles in the population based on rates derived from Collins' distribution, however, those rates were altered to come more nearly to representing the true situation with respect to immunity in the population, would be well below that derivable from the reported cases.

The evidence is as a whole indicative of a value for m in the neighborhood of $5^1/_2$ years of recruits.

[4] Peak cases as reported may be estimated by inspection as slightly more than half the cases in the highest month. The difference between 30 and 31 day months is small

enough to be disregarded in view of statistical fluctuations and the inherent inaccuracies of estimating the factor of reporting. The short month February may better be adjusted by taking from January and March an allowance for the last day of January and the first day of March, leaving these as 30 day months, in ordinary years, with slightly different allowances in leap years. If it be assumed that the case rate is parabolic for the three highest months and that the cases are, respectively, k_{-1}, k_0, k_1 in sequence, then on the assumption that τ is half a month

$$C_{0T} = \text{"peak cases"} = \frac{26k_0 - k_{-1} - k_1}{48} + \frac{(k_1 - k_{-1})^2}{16(2k_0 - k_{-1} - k_1)}.$$

As the theoretical solution in the absence of recruits ($A = 0$) for the epidemic curve is the sech^2 curve, and as the effect of the recruits is probably small in the three highest months of the epidemic, and finally as the probability and sech^2 curves have been used more for fitting observed cases than the parabola, it might be better to estimate peak cases by fitting a sech^2 curve instead of a parabola; we have indeed used the sech^2 curve in a number of cases but have come to the conclusion that the extra work involved is not justified by the slight increase in accuracy which may thereby be obtained.

[5] It is interesting to make some calculations for the data Soper gave for Glasgow for the years 1901–1916 which he seemed to think were of the forty years with which he worked those best suited to test his theory. One noticeable difference between his data and Hedrich's for Baltimore or that for other American cities is the infrequency with which clear-cut epidemics, which rise from few cases in one summer and die away to few cases in the next, are found in Glasgow. For five epidemic years we find, however,

YEAR	TOTAL CASES	PEAK	A	φ	P	$c\omega_0$	f	p
03–04	13,544	1177	800	0.55	1.2	2.7	6.6	4.8
07–08	21,484	1988	800	0.55	1.9	4.5	7.3	2.3
09–10	22,533	3153	800	0.55	2.0	7.2	8.5	2.9
11–12	18,176	2158	800	0.55	1.6	4.9	7.5	3.4

In making the calculations we have taken Soper's estimate of births as 25,500 and reduced it by about 20 per cent to allow for deaths at early ages, leaving recruits as $A = 800$ per fortnight. We have taken φ as 0.55 because Soper estimated the reporting as a little less than 50 per cent of births. Further we have taken m as 4 years of recruits in place of $5^1/_2$ years because he mentions that the average age of measles in one epidemic was $4^1/_2$ years. The formula for p then becomes $104(440\ f/\text{total cases})^2$. The average value of P is 1.7 years which corresponds fairly well to his figure of a little under two years, but the values of p are certainly not reconcilable with the assumption $p = 1$. Soper appeared to take m as 2 years of births (recruits?) and this would cut our values of p in two. He did, however, note that m would appear to be 3.87 years of births (something probably in excess of four years of recruits) judged from the average age of cases but rejected this estimate for reasons that seem to us of doubtful cogency.