

DETERMINISTIC AND STOCHASTIC MODELS FOR RECURRENT EPIDEMICS

M. S. BARTLETT

UNIVERSITY OF MANCHESTER, ENGLAND

CONTENTS

1. Introductory remarks	81
2. Simple stochastic models	83
3. Deterministic approximations	85
4. Miscellaneous modifications	87
5. Properties of stochastic models	90
6. Artificial epidemic series	93
7. The chance of avoiding extinction	97
8. The spatial or topographical factor	100
9. Specification in terms of point processes	104
10. Concluding remarks	107

1. Introductory remarks

In spite of the brilliant pioneering work of Farr, Hamer and Ross, and of important later studies by Soper, Greenwood, McKendrick, E. B. Wilson and others (see references and also the historical survey by Serfling [31]), a quantitative theory of epidemics in any complete sense is still a very long way off. The well-known complexity of most epidemiological phenomena is hardly surprising, for not only does it depend on the interactions between "hosts" and infecting organisms, each individual interaction itself usually a complicated and fluctuating biological process, but it is also, and this is a further point to be stressed, a struggle between opposing populations, the size of which may play a vital role. This last aspect is essentially one that can only be discussed in terms of statistical concepts. Greenwood (see p. 15, [16]) has remarked that "the epidemiologist's unit is not a single human being, but an aggregate of human beings"; however, even this remark omits to stress the second population of infecting virus or other parasitic invaders, and a much more comprehensive statement by Greenwood and his co-authors will be found in *Experimental Epidemiology* (see pp. 7-11, [17]). From the time of Ross at least, the importance of studying the nature, density and mode of transmission of the infecting agent has been recognized, although reliable information of this kind is often comparatively meagre. It should also be realized that the virus or bacterial populations may be in a continuous genetic or other biological state of flux. One need merely recall, for example, the existence of different strains of influenza virus, or the evidence for strains of different virulence in experimental epidemiological studies (see section 6, [17]). Considerable care is of course necessary not to confuse such variation in the virus with variation in resistance of the susceptible population, or with variation in the facility of transmission, especially when one remembers the severity of, say, a first epidemic of measles introduced into an isolated community, or asks what unambiguous evidence there is for intrinsic rise or

The presentation of this paper was supported (in part) by funds provided under Contract AF-18(600)-958 with the Air Research and Development Command.

fall in virulence during the course of single epidemics. Brownlee in particular appeared at times too ready to identify the cycles and waves he isolated from observed epidemics with an intrinsic variability in the virus, without always fully considering how far the phenomena he found could be, in part at least, purely statistical properties of the system under observation.

Any complete quantitative theory, in so far as it is realizable at all, would be based on hypothetical models or systems depending on a few parameters whose values could be determined from observations, perhaps of more local or isolated units. For example, in the case of infection from person to person, some of these parameters specify the nature of the incubation and infectivity periods, and the probability of transmission of infection. Other variables to be considered are the size and structure of the susceptible population, and the nature of, and changes in, immunity to the disease. The accurate determination of these parametric values from appropriate statistical data is thus of the utmost importance. (For recent investigations of this kind see, for example [4], [5], [18], and [19].) However, such information by itself does not automatically lead to an understanding of the behaviour of the population as a whole, and the justification of further theoretical discussion, as attempted in the present paper, is that mathematical formulations of typical epidemiological situations indicate (in so far as the equations can be handled) quite complex consequences even on the simplest assumptions. Until these have been studied and understood it would seem premature to embark on possibly more realistic, but even more complex, theories.

In most of these mathematical models the problem of the characteristics of the invading population of infecting virus has been largely shelved by the introduction of simple divisions of the human population into susceptible but uninfected, infected, recovered and immune, and so on. This procedure, while it may be criticized, has been retained below, as it is not unreasonable to see how far we can get with this approach. (It might be noticed that in Ross' formulation of the epidemiology of malaria, it was necessary to include also the population of mosquito vectors which transmit the infection.) In view of the simplifications inevitable at present in any theoretical discussion, it is clear that no detailed statistical agreement with observation, in the large-scale sense to be considered here, can be expected at this stage. But this situation is quite analogous to others where complex and interacting statistical systems are under consideration, such as some of the more complicated models of physical statistical systems, or of genetically heterogeneous animal populations. What should be looked for is a comparable overall pattern of predicted events; only when this seems a consequence of the assumed model does the model become a promising one for further study or elaboration.

One contrast of my own approach with that of many earlier theoretical studies is that complete probabilistic or stochastic formulations have always been in mind.¹ This enables the status of previous "deterministic" formulations, as approximations valid to some extent in the case of large numbers, to be examined. It will be shown that in some respects, as (i) epidemics always begin with only one or two infected persons, (ii) local units even of a large population are still small, the neglect of the

¹ Note, however, the important but largely overlooked paper by McKendrick [27], in which complete probabilistic models in continuous time are discussed. The probabilistic approach is also necessary in testing theoretical fits to local units such as households, being first developed by Greenwood and others for discrete generation or "chain" models; for a recent survey of this work see [5].

random or chance factor can be quite misleading. I have been interested in particular in possible mechanisms for *recurrent* epidemics, when the susceptible population is in one way or other replenished. Measles, with children continually growing up into the critical age period, has been the explicit infectious disease usually in mind. For this disease the main features of any model, while far from completely understood, are perhaps, owing to the work of Hamer, Soper, and others, as well accepted as any, and while the theoretical equations and techniques developed are obviously applicable to epidemic models in general, it is useful to avoid over-vague generalization and to relate the general theory to specific problems.

Some of my work along these lines has already been briefly referred to ([7] through [10]), and needless repetition has been avoided as far as possible. Moreover, for reasons of space some minor or incidental details have still been omitted. Otherwise the discussion below has been made, as far as practicable, self-contained.

2. Simple stochastic models

In the simple stochastic formulation of the Hamer-Soper model [32] of measles epidemics previously proposed [7], it was assumed that at any time t , S_t individuals were susceptible to the disease by transmission of infection from infected persons (I_t in number). It was assumed further, as typical of this particular disease, that recovered individuals were permanently immune, and were similar to isolated (or dead) individuals in not giving rise to new infections. The "transitions" that can occur during a small enough interval δt , from any given state i, s of this mixed population of susceptible and infected persons, are assumed proportional to δt , and to have probabilities (independently of previous states of the population)

- (i) $\lambda i s \delta t$ for $s \rightarrow s - 1, i \rightarrow i + 1$
- (ii) $\mu i \delta t$ for $i \rightarrow i - 1$
- (iii) $\nu \delta t$ for $s \rightarrow s + 1$.

Then if the simultaneous probability of $I_t = i, S_t = s$ is p_{is} , and

$$(2.1) \quad \pi_t(w, z) \equiv \sum_{i,s=0}^{\infty} p_{is} w^i z^s,$$

the probability-generating function π_t satisfies the partial differential equation

$$(2.2) \quad \frac{\partial \pi}{\partial t} = \lambda(w^2 - zw) \frac{\partial^2 \pi}{\partial z \partial w} + \mu(1 - w) \frac{\partial \pi}{\partial w} + \nu(z - 1)\pi.$$

This equation is valid even if λ, μ, ν are dependent on t , but (unless otherwise stated) these coefficients are here assumed constant. Special cases of (2.2) are

- (a) $\mu = \nu = 0$, so that $S + I$ is constant, n , say.
- (b) $\nu = 0$.

However, such cases, while already of some mathematical intractability,² represent theoretical examples of single epidemics, and the condition $\nu > 0$ (if the assumption of permanent immunity of recovered individuals is maintained) is necessary to ensure the possibility of recurrent epidemics. Of course if no initial infection is present the solution of (2.2) is independent of w and is

$$(2.3) \quad \pi_t = z^n \exp \{ \nu t(z - 1) \};$$

² Their detailed stochastic solution has been discussed by Bailey [2], [3]; and see [27] and [35] also. An ingenious approximation for case (b), based partly on the deterministic and partly on the stochastic model, has recently been developed by Kendall [23].

(2.3) also indicates the nature of the ultimate general solution of (2.2), for n can more generally be interpreted as the number of susceptibles when infection has become extinct (a contingency not included in the deterministic formulation and discussed in further detail later) and t reckoned from this extinction time as origin. To avoid this trivial situation, it will sometimes be more convenient to replace equation (2.2), which will be denoted by

$$(2.4) \quad \frac{\partial \pi}{\partial t} = H\pi,$$

where H is a particular operator, by the augmented equation

$$(2.5) \quad \frac{\partial \pi}{\partial t} = H'\pi,$$

where

$$(2.6) \quad H' \equiv H + \epsilon(w - 1).$$

The extra term represents the probability $\epsilon \delta t$ for the entry of a new infected person from outside in the interval δt , and ensures that infection never permanently dies out.

An equation of the type (2.4) can always be solved formally by writing

$$(2.7) \quad \pi_t = e^{Ht} \pi_0,$$

but this is most relevant for providing short-term solutions by expansion in powers of t (compare [35]), and does not seem particularly useful for studying the long-term behaviour. For seeking approximate solutions, it is convenient to write

$$(2.8) \quad \pi_t(w, z) \equiv M(\log w, \log z),$$

where

$$(2.9) \quad M(\theta, \phi) \equiv E\{e^{I\theta + S\phi}\},$$

so that (2.5) becomes

$$(2.10) \quad \frac{\partial M}{\partial t} = \lambda(e^{\theta - \phi} - 1) \frac{\partial^2 M}{\partial \theta \partial \phi} + \mu(e^{-\theta} - 1) \frac{\partial M}{\partial \theta} + \nu(e^{\phi} - 1)M + \epsilon(e^{\theta} - 1)M.$$

For a change of variables from I, S to $I/m, S/n$, where $m = \nu/\mu, n = \mu/\lambda$, we have for the new moment-generating function M'

$$(2.11) \quad \frac{1}{\mu} \frac{\partial M'}{\partial t} = m(e^{\theta/m - \phi/n} - 1) \frac{\partial^2 M'}{\partial \theta \partial \phi} + m(e^{-\theta/m} - 1) \frac{\partial M'}{\partial \theta} + m(e^{\phi/n} - 1)M' + \frac{\epsilon}{\mu} (e^{\theta/m} - 1)M'.$$

For m and n (but not ϵ/μ) large, the first-order approximation to (2.11) is

$$(2.12) \quad \frac{1}{\mu} \frac{\partial M'}{\partial t} = \left(\theta - \frac{m}{n} \phi \right) \frac{\partial^2 M'}{\partial \theta \partial \phi} - \theta \frac{\partial M'}{\partial \theta} + \frac{m}{n} \phi M'.$$

The solution of (2.12) is

$$(2.13) \quad M' = e^{\theta i_t + \phi s_t},$$

where

$$(2.14) \quad \frac{1}{\mu} \frac{di_t}{dt} = i_t(s_t - 1), \quad \frac{1}{\mu} \frac{ds_t}{dt} = \frac{m}{n} (1 - i_t s_t).$$

Equations (2.13) and (2.14) represent a deterministic (that is, nonrandom) solution, and provide one justification of the equations (2.14), which are equivalent to a deterministic model discussed by Soper, [32]. Their solution is developed somewhat further below (see [7], [10]).

3. Deterministic approximations

The equilibrium values of i_t and s_t in (2.14) are obviously unity on their new scale. To investigate small oscillations about equilibrium, put $i_t \equiv 1 + u$, $s_t \equiv 1 + v$, and write (2.14) as

$$(3.1) \quad \begin{aligned} \frac{1}{\mu} \frac{du}{dt} &= v(1 + u), \\ \frac{1}{\mu} \frac{dv}{dt} &= -\frac{m}{n} (u + v + uv). \end{aligned}$$

The first-order solutions (uv being neglected) are, for an appropriate choice of time origin,

$$(3.2) \quad \begin{aligned} u &\sim u_0 \exp(-\frac{1}{2}t/\sigma) \cos \xi t, \\ v &\sim u_0 \sqrt{\beta} \exp(-\frac{1}{2}t/\sigma) \cos(\xi t + \psi), \quad (0 \leq \psi \leq \pi) \end{aligned}$$

where

$$\sigma = \mu/(\nu\lambda), \quad \tau = 1/\mu, \quad \beta = \tau/\sigma, \quad \cos \psi = -\frac{1}{2}\sqrt{\beta},$$

and

$$(3.3) \quad \xi^2 = \frac{1}{\sigma\tau} - \frac{1}{4\sigma^2}.$$

For larger oscillations, the nonlinear character of (3.1) affects the shape of the waves. If, however, we write up to the second-order terms

$$(3.4) \quad \begin{aligned} u &= u_1 + au_1^2 + bu_1v_1 + cv_1^2 + \dots, \\ v &= v_1 + du_1^2 + eu_1v_1 + fv_1^2 + \dots, \end{aligned}$$

where u_1, v_1 are the first-order solution (3.2), the coefficients a, b, c, d, e, f are found by a straightforward investigation [10]. For β small (so that $\psi \sim 90^\circ$), a further simplification is possible and the solution becomes

$$(3.5) \quad \begin{aligned} u &\sim u_1 + \frac{1}{3}u_0^2 e^{-t/\sigma} \cos 2\xi t, \\ v &\sim v_1(1 + \frac{1}{3}u_1), \end{aligned}$$

where

$$v_1 \sim -u_0 \sqrt{\beta} \exp(-\frac{1}{2}t/\sigma) \sin \xi t.$$

In the case of measles, Soper took τ equal to the incubation period of two weeks, and estimating σ for London as 68.2³ calculated from the solution (3.2) a period of $2\pi/\xi = 73.7$ weeks, and a damping factor from peak to peak of $e^{-\pi/(\xi\sigma)} = 0.58$. The damping coefficient has a comparatively small influence on the period, which is approximately proportional to $\sqrt{\sigma\tau} = 1/\sqrt{\nu\lambda}$, that is, is affected mainly by the influx rate of susceptibles and the infectivity coefficient λ .

The actual solution of the equations (2.14) can of course always be calculated numerically. Some care is necessary with step-by-step numerical methods of solution if these are to be sufficiently accurate; otherwise (as occurred with Soper's calculations) fallacious conclusions on the damping effect may be reached. The numerical method used consisted in integrating (2.14), or rather the equivalent equations for S_t, I_t , to

$$\begin{aligned}\Delta I_t &= \lambda \int_t^{t+1} I_u S_u du - \mu \int_t^{t+1} I_u du, \\ \Delta S_t &= -\lambda \int_t^{t+1} I_u S_u du + \nu,\end{aligned}$$

and writing $f_t = (1 - \nabla)^{-1} f_0$, where ∇f_t is the "backward first difference" $f_t - f_{t-1}$, so that

$$\begin{aligned}\int_t^{t+1} f_u du &= \int_0^1 [f_t + u\nabla f_t + \frac{1}{2}u(u+1)\nabla^2 f_t + \dots] du \\ &= f_t + \frac{1}{2}(f_t - f_{t-1}) + \frac{5}{12}(f_t + f_{t-2} - 2f_{t-1}) + \dots \\ &\sim \frac{23}{12}f_t - \frac{4}{3}f_{t-1} + \frac{5}{12}f_{t-2}.\end{aligned}$$

To start the solution, the values for the first and second weeks were obtained by the cruder difference method obtained by replacing df by Δf in (2.14), using half-weekly steps; the solution could then be continued by the above method, with weekly steps.

The approximate solution from (3.5) was compared, for the above values of σ, τ (for which $\beta = 0.0293$) and initial conditions $I_0 = 13,000, S_0 = 150,000$, with such a step-by-step solution and appeared reasonable in spite of u_0 in this case being as large as 1.348 (the values for I_t are shown in figure 1).

The damping of the oscillations for arbitrary initial amplitudes may conveniently be depicted in terms of the path traced by a point with coordinates S_t, I_t . Such a path is shown in figure 2. This type of diagram is referred to again later (figure 8), as it also provides a convenient method of contrasting with this deterministic solution actual paths realized with the stochastic model. It might be noted that the nonlinear character of these (deterministic) oscillations may be shown theoretically not to affect the tendency to damping, the argument⁴ running as follows. In equations (2.14), put for convenience $\mu t = T$ and $m/n = c$, and consider the function

$$(3.6) \quad f(i_t, s_t) = c(i_t - \log i_t) + (s_t - \log s_t).$$

³ From previous data of Hamer, with $\nu = 2,200, n = 150,000$. While these figures are now somewhat obsolete, it will sometimes be convenient to retain them for illustrative calculations.

⁴ It is due to Mr. G. E. H. Reuter, to whom I am consequently indebted.

Along the path traced by any solution, we easily find that

$$(3.7) \quad \frac{df}{dt} = \frac{-c(s_t - 1)^2}{s_t} \leq 0,$$

so that f decreases along such a path as T increases. As $f \geq 1 + c$ always, f tends to a finite limit $f_0 \geq 1 + c$ as $T \rightarrow \infty$. The curves $f = k$ are closed, surround the point $(1, 1)$ and shrink down as k decreases to $1 + c$. The argument may be completed by showing (for example, by *reductio ad absurdum*) that the limit f_0 is in fact $1 + c$, so that $(i_t, s_t) \rightarrow (1, 1)$.

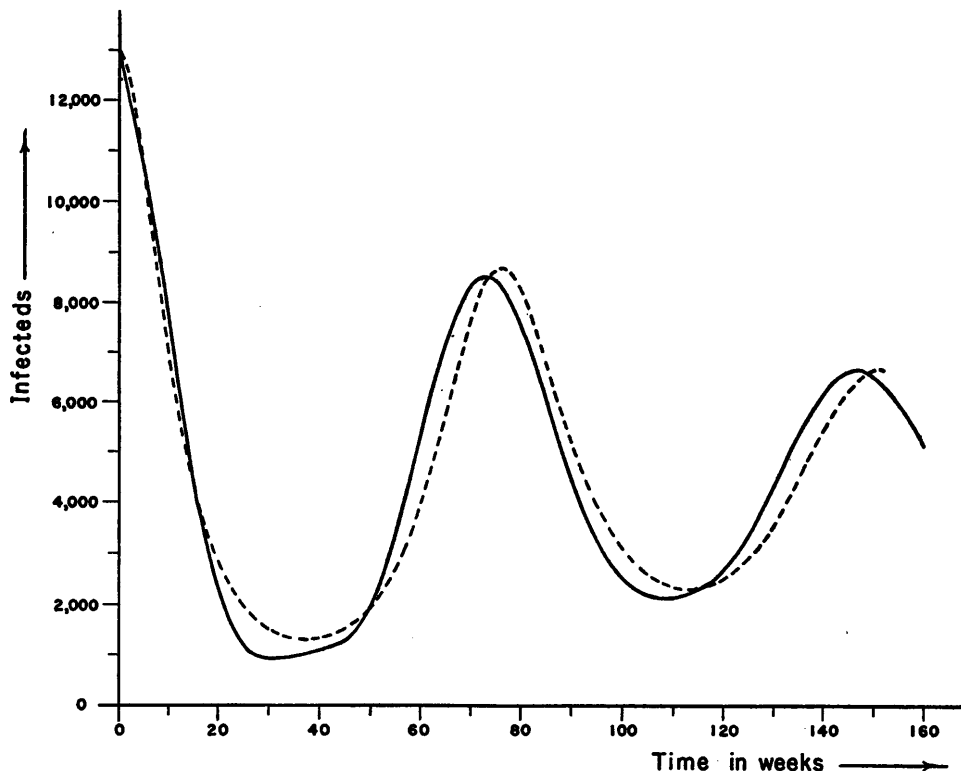


FIGURE 1

Deterministic model. Comparison of step-by-step calculation (thick line) with approximating curve (dotted line).

4. Miscellaneous modifications

The stochastic formulation (2.2), while typical of a certain class of epidemiological situation, is of course only one of a number of variants which might be more typical of modified situations, but there is no difficulty of principle in the formulation of such variants. The use of deterministic models, while subject to limitations to be discussed presently, has been seen in section 2 (see also [23]) to be a possible approximating procedure under appropriate conditions, and it is often useful to consider directly such deterministic formulations, without the full stochastic model being first formulated. Some of these deterministic models have been specified in

some detail by Kermack and McKendrick [25], and I shall not consider them in any generality here, but merely make one or two remarks with particular reference to measles. With this illness it is known that there is a fairly restricted interval of a few days towards the end of the incubation interval of about a fortnight when the infection is actually transmitted (see, for example, [18], [4], and [5]). Two

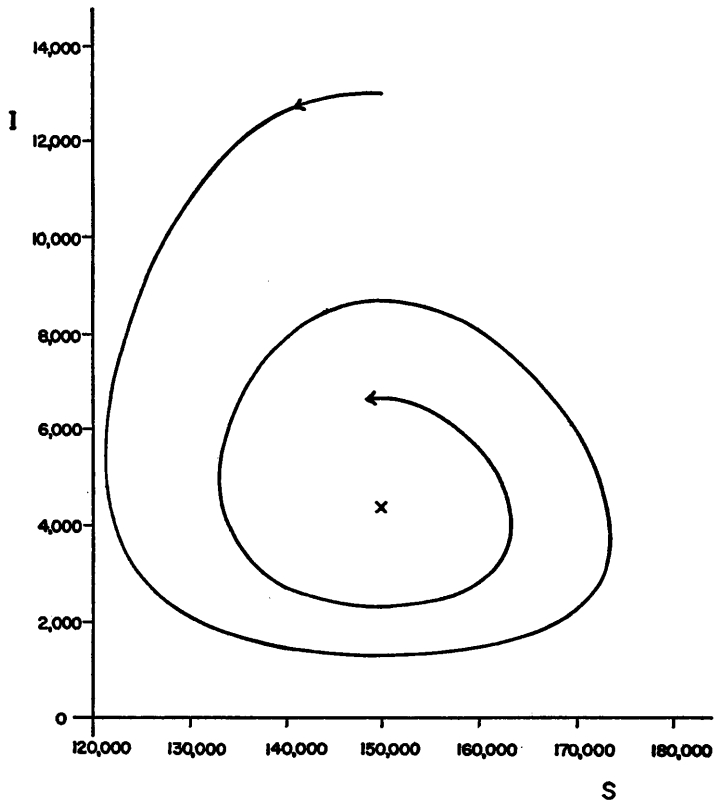


FIGURE 2

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

methods of modifying the assumption of a simple stochastic rate λ of infection are possible.

(i) The infectivity rate λ may be made a function $\lambda(x)$ of the "age" of the infected person from the instant of infection. This is not perhaps too complicated in deterministic formulations, leading to differential-integral equations, or in special cases,⁵ but in general it transforms the process to a nonlinear point process, methods of dealing with which (to be referred to later in connection with spatial effects) are not yet very far developed.

(ii) As a somewhat simpler device, a number of substates may be introduced, some of which have to be traversed before the incubation interval without infectivity is passed and the rest before infectivity ends. Such a device allows a more representative probability distribution of the time of infection to be achieved. In a

⁵ Such age-dependent models include those developed by Bellman and Harris (for an account of which, see [11]).

limiting case in which the first set of substates becomes very large and the rate of infection in one final state very large, corresponding to the hypothesis of a very short infectivity interval (this is the discrete generation or "chain model," and is referred to again in section 6), Soper was misled by his approximate numerical calculations into thinking that the damping coefficient was reduced to zero (see

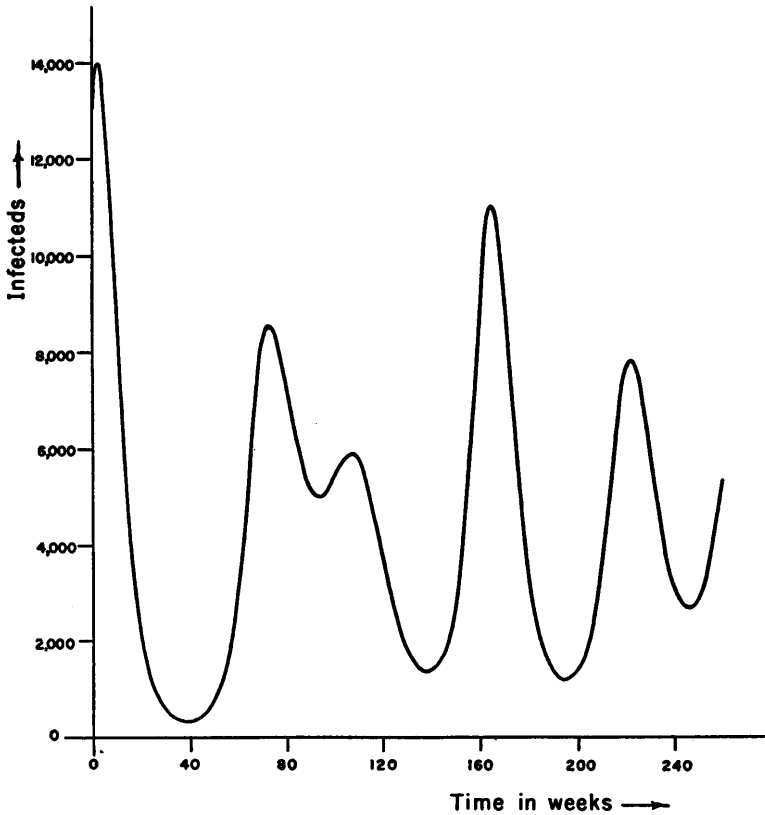


FIGURE 3

Deterministic model. Step-by-step calculation in the case of 10 per cent seasonal variation in infectivity.

section 3), whereas it may be shown to be merely halved, a conclusion first noted by Wilson and Worcester [38].

Seasonal changes. It is well known that many epidemics exhibit a seasonal variation in their average intensity, and this implies that the *effective* intensities, in particular the infectivity coefficient λ , may be periodic functions of the time. Whether such changes are due to atmospheric changes in the facility of transmission of the infectious disease, or seasonal changes in the viability or virulence of the infecting virus or other organism, or due merely to more artificial causes such as dispersion and subsequent reassembly of school children during the summer, is largely irrelevant here. In the case of measles there is a comparatively well-defined seasonal change of incidence (for example, for Manchester for the years 1917-1951 the change is 60 per cent above average at the beginning of a calendar year to 60 per cent below in the later summer), but no very direct evidence of seasonal change in the

infectivity coefficient λ [19] so that the effect of dispersal and reassembly at school might well be the most important contributor to the over-all seasonal effect.

The simplest theoretical modification to make is the substitution for λ of $\lambda' \equiv \lambda + \lambda_1 \cos \omega t$, where ω corresponds to an annual seasonal cycle. The deterministic equations then give, for small $\rho = \lambda_1/\lambda$, forced annual oscillations, for the incidence rate

$$(4.1) \quad \lambda' I_i S_i \sim \lambda' mn(1 + u + v),$$

with an amplitude

$$(4.2) \quad \frac{\rho[\{\omega^2(2\mu\sigma - 1) - \mu^2\}^2 + \omega^2\sigma^2\{\omega^2(1 - \mu\sigma) + \mu^2\}^2]^{\frac{1}{2}}}{(\mu - \sigma\omega^2)^2 + \omega^2}.$$

For $\mu = \frac{1}{2}$, $\omega = 2\pi/52$, $\sigma = 68.2$, this last expression becomes 7.9ρ , showing that a 10 per cent variation in λ' (as assumed by Soper) is sufficient on the above calculations to lead to about 80 per cent seasonal variation in incidence, which is of the right order of magnitude for measles. For larger initial oscillations the nonlinear character of the system does not appear to affect this conclusion; a step-by-step calculation is depicted in figure 3, where the gradual transition of the initial damped oscillations into the permanent forced annual oscillations is apparent.

Temporary immunity. With some infections, such as colds and influenza, immunity is only temporary, and the assumptions must be modified accordingly. A different kind of temporary immunity, following infection of a subclinical type, was suggested for measles by Stocks and Karn [33]. With other infections, for example, polio, even lasting immunity may result in this way, but what evidence there is for this kind of effect in the case of measles does not seem to me to suggest that it is an important factor in determining the nature of measles incidence in the community at large, and I shall not consider it further here.

5. Properties of stochastic models

In spite of the success of the Hamer-Soper model in predicting a cycle from simple assumptions which do not explicitly include it, the damping of the waves leads to an apparent difficulty in seeking for an explanation of recurrent epidemics on this basis. The problem has been clearly stated by Wilson and Worcester [38], when they remark

“... it must be admitted that the phenomenon of recurrent measles epidemics gives no clear evidence of any damping. This creates something of a difficulty with the theory in respect to the prediction of damping and throws some doubt on the reality of periods; it is possible that measles simply dies out and then returns and under such a hypothesis there would seem to be no reason to expect either definite periods or damping to be observable by comparing successive epidemics.”

Before this dilemma is accepted as a failure of the theory, the latter must be examined in regard to each (and both simultaneously) of two possible ways of escape:

- (a) the effect of the more complete stochastic formulation;
- (b) the spatial or topographical factor.

That effect (a) may be important is most strikingly realized by contrasting the endemic stable equilibrium level which the deterministic theory predicts with the

instability of this average level on the stochastic model. Consider equation (2.2) again in the critical situation when only a few infected individuals are present in a much larger initial population S_0 of susceptibles. For the crucial interval during which the fate of the majority of the susceptibles hangs in the balance, their number will remain approximately constant (reduction of their number by infection will diminish still more any chance of extinction of the infected individuals calculated with neglect of such reduction). With this approximation, if the number of susceptibles at time t is written S_t , the chance P of extinction is available at once from the theory of the "generalized birth-and-death process" [21] and is $[J/(1 + J)]^j$, where j is the initial number of infected, and

$$(5.1) \quad J = \int_0^\infty \mu \exp \left\{ \int_0^t (\mu - \lambda S_u) du \right\} dt .$$

Strictly in (2.2) S_t is increasing according to a Poisson increment of mean νt , and the above solution for P would require to be averaged for variation in S_t ; however, an equally relevant case is that of S_t increasing regularly or deterministically, so that

$$(5.2) \quad \int_0^t (\mu - \lambda S_u) du = (\mu - S_0 \lambda) t - \frac{1}{2} \lambda \nu t^2 .$$

The value of J may then be written in the form

$$(5.3) \quad J = \eta e^{\frac{1}{2} \eta^2 (f-1)^2} \int_{\eta(f-1)}^\infty e^{-\frac{1}{2} u^2} du ,$$

where $\eta = \mu/\sqrt{\lambda\nu}$, $f = S_0\lambda/\mu$. For values of λ , μ , ν previously used (section 2), $\eta = 5.84$, and the values of $1 - P(f)$ for $j = 1$ are shown in figure 4. In the alternative important case of a "closed" population ($\nu = 0$), the chance $P_0(f)$ becomes the extinction probability for a simple birth-and-death process with birth-rate $S_0\lambda$, and is, for $j = 1$ (see, for example, p. 71, [10]),

$$(5.4) \quad 1, f \leq 1; \quad 1/f, f > 1 .$$

For large f , the chance $P(f)$ may be written asymptotically as

$$(5.5) \quad P(f) \sim \frac{1}{f} - \frac{1}{f^2(f-1)\eta^2} .$$

This extinction phenomenon shows that in a small isolated community the deterministic solution is quite unrepresentative of what will occur on the epidemic model assumed. After a major epidemic the infected number drops to very few, and consequently, as the susceptible population is also temporarily small, the infection will disappear, as remarked in section 2. However, if it is allowed to reenter, as envisaged in the extra term of equation (2.6), a fresh epidemic can break out when the susceptible population has again become large enough.

It is possible to investigate this stochastic mechanism somewhat further in this case of uniform (random) rate of entry of infectives. For if the outbreak of an epidemic is defined from the date of entry of infection which does not become extinct,

and, as suggested above, the ultimate reduction in number of susceptibles when an epidemic gets started is neglected, then the effects of each new infection are independent, and the chance of no epidemic from $t = 0$ up to time t may be written

$$(5.6) \quad \prod_{0 < u < t} [1 - \epsilon du + \epsilon du P(f_u)] = \exp \left\{ -\epsilon \int_0^t [1 - P(f_u)] du \right\},$$

where $P(f_u)$ denotes the chance of extinction with one new infection at time u when the number of susceptibles is S_u , and it is assumed that there is no infection

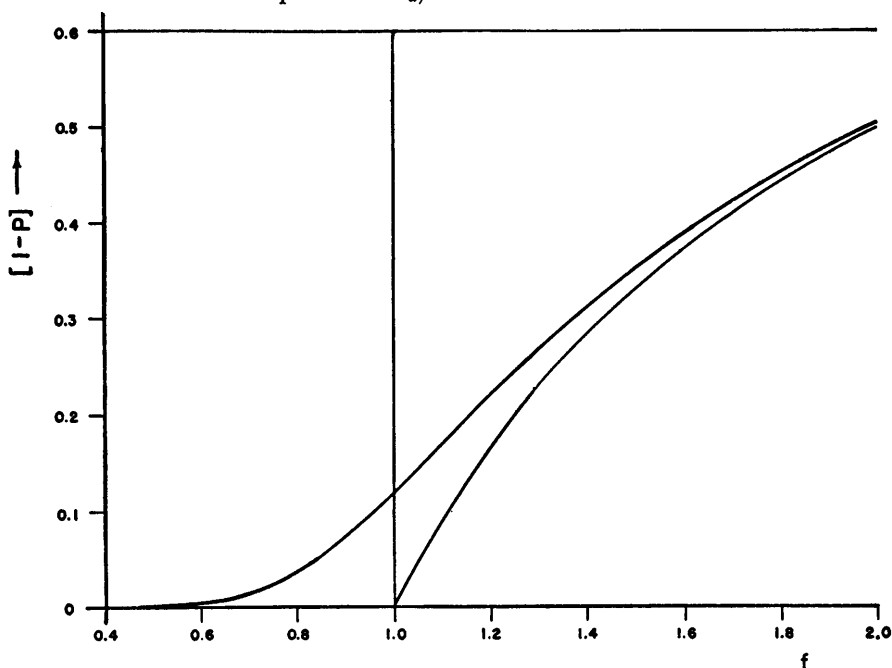


FIGURE 4

Stochastic model. The chance of a "major epidemic" plotted against $f = S_0 \lambda / \mu$. This chance is $1 - P$, where P is the extinction probability for a steady influx of susceptibles ($\eta = 5.84$). The chance $1 - P_0$ for a closed population is also shown for comparison (it is zero for $f \leq 1$).

present at $t = 0$. If in this rather rough argument the chance $P(f_u)$ is replaced by the further approximation $P_0(f_u)$, where P_0 neglects the effect on P of the continued influx of susceptibles after time u , it is possible to obtain an explicit solution [10]. The equivalent frequency law, in terms of $T = \lambda \nu t / \mu$, is

$$(5.7) \quad r(T - 1) T^{r-1} e^{-r(T-1)} dT \quad (r = \epsilon \mu / (\lambda \nu)),$$

with a mode at $T_m = 1 + 1/\sqrt{r}$. The mean value of T for the distribution (5.7) is shown in figure 5, and will be seen to be relatively insensitive to values of ϵ until ϵ becomes very small, when considerable delay in an outbreak may of course arise from the long intervals between successive entries of new infection. The mean value of t is directly proportional to $\sigma = \mu / (\lambda \nu)$ (being, for example, 2.25σ when $r = 2$, say), in contrast to the deterministic model with period approximately proportional to $\sqrt{\sigma \tau}$.

The interest of this stochastic model is in providing a mechanism which does generate a permanent and undamped series of outbreaks, with a period which, although not completely regular, will show a marked tendency to keep within a certain range of values. The time of an outbreak is associated closely with the epidemiologists' notion of a "threshold" of susceptible density, but this threshold (at $S_0 = \mu/\lambda$) can in the stochastic model represent a genuine discontinuity of effect, particularly if several infectives are introduced simultaneously into a temporarily closed population of susceptibles.

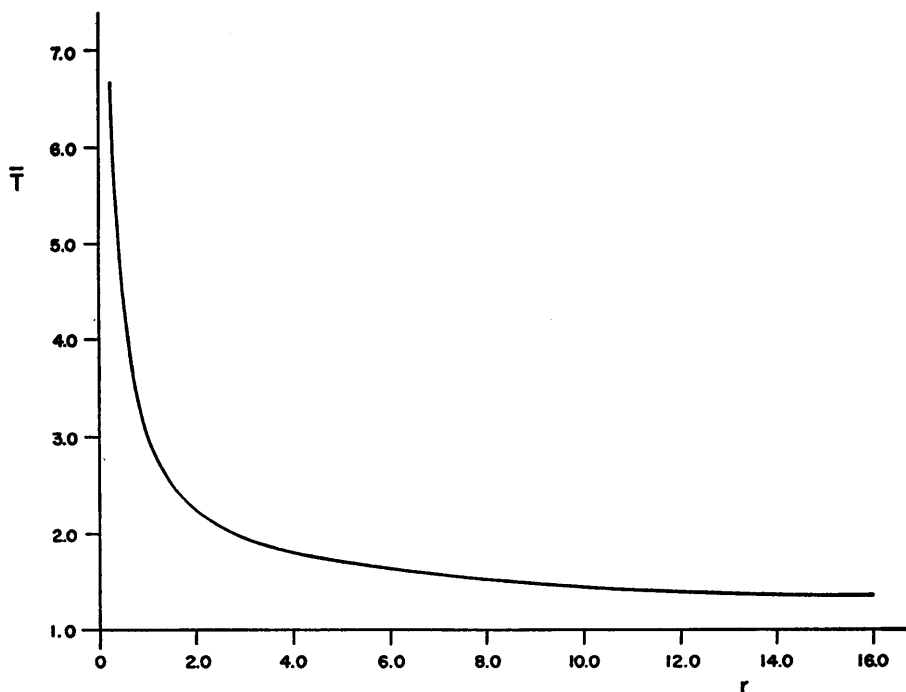


FIGURE 5
Stochastic model. The mean value \bar{T} of the delay before a new epidemic for the distribution of equation (5.7).

The contrast of the stochastic behavior of the model with the previous deterministic treatment is well illustrated on the I, S path diagram. In the deterministic model the line $I = 0$ is a possible path, but one never reached from inside the positive I, S quadrant. In the stochastic model, on the other hand, the actual path is extremely liable to drop to $I = 0$ when S drops below its threshold value (and the path then proceeds along $I = 0$ until new infection enters); unless m is large or other conditions (see the further discussion in section 7) are present favoring a small chance of extinction, the path is thus unlikely to get beyond the deterministic first epidemic cycle before degenerating to the line $I = 0$.

6. Artificial epidemic series

The type of epidemic behaviour discussed in the last section has been demonstrated by means of mock series generated with the aid of random numbers. Such demonstrations are useful in view of the difficulty of complete solution of the sto-

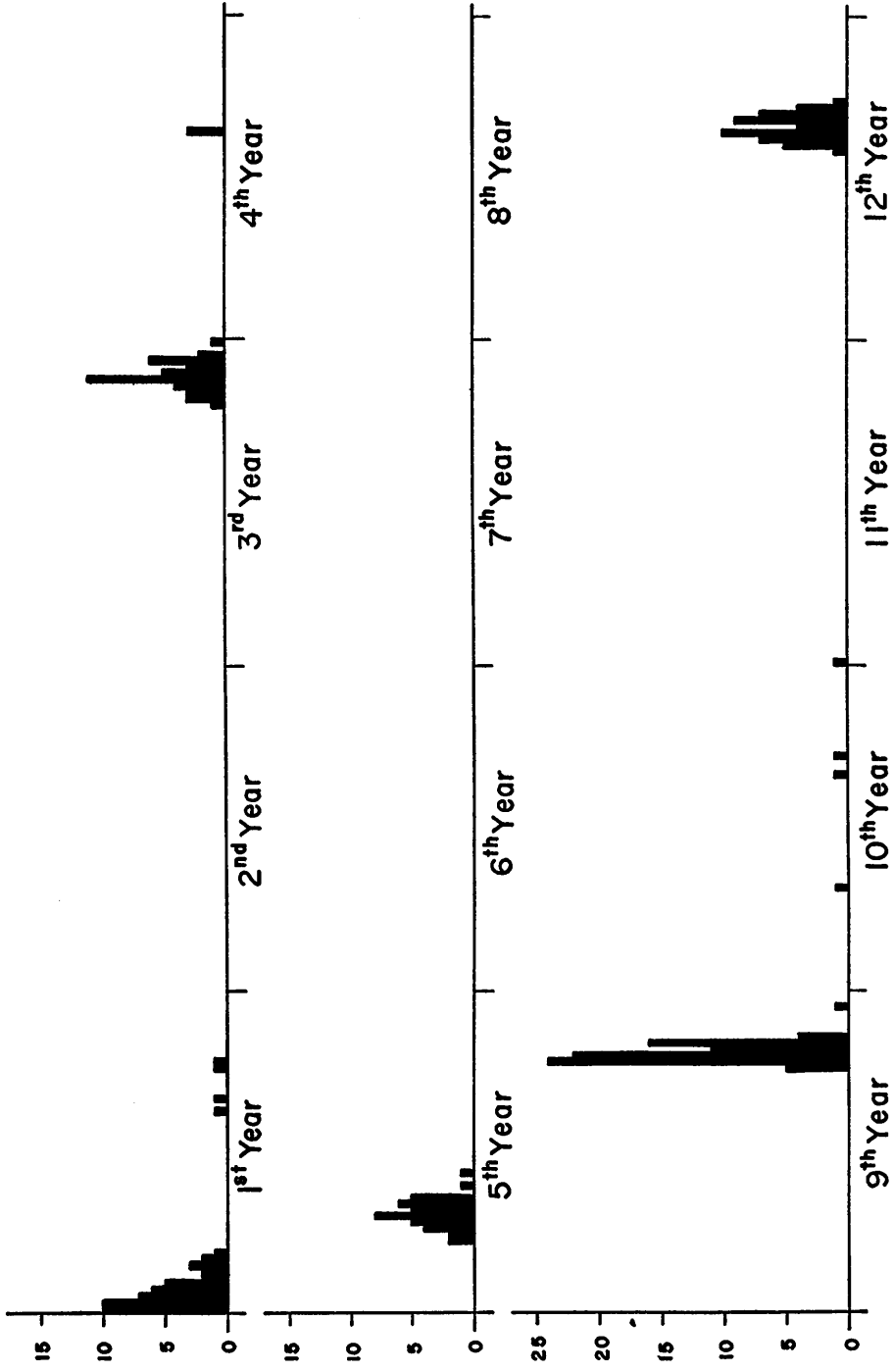


FIGURE 6

Stochastic model. Weekly notifications for artificial series (with $\mu = 1/2$, $\lambda = 0.02$, $\nu = 0.3525$, $\epsilon = 0.0225$, $I_0 = 3$, $S_0 = 50$).

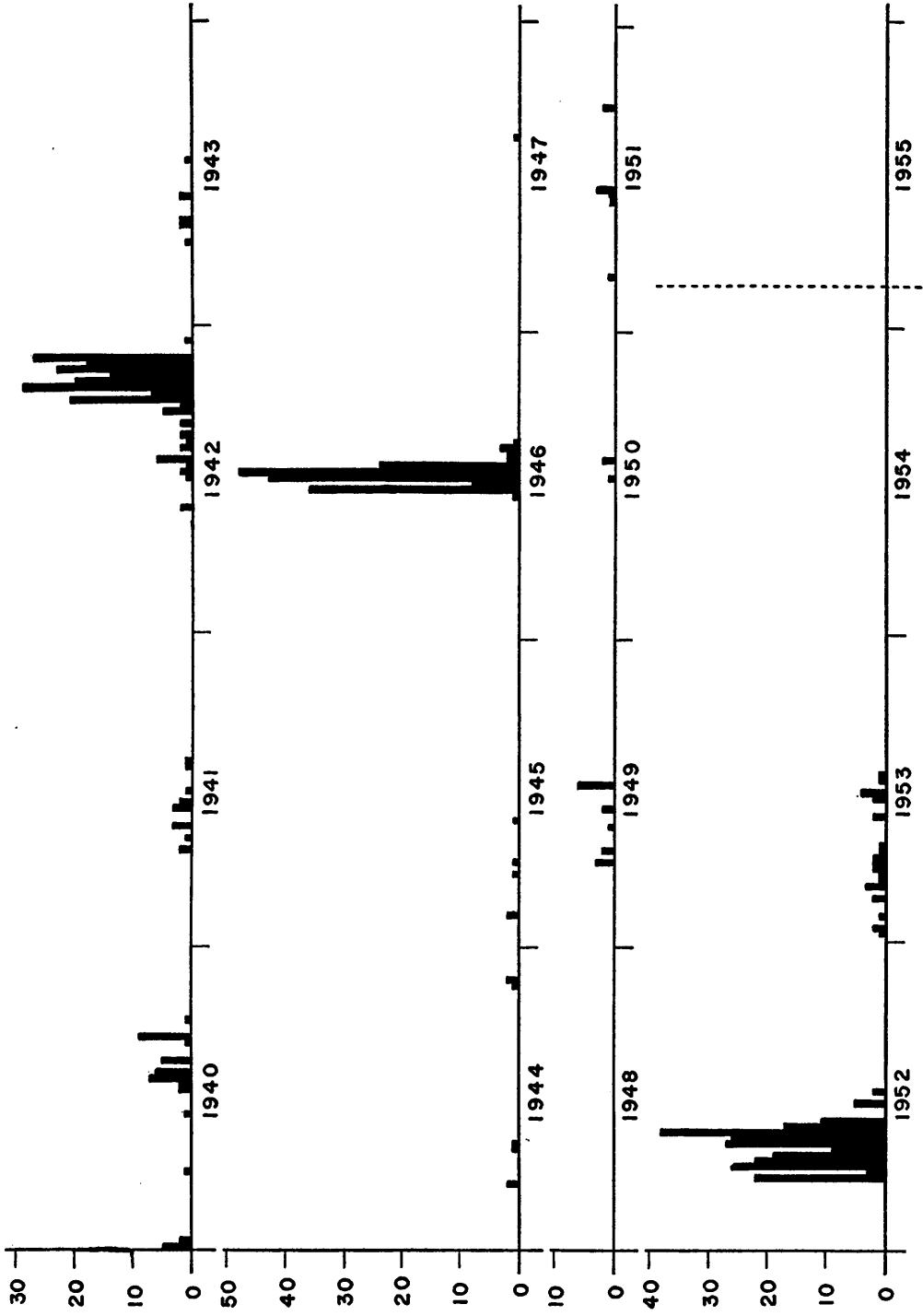


FIGURE 7

Actual epidemic series for measles (weekly notifications in Ffestiniog, Wales, from 1940 to 1955).

chastic equations. The method of constructing these artificial series has been previously indicated [8], the recurrent series there reported being intended to simulate successive measles outbreaks in a boarding school. The "notifications" in another series of this type are shown in figure 6. For convenience a random entry of susceptibles has been reverted to, and the simple model represented by equation (2.5) has been used. Standard random intervals with unit mean were listed (the method used was to take $T = \frac{1}{2}(X^2 + Y^2)$, where X and Y are standardized normal variables tabulated by Wold [40]), and as for any state I_t, S_t the interval before a new occurrence has an exponential distribution with mean $I/(\lambda I \nu S_t + \mu I_t + \nu + \epsilon)$, these random intervals were at each stage scaled accordingly. Moreover, when such an

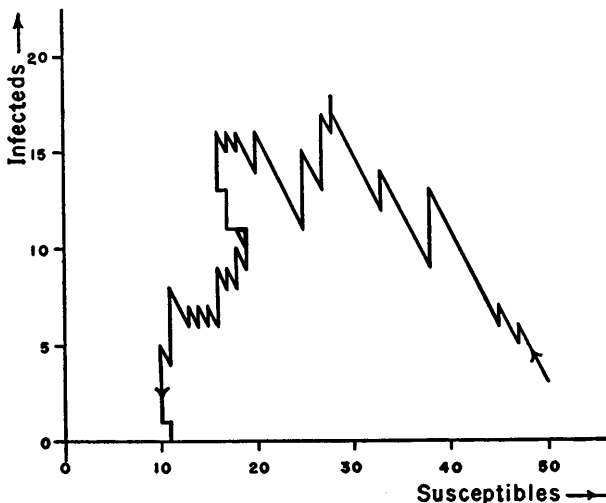


FIGURE 8

Stochastic model. Portion of I, S curve for the artificial series shown in figure 6.

event occurs, the relative chances of the four possible transitions (i) $I \rightarrow I + 1, S \rightarrow S - 1$, (ii) $I \rightarrow I - 1, S \rightarrow S$, (iii) $I \rightarrow I, S \rightarrow S + 1$, (iv) $I \rightarrow I + 1, S \rightarrow S$, are $\lambda IS : \mu I : \nu : \epsilon$, so that the actual event could also be determined with the aid of a table of random numbers [24].

The numerical constants for the series shown were $\mu = 1/2, \lambda = 0.02, \nu + \epsilon = 0.375, \epsilon = 0.0225$ (and $I_0 = 3, S_0 = 50$). These values give $\sigma = 70.9$, of the same order as before; but as ϵ is fairly low ($r = 1.60$), there may sometimes be delays before epidemics break out. The situation may be compared with that in a small and largely isolated community, and it is relevant to note that measles epidemics for such isolated communities manifest similar characteristics. (The measles notifications from 1940 to 1955 in the Welsh district of Ffestiniog, an isolated urban center with a total population of about 7,000, are shown for comparison in figure 7). The mock series is not extended enough for any precise comparison with the theoretical distribution (5.7) of the previous section, but the intervals between the mock epidemics appear reasonably consistent with it. (The distribution (5.7) is based on the approximate chance of extinction $P_0(f)$. The somewhat more accurate distribution based on $P(f)$ has subsequently been calculated by Miss Joyce Almond, and has, as might be expected, not quite such an abrupt beginning.)

A disadvantage with the continuous-time model is the rather laborious computation needed for the generation of these mock series, and in some cases it is more convenient to construct series based on the discrete time or chain model. This is in principle readily achieved in exactly the same way as for closed populations, the influx of susceptibles (and, if required, of infectives) being arranged at the end of each time unit. Such a model seems no easier to handle theoretically, but the computation of artificial series (with the aid of the National Bureau of Standards tables of the binomial probability distribution) is much more rapid. An example is given in the next section.

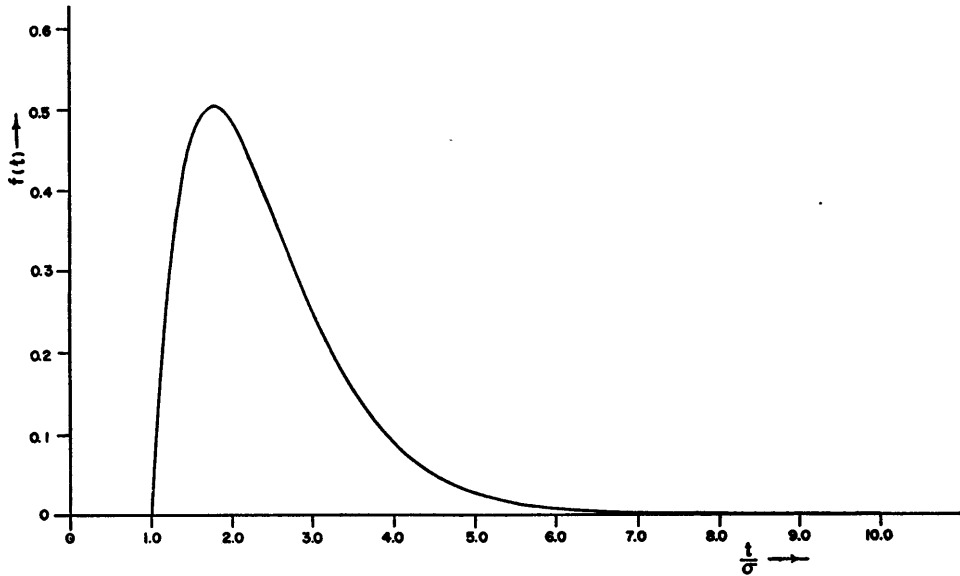


FIGURE 9
Frequency curve for the distribution of equation (5.7), with $r = 1.60$.

7. The chance of avoiding extinction

The new stochastic phenomenon of extinction or fade-out of infection is important, though at first sight this effect might diminish if large enough units or groups are considered, that is, if instead of boarding schools or small towns, larger urban districts are taken as units. This notion of a critical size is difficult to discuss quantitatively, but the following very crude argument may have some relevance. From the approximate extinction probability evaluated in section 5 (allowing for further influx of susceptibles) the order of magnitude of the number of infectives required to give not more than a 50 : 50 chance of extinction after half the epidemic cycle (for $\eta = 5.84$) is

$$\begin{aligned} f = 0, & & I = 2 \times 10^8, \\ f = \frac{1}{2}, & & I = 700, \\ f = \frac{2}{3}, & & I = 70, \end{aligned}$$

where f is the ratio of the number of susceptibles to its average value. As no reliable theoretical value for the swing in f is available, appeal at this point must be made to experience, which suggests that for measles the range is not more than $1\frac{1}{2}$ to $\frac{1}{2}$

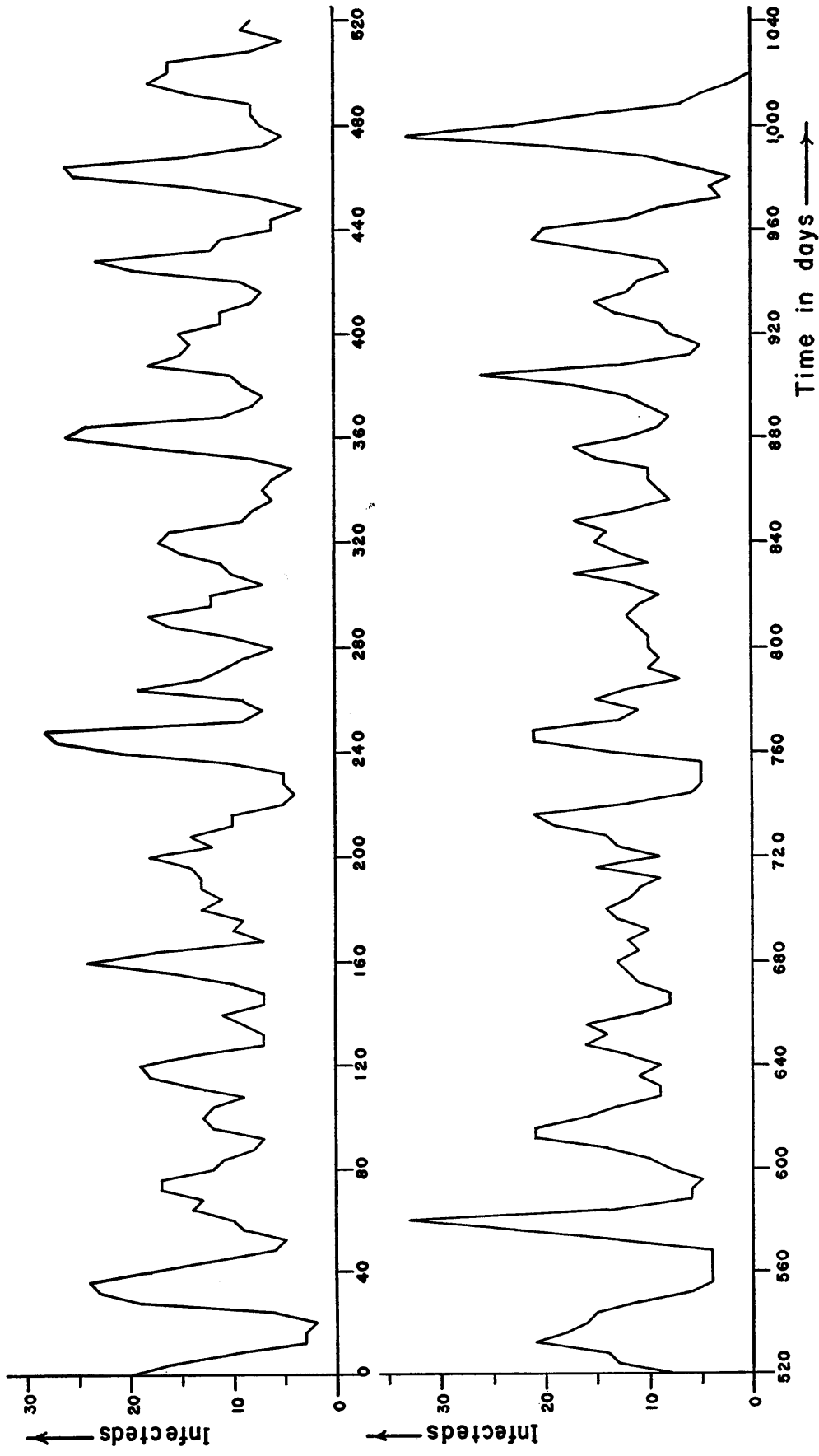


FIGURE 10

Stochastic model. Artificial series of numbers of infected (constructed on a "chain" basis with an incubation period of four days as time unit, at the end of which interval the chance of infection per infected person is 0.04, and influx of susceptibles 12; $S_0 = 25$, $I_0 = 20$).

(a 3 : 1 ratio). In this case as many as 700 infectives appear needed. Owing to the phase difference in the deterministic cycle between numbers of susceptibles and infectives, this number should be taken as the *average* number m of infectives, but even so would imply a rather larger town than Manchester, for which m appears to be rather less than 400. It is interesting to find that notifications do fade out over larger intervals than a fortnight for smaller towns like Preston or for subareas of Manchester, but for Manchester as a whole no extinction occurred throughout the whole of the period 1917–1951 for which the statistics were available. Thus these very rough quantitative estimates are perhaps partially, but not completely, confirmed by observation. The necessity of treating large-scale units adequately, as mosaics of hundreds of smaller units each with their own local epidemiological histories, becomes apparent.

Before this is attempted, however, the theoretical value of the above ideas will be demonstrated by a deliberate change of constants to facilitate the recurrence of epidemics without the need for re-entry of infection. Clearly we must reduce the value of η if the chance of extinction during each epidemic cycle is to be made small. It is not too easy to do this without imposing such heavy damping that an endemic equilibrium level results, for the damping factor from peak to peak tends to be inversely dependent on η . However, the compromise value of $\eta = 1.44$ was selected.

The new conditions do not of course any longer represent measles, but to render them of some interest they were chosen in relation to the observed epidemic series for ectromelia in mice reported in *Experimental Epidemiology* (see p. 70, [17]). The intake of susceptibles ν was 3 per day, the incubation period was 4 days, so that $\mu = 1/4$. The value of λ was taken as 0.01, in order to give a (deterministic) period of about 30 days, comparable to that observed. S_0 was 25, I_0 was 20, as in the actual series. The average susceptible population in the artificial series was then about 30; in the actual series the *susceptible* population, in contrast with the *total* population, was unknown, but probably rather higher (see p. 75). No very close agreement would be expected, but the authors' apparent rejection of the Soper model (at the beginning of p. 76), on the grounds that no oscillatory movement with the right period occurred, seems unjustifiable and at variance with their own figure 8 (opposite p. 70).

The artificial series, which was for convenience constructed on a "chain" basis with the incubation period of four days (see preceding paragraph for other constants) as time unit, is shown (for I_t) in figure 10. It ran for nearly three "years," or *about 30 cycles*, before extinction of infection occurred.

When conditions permit recurrent epidemics, either by (a) reentry of infection, or by (b) a quasi-stationarity, as in figure 10, before extinction finally occurs, it is evident that stochastic variability and the damping tendency determine together a mean amplitude of fluctuations, just as in more orthodox linear time-series theory. Under (a) it should be possible in principle to determine the equilibrium distribution of I and S from equation (2.5), for on heuristic grounds it obviously exists and satisfies the equation $H'\pi = 0$. However, even the moments seem difficult to find unless some further approximating assumption of near-normality is made. This can only be justified under the limiting conditions of m, n large, and small stochastic oscillations about the deterministic mean level, with the chance of extinction unimportant. It is perhaps worth stressing the orthodox time-series aspect of this limiting case, by the following (formal) derivation.

In place of the deterministic equations (2.14), we may write for the stochastic changes in a small time Δt

$$(7.1) \quad \begin{aligned} \Delta I_t &= -\mu I_t \Delta t + \lambda I_t S_t \Delta t - \Delta Z_1 + \Delta Z_2, \\ \Delta S_t &= \nu \Delta t - \lambda I_t S_t \Delta t - \Delta Z_2 + \Delta Z_3, \end{aligned}$$

where ΔZ_1 , ΔZ_2 , ΔZ_3 are Poisson variables with adjusted zero means and variances $\mu I_t \Delta t$, $\lambda I_t S_t \Delta t$ and $\nu \Delta t$ respectively. In terms of the reduced variables $i_t \equiv I_t/m$, $s_t \equiv S_t/n$, with $\tau \equiv \nu t$ (and $\Delta \tau$ for Δt), equations (7.1) become

$$(7.2) \quad \begin{aligned} m \Delta i_t &= -i_t \Delta \tau + i_t s_t \Delta \tau - \Delta Z_1 + \Delta Z_2 \\ n \Delta s_t &= \Delta \tau - i_t s_t \Delta \tau + \Delta Z_3 - \Delta Z_2, \end{aligned}$$

where the variances of ΔZ_1 , ΔZ_2 and ΔZ_3 are now $i_t \Delta \tau$, $i_t s_t \Delta \tau$ and $\Delta \tau$ respectively. These *nonlinear* stochastic equations are approximated as in the deterministic model, with $i_t \sim 1 + x_t$, $s_t \sim 1 + y_t$. Then the first approximation gives

$$(7.3) \quad \begin{aligned} m dx_t &= y_t d\tau - dZ_1 + dZ_2, \\ n dy_t &= -(x_t + y_t) d\tau + dZ_3 - dZ_2, \end{aligned}$$

orthodox bivariate linear time-series (with the variances of dZ_1 , dZ_2 and dZ_3 all $d\tau$). This readily yields the autocorrelations and cross-correlations of x_t and y_t in terms of the damped deterministic linear approximation given earlier. The equilibrium variances and covariance of x and y are moreover obtained by squaring and cross-multiplication of the equations equivalent to (7.3)

$$(7.4) \quad \begin{aligned} mx_{t+d\tau} &= mx_t + y_t d\tau - dZ_1 + dZ_2, \\ ny_{t+d\tau} &= y_t(n - d\tau) - x_t d\tau + dZ_3 - dZ_2 \end{aligned}$$

and averaging. This yields three equations for σ_x^2 , σ_y^2 and $\text{cov}(x, y)$, whence

$$(7.5) \quad \sigma_x^2 \sim \frac{n}{m^2} + \frac{1}{m}, \quad \sigma_y^2 \sim \frac{1}{n} + \frac{1}{m}, \quad \text{cov}(x, y) \sim -\frac{1}{m}.$$

These results will of course only be applicable under conditions for which σ_x^2 , σ_y^2 remain small.

8. The spatial or topographical factor

Coming now to the spatial effect, theoretical progress with this problem is hardly likely to be rapid when even the simpler case of a small unit has proved so intractable, but two questions at least demand an answer. The first is whether the tendency to a stable endemic level on the simpler deterministic theory still holds; if it does, the second question is to what extent the stochastic instability previously discussed now modifies the deterministic solution.

To obtain some information on the first question, the case of two interacting groups is worth consideration. The deterministic equations will be written down at once,⁶ though there is of course no difficulty in formulating the corresponding stochastic model. The interaction can be supposed due to actual diffusion or migration of susceptibles or infectives from one group to the other, or to a chance of infection to a susceptible of one group from an infective of the other, for example, by the infective making a visit to the other group and then returning. There does not appear to be any very vital difference between these two hypotheses, and for the time being both will be included.

The changes in the numbers x_1, y_1 of the infectives and susceptibles in the first group are thus supposed determined by the equations ($D \equiv d/dt$)

$$(8.1) \quad \begin{aligned} Dy_1 &= -(\lambda_1 x_1 + \lambda_2 x_2)y_1 + \nu + \beta(y_2 - y_1), \\ Dx_1 &= (\lambda_1 x_1 + \lambda_2 x_2)y_1 - \mu x_1 + \alpha(x_2 - x_1), \end{aligned}$$

and similarly for the second group, where β and α are diffusion or emigration rates for susceptibles and infectives. The equilibrium values of x_1, x_2, y_1, y_2 , are

$$(8.2) \quad m_1 = m_2 = \nu/\mu, \quad n_1 = n_2 = \mu/(\lambda_1 + \lambda_2);$$

and, if $x_r = m_r(1 + u_r), y_r = n_r(1 + v_r), r = 1, 2$, then for small fluctuations about equilibrium it will be found that

$$(8.3) \quad \begin{aligned} \left(D + \frac{1}{\sigma} + \beta \right) v_1 + \frac{\rho}{\sigma} u_1 - \beta v_2 + \frac{1 - \rho}{\sigma} u_2 &= 0 \\ -\mu v_1 + (D + \mu[1 - \rho] + \alpha)u_1 - [\mu(1 - \rho) + \alpha]u_2 &= 0, \end{aligned}$$

and similarly with the suffices 1 and 2 interchanged, where $\sigma = \mu/[\nu(\lambda_1 + \lambda_2)], \rho = \lambda_1/(\lambda_1 + \lambda_2)$. These linearised equations yield the same equations as before for $v = v_1 + v_2, u = u_1 + u_2$, with $\lambda = \lambda_1 + \lambda_2$. For $v' = v_1 - v_2, u' = u_1 - u_2$, moreover, the equations are

$$(8.4) \quad \begin{aligned} \left(D + \frac{1}{\sigma} + 2\beta \right) v' - \frac{1 - 2\rho}{\sigma} u' &= 0 \\ -\mu v' + [D + 2\mu(1 - \rho) + 2\alpha]u' &= 0 \end{aligned}$$

with determinantal equation in D

$$(8.5) \quad D^2 + D \left(\frac{1}{\sigma} + 2\mu(1 - \rho) + 2\alpha + 2\beta \right) + \frac{\mu}{\sigma} + \frac{2\alpha}{\sigma} + 4\mu(1 - \rho)\beta + 4\alpha\beta = 0.$$

As all the terms in the coefficient of D are positive, the damping of the differences v', u' will be greater than the damping of v, u .

⁶ See a recent paper by Rushton and Mautner [30] on a deterministic model for several communities. The different and more restricted conditions they assumed ensure, however, no overlap with my own discussion.

This result indicates that a similar result should follow for several interconnected units, but it is perhaps advisable to check this in the extreme case when the units have become relatively so small that the system spatially may be considered continuous. In this case the equations (8.1) are assumed replaced by

$$(8.6) \quad \begin{aligned} Dy(\mathbf{r}, t) &= -\lambda y(x + \rho \nabla^2 x) + \nu + \beta \nabla^2 y, \\ Dx(\mathbf{r}, t) &= \lambda y(x + \rho \nabla^2 x) - \mu x + \alpha \nabla^2 x, \end{aligned}$$

where \mathbf{r} stands for the spatial coordinates (ξ, η) , x and y are now spatial "densities" of infected and susceptible, and $\nabla^2 \equiv \partial^2/\partial\xi^2 + \partial^2/\partial\eta^2$. (The same letters ρ , α and β are used as in (8.1), but of course with different interpretations; the assumption underlying the appearance of the operator ∇^2 is that infection and migration are local and isotropic in character. Homogeneity in space is implied by the coefficients λ , ρ , etc., being independent of \mathbf{r} .) The linearised form of these equations is

$$(8.7) \quad \begin{aligned} \left(D + \frac{1}{\sigma} - \beta \nabla^2 \right) v(\mathbf{r}, t) + \frac{1}{\sigma} (1 + \rho \nabla^2) u(\mathbf{r}, t) &= 0 \\ -\mu v(\mathbf{r}, t) + (D - [\alpha + \mu\rho] \nabla^2) u(\mathbf{r}, t) &= 0. \end{aligned}$$

Equations of this type have been studied by Turing [34] as possible models for biological growth and form, and in general may lead to undamped waves in space and time, but it is not difficult to show that the particular equations (8.7) still yield waves damped in time. Consider a component of the general solution of (8.7) of the form $\exp(at + ib\xi + ic\eta)$, where $i \equiv \sqrt{-1}$, and b and c are real. Then from (8.7)

$$(8.8) \quad \left[a + \frac{1}{\sigma} + \beta(b^2 + c^2) \right] [a + (\alpha + \mu\rho)(b^2 + c^2)] + \frac{\mu}{\sigma} [1 - \rho(b^2 + c^2)] = 0.$$

If $a = a_1 + ia_2$, the imaginary part of this equation yields

$$(8.9) \quad a_2 \left[\frac{1}{\sigma} + \beta(b^2 + c^2) + (\alpha + \mu\rho)(b^2 + c^2) \right] + 2a_1 a_2 = 0,$$

or (for $a_2 \neq 0$) $b^2 + c^2 \geq 0$ provided

$$(8.10) \quad \frac{1}{\sigma} + 2a_1 \leq 0,$$

or

$$(8.11) \quad a_1 \leq -\frac{1}{2\sigma},$$

that is, the damping in time is at least as great as in the corresponding nonspatial model.

A particular situation of some importance occurs when infection is introduced at one point in an area where it has previously been extinct. In this case, the equation for x in (8.3) may be written

$$(8.12) \quad Dx = (\lambda y - \mu)x + (\alpha + \lambda\rho y) \nabla^2 x,$$

where y is treated as temporarily constant. This is a standard diffusion equation whose solution has Fourier transform

$$(8.13) \quad \begin{aligned} M &= \iint e^{i(\theta_1 \xi + \theta_2 \eta)} x \, d\xi \, d\eta \\ &= \exp \{ (\lambda y - \mu)t - (\alpha + \lambda \rho y)(\theta_1^2 + \theta_2^2)t \} . \end{aligned}$$

This solution represents a Gaussian distribution normalized to $\exp (\lambda y - \mu)t$. The logarithm of

$$(8.14) \quad x_R = \int_{\xi^2 + \eta^2 \geq R^2} x \, d\xi \, d\eta$$

is, apart from a constant,

$$(8.15) \quad -\frac{1}{2} \frac{R^2}{\gamma^2 t} + (\lambda y - \mu)t \quad \left(\frac{1}{2} \gamma^2 = \alpha + \lambda \rho y \right) ,$$

which corresponds to any arbitrarily chosen but constant number when $R^2/t^2 - 2(\lambda y - \mu)\gamma^2 = 0$ except for an amount which decreases to zero as t increases. In this sense [14] the limiting velocity of propagation of infection is $\gamma\sqrt{2(\lambda y - \mu)}$. For example, if $f = \lambda y/\mu = 2$, $u = \frac{1}{2}$, this velocity is just γ . In the discussion on Soper's paper [32], an interval of 24 weeks was suggested by Dr. Halliday as the time for measles to spread across Glasgow. If the relevant area is, say, five square miles, this implies (in so far as these rough calculations are applicable) an effective diffusion of infectives with parameter γ (standard deviation) ~ 0.05 miles $/\sqrt{\text{week}}$

Two limitations of the above theoretical discussion are (a) its neglect of stochastic fluctuations, (b) its applicability is in any case restricted to the initial spread of infection. On this last point some rather cumbrous numerical calculations on the deterministic basis were carried out (for the one-dimensional case), but these calculations will not be considered here in any detail, especially as the deterministic approach becomes even less appropriate for a finite model extended in space. It might, however, be noted that the propagation velocity of the initial infection wave agreed reasonably well with the theoretical formula.

When the full stochastic model is considered, there is the same alternative as with the deterministic model of (i) taking an arrayed set of discrete units, or (ii) of formulating the equations precisely for a continuous area. The first choice may often be no more an approximation than the second, for epidemics are sometimes accelerated by the concentration of susceptibles in schools or other collective units, whereas with the second method (to be discussed in section 9) a mathematical model will usually have to be made spatially homogeneous. Mock stochastic epidemic series are most conveniently computed for (i), and several small-scale series in one space-dimension have been constructed. One of these was designed to check the adequacy of the deterministic formula for the propagation velocity of infection, and gave an average velocity of 0.24, against the theoretical value of 0.28. (For $S_0 = 100$ per unit, $\lambda = 0.01$, $\nu = 0$ and a diffusion coefficient for infectives between neighbouring units of 0.08, cross-infection being only possible by such diffusion.)

The theoretical analysis at the end of section 7 on small-scale stochastic fluctuations has also been extended to the case of several interconnected regions, and, as

might be expected, the equilibrium variances of fluctuations in the total numbers of infected and susceptibles are increased by diffusion between the regions. However, the neglect of extinction for the individual regions makes such an analysis, at least for conditions simulating measles, seem of doubtful practical interest, and the more promising project at present for studying the long-term time behaviour of epidemic series extended over space is by more elaborate Monte Carlo computations (such a project was planned some time ago for the electronic computer at Manchester, but unfortunately it has not yet been possible to have any results available).

9. Specification in terms of point processes

In spite of this difficulty of studying long-term behaviour, a precise point-process specification of the complete spatial problem will now be given. It permits further generalization at least of the approximate solution for the initial spread of infection.

The theoretical technique employed [9] is that of a "point" stochastic process $N(\mathbf{r}, t)$ depending on both the time and the spatial coordinates $\mathbf{r} \equiv (\xi, \eta)$, and of its probability-generating functional

$$(9.1) \quad \pi(z(\mathbf{r}), t) \equiv E \left\{ \exp \left[\int_{\mathbf{r}} \log z(\mathbf{r}) dN(\mathbf{r}, u) \right] \right\}.$$

For both infectives and susceptibles, the *two* auxiliary variables w, z in $\pi(z, w)$ are replaced by the two functions $w(\mathbf{r}), z(\mathbf{r})$. The possible transitions will now be assumed to be represented by the scheme

<i>Transition</i>	<i>Rate</i>	<i>Operator</i>
$w(\mathbf{r}) \rightarrow 1$	μ	$\frac{\partial}{\partial w(\mathbf{r}) d\mathbf{r}}$
$1 \rightarrow z(\mathbf{r})$	$\nu d\mathbf{r}$	1
$w(\mathbf{s})z(\mathbf{r}) \rightarrow w(\mathbf{s})w(\mathbf{r})$	λ	$\frac{\partial^2}{\partial w(\mathbf{s})\partial z(\mathbf{r}) d\mathbf{r} d\mathbf{s}}$
$w(\mathbf{r}) \rightarrow w(\mathbf{s})$	$\alpha d\mathbf{s}$	$\frac{\partial}{\partial w(\mathbf{r}) d\mathbf{r}}$
$z(\mathbf{r}) \rightarrow z(\mathbf{s})$	$\beta d\mathbf{s}$	$\frac{\partial}{\partial z(\mathbf{r}) d\mathbf{r}}$

with an entry of new infection, if also included, given by

$$1 \rightarrow w(\mathbf{r}) \quad \epsilon d\mathbf{r} \quad 1.$$

The notation for the functional derivatives follows Hopf [20], and indicates the assumed order of magnitude of the associated moment densities, for example, for one quantity $N(\mathbf{r}, t)$,

$$(9.2) \quad E\{N(\mathbf{r}, t)\} = \int_{\mathbf{r}} E\{dN(\mathbf{r}, t)\} = \int_{\mathbf{r}} f_1(\mathbf{r}, t) d\mathbf{r} \\ = \int_{\mathbf{r}} \left[\frac{\partial \pi(z(\mathbf{u}), t)}{\partial z(\mathbf{r}) d\mathbf{r}} \right]_{z(\mathbf{u})=1} d\mathbf{r}.$$

The quantities μ , ν (and ϵ) are now in general functions of t and \mathbf{r} , and λ , α and β of t , \mathbf{r} and \mathbf{s} , but it will be assumed in particular that the process is both temporally and spatially homogeneous, so that μ , ν (and ϵ) are constants, and λ , α , β functions of the vector displacement $\mathbf{r} - \mathbf{s}$. For simplicity the process will also when necessary be assumed isotropic, so that only the magnitude $|\mathbf{r} - \mathbf{s}|$ of the displacement is involved. It might be noticed that if the further complication to rates depending on the "ages" of the various infections to allow, for example, for incubation periods, is required, this can also be included in the same general formalism. The formal equation for the above scheme (with $\epsilon = 0$) is then immediately written down as

$$\begin{aligned}
 (9.3) \quad \frac{\partial \pi}{\partial t} &= \int \mu [1 - w(\mathbf{r})] \left[\frac{\partial \pi}{\partial w(\mathbf{r})} \frac{d\mathbf{r}}{d\mathbf{r}} \right] d\mathbf{r} + \int \nu (z(\mathbf{r}) - 1) \pi d\mathbf{r} \\
 &+ \iint \lambda (\mathbf{r} - \mathbf{s}) w(\mathbf{s}) [w(\mathbf{r}) - z(\mathbf{r})] \left[\frac{\partial^2 \pi}{\partial w(\mathbf{s}) \partial z(\mathbf{r})} \frac{d\mathbf{s} d\mathbf{r}}{d\mathbf{s} d\mathbf{r}} \right] d\mathbf{s} d\mathbf{r} \\
 &+ \iint \alpha (\mathbf{r} - \mathbf{s}) [w(\mathbf{s}) - w(\mathbf{r})] \left[\frac{\partial \pi}{\partial w(\mathbf{r})} \frac{d\mathbf{r}}{d\mathbf{r}} \right] d\mathbf{s} d\mathbf{r} \\
 &+ \iint \beta (\mathbf{r} - \mathbf{s}) [z(\mathbf{s}) - z(\mathbf{r})] \left[\frac{\partial \pi}{\partial z(\mathbf{r})} \frac{d\mathbf{r}}{d\mathbf{r}} \right] d\mathbf{s} d\mathbf{r} .
 \end{aligned}$$

As in the nonspatial formulation, progress with this equation is possible in the initial stages of infection in an area, when the number of susceptibles may be treated as temporarily constant. In this case π is taken to be a functional of $w(\mathbf{r})$ only, and the actual number of susceptibles is replaced by a constant (approximating) density $n(\mathbf{r})$, which will further be assumed uniform, that is, $n(\mathbf{r}) = n$. Equation (9.3) is then replaced by the stochastic analogue of (8.2), namely,

$$\begin{aligned}
 (9.4) \quad \frac{\partial \pi}{\partial t} &= \int \mu [1 - w(\mathbf{r})] \left[\frac{\partial \pi}{\partial w(\mathbf{r})} \frac{d\mathbf{r}}{d\mathbf{r}} \right] d\mathbf{r} \\
 &+ n \iint \lambda (\mathbf{r} - \mathbf{s}) w(\mathbf{s}) [w(\mathbf{r}) - 1] \left[\frac{\partial \pi}{\partial w(\mathbf{s})} \frac{d\mathbf{s}}{d\mathbf{s}} \right] d\mathbf{s} d\mathbf{r} \\
 &+ \iint \alpha (\mathbf{r} - \mathbf{s}) [w(\mathbf{s}) - w(\mathbf{r})] \left[\frac{\partial \pi}{\partial w(\mathbf{r})} \frac{d\mathbf{r}}{d\mathbf{r}} \right] d\mathbf{s} d\mathbf{r} .
 \end{aligned}$$

The process as formulated in (9.4) is of "multiplicative" type, and hence if $\pi(\mathbf{v})$ denotes the probability-generating functional resulting from *one* infected individual at \mathbf{v} at $t = 0$ (and none elsewhere), $\pi(\mathbf{v})$ satisfies also the equation

$$\begin{aligned}
 (9.5) \quad \frac{\partial \pi(\mathbf{v})}{\partial t} &= \mu [1 - \pi(\mathbf{v})] + n \int \lambda (\mathbf{u} - \mathbf{v}) \pi(\mathbf{v}) [\pi(\mathbf{u}) - 1] d\mathbf{u} \\
 &+ \int \alpha (\mathbf{v} - \mathbf{u}) [\pi(\mathbf{u}) - \pi(\mathbf{v})] d\mathbf{u} .
 \end{aligned}$$

This last equation is useful in yielding the chance of extinction $p_t(\mathbf{v})$ of infection,

for putting $w(\mathbf{r}) \equiv 0$, it gives

$$(9.6) \quad \frac{\partial p_t(\mathbf{v})}{\partial t} = \mu[1 - p_t(\mathbf{v})] + n \int \lambda(\mathbf{u} - \mathbf{v}) p_t(\mathbf{v}) (p_t(\mathbf{u}) - 1) d\mathbf{u} \\ + \int \alpha(\mathbf{v} - \mathbf{u}) [p_t(\mathbf{u}) - p_t(\mathbf{v})] d\mathbf{u},$$

or, with the homogeneity assumption $p_t(\mathbf{v}) = p_t(\mathbf{u}) = p_t$,

$$(9.7) \quad \frac{\partial p_t}{\partial t} = \mu(1 - p_t) + n p_t (p_t - 1) n \int \lambda(\mathbf{u}) d\mathbf{u} \\ = \mu(1 - p_t) + n p_t (p_t - 1) \Lambda_0,$$

say. This equation is the equation for the chance of extinction in a simple birth-and-death rate with death-rate μ and birth-rate $n\Lambda_0$, and has solution

$$(9.8) \quad p_t = \frac{\mu[e^{(n\Lambda_0 - \mu)t} - 1]}{n\Lambda_0 e^{(n\Lambda_0 - \mu)t} - \mu}.$$

Equation (9.4) may be used to generate and solve equations for the moment densities. In particular, if

$$(9.9) \quad f_1(\mathbf{r}, t) d\mathbf{r} \equiv E\{dI(\mathbf{r}, t)\},$$

where $dI(\mathbf{r}, t)$ is the random or stochastic number of infectives at time t in the infinitesimal element $d\mathbf{r}$,

$$(9.10) \quad \frac{\partial f_1(\mathbf{r})}{\partial t} = -\mu f_1(\mathbf{r}) + n \int \lambda(\mathbf{r} - \mathbf{s}) f_1(\mathbf{s}) d\mathbf{s} - f_1(\mathbf{r}) \int \alpha(\mathbf{r} - \mathbf{s}) d\mathbf{s} + \int \alpha(\mathbf{s} - \mathbf{r}) f_1(\mathbf{s}) d\mathbf{s}.$$

If

$$(9.11) \quad M_1(\boldsymbol{\theta}) = \iint e^{i(\boldsymbol{\theta}_1 x_1 + \boldsymbol{\theta}_2 x_2)} f_1(\mathbf{r}) dx_1 dx_2,$$

where $\mathbf{r} \equiv (x_1, x_2)$, and correspondingly Λ, A are the Fourier transforms of λ, α , then the Fourier transform of equation (9.10) gives

$$(9.12) \quad \frac{\partial M_1(\boldsymbol{\theta})}{\partial t} = [-\mu + n\Lambda(\boldsymbol{\theta}) - A_0 + A(-\boldsymbol{\theta})] M_1(\boldsymbol{\theta}),$$

where $\Lambda_0 = \Lambda(\boldsymbol{\theta}), A_0 = A(\boldsymbol{\theta})$, with solution

$$(9.13) \quad M_1(\boldsymbol{\theta}) = \exp \{ [n\Lambda(\boldsymbol{\theta}) - \mu - A_0 + A(-\boldsymbol{\theta})] t \}.$$

In the particular case of purely local (isotropic) infection and diffusion, this solution corresponds to the deterministic solution (8.12).

It should be noticed that while $f_1(\mathbf{r})$ is an average density which includes the zero contribution from cases where total extinction of infection has occurred, the average density f'_1 conditional on no total extinction up to time t is given by

$$(9.14) \quad f'_1 = f_1 / (1 - p_t)$$

and the "velocity of propagation" associated with f'_1 will ultimately be the same as that referred to in section 8.

The long-term stability of the system in the sense of avoiding total extinction of infection cannot unfortunately be discussed in terms of the chance p_i above, which applies only to the initial stages of an epidemic. (Moreover, the system treated would have to be finite and hence no longer homogeneous.) In the nonspatial model, a further rough argument was used to follow the probable course of events developing from the situation *after* an epidemic had occurred, when the susceptibles were at their lowest value. Such an argument can hardly be extended to the more general spatial model, for the epidemic will show some heterogeneity of phase in different places, and the only immediate conclusion appears to be the qualitative one that such heterogeneity should presumably facilitate the preservation of infection.

10. Concluding remarks

When the theoretical results so far reached are reviewed, it can hardly be claimed that any complete quantitative picture has yet been achieved; but I suggest that, in addition to the more classical (deterministic) results, some important further features have begun, at least qualitatively, to emerge. The phenomenon of extinction or fade-out of infection (impossible to consider in deterministic models) largely decides whether the deterministic approximation has any relevance or not to the quasiperiodicity of recurrent epidemics. This chance of extinction, which alters with the characteristics of the model, is high for models appropriate to measles in comparatively small communities. This implies that the "two-year cycle" sometimes claimed as an inherent feature of measles, and an observed fact for many large urban areas, will be replaced by a longer average interval between epidemics for smaller communities (also an observed fact,⁷ familiar of course to many epidemiologists).

The dangers of comparing small-scale experimental studies with deterministic models (or, for that matter, with large-scale field studies) are now seen to be even greater than I envisaged in 1949 (see p. 227, [7]); for example, the effect of scale will be more important for infections liable to extinction than for others, and hence very dependent on the particular type of infection under study.

In spite of the availability of a precise specification, the difficulties of a complete theoretical treatment, including spatial effects, for recurrent epidemics in urban areas (with interacting units in and out of phase, some with infection present and some not) need not be further emphasized. Like the many other problems arising in epidemic theory, it remains a very obstinate, but, one hopes, not a permanently intractable, problem.

REFERENCES

- [1] P. ARMITAGE, Bibliography on "Medical Statistics and Epidemiology," *Jour. Roy. Stat. Soc.*, Ser. A, Vol. 114 (1951), pp. 533-535.
- [2] N. T. J. BAILEY, "A simple stochastic epidemic," *Biometrika*, Vol. 37 (1950), pp. 193-202.
- [3] ———, "The total size of a general stochastic epidemic," *Biometrika*, Vol. 40 (1953), pp. 177-185.

⁷ A more quantitative observational study of the relation between measles periodicity and community size is in hand.

- [4] N. T. J. BAILEY, "A statistical method of estimating the periods of incubation and infection of an infectious disease," *Nature*, Vol. 174 (1954), pp. 139-140.
- [5] ———, "Some problems in the statistical analysis of epidemic data," *Jour. Roy. Stat. Soc.*, Ser. B, Vol. 17 (1955), pp. 35-68.
- [6] M. S. BARTLETT, *Stochastic Processes*, (mimeographed notes on a course given at the University of North Carolina in the Fall Quarter, 1946).
- [7] ———, "Some evolutionary stochastic processes," *Jour. Roy. Stat. Soc.*, Ser. B, Vol. 11 (1949), pp. 211-229.
- [8] ———, "Stochastic processes or the statistics of change," *Applied Statistics*, Vol. 2 (1953), pp. 44-64.
- [9] ———, "Processus stochastiques ponctuels," *Ann. de l'Inst. Henri Poincaré*, Vol. 14 (1954), fasc. 1, pp. 35-60.
- [10] ———, *An Introduction to Stochastic Processes*, Cambridge, Cambridge University Press, 1955.
- [11] A. T. BHARUCHA-REID, *An Introduction to the Stochastic Theory of Epidemics and Some Related Statistical Problems*, to be published by the U. S. Air Force School of Aviation Medicine, Randolph Field, Texas.
- [12] J. BROWNLEE, "An investigation into the periodicity of measles epidemics in London in 1703 to the present day by the method of the periodogram," *Phil. Trans. Roy. Soc. Lond.*, Ser. B, Vol. 208 (1917), pp. 225-250.
- [13] E. A. CHEESEMAN, *Epidemics in Schools, Medical Research Council Special Report Series No. 271* (1950).
- [14] R. A. FISHER, "The wave of advance of advantageous genes," *Annals of Eugenics*, Vol. 7 (1937), pp. 355-375.
- [15] F. G. FOSTER, "A note on Bailey's and Whittle's treatment of a general stochastic epidemic," *Biometrika*, Vol. 42 (1955), pp. 123-125.
- [16] M. GREENWOOD, *Epidemics and Crowd Diseases*, London, Williams and Norgate, Ltd., 1935.
- [17] M. GREENWOOD, A. BRADFORD HILL, W. W. C. TOPLEY and J. WILSON, *Experimental Epidemiology, Medical Research Council Special Report Series No. 209* (1936).
- [18] R. E. HOPE SIMPSON, "The period of transmission in certain epidemic diseases," *The Lancet*, Vol. 255 (1948), pp. 755-760.
- [19] ———, "Infectiousness of communicable diseases in the household," *The Lancet*, Vol. 263 (1952), pp. 549-554.
- [20] E. HOPF, "Statistical hydromechanics and functional calculus," *Jour. Rational Mech. and Anal.*, Vol. 1 (1952), pp. 87-124.
- [21] D. G. KENDALL, "On the generalized birth-and-death process," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 1-15.
- [22] ———, "An artificial realization of a simple 'birth-and-death' process," *Jour. Roy. Stat. Soc.*, Ser. B, Vol. 12 (1950), pp. 116-119.
- [23] ———, "Deterministic and stochastic epidemics in closed populations," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, Berkeley and Los Angeles, University of California Press, 1956, pp. 149-165.
- [24] M. G. KENDALL and B. BABINGTON SMITH, *Tables of Random Sampling Numbers, Tracts for Computers No. XXIV*, Cambridge, The University Press, 1951.
- [25] W. O. KERMACK and A. G. MCKENDRICK, "Contributions to the mathematical theory of epidemics," *Proc. Roy. Soc. Lond.*, Ser. A, Vol. 115 (1927), pp. 700-721; Vol. 138 (1932), pp. 55-83 and Vol. 141 (1933), pp. 94-122.
- [26] L. LINNERT, "A statistical report on measles notifications in Manchester, 1917-1951," unpublished University of Manchester report.
- [27] A. G. MCKENDRICK, "Applications of mathematics to medical problems," *Proc. Edin. Math. Soc.*, Vol. 44 (1926), pp. 98-130.
- [28] *Tables of the Binomial Probability Distribution*. National Bureau of Standards Applied Mathematics Series No. 6 (1950).
- [29] A. RAPOPORT, "Nets with distance bias," *Bull. Math. Biophysics*, Vol. 13 (1951), pp. 85-91.
- [30] S. RUSHTON and A. J. MAUTNER, "The deterministic model of a simple epidemic for more than one community," *Biometrika*, Vol. 42 (1955), pp. 126-132.

- [31] R. E. SERFLING, "Historical review of epidemic theory," *Human Biology*, Vol. 24 (1952), pp. 145-166.
- [32] H. E. SOPER, "The interpretation of periodicity in disease prevalence," *Jour. Roy. Stat. Soc.*, Ser. A, Vol. 92 (1929), pp. 34-61.
- [33] P. STOCKS and M. KARN, "A study of the epidemiology of measles," *Annals of Eugenics*, Vol. 3 (1928), pp. 361-398.
- [34] A. M. TURING, "The chemical basis of morphogenesis," *Phil. Trans. Roy. Soc. Lond.*, Ser. B, Vol. 237 (1952), pp. 37-72.
- [35] P. WHITTLE, "Certain nonlinear models of population and epidemic theory," *Skand. Aktuar.*, Vol. 35 (1952), pp. 211-222.
- [36] ———, "The outcome of a stochastic epidemic—a note on Bailey's paper," *Biometrika*, Vol. 42 (1955), pp. 116-122.
- [37] E. B. WILSON, C. BENNETT, M. ALLEN and J. WORCESTER, "Measles and scarlet fever in Providence, R. I., 1929-1934, with respect to age and size of family," *Proc. Amer. Phil. Soc.*, Vol. 80 (1939), pp. 357-476.
- [38] E. B. WILSON and J. WORCESTER, "Damping of epidemic waves," *Proc. Nat. Acad. Sci.*, Vol. 31 (1945), pp. 294-298.
- [39] ———, "The spread of an epidemic," *Proc. Nat. Acad. Sci.*, Vol. 31 (1945), pp. 327-333. (Further epidemiological papers by Wilson and Worcester are listed in [1].)
- [40] H. WOLD, *Random Normal Deviates, Tracts for Computers*, No. XXV, Cambridge, The University Press, 1948.