EPIDEMICS

Maurice S. Bartlett *Oxford University*

[*Editorial comment.* Many advisors have suggested that this volume include an application in which a statistical or stochastic model is especially constructed to fit a real world process. These advisors want the volume to show the reader the model builder at work, to show how the model is tested, and to explore the new things it demonstrates. To appreciate the whole process requires some mathematics. We also might have to see the worker's wastebasket to appreciate how much effort may go into attempts he found unsatisfying.

Even the reader who has little mathematical equipment can gain considerable insight into the mathematical study of the birth, death, and maintenance of epidemics by skipping the harder mathematics in the piece Professor Bartlett has so kindly provided. For those who wish to skip along, the more mathematical parts have been set off and indented, and a few words, sometimes redundant, of transition have been inserted.—J.T.]

A CLASSIC book by the late Professor Greenwood, a medical statistician who was an authority on epidemics, has the title *Epidemics and Crowd Diseases,*

which emphasizes the relevance of the population, or community, in determining how epidemics arise and recur. One of the first to realize the need for more than purely empirical studies of epidemic phenomena was Ronald Ross, better known for his discovery of the role of the mosquito in transmitting malaria to human beings. Since those early years, the mechanism and behavior of epidemics arising from infection that spreads from individual to individual, either directly or by an intermediate carrier, have been extensively studied,[1] but it is perhaps fair to say that only in recent years have some quantitative features become understood and even now much remains to be unraveled.

MATHEMATICAL MODELS AND DATA

The statistical study of epidemics, then, has two aspects—on one hand, the medical statistics on some infectious disease of interest and, on the other hand, an appraisal of the theoretical consequences of the mathematical model believed to be representative, possibly in a very simplified and idealized fashion, of the actual epidemic situation. If the consequences of the model seem to agree broadly with the observed characteristics, there is some justification for thinking that the model is on the right lines, especially if it predicts some features that had been unknown, or at least had not been used, when the model was formulated.

How do we build a mathematical model of a population under attack by an infectious agent? There is no golden rule for success. Some feel that everything that is known about the true epidemic situation should be set down and incorporated into the model. Unfortunately, this procedure is liable to provide a very indigestible hotchpotch with which no mathematician can cope, and while in these days of large-scale computers there is much more scope for studying the properties of these possibly realistic, but certainly complicated, models, there is still much to be said for keeping our model as simple as is feasible without too gross a departure from the real state of affairs. Let us begin then at the other extreme and put in our ingredients one at a time.

We start with a community of individuals susceptible to the infection, let us say a number S of them. We must also have some infection, and we will confine our attention to the situation in which this can be represented by a number of individuals already infected and liable to pass on the disease. The astute reader will notice that even if we are concentrating on epidemic situations with person-to-person infection, we perhaps ought not to amalgamate *infected* and *infective* persons. Some people may be already infected, but

[1] Other pioneering workers in this field include W. Hamer, A. G. McKendrick, H. E. Soper, and E. B. Wilson.

not yet infective; some might not be infected, at least visibly, and yet be infective—so-called *carriers*. As we are considering the simplest case, however, we merely suppose there is a number I, say, of infective persons. When, as is to be hoped in real life, these persons recover, they may become resusceptible sooner or later (as for the common cold) or permanently immune, as is observed with a very high proportion of people in the case of measles.

While these ingredients for our epidemic recipe are very basic and common to many situations, there is a better check on our model if we are more definite and have one disease in mind, so let us consider only measles from now on. This is largely a children's complaint, mainly because most adults in contact with the virus responsible have already become immune and so do not concern us. Measles is no longer as serious an illness as it used to be (even less now that there is a preventive vaccine), but is a convenient one to discuss because many of its characteristics are fairly definite: the permanence of subsequent immunity, the incubation period of about a fortnight, and the requirement of notifiability in several countries, including the U.S., England, and Wales. The last requirement ensures the existence of official statistics, though it is known that notifications, unfortunately, are far from complete. Provided we bear this last point in mind, however, and where necessary make allowance for incomplete notification, it should not mislead us.

A DETERMINISTIC MATHEMATICAL MODEL

To return to our mock epidemic, we next suppose that the infective persons begin to infect the susceptibles. If the infectives remain infective, all the susceptibles come down with the infection eventually, and that is more or less all there is to be said. A more interesting situation arises when the infective persons may recover (or die, or be removed from contact with susceptibles) because then a competition between the number of new infections of susceptibles and the number of recoveries of infectives is set up. At the beginning of the epidemic, when there may be a large number of susceptibles to be infected, a kind of chain reaction can occur, and the number of notifications of new infected persons may begin to rise rapidly; later on, when there are fewer susceptibles, the rate of new notifications will begin to go down, and the epidemic will subside.

If we drew a graph of the number of infectives I against the number of susceptibles S at each moment, it would look broadly like the curve in Figure 1. The precise path, of course, will depend on the exact assumptions made on the overall rate of infection, on whether this is strictly proportional to the number I, and on whether also proportional to S, so that the rate at any moment is, say, calculated from the formula aIS where a is a constant. The path will depend also on the rate of recovery of the infected population,

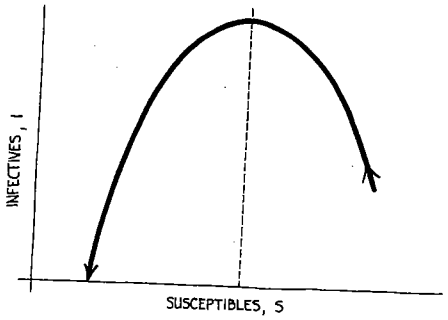

FIGURE 1

General path of an epidemic beginning with many susceptibles S, increasing at first the number of infectives I, then decreasing

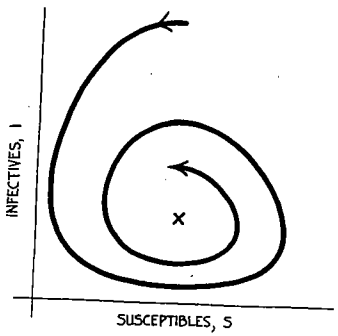

FIGURE 2

Deterministic model. Approach to equilibrium point (at cross) of I, S curve

likely to be proportional to I and given by bI, say, where b is a constant. Without going into too much detail yet, we can note one or two distinct features in the figure: (1) the susceptible population may not be reduced to zero at the time that the sources of infection have been eliminated; (2) because the path has no "memory," we could start at any point on it and proceed along the same curve—if our starting point were to the right of the maximum point, our infectives I would rise, if to the left, they would fall.

Actually, our more detailed assumptions are equivalent to the pair of recurrence relations for calculating I_{t+1}, S_{t+1} at any time $t + 1$ in terms of the values I_t, S_t at the previous time t (the unit of time should be small, say, a day or week):

$$S_{t+1} = S_t - aI_tS_t, \qquad I_{t+1} = I_t + aI_tS_t - bI_t.$$

We notice that $I_{t+1} - I_t$ is positive or negative according to whether $aS_t - b$ is positive or negative. The value b/a for S (on our assumptions it is independent of I) is called the *critical threshold* for reasons that will become clear.

Thus far we have a model for a "closed epidemic," which terminates when I is zero. Can we turn it into a model for measles, which has been claimed to come in epidemics about every two years? One ingredient still missing is an influx of susceptibles, due, in the case of childhood illnesses, to births within the community.

Let us, therefore, add a term c, say, to the right-hand side of the first equation above. If we follow the course of events in Figure 1, the path will start turning

right before it reaches the axis $I = 0$ and can be shown to proceed in an ever-decreasing spiral (Figure 2) till it finally arrives at an equilibrium point, which is determined by the equations

$$c - aIS = 0, \qquad aIS - bI = 0.$$

The second of these gives $S = b/a$ (the critical threshold value), and the first then yields $I = c/b$. These results are partly encouraging, but partly erroneous. The encouraging feature is the tendency to recurrent epidemics; we can even find the period of the cycle, which is found to be approximately

$$\frac{2\pi}{\sqrt{ac - \tfrac{1}{4}a^2c^2/b}}.$$

Sir William Hamer, who first put forward the above model for measles in 1906, took $b = \frac{1}{2}$ when t is reckoned in weeks, corresponding to an average incubation period of a fortnight, and c for London at that time as 2200. The value of a is more uncertain, but one method of arriving at it is to note that the *average* number of susceptibles, which was put at 150,000, should be around the theoretical equilibrium value b/a giving $a = 1/300{,}000$. Notice that c will tend to be proportional to the size of the community, so that if ac is to remain constant, a must be inversely proportional to the population, but this is not an unreasonable assumption; it is consistent, for example, with effective infectivity over a constant urban area, the entire town being regarded as an assemblage of such units.

The introduction of an influx of susceptibles showed that instead of following the simple curved path of Figure 1, an epidemic might follow a spiral until it finally settled down with a particular number of susceptibles. The time to go around the spiral once, called the *period,* is estimated for London data at 74 weeks, in reasonable agreement with the average period of somewhat less than two years that has been observed for large towns in England and Wales (see Table 2), the U.S. and comparable countries in the present century. We would hardly expect the epidemic pattern to remain precisely the same under very different social conditions, though the annual measles mortality figures for London quoted from John Graunt for the seventeenth century (Table 1) suggest a similar pattern even then (with perhaps a slightly longer average period of 2 to 3 years).

TABLE 1. Deaths from Measles in London in the Seventeenth Century

1629	1630	1631	1632	1633	1634	1635	1636	1637–46	1647	1648
41	2	3	80	21	33	27	12	Not recorded	5	92

1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660
3	33	33	62	8	52	11	153	15	80	6	74

UNSATISFACTORY FEATURES OF THE MODEL

The erroneous feature of the improved model is that actual measles in London or other large towns recurs in epidemics without settling down to a steady endemic state represented by the theoretical equilibrium point. What aspects of our model must we correct? There are some obvious points to look at:

(1) Our assumptions about the rate of recovery correspond to a gradual and steadily decreasing fraction of any group infected at the same time, whereas the incubation period is fairly precise at about two weeks, before the rash appears and the sick child is likely to be isolated (this being equivalent to recovery).

(2) We have ignored the way the children are distributed over the town, coming to school if they are old enough or staying at home during a vacation period.

(3) Measles is partly seasonal in its appearance, with a swing in average notifications from about 60% below average in summer to 60% above average in winter.

INTRODUCING CHANCE INTO THE MODEL

We will consider these points in turn. The effect of point (1) is to lessen the "damping down" to the equilibrium level, but not, when correctly formulated, to eliminate it. Point (2), on the movement over the town, raises interesting questions about the rate of spread of infection across different districts, but is less relevant to the epidemic pattern in time, except for its possible effect on (3). If we postulate a $\pm 10\%$ variation in the "coefficient of infectivity" a over the year, it is found to account for the observed $\pm 60\%$ or so in notifications. There seems to be little evidence of an intrinsic change in a due, say, to weather conditions, and it may well be an artifact arising from dispersal for the long summer vacation and crowding together of children after the holidays. Whatever its cause, it does not explain the persistence of a natural period; only the seasonal variation would remain and give a strict annual period, still at variance with observation.

To proceed further, let us retrace our steps to our closed epidemic model of Figure 1. To fix our ideas, suppose we initially had only one infective individual in the community. Then the course of events is not certain to be as depicted; it may happen that this individual recovers (or is isolated) before passing on the infection, even if the size of susceptible population is above the critical threshold. This emphasizes the chance element in epidemics, especially at the beginning of the outbreak, and this element is specifically introduced by means of probability theory. To examine the difference it makes, let us suppose

the *chance* P of a new infection is now proportional to aIS and the chance Q of a recovery proportional to bI. Denote the chance of the outbreak ultimately fading out without causing a major epidemic by p. We shall suppose also that the initial number S_0 of susceptibles is large enough for us not to worry about the proportionate change in S if the (small) number of infective persons changes. Under these conditions two infective persons can be thought of as acting independently in spreading infection, so that the chance of the outbreak fading out with *two* initial infective persons must be p^2.

Now consider the situation after the first "happening." Either this is a new infection or a recovery, and the relative odds are $P/Q = aS_0/b$. If it is a new infection, I changes from 1 to 2, and the chance of fade-out is p^2 from now on, or I drops to zero, and fade-out has already occurred. This gives the relation

$$p = \frac{P}{P+Q}p^2 + \frac{Q}{P+Q},$$

whence either $p = 1$ or $p = Q/P = b/(aS_0)$. If $b \geq aS_0$ (that is, if we are below the critical threshold) the only possible solution is $p = 1$, implying as expected that the outbreak certainly fades out. However, the ultimate probability of fade-out can be envisaged as the final value reached by the probability of fade-out up to some definite time t, this more general probability steadily increasing from zero at $t = 0$ to its limiting value, which therefore will be the *smaller* of the roots of the above quadratic equation. This is $b/(aS_0)$ if this value is less than one, providing us with a quantitative (nonzero) value of the chance of fade-out *even if the critical threshold is exceeded* and stressing the new and rather remarkable complications that arise when probabilistic concepts are brought in.

The mathematics shows that if the initial number of susceptibles is smaller than a value determined by the ratio of some rates used in the model, then the epidemic will certainly fade out. If it is larger than this critical value, then there is still a positive probability that the epidemic will fade out.

When new susceptibles are continually introduced, represented by c, the complications are even greater. For small communities, however, the qualitative features can be guessed. Once below the threshold, the number of infectives will tend to drop to zero, and though the susceptibles S can increase because of c, it seems unlikely that the number will pass the threshold before I has dropped to zero. The epidemic is now finished, and cannot re-start unless we introduce some *new* infection from outside the community. This is exactly what is observed with measles in a small isolated community, whether it is a boarding school, a rural village, or an island community. For such communities, the period between epidemics depends partly on the rate of immigration of new infection into the area and not just on the natural epidemic cycle. Moreover, when new infection enters, it cannot take proper hold if the susceptible population is still below the threshold, and even if

TABLE 2. Measles Epidemics for Towns in England and Wales (1940–56)

TOWN	POPULATION (THOUSANDS)	MEAN PERIOD BETWEEN EPIDEMICS (WEEKS)	TOWN	POPULATION (THOUSANDS)	MEAN PERIOD BETWEEN EPIDEMICS (WEEKS)
Birmingham	1046	73	Newbury	18	92
Manchester	658	106	Carmarthen	12	79
Bristol	415	92	Penrith	11	98
Hull	269	93	Ffestiniog	7.1	199
Plymouth	180	94	Brecon	5.6	149
Norwich	113	80	Okehampton	4.0	105
Barrow-in-Furness	66	74	Cardigan	3.5	>284
Carlisle	65	75	South Molton	3.1	191
Bridgwater	22	86	Llanrwst	2.6	>284
			Appleby	1.7	175

Source: Bartlett (1957), Tables 1 and 2.

above, new infection may have to enter a few times before a major outbreak occurs. The average period tends to be above the natural period for such communities. If we assume that the rate of immigration of new infection is likely to be proportional to the population of the community, the average period between epidemics will tend to be larger the smaller the community, and this again is what is observed (Table 2).

Consider now a larger community. We expect random effects to be proportionately less; there is still, however, the possibility of extinction when the critical threshold is not exceeded. Nevertheless, before all the infectives have disappeared, the influx c of new susceptibles may have swung S above the threshold, and the stage is set for a new epidemic. Under these conditions (and provided ac remains constant, as already assumed), the natural period will change little with the size of community.

CRITICAL SIZE OF COMMUNITY

How large does the community have to be if it is to begin to be independent of outside infection and if its epidemic cycle is to be semipermanent? Exact mathematical results are difficult to obtain, but approximate solutions have been supplemented by simulation studies of the epidemic model, using computers. An example of one such series plotted to extinction of infection after four epidemics (an interval representing nearly seven years) is shown in Figure 3. This particular series has no built-in seasonal incidence, but some internal

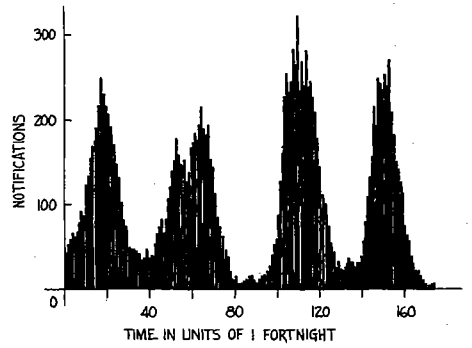

FIGURE 3
Results of simulation of four epidemics of measles over a seven-year period for a town whose average susceptible population is 3700

migration within its population boundaries; its average susceptible population is 3700. It appears from all such results that the critical size of the susceptible population is, for the measles model, of the order of 7000, or if the factor of 40 estimated for Manchester, England, between total and susceptible population is used, over a quarter of a million people in the community.

Now we do not need to use this theoretical figure as more than an indication of what to look for. By direct examination of measles notifications for any town, we can see whether notifications have been absent for more than two or three weeks. In view of the rather well-defined incubation period, we would infer from this lack of notifications that the infection had disappeared if we knew that notifications were complete. Incomplete notification is a complication, but not one that is likely to affect these quantitative conclusions very much, for fade-out of infection is found to increase rather rapidly as the community size decreases and soon becomes quite recognizable from the detailed statistics. In this way, it was ascertained that in England and Wales, during the period 1940–56, cities of critical size were Bristol (population about 415,000) and Hull (269,000). This investigation was supplemented by an examination of U.S. statistics for the period 1921–40, from which it was found that some comparable North American cities were Akron (245,000), Providence (254,000), and Rochester (325,000). Therefore, there is an observed critical community size of around 300,000, in reasonable agreement with what we were expecting.

Of course, towns of such size are not completely isolated from other communities as assumed in our model; this could tend to lessen the observed critical size, especially if the isolation is comparatively slight. In Table 3 the fade-out effect is shown for aggregates of individual "wards" in Manchester to demonstrate how it decreases with the population aggregate. The critical size (defined precisely in terms of 50% probability of fade-out after an epidemic) is, again as expected, smaller than for complete towns due to the

TABLE 3. Observed (Aggregate) Fade-Out Effect in Manchester Wards

WARDS	CUMULATIVE POPULATION (THOUSANDS)	NUMBER OF EPIDEMICS FOLLOWED BY FADE-OUT	PROBABILITY OF FADE-OUT (%)
Ardwick	18.4	12	100
St. Mark's	38.2	12	100
St. Luke's and New Cross	71.8	9	75
All Saints, Beswick, and Miles Platting	140.8	4	33
Openshaw, Longsight, N. and S. Gorton, Bradford, and St. Michaels	254.1	1	8
Medlock St., W. and E. Moss Side, Rusholme, Newton Heath, Collyhurst, Harpurhey, and Cheetham	419.3	0	0

Source: Bartlett (1957), Table 3.

extensive migration across the ward boundaries; it is estimated to be 120,000 total population living within the area.

CONCLUSIONS

If we review these results, we may justifiably claim that our theoretical model for measles, idealized though it inevitably is, has achieved some fair degree of agreement with the observed epidemic pattern. In particular:

(1) It predicts a "natural" period between epidemics of rather less than two years.
(2) A small ($\pm 10\%$) seasonal variation in infectivity (whether or not an artifact of seasonal pattern in school-children's movements) accounts for the larger ($\pm 60\%$) observed seasonal variation in notifications.
(3) It predicts extinction of infection for small communities, with consequent extension (and greater variability) of periods between epidemics.
(4) It predicts a critical community size of over a quarter of a million necessary for the infection to remain in the community from one epidemic to the next.

Epidemic patterns, of course, will be very sensitive to changing customs and knowledge; and the introduction of a vaccine for measles will inevitably change its epidemic pattern, and perhaps in time eliminate the virus completely. However, the greater understanding of epidemics that follows from appropriate models may be applied to other epidemic infections, and should assist in predicting and assessing the consequences of any changed medical practice or social customs even for measles.

PROBLEMS

1. Explain the difference between a deterministic mathematical model and a chance model.

The following problems refer to the material in small print.

2. Explain what is meant by "critical community size."

3. The deterministic model in Figure 2 predicts that there is an equilibrium point of I,S; i.e., the epidemic will never fade out. What is the concept that had to be introduced into the model to alter this prediction and thus better explain the data in Table 3 where we notice that in some communities nearly all the epidemics eventually fade-out?

4. Stated in words the formula $S_{t+1} = S_t - aI_tS_t$ says that "the number of susceptibles at the time $t+1$=number of susceptibles at time t–_____ at time t.

5. What is the equilibrium point (S_e, I_e) for the deterministic model shown in Figure 2?

6. Check by substitution that $p = 1$ and $p = Q/P$ are the solutions of the equation $p = p^2P/(P+Q) + Q/(P+Q)$.

REFERENCES

M. S. Bartlett. 1957. "Measles Periodicity and Community Size." *Journal of the Royal Statistical Society,* A120: 48–70.

J. Graunt. 1662. Reprint, 1939. *Natural and Political Observations upon Bills of Mortality.* Baltimore: Johns Hopkins University Press.

M. Greenwood. 1935. *Epidemics and Crowd Diseases.* London: William and Norgate.

R. Ross. 1910. *The Prevention of Malaria,* Second Edition. London: J. Murray.