Chaos, Population Biology, and Epidemiology: Some Research Implications

P. PHILIPPE[1]

Abstract In this article I aim to provide some feeling of the new paradigm of disease causation (chaos) as it applies to the field of population biology and epidemiology. A secondary objective is to show, with the aid of qualitative methods, how one can approach chaos in time-series data. The multifactorial stochastic paradigm of causation is contrasted with the new deterministic approach. This approach is embedded in the theory of nonlinear system dynamics. Chaos implies that randomness is intrinsic to a nonlinear deterministic system; this is true despite the extent of knowledge of the intervening causes and, ultimately, despite determinism. Three research avenues are discussed in depth from the standpoint of chaos theory. First, the topic of sporadic epidemics is dealt with. I argue that the space-time clustering of cases from a starting epidemic is due to a sudden and high increase of the contact rate beyond a threshold. Interaction rather than main effects and nonlinear rather than linear dynamics are involved. Second, the incubation period of disease is studied. I advocate that an individual-level deterministic process underlies Sartwell's model of the incubation period. This accounts for the robustness of the model vis-à-vis confounding variables. Third, monozygotic twinning is analyzed. Assumed by some to be a random process, monozygotic twinning proves to be dynamically different from dizygotic or single-maternity processes; its dynamics can actually be chaotic. Throughout the provided examples, the point is made that chancelike phenomena are primarily concerned with chaos theory. For biological problems showing recurrent inconsistencies by stochastic modeling, dynamic modeling should be envisaged. Inconsistencies can suggest that the relevant factors are out of the model and that they are related deterministically. Finally, spectral analysis and attractors in the phase space are presented; these tools can aid the population biologist in tracing out chaos from time-series data sets. Several time-series data sets are simulated according to a simple nonlinear difference equation that bears some relationship to the basics of the dynamics of infections in the population. I show how the series can be analyzed and interpreted. Much research remains to be carried out until the nonlin-

[1]Department of Social and Preventive Medicine, Faculty of Medicine, University of Montreal, Pav. M. d'Youville, C.P. 6128, Succ. "A," Montreal H3C 3J7, QC, Canada.

Human Biology, August 1993, v. 65, no. 4, pp. 525–546.

ear effects of risk factors can be validated. The undertaking is worth the effort, as a new paradigm of causation is at stake.

Variation is well known to human biologists. Variation is usually investigated with the aid of stochastic models. A major feature of multivariate stochastic modeling is that effects that cannot be explained by independent variables are ascribed to random variables. This is the so-called multifactorial paradigm. This paradigm is also well known in epidemiology, wherein multivariate models (e.g., logistic regression) are used extensively and a large portion of the unexplained variation is ascribed to unidentified independent risk factors. Reality, it is surmised, can be known by stochastic models; this contention is directly related to the random variation generated by the large number of unknown causes of the origin of the population phenomenon. Philosophically, the approach is based on the premise that, if the multiple sources of variation were known, it would be possible to predict the future state of (and to explain) the phenomenon at any time, thus leaving nothing to chance. This pretense is not recent; it dates back to the eighteenth century and is typified by Laplace's contention that a complete knowledge of the universe is awaiting the identification of its various causes (Crutchfield 1986).

The multifactorial paradigm is secure for biologists and epidemiologists because it suggests that the current indeterminacy is temporary. Indeed, it is believed that random variation results from the multiple independent factors that the researcher cannot master properly. The factors escape either because they are multiple and their effect is small or because they are unidentified or, yet, because of the postulated absence of covariation of the factors with the studied variables. Notoriously, in the presence of random variation, confounding is absent and the multiple independent chance factors do not alter the main pathway between the cause and the effect. If, on the other hand, randomness invades the studied relationship or, similarly, if the multiple independent factors are numerous, there is the drawback that only a small fraction of the variation of the effect is explained by the studied factors. Be this as it may, a fraction of the variation of the effect is under control, and this provides confidence that identifying the multitude of other factors is only a matter of time or technology.

New Paradigm of Causation

Is there another way to approach the study of variation that can provide the advantage of parsimony and generality? A different paradigm has emerged lately. It is interesting because it proposes a new way to

look at randomness. The new paradigm confounds the classical multifactorial view of external causation by embedding its own apparent randomness. The new paradigm posits that the disease process is due to few causes, that the causes are internal to the system, that the system is deterministic and nonlinear (as opposed to stochastic and linear), and that the system is sensitive to its initial conditions (Glass and Mackey 1988). These features are the main components of a definition of chaos. The essence of chaos is obviously in contradistinction to the prevailing multifactorial paradigm. Specifically, simple and closed deterministic systems featured by nonlinear components can, under appropriate conditions, mimic randomness. In other words, the chancelike variations surrounding disease incidence in time (e.g., time series of infections) can be traced directly to the deterministic and nonlinear components of the model (and no more to extrinsic causes). Furthermore, the components are few and, accordingly, there is no need to call for many different extrinsic causes to account for what appears to be noise (i.e., the time-series random variation).

This is admittedly a new approach and, to say the least, no less than a paradox. In particular, how can a deterministic system generate noise? The new paradigm has far-reaching consequences for the theory of causation. For one, the epidemiological phenomenon is said to be determined; that is, the model does not include the well-known e (error) term standing for the independent random factors of variation; it is straightforward with no possibility of alteration of the output. Second, only a few factors are required to explain the phenomenon; that is, all the important control mechanisms hold in a few interdependent parameters. Although the time series can be modeled more accurately by complex deterministic equations, one differential or difference (in discrete instances) equation of first order is often sufficient to fit the observation. Third, the deterministic function behaves as though the phenomenon were driven by stochastic noise. At face value, it is no less predictable than when extrinsic noise is involved, and the phenomenon can altogether be investigated by statistical methods. But most surprising, the chaotic phenomenon is structured and thus is amenable to forecast, at least in the short run. Fourth, the deterministic approach implies that the phenomenon is controlled by the initial conditions that induced it. The initial conditions are all the small differences that impinged on the signal (the true information at the origin of the first values of the series) from the start. Because of nonlinearity, the initial conditions acquired an exponential influence over time. Therefore the observed chancelike variations represent disinformation traceable to the enhancement of measurement errors and small perturbations from the far past of the series. Because the remote conditions remain unknown for most complex systems, it follows that the phenomenon also remains unpredictable in the long run,

thus the need for statistical tools to characterize the deterministic system. From a practical standpoint forecast of a chaotic series is currently impossible except for short-run events; this is true despite the extent of our knowledge of the intervening causes and, ultimately, despite determinism! Indeterminacy is no more temporary here; it is inherent and unavoidable.

The point I wish to raise is the following: Biological phenomena that were ascribed in the past to the action of external and random factors rather can be explained by fewer deterministic internal variables with nonlinear effects. This is not to mean that extrinsic noise is ruled out of statistical observations; it is simply that it should occupy a different position in the assessment of causation. This new paradigm can drastically alter our view of causality.

Applications of Chaos to Biology

Origin of Chaos Theory. Chaos theory constitutes a new paradigm of causation. The new approach is currently entrenched in the theory of dynamic systems. The theory is akin to the study of unpredictable phenomena, such as the mechanics of turbulent fluids and meteorology, well known for its hardly credible forecasts. Chaos theory has been introduced in ecology to explain the fluctuations of animal populations over time and in economics to seize the day-to-day variation of economic cycles. More recently, chaos has found its way into medicine. It has been applied in the field of cardiology to distinguish normal from abnormal electrocardiograms. Attempts have also been found in the study of electroencephalograms to "quantitatively" assess the diagnosis of various disorders. The use of nonlinear dynamics is now ubiquitous, and the field of population biology has not escaped it (Krasner 1990). Some investigators have tried to apply chaos theory to the epidemics of common childhood infections (Schaffer and Kot 1985; Olsen et al. 1988; Olsen and Schaffer 1990).

Time-Series Data. Most attempts at deterministic modeling revolve around the analysis of time-series data, as underscored by the previous examples. But it is in no way exclusive, although analyses of cross-sectional data are less developed (Montroll and Schlesinger 1982; West and Schlesinger 1990; Bak and Chen 1991). Box-Jenkins modeling of time-series data is forthcoming in population biology, but the model is stochastic. Chronological series of infections (Anderson and Grenfell 1984; Helfenstein 1986), injuries (Martinez-Schnell and Zaidi 1989), clinical parameters (Crabtree et al. 1990), environmental pollution (Schwartz and Marcus 1990), and single and multiple maternities (Philippe 1991) were

recently studied because they were available for a few large populations. However, the parameters of stochastic modeling have no particular significance except to describe the dependencies embedded in the series. On the other hand, deterministic models readily yield interpretable parameters, as they are incorporated in the model *because* of their unique meaning. A major field of recent population biology investigations has been deterministic modeling of infectious diseases.

Infectious Disease Processes. Throughout time the world has experienced various types of epidemics that have had their own specific dynamics (Anderson 1982). Some epidemics, such as common infections (chicken pox, measles, mumps, and rubella), come periodically with cycles of peaks and troughs; others, when installed, root themselves and become endemic. Cholera, well known for its devastation of Africa, India, and lately Peru, is a case in point. Other epidemics, such as AIDS, occur once (although it has not been proved that similar epidemics have not spread in populations in the far past) and nearly vanish either with the population itself (e.g., smallpox brought from abroad to Early Americans and Polynesians) or because of drastic intervention (e.g., kuru in New Guinea). Finally, other epidemics occur sporadically as though the conditions of their appearance, largely unknown, had fallen together randomly. These include illnesses such as various toxic shocks (streptococcal or staphylococcal in origin) and recent aggregates of meningococcemia in the Montreal South Shore region (Philippe 1992). What are the dynamics of infectious diseases? What are the forces at the origin of an epidemic? Before presenting a first case study, it is worthwhile to briefly set forth the theory of epidemics.

Many investigators have attempted to model the evolution of epidemics in time and space (Anderson 1982). The susceptible-exposed-infected-recovered (SEIR) model is an example of such an attempt for time series of common infections. It is represented by four differential equations specifying the state of the system at any time:

$$dS(t)/dt = m[N - S(t)] - bS(t)I(t), \tag{1}$$

$$dE(t)/dt = bS(t)I(t) - (m + a)E(t), \tag{2}$$

$$dI(t)/dt = aE(t) - (m + g)I(t), \tag{3}$$

$$dR(t)/dt = gI(t) - mR(t), \tag{4}$$

where N is the population size and t is time. Equation (1) states that individuals enter the population as susceptible individuals (S) (at risk of being diseased) by birth (m is the birth or death rate, postulated to be equal) and leave by death (mS) and through contacts (b) with infected individuals (I). Equation (2) states that the number of exposed individ-

uals (E) (not yet infectious) increases by contacts of susceptible individuals with infected individuals but is depleted through death and because the exposed subjects become infectious (at rate a). Equation (3) states that the number of infected individuals depends on the turnover of exposed subjects with account being taken of the loss of infected individuals by acquired immunity (at rate g) and a death component. Last, Eq. (4) states that recovered individuals (R) are a function of newly immunized subjects and death.

A further component (not shown here) is usually inserted into the model; it is the seasonal variation in the contact rate. Parameter values (a, g, and m) are provided by statistical estimates on real populations. The estimate of the contact rate (actually a proportion) is less straightforward (Anderson 1982). As can be seen, an epidemic is a function of several well-identified, dynamically related compartments (the population subgroups). The dynamics of a given disease are up to its specific parameter value. This model (including the seasonal component) yields a good fit to epidemics of common infections when simulated according to the relevant parameter values. In fact, the key parameter is the contact rate; that is, the proportion of contacts an infected individual has with susceptible individuals. Readers interested in applications to common infections can consult Schaffer and Kot (1985), Olsen et al. (1988), and Olsen and Schaffer (1990) for thorough analyses.

A more anecdotal example highlights the significance of the contact rate for the determination of the dynamics. Several months ago, four cases of meningococcemia were diagnosed in young people in a well-delimited geographic area of the Montreal metropolitan area (Philippe 1992). Two of the patients died. The four individuals had attended the same disco bar but were unknown to each other and were infected by the same strain of bacteria. The cases were followed in the few following months by several additional foci of cases over a large geographic area. It was far from clear that contacts had been established between these highly distant regions. The epidemic nevertheless grew to the point that a vaccination program had to be set up in the highest-risk regions. According to chaos theory, the newly arising cases from this uncommon age group were not sporadic (although a few cases were expected in a large geographic area as a result of random infection) but the first identified cases of a building epidemic. The between-people contact rate in disco bars rapidly increased to the point that dissemination of the bacteria quickly reached the threshold for a new established epidemic process. Two mechanisms explain the time-space aggregation: new foci of infection (other disco bars, schools), as if the state conditions for chaos simultaneously came true in different distant regions; and contact with infectious subjects.

To be entertained, an infectious process must be based on a minimal (threshold) contact rate. The contact rate can be seen as a tuning parameter. Below the threshold the disease disappears from the population. Beyond the threshold but within given limits the contact rate elicits cases at a periodic rate (such as the repetitive autumn peaks of measles). But if the contact rate becomes suddenly high, the phenomenon acquires a new dynamics, that of chancelike epidemics, that is, an unpredictable aggregation of cases in time and space. According to chaos theory, the concept of random occurrence of epidemics should be given up. The unpredictability of apparently sporadic events is not due to chance but is determined by a dynamics that in many cases (besides the field of infections) has not been properly investigated. The French mathematician J.-H. Poincaré had intuitively grasped, as early as 1903, the causal basis of apparently random events when he wrote that in the long run "small causes which escape our notice determine a considerable effect that we cannot fail to see, and then we say that the effect is due to chance" [cited by Crutchfield (1986, p. 48)].

The time-space clustering of sporadic cases of diseases has long been observed and scrutinized by epidemiologists. Several investigators have suggested a threshold contact rate to explain the observation. For example, leukemia was thought to be related to the mixing of unrelated populations that had come into contact in a short time-space dimension. However, the relevant explanation fell behind until chaos theory had its first applications. One can speculate that the threshold contact rate uncovers dormant microorganisms that become activated or that the susceptible individual's defenses are overwhelmed by the new antigenic stimulation or by the interaction of multiple incompatible infectious agents. As far as new epidemics are concerned, energy should be spent attempting to find the factors at the origin of the pushing of the tuning parameter. High person-to-person contact rates can have a bearing on the microorganisms' ecology, thus altering the dynamics.

The AIDS epidemic has probably grown in a similar way. AIDS may have occurred in the distant past as different populations came into contact and as the interactions of various infectious agents were multiplied. Similarly, the outbreak of AIDS epidemics from two independent African foci calls for the linking of new interdependent conditions that may have arisen from time to time in the far past. This emphasizes the role of key events for the changing dynamics of infectious processes, which can push the tuning parameter into the chaotic realm. Analogously, the contribution of opportunistic infections such as the herpes simplex virus (HSV) to AIDS diagnosis may also be a case in point here. Treatment of HSV infections could put off diagnosis. Clearly, interactions rather than main effects and nonlinear rather than linear dynamics are to be privileged.

Incubation Period of Diseases. I now illustrate how nonlinear dynamics at the individual level can present itself under the cover of stochasticity. Because nonlinear deterministic systems can generate, under appropriate circumstances, their own chancelike variation, they can be studied by statistical methods. This property of dynamic systems is termed ergodicity. Suppose that an individual cellular process depends on nonlinear deterministic dynamics. Furthermore, suppose that the dynamics evolve progressively from a steady state (the initial conditions of the system) to chaos and then to the extinction of the system. It has been shown by Lasota and Mackey (1980) that the waiting time from exposure to the initial conditions to the process extinction has a stable invariant distributional law. This is true despite the wide variability of the initial conditions. In other words, the distribution of the waiting times is independent of the distribution of the initial conditions that elicited the extinction times. This should be expected because the significant information at the start is progressively replaced by chancelike variation during the process evolution. Lasota and Mackey (1980) simulated the dynamics of leukemic white blood cells. The distribution of the simulated patients' survival time was well fitted by a Weibull law; incidentally, the estimated Weibull parameters of the simulated data were in line with those of observed leukemic patient survival times.

One may then wonder whether the ergodic theory can also be called on to deal with incubation periods. Sartwell (1950) proposed a method to study the incubation period of infectious diseases. He noticed that the incubation period of common-vehicle single-exposure infectious diseases presented with positive skewness and fitted a log-normal distribution. Sartwell's model is notoriously robust and apparently free from important sources of confounding, such as the type of design, age at exposure, dose, and measurement errors. Most surprising, the method is also robust with respect to the population age distribution in age-of-onset studies (Horner 1988). Goodness of fit to various infectious and even to chronic disease incubation periods was put forth [for a recent account, see Philippe (1990)]. Therefore the conceptual difficulties were engendered by the nearly generalized goodness of fit to several empirical distributions despite potentially confounding factors. A further point is the etiological significance to attach to the dependence on the log-normal distributional law. The explanation drawn from chaos theory is the following. The infectious process, from exposure to disease onset, is composed of two steps. The first step is unobserved. It covers a short time lapse between exposure and a threshold that sets the magnitude of individual susceptibility. The threshold is the point of no return after which the process is entertained by itself. The second step is observed and goes from the threshold to disease onset. This is the incubation period per se, and from this time on the process is deterministic. In other words, once the in-

dividual infectious process is engaged (at the threshold), it cannot stop by itself; it follows a given pathway until symptoms appear or the disease is diagnosed. My contention is that the process eventually becomes chaotic and finally elicits clinical disease (corresponding to the process extinction). Accordingly, the incubation period displays *apparent* random variation (dispersion of the time lapse from exposure to diagnosis). This is because of the wide distribution of initial conditions (at the threshold) among individuals; the nonlinear deterministic evolution finally yields an unpredictable time of diagnosis. The assumption of nonlinearity leads to unpredictable fates given certain values of the dynamics (kinetics of the tuning parameter) of the process. Disease symptoms or diagnosis therefore occurs at randomly appearing times.

Now, why should the distribution be log-normal? Because the function is independent of the distribution of the initial conditions, it should depend on the particulars of the motion law embedded in the nonlinear model. It follows that specific models provide typical distributions of waiting times, some of which fit the log-normal. The point of major interest, for the time being, is why there is a distribution at the outset of a deterministic system and why it appears free from confounders. The answer lies in the chaotic dynamics of the process. How? For subjects who have nearly identical threshold liabilities, small initial differences or random impulses near the threshold grow exponentially and elicit large variations in time of diagnosis (and vice versa). Although the evolutionary process is entirely deterministic and leaves nothing to chance, the incubation period nevertheless registers among-patient differences because of the random initial liabilities. This is due to ergodicity. And it supports the nonlinear dynamics of disease processes that show inordinate resistance to the usual distorting factors. Insensitivity to the population age distribution in studies of age of disease onset is not new and has been the subject of unexpected observations, such as the simultaneous diagnosis of a rare condition in sibs of widely different ages.

A secondary aspect of the incubation period is the time scale of the etiologic chaotic process. Incidentally, the duration of the incubation period is not an asset to establish or even to suggest the type of dynamics involved. Short incubation periods are not generally more linear or more stochastic than lengthy ones, although it should be easier in principle to recognize a small initial influence in a short period because the chaotic process allows prediction in the short run. However, even for infections with short incubation periods, small influences can have large effects in a short time span depending on the kinetics of the nonlinear dynamics. The key factor is the chaotic level, that is, the kinetics of the process, rather than the duration of the incubation period. With lengthy induction periods, even large differences are liable to remain undetected as etiologic influences. This surmise has far-reaching consequences for the ep-

idemiology of chronic diseases, which are well known for their long-lasting induction period. Chaotic dynamics militates against the finding of causes of diseases because disinformation accrues from the beginning of the process. However, the history of epidemiology testifies that well-validated etiologic relationships were discovered, ensuring the existence of either true linear effects in pathways between the cause and effect in some chronic disease processes or potent risk factors. Presumably, potent risk factors would have been discovered easily. They were indeed pointed out in the case of cigarette smoking and lung cancer. But few are as potent. A further possibility is a reduced number of steps between the action of the cause and the observation of the effect, thus allowing weak risk factors to be detected in spite of chaos. Presumably, with many small influences acting sequentially, multifactorial diseases are bound to remain ill-understood. This might account for the paucity of consistent results on the role of risk factors in etiologic relationships involving the most popular contemporaneous chronic diseases. One may thus wonder whether nonlinear effects are not ubiquitous.

In addition, many different disease processes could have the same dynamics and therefore could be structured by the same type of differential equation(s). Accordingly, different epidemiological phenomena should behave in the same way. This would suggest that some basic pathological processes are based on both simple and universal models, a conjecture that, if true, could bring together the evolution of different and now postulated independent disease processes. The similar pathological brain findings for Down's syndrome and Alzheimer's disease are a case in point (Rumble et al. 1989). Furthermore, well-established distinct disease entities can differ by small differences in the level of their risk factors, whether environmental or genetic, that elicited them from the start. Those trivial differences are rapidly absorbed by the evolving dynamics of the nonlinear system, which in the end keeps no trace of the initial difference but produces entirely different diseases. The Li-Fraumeni cancer syndrome, the familial pattern of which includes many different kinds of neoplasms, seems to be caused by a single family-limited anti-oncogene defect (Li et al. 1988). How this single defect can be responsible for the neoplastic variability is intriguing. It is likely that a given risk factor, processed through the irreducible dynamics of various environments or inborn host susceptibilities, will yield different disease entities.

Time Series of Twins. Chaos theory suggests that what appears to be random in a time series could rather be the result of nonlinear determinism. Classically, monozygotic (MZ) twinning has been considered a random phenomenon; this position can now be questioned. The Japanese high prevalence of MZ twinning in comparison with Western countries

and the concomitant opposite MZ and DZ twinning recent evolution reported from several countries are intriguing observations with no clear explanation (Bressers et al. 1987). Chance or determinism? A biologic problem worthy of study by the techniques of nonlinear dynamics is therefore set forth. It might be that the reasons for the current variation in MZ twinning prevalence are due to factors from the distant past of the series. Incidentally, the parameters of change can be part of the system much in the same way that a current epidemic reflects the past dynamic interaction of its system compartments, namely, the infected, susceptible, exposed, and recovered subjects. The difference with infections is that the laws of motion of the twinning process are largely unknown and their compartments are related over a much longer time span. A likely possibility is that the MZ-twin-prone mothers compartment is not independent of that of DZ-twin-prone women and, of course, of single maternities. Accordingly, any change in the prevalence of MZ twinning would result from the lagged effects of the constitutive key system components; in such a framework no extrinsic risk factor to account for the dynamics of the system need be involved.

Such a perspective is in line with some observations. For instance, populations of seemingly similar socioeconomic status or of the same level of industrialization and expected (based on the multifactorial model) to start a decline in DZ twinning at the same time have proved to maintain their own specific trend. Sweden and Finland are cases in point: Sweden started its decline in the nineteenth century and Finland only since 1960. These differences cannot be ascribed to any of the known confounders, according to Doherty and Lancaster (1986). On the other hand, populations sharing similar initial conditions (but actually distinct so far as risk factors are concerned) behave similarly. For instance, Australia is closer to Finland and southern Italy over time than the last two are to Sweden and northern Italy, respectively. Therefore a simple fit with industrialization is patently absent. Again, the observation resists the multifactorial model; things go as if the causes of the discrepant behavior were out of reach of the classical explanation.

I have investigated MZ, DZ, and single-maternity time series in the framework of nonlinear dynamics (unpublished data, 1992). Analyzed data were from Australia (Doherty and Lancaster 1986). Although the data are sparse, an interesting pattern emerged. I found long-term periodic cycles in the singleton and DZ twin series; this finding suggests that periodicity recurs according to long periods not even available for observation. This is in line with the expected long-term lag effects of the dynamic system components. If this is the case, the postulated risk factors of industrialization, oral contraceptive use, etc. should have no significant bearing on the observed time trend. Another point worth stressing is that the DZ and MZ twinning fundamental periodicity un-

derscores the likely possibility that both series share some common factor. This is not new but nevertheless is in line with a deterministic model. Furthermore, the MZ twin series seems to have its own specific secondary cycles superimposed on the fundamental frequency that provide the series with a particular configuration over time. In other words, the fundamental signal (shared by the other two series) is altered by secondary cycles. The secondary cycles of the MZ twin series are responsible for the particulars of its dynamics, at variance with the related pattern of the DZ twin series. The MZ twinning dynamics thus appear to be slightly chaotic; this feature is not shared by the other two series. On the whole, it appears that the MZ twinning dynamics should be ascribed to yet unidentified laws of motion of the involved compartments. Explanations for the current variations in the MZ twin series are likely to be found in the history of the interacting series, such as those of DZ twin and single maternities. A model of the twinning process analogous to the SEIR model might be useful to set out the details of the dynamics. The model might be tested with the basic techniques of nonlinear dynamics.

Qualitative Approach to Chaos

A brief account of techniques used to unravel the dynamics of a time series might be useful. This account is qualitative, that is, not too quantitative. A complex armamentarium to deal with chaotic systems is available (including the mandatory specialized software) from the relevant literature. I attempt to provide an account of chaos with the aid of the basic techniques included in most common software. I do not rely on the analysis of databases because they are not readily available (or have already been analyzed) but rather on the simulation of one simple difference equation that anyone can try on a personal computer. This will help to assess the value and interest of a field of investigation that might well become important in the future.

I start with a simple difference (recurrence) equation (Lebreton and Millier 1982) of the form

$$I_{t+1} = I_t e^{[r(1 - I_t)]}. \tag{5}$$

To ease things, one may think of the equation parameters in terms of the population biology of infections. The equation involves nonlinear (exponential and interactive) terms wherein r is the tuning parameter [i.e., the rate at which the infection grows (the contact rate in the SEIR model)] and I is the proportion of infected individuals and varies between 0 and 1. It is clear from the equation that the number of infected individuals controls itself: When it gets high, its rate of increase drops, and vice versa. Obviously, the model is not sophisticated. Notably, it does not

account for the time-dependent contact rate included in the original SEIR model; nor does it provide a specific term for the number of exposed individuals or susceptible individuals, to mention a few changes. Nevertheless, the pushing in the chaotic region of this simplistic model confirms its appropriateness for the study of the dynamics of real infections. The interpretation of Eq. (5) in terms of infections is not intended to replace the SEIR model. The SEIR model involves as many as four similar equations whose sophistication makes it the most relevant approach to infectious processes in populations (Anderson 1982).

The difference Eq. (5) is purely deterministic; it is used to simulate the population of infected individuals over time; that is, as time proceeds, I_{t+1} is reintroduced in the equation to derive I_{t+2} and so on. For a given simulated time series, r and I_0 (the starting number of infected individuals) are constant. I simulated four distinct time series involving 534 data points each (the reason for the sample size will become clear later). The time series differ by the value of their tuning parameter only. By increasing r, the process moves from a steady state to successive period-doubling bifurcations (corresponding to a change in the dynamics) and finally to chaos. The selected r values typify some of the expected bifurcations the system should go through.

The study of chaos in time series involves the use of several qualitative and quantitative tools (Schaffer et al. 1990). I have limited my approach to the qualitative tools, recalling that this introduction to nonlinear dynamics is to give the biologist some feeling for the method and for its expected results. The simulated time series are therefore analyzed by using spectral analysis and by plotting attractors.

Simulated Series. The simulated chronological series are depicted in Figures 1A–D. To gain insight into the series, only the first 100 data points are plotted. The first series (Figure 1A) points to a steady-state process ($r = 1.5$). The second series (Figure 1B) is periodic ($r = 2.6$); it has in fact two periods repeating themselves indefinitely. The first two series have a predictable number of infected individuals at any time, whatever the initial conditions. The third and fourth series (Figures 1C and 1D, respectively) are chaotic in that they have unpredictable chance-like variations. At face value they can be taken for white noise series with random spikes of infected individuals. However, they also could be true monthly observations of measles from New York City (Figure 1E) (Yorke and London 1973); one can note the impressive qualitative similarity between the measles series and any of the last two simulations. An even better similarity can be reached with a more precise r value or a more complex model (such as the SEIR model).

Spectral Analysis. Spectral analysis (Gottman 1981), a complex mathematical device, is a method that detects time-series cycles and

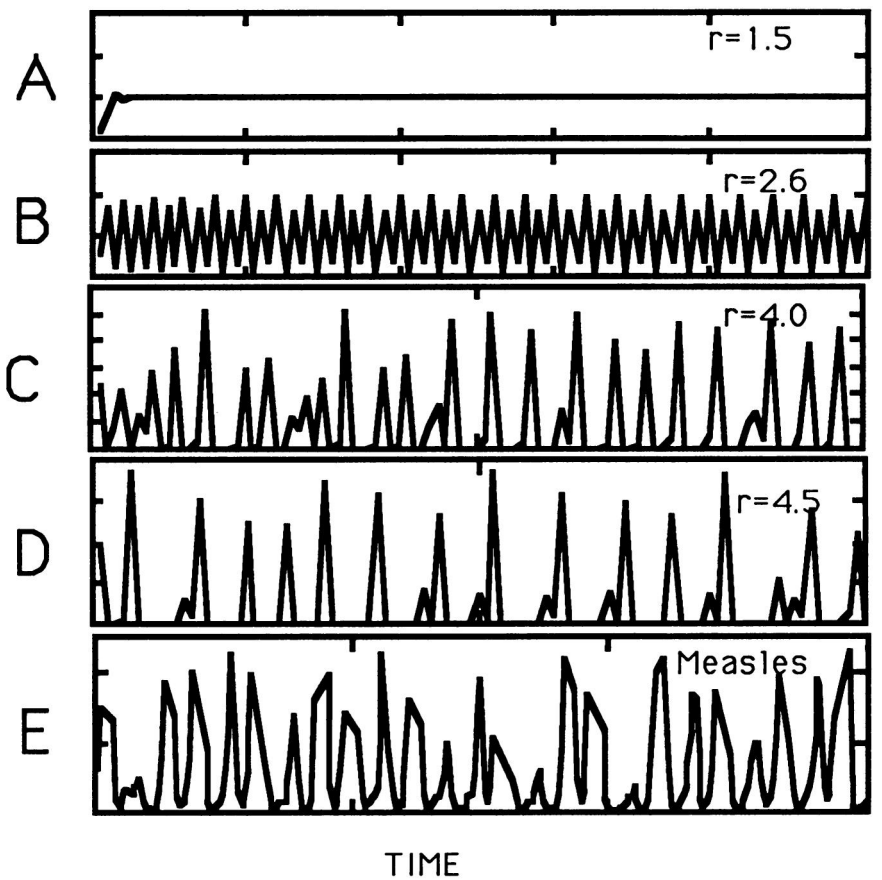

Figure 1. (A–D) Simulated time-series data according to various values of the tuning (contact rate) parameter and (E) observed monthly infections of measles in New York City.

quantifies their relative importance. This is provided by the spectrum that gives the variance ascribed to the identified cycles. Only three different patterns of spectrum are commented on here; needless to say, they do not exhaust all the possibilities liable to be found in real data sets. I consider them here because all three models are simulated and one can rely on them to distinguish chaotic, periodic, and random time series; they are therefore interpreted at face value.

First, a time-series data set may disclose white noise only; this means that the series embeds no particular cycle(s) besides random frequency variation, independent of time. Second, the data set may involve one or two distinct cycles that occur periodically and can be distinguished according to their frequency, as opposed to random variation. In this case

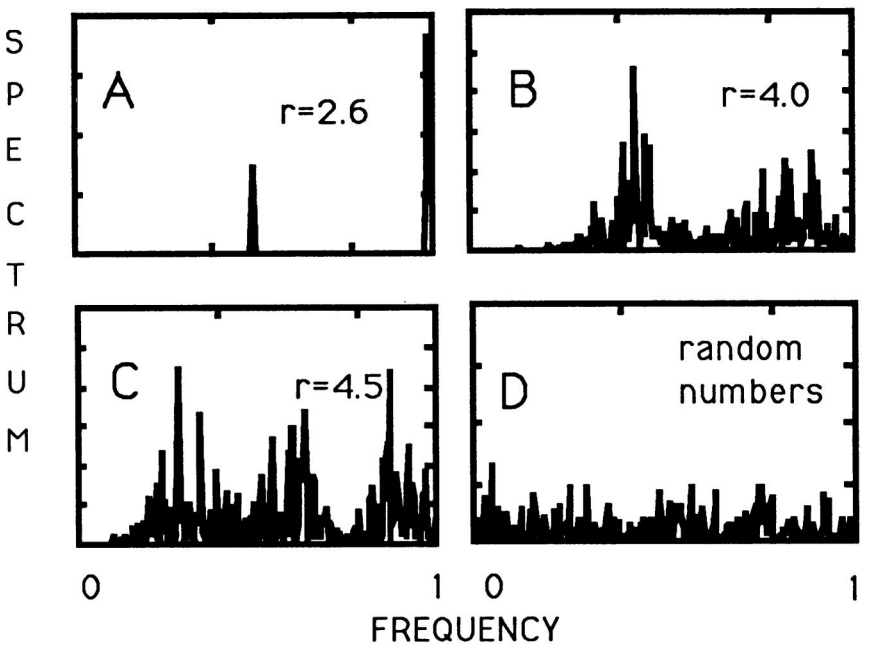

Figure 2. Spectral density of simulated time-series data differing by the value of their tuning (contact rate) parameter.

the spectrum has one or two distinct spikes corresponding to specific frequencies. Third, the time series may have an infinity of superimposed cycles (different scales), a sign of the presence of chaos. The spectral pattern in this case, a distribution of many different scales (or frequencies), is different from that of white noise. Incidentally, simply examining a crude series of white noise and one resulting from chaos cannot differentiate them.

Figure 2 displays the spectra of the simulated series. The steady-state series (corresponding to Figure 1A) is not shown because its spectrum is flat. The periodic series in Figure 2A (corresponding to Figure 1B) does not actually distinguish between the two periods because they are of nearly identical frequency. If the data points of the original time series are taken as the number of infected individuals per year, the two cycles of Figure 1B have a frequency of approximately 0.5 (i.e., the repeat waiting time of the periods is 2 years). The two chaotic series (Figures 2B and 2C, corresponding to Figures 1C and 1D, respectively) show, as expected, multiple superimposed cycles, although the series with $r = 4.5$ is more chaotic than the series with $r = 4.0$. The chaotic spectral density displays no privileged scale (rather multiple frequencies are present), a feature of chaotic phenomena. Last, the series

of computer-generated random numbers (Figure 2D) puts forth no particular cycle (no large spike); this is white noise.

Phase Space Attractor. Attractors represent a three-dimensional display of the trajectories of the series, where I_t is plotted along with I_{t+3} and I_{t+6}. The attractor can be thought of as a spatial asymptotic or parametric geometric object. The plot provides for the trajectories of the process and is an analogue of the phase space. Because the number of factors at the origin of a dynamic phenomenon is ordinarily high and impossible to master for complex systems, plotting the lagged values in an appropriate number of dimensions gives an unbiased representation of the original phase space. This technique is typical of the methodology of the study of dynamic systems (Glass and Mackay 1988). It can distinguish, at face value, between random, systematic, and chaotic processes, much in the same way that spectral analysis does. The rationale is the following. If one tries to display random numbers in one dimension, say, on a line from 0 to 1, they will occupy the whole line. If, on the other hand, one tries to correlate random numbers on a surface, the cloud of points will occupy two dimensions. Finally, if three dimensions are available, the whole box will be filled with points. Conversely, if the time-series pattern is somewhat systematic, only a fraction of the box will be filled.

Three distinctive patterns (which do not exhaust the number of possibilities) can therefore be obtained. First, the trajectories of the series may be attracted toward a single point in the phase space; this represents a steady-state process (i.e., a stable incidence of infected individuals over time). Second, the attractor may look periodic, tracing a circle (or two) or a similar form, each trajectory being near the other; this would stand for a periodic phenomenon (a purely seasonal infection, for example). Third, the attractor could have a deformed image, folding and stretching in many ways and occupying less than the three available dimensions. This is a strange attractor. It is also called a fractal because it has a noninteger number of dimensions. Fractal geometry is the signature of chaos; it results from the interplay of determinism entrenched in the equations (and constrained to the phase space by folding) and unpredictability because of the initial conditions of the system (resulting in the variation of the trajectories through stretching) (Ruelle 1980).

Figure 3 displays the attractors. From the three-dimensional perspective (a front view of the meeting point of the three vertices), the space occupied by the random number series (Figure 3D) readily distinguishes it from the other three spatial representations. The periodic series (Figure 3A) is different from the chaotic series (Figures 3C and 3D) in that the trajectories clearly alternate between two different basins of attraction corresponding to the two frequencies. The trajectories follow the

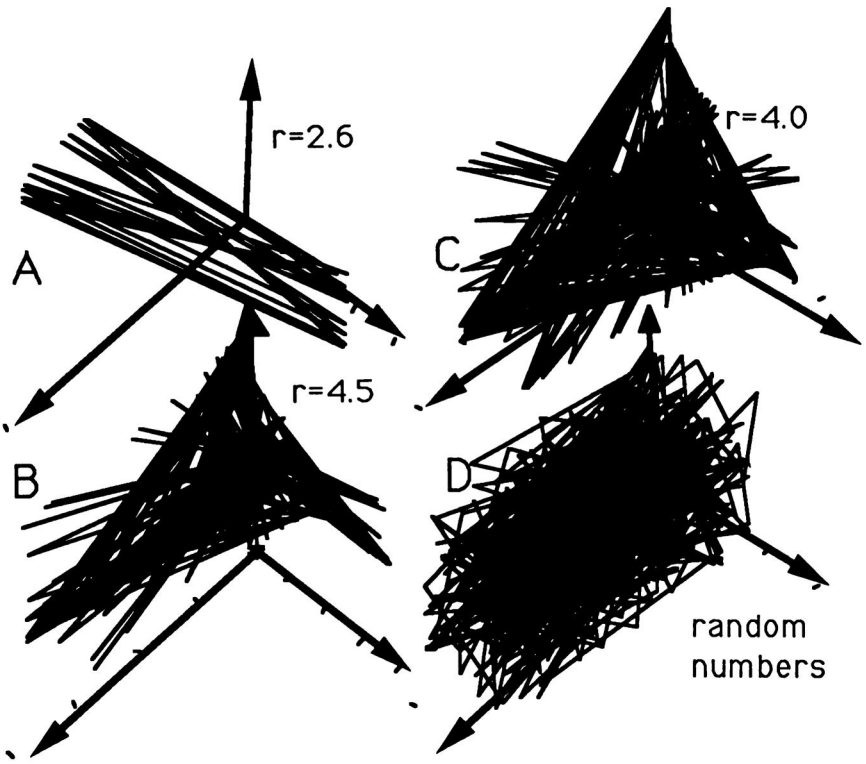

Figure 3. Phase space attractors of simulated time-series data differing by the value of their tuning (contact rate) parameter.

same paths indefinitely and therefore yield the same periodic number of infected individuals. For the chaotic series the attractors are highly intertwined and are said to be strange. Because the trajectories of the system are unpredictable, the attractor never crosses the same path twice (stretching in multiple directions is followed by folding back), a feature responsible for the impression of piling up of an infinity of thin sheets. Actually, magnification of a section of a strange attractor shows the endless repetition at smaller scales of its macroscopic shape; such a pattern is scale invariant or fractal. A simple fundamental signal (as exemplified by the deterministic equation) elicits the macroscopic aperiodic observable phenomenon, bypassing multiple different increasing scales through the exponential amplification of its own small differences.

I have mentioned that a small difference (a measurement error, for instance) in the initial conditions can have a dramatic effect on the observable phenomenon. Figure 4 presents the first 100 data points of the

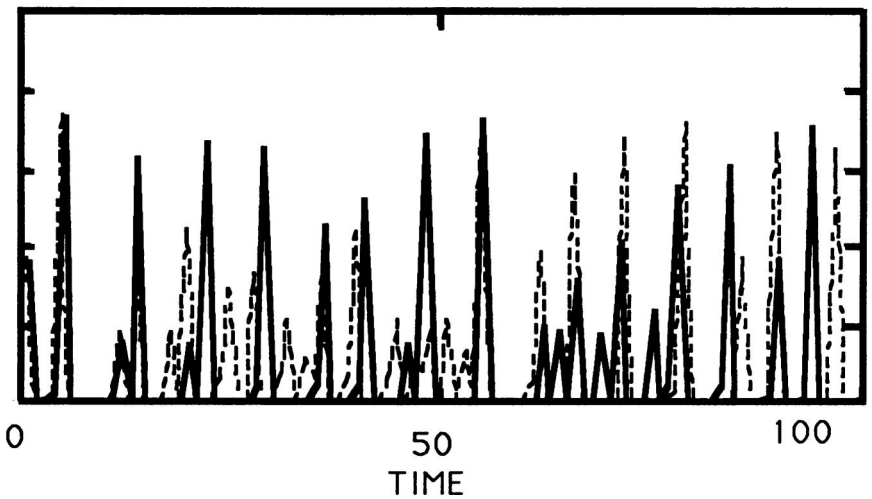

Figure 4. Simulated time-series data according to a small difference in their starting value (solid line: $I_0 = 0.50$; dashed line: $I_0 = 0.51$).

chaotic series ($r = 4.5$) for $I_0 = 0.50$ and $I_0 = 0.51$ (the dashed line). This shows how a trivial difference (2% in the number of infected individuals at the start of the epidemic) can allow, even in the short run (less than 10 time units after the beginning), the two series to diverge exponentially. It also exemplifies the source of macroscopic unpredictability for a deterministic process. More generally, a deterministic series will remain unpredictable simply because of measurement error made at the start (and at any other point of the process). On the contrary, steady-state or periodic time series are independent of the initial conditions; that is, whatever the starting values, the phase space trajectories are identical.

Finally, Figure 5 displays the spectral analysis and the phase space representation of the trajectories for the New York City series of measles (Figure 5A) and chicken pox (Figure 5B) infections. The original data, 534 monthly observations published by Yorke and London (1973), were analyzed exhaustively (Olsen and Schaffer 1990). I nevertheless repeated part of the investigation. The spectral analysis of measles (Figure 5A) shows few peak frequencies corresponding to superimposed cycles. The attractor is strange and fills a small part of the three-dimensional space; the trajectories (submitted to a three-point smoothing) exhibit obvious stretching and folding. Chicken pox, on the contrary, is featured by a single peak spectrum, and although it has a structure in the phase space, the configuration is elliptic, as one would expect for a periodic process. The thickness of the regular attractor is due to the noise associated with

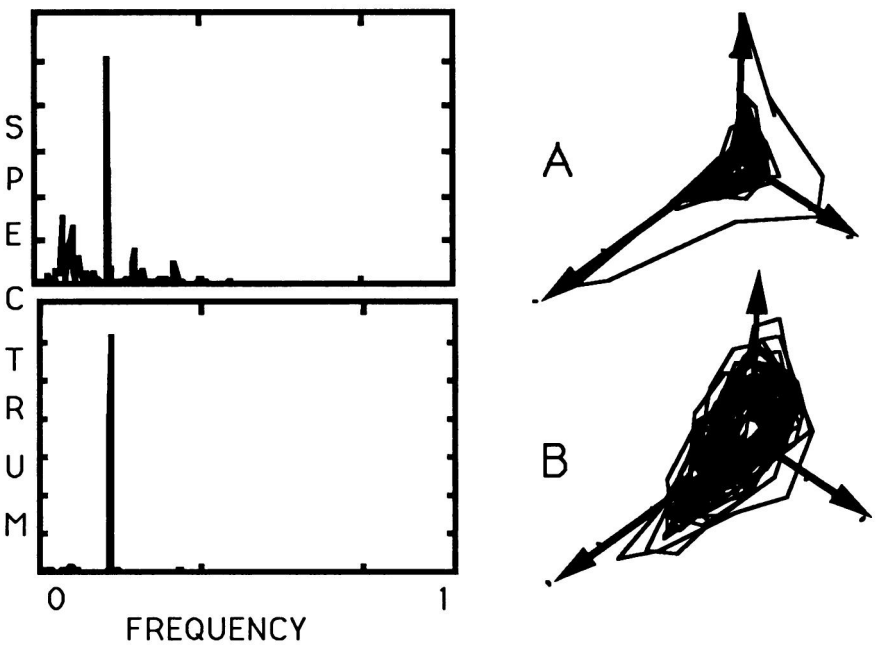

Figure 5. Spectral density and phase space attractor of the New York City time series of (A) measles and (B) chicken pox case declarations.

its specific period. Notably, both diseases reoccur with a yearly cycle, but their dynamics are fundamentally different (Olsen and Schaffer 1990).

Conclusion

Only a few tools have been introduced here that can help disclose chaos from time-series data sets. The literature is full of various quantitative techniques that, for the time being, are not readily available to epidemiologists and biologists. One should preferentially consult Schaffer et al. (1990) for the state-of-the-art in time-series analyses. My objective was to provide a rather basic account of chaos theory and applications. Complexities can be found in the referenced literature.

Several features of the new paradigm have been expounded, and three in-depth applications have been discussed: apparently sporadic epidemics, the incubation period, and the twinning process. From a practical viewpoint, what can one do to investigate epidemiological phenomena? Obviously, observed time-series data can be studied from the qualitative standpoint presented here. Lengthy epidemiological time series are needed in the worst way. They may be either population-level

or individual-level series of biological parameters. A second approach would involve formulating nonlinear differential equations related to the studied epidemiological phenomenon. The MZ twin series is a case in point. Identification of the main system components and of laws of motion is a prerequisite. In this undertaking a mathematician with knowledge of dynamic modeling should be recruited (Huckfeldt et al. 1982). Once equations are available, time series can be simulated. The tuning parameter is the key factor in chaotic dynamics. A simulated epidemiological series can be generated by the forcing of its appropriate differential or difference equations. The behavior of the system can be explored in detail with the aid of a bifurcation diagram, which traces the dynamics according to different values of the tuning parameter (Glass and Mackey 1988). For instance, chaotic dynamics can be reversed by reducing the growth rate. It can be reset at a periodic value, depending on the efficiency of the intervention. The disease process can also be eradicated (at least on paper). But, by no means can the phenomenon be predicted well in advance so long as it is in the chaotic range. At any rate, techniques are being developed that will allow short-term predictions. This can be achieved partly because of intrinsic determinism (Sugihura and May 1990).

The study of chaos needs so long a series that it is difficult to obtain human data commensurate with the requirements of this new type of methodology (500 data points is a minimum and this is considered a short series). However, new techniques will soon be available and will allow shortcuts through the current requirements (Eckmann and Mashaal 1991). An empirical investigation of less than 100 data points was nevertheless successful (Schaffer 1984). According to some, the outcome of the study is up to the kind of dynamics involved; a relatively simple chaotic dynamics from a relatively short series can yield interpretable results (Albano et al. 1987). Given that the new technology will make the criteria of sample size less stringent, it will be possible to study the natural history of diseases, such as the evolving episodes of multiple sclerosis, or of chronic conditions, such as diabetes. The new approach will also help to characterize the heterogeneity of chronic diseases.

If the number of dimensions of the attractor is low (e.g., measles with less than three dimensions), a complete account of the dynamics rests with a rather low number of parameters. The number of dimensions can be computed as the fractal or Hausdorff's dimension (Glass and Mackey 1988). Such results are encouraging so far as intervention is concerned. They are even more relevant if any previous modeling has identified the parameters at stake along with a good fit to the observed process.

The new paradigm of chaos undoubtedly can contribute a new approach to disease causation. The domain is primarily sporadic phenomena. Sporadic cases of disease, noisy time series, and robust distributions

may never be viewed in the same manner from now on. In addition, for biological problems showing recurrent inconsistencies by stochastic modeling, dynamic modeling might be more to the point. Inconsistencies can suggest that the relevant factors are out of the model and related deterministically. Nevertheless, much research remains to be carried out until the nonlinear effects of risk factors can be taken for granted; but the undertaking is worth the effort because a new paradigm of causation is at stake.

Received 13 April 1992; revision received 23 November 1992.

Literature Cited

Albano, A.M., A.I. Mees, G.C. de Guzman, and P.E. Rapp. 1987. Data requirements for reliable estimation of correlation dimensions. In *Chaos in Biological Systems*, H. Degn, A.V. Holden, and L.F. Olsen, eds. New York: Plenum Press, 207–220.

Anderson, R.M. 1982. *Population Dynamics of Infectious Diseases: Theory and Applications*. London, England: Chapman and Hall.

Anderson, R.M., and B.T. Grenfell. 1984. Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: Time series analysis. *J. Hyg. Cambridge* 93:587–608.

Bak, P., and K. Chen. 1991. Self-organized criticality. *Sci. Am.* 260(January):46–53.

Bressers, W.M.A., A.W. Eriksson, P.J. Kostense, and P. Parisi. 1987. Increasing trend in the monozygotic twinning rate. *Acta Genet. Med. Gemellol.* 36:397–408.

Crabtree, B.F., S.C. Ray, P.M. Schmidt, P.J. O'Connor, and D.D. Schmidt. 1990. The individual over time: Time series applications in health care research. *J. Clin. Epidemiol.* 43:241–260.

Crutchfield, J.P. 1986. Chaos. *Sci. Am.* 255(December):46–57.

Doherty, J.D.H., and P.A.L. Lancaster. 1986. The secular trend of twinning in Australia, 1853–1982. *Acta Genet. Med. Gemellol.* 35:61–76.

Eckmann, J.-P., and M. Mashaal. 1991. La physique du désordre. *La Recherche* 232:554–564.

Glass, L., and M.C. Mackey. 1988. *From Clocks to Chaos*. Princeton, NJ: Princeton University Press.

Gottman, J.M. 1981. *Time Series Analysis: A Comprehensive Introduction for Social Scientists*. New York: Cambridge University Press.

Helfenstein, U. 1986. Box-Jenkins modeling of some viral infectious diseases. *Stat. Med.* 5:37–47.

Horner, R.D. 1988. Re: "Age at onset of Alzheimer's disease: Clue to the relative importance of etiologic factors?" *Am. J. Epidemiol.* 128:443–444.

Huckfeldt, R.R., C.W. Kohfeld, and T.W. Likens. 1982. *Dynamic Modeling: An Introduction*. Beverly Hills, CA: Sage Publications.

Krasner, S., ed. 1990. *The Ubiquity of Chaos*. Washington, DC: American Association for the Advancement of Science.

Lasota, A., and M.C. Mackey. 1980. The extinction of slowly evolving dynamical systems. *J. Math. Biol.* 10:333–345.

Lebreton, J.D., and C. Millier, eds. 1982. *Modèles Dynamiques Déterministes en Biologie*. Paris: Masson.

Li, F.P., J.F. Fraumeni, Jr., J.J. Mulvihill, W.A. Blattner, M.G. Dreyfus, M.A. Tucker, and R.W. Miller. 1988. A cancer family syndrome in twenty-four kindreds. *Cancer Res*. 48:5358–5362.

Martinez-Schnell, B., and A. Zaidi. 1989. Time series analysis of injuries. *Stat. Med*. 8:1497–1508.

Montroll, E.L., and M.F. Schlesinger. 1982. On 1/f noise and other distributions with long tails. *Proc. Natl. Acad. Sci. USA* 79:3380–3383.

Olsen, L.F., and W.M. Schaffer. 1990. Chaos versus noisy periodicity: Alternative hypotheses for childhood epidemics. *Science* 249:499–504.

Olsen, L.F., G.L. Truty, and W.M. Schaffer. 1988. Oscillations and chaos in epidemics: A nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor. Popul. Biol*. 33:344–370.

Philippe, P. 1990. Twinning causative origin investigated by Sartwell's biometrical method. *Am. J. Hum. Biol*. 2:107–115.

Philippe, P. 1991. Box-Jenkins modeling of chronological series of twin births and single births. *Hum. Biol*. 63:367–387.

Philippe, P. 1992. Chaos and public health: Implications for an epidemic. *Can. J. Public Health* 83:165–166.

Ruelle, D. 1980. Les attracteurs étranges. *La Recherche* 11:132–144.

Rumble, B., R. Retallack, C. Hilbich, G. Simms, G. Multhaup, R. Martins, A. Hockey, P. Montgomery, K. Beyreuther, and C.L. Masters. 1989. Amyloid A4 protein and its precursor in Down's syndrome and Alzheimer's disease. *New Engl. J. Med*. 320:1446–1452.

Sartwell, P.E. 1950. The distribution of incubation periods of infectious disease. *Am. J. Hyg*. 51:310–318.

Schaffer, W.M. 1984. Stretching and folding in lynx fur returns: Evidence for a strange attractor in nature? *Am. Natur*. 124:798–820.

Schaffer, W.M., and M. Kot. 1985. Nearly one-dimensional dynamics in an epidemic. *J. Theor. Biol*. 112:403–427.

Schaffer, W.M., L.F. Olsen, G.L. Truty, and S.L. Fulmer. 1990. The case for chaos in childhood epidemics. In *The Ubiquity of Chaos*, S. Krasner, ed. Washington, DC: American Association for the Advancement of Science, 138–166.

Schwartz, J., and A. Marcus. 1990. Mortality and air pollution in London: A time series analysis. *Am. J. Epidemiol*. 131:185–194.

Sugihura, G., and R.M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344:734–741.

West, B.J., and M. Schlesinger. 1990. The noise in natural phenomena. *Am. Sci*. 78:40–45.

Yorke, J.A., and W.P. London. 1973. Recurrent outbreaks of measles, chickenpox, and mumps. II. Systematic differences in contact rates and stochastic effects. *Am. J. Epidemiol*. 98:469–482.