AMERICAN
Journal of Epidemiology
Formerly AMERICAN JOURNAL OF HYGIENE

VOL. 106 AUGUST, 1977 NO. 2

Reviews and Commentary

A COMMENTARY ON THE MECHANICAL ANALOGUE TO THE REED-FROST EPIDEMIC MODEL[1]

PAUL E. M. FINE[2]

Recent years have witnessed a considerable increase in the development and use of mathematical and computer-based models in infectious disease epidemiology. As a measure of this increase, Bailey (1) finds that the literature in this field has grown at a greater-than-exponential rate over the past two decades. Although one may suspect that some of the published models have greater value as exercises in mathematics than as contributions to "real disease" epidemiology, this is certainly not true of all of them; and many examples can be cited of major contributions of mathematical models in revealing relationships underlying complex epidemiologic phenomena, in hypothesis testing, and in the rational formulation of public health policy (1–4).

A further motive for the development of models, and one which is sometimes overlooked, relates to their usefulness in classroom teaching. Indeed, it might even be argued that it is through this medium that models have had their greatest impact upon the practice of epidemiology today. It is probable that discussions of Muench's (5) catalytic models (especially at Harvard University) and of Macdonald's (6) malaria equations (especially at the School of Hygiene and Tropical Medicine in London) have been of major and enduring value to many epidemiologists and public health officers, by introducing them clearly and forcefully to the quantitative subtleties of epidemiologic processes. In this context, it is certain that no epidemiologic model has enjoyed such widespread and lasting success in the teaching milieu as has that originally developed at the Johns Hopkins University by Wade Hampton Frost and Lowell Reed, and which has come to be known familiarly as the Reed-Frost model.

History of the Reed-Frost Model

The story of the development of the Reed-Frost model is both well and ill known. Though mentioned by Frost in a Cutter lecture at Harvard in 1928, the

[1] From the Department of Biomedical and Environmental Health Sciences, School of Public Health, University of California, Berkeley.

This research was sponsored in part by General Research Support Grant 5-SO1-RR-05441 from the National Institutes of Health.

[2] Present address (for reprint requests): Ross Institute, School of Hygiene and Tropical Medicine, Keppel Street, London WC1, England.

This commentary is dedicated to Lowell Reed, to Wade Hampton Frost and to the many other workers who have taught and published on this ingenious model. The author is especially indebted to Dr. Chin Long Chiang for his encouragement and suggestions in the preparation of this paper. Drs. Lila Elveback, Abraham M. Lilienfeld, Margaret Merrell and Philip E. Sartwell also made helpful comments on the manuscript.

model was never discussed in publications by its authors (7, 8). It is indeed ironic that one of the more fertile ideas in twentieth century epidemiology should have been considered by its authors as "too slight a contribution" for publication (9). But other workers (beginning with Zinsser and Wilson, in 1932) were impressed with the model's potential usefulness for the investigation of a variety of epidemiologic problems. And it has since served as the basis of several important contributions, notably by Elveback, Fox, and their colleagues (10, 11).

The beauty and strength of the Reed-Frost model lie in the simplicity and versatility of its algebraic formulation. The ease with which it is converted from deterministic to stochastic form makes it ideal in the context of teaching biomedical students, for whom stochastic theory may appear somewhat threatening. Indeed, it is with reference to this quality of their model that Reed and Frost showed a special streak of their genii, in the development of a mechanical analogue to illustrate the model's stochastic properties.

This mechanical model – based upon the probabilists' traditional container full of colored balls – was developed at Johns Hopkins about 1930, and has since been used in teaching laboratories in many institutions both in the United States and abroad. Invented before the age of computers, this apparatus probably provided the first technique for the stochastic simulation of epidemiologic phenomena with non-biological material (we recall that there was considerable interest in animal-model "experimental epidemiology" at that time in history). The mechanical model has undoubtedly undergone transformations in its passage from one institution to another; but none of these transformations – let alone the basic model itself – has ever been fully described in publication. This is surprising, as the relationship between the algebraic formulation and the mechanical analogue of the Reed-Frost model is both interesting and subtle. Some aspects of this relationship may be overlooked in the epidemiology classroom; but others frequently arise in discussion of the model's properties. It then becomes of interest to explore different methods of treating the relationship between the algebraic and mechanical models.

It is in recognition of the historical importance, the intrinsic subtlety and the continued value of the Reed-Frost model in epidemiology teaching, that several variants of its mechanical analogue are discussed here.

THE MODEL

The basic deterministic formulation

The Reed-Frost model was originally designed to describe the epidemic pattern of an acute, contagious infection after its introduction into a closed population. The model's assumptions were outlined by Abbey (12), in the following way:

> The infection is spread directly from infected individuals to others by a certain kind of contact (adequate contact) and in no other way.
>
> Any non-immune individual in the group, after such contact with an infectious person in a given period, will develop the infection and will be infectious to others only within the following time period, after which he is wholly immune.
>
> Each individual has a fixed probability of coming into adequate contact with any other specified individual in the group within one time interval, and this probability is the same for every member of the group.
>
> The individuals are wholly segregated from others outside the group.
>
> These conditions remain constant during the epidemic.

Though stringent, these assumptions may provide a reasonable description of the processes underlying outbreaks of acute infections within institutions (e.g., measles within schools). In this case, the model's "time period" is taken to correspond to the latent period of the infection, the time between acquisition of the infection and maximum infectiousness of the

case. The traditional notation for the Reed-Frost model is as follows:

S_t, C_t = The numbers of susceptibles, and cases, in the population during time interval t.

S_{t+1}, C_{t+1} = The numbers of susceptibles, and cases, in the population during the next time period, $t + 1$.

$p = (1 - q)$ = The probability that *any* two individuals (selected at random from the population) come into "effective contact" (i.e., contact sufficient for the transfer of the infectious agent) during one time period. (Some teachers of the model have found it helpful to introduce another parameter here, in order to clarify the definition of this probability of effective contact. If there are M individuals in the entire population, then $p(M - 1) = K$ = the expected number of contacts experienced by each individual during one time period.)

We argue that, during time period t, the probability that a susceptible individual comes in contact with none of the cases is q^{C_t}. The complement of this term describes the probability that a susceptible individual contacts at least one case, and hence contracts the infection.

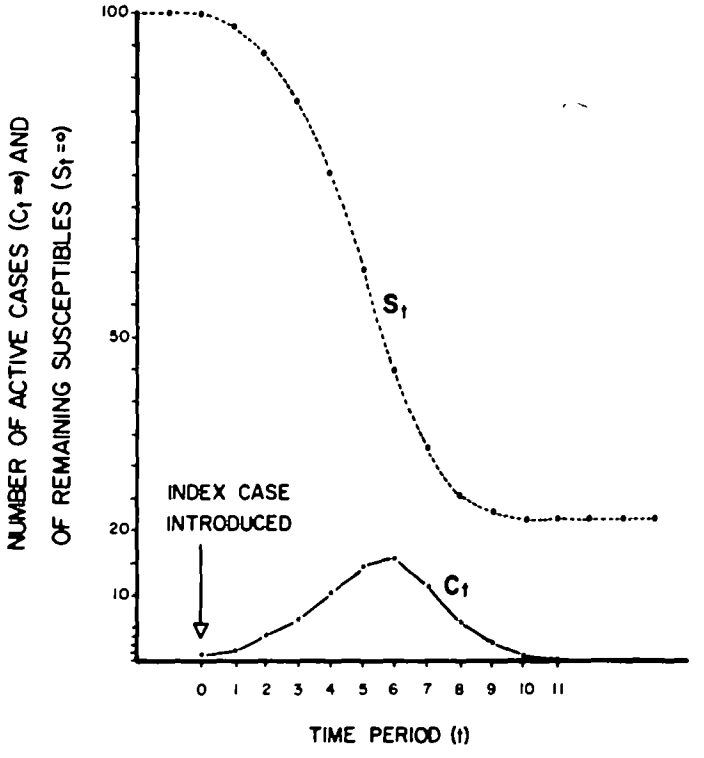

FIGURE 1. Epidemic course predicted by the basic Reed-Frost model, assuming $S_0 = 100$, $C_0 = 1$ and $p = 0.02$. The successive numbers of cases (solid line) and numbers of remaining susceptibles (dotted line) were obtained by iteration of equations 2 and 3 in the text. The expected numbers of cases were rounded off to their nearest integer values. Given these conditions, a total epidemic size of 79 cases is predicted by the deterministic Reed-Frost model, and 22 susceptibles remain still unaffected at its conclusion.

Probability a susceptible contacts at least one case during time period t

$$= (1 - q^{C_t}). \quad (1)$$

The expected number of cases in the next time interval is thus defined deterministically as:

$$C_{t+1} = S_t(1 - q^{C_t}). \quad (2)$$

And, of course:

$$S_{t+1} = S_t - C_{t+1} \quad (3)$$

These two equations comprise the basic Reed-Frost model. Iteration of these equations for successive time periods allows the prediction of an entire epidemic, as illustrated in figure 1.

The implications of this formulation have been investigated by a number of authors. Notable are Wilson and Burke's (9) comparison of the model with the earlier "mass action" formulations of Hamer (13) and Soper (14), Costa Maia's (15) investigation of the implications of adding new susceptibles, and Zinsser and Wilson's (4) examination of the effect of variations in virulence upon the apparent case fatality rates in such theoretical epidemics.

The stochastic formulation

The conversion from deterministic to stochastic formulation is easily illustrated by comparing a Reed-Frost epidemic process to a series of binomial trials. In each time period, each of the S_t susceptible individuals "has" a probability of becoming infected equal to $(1 - q^{C_t})$. We can thus imagine the events of one time period to be analogous to the tossing of S_t "coins," each of which has a probability of $(1 - q^{C_t})$ of falling "cases up." And the probability that exactly r of the susceptibles ("coins") become infected (fall "cases up") is thus given by the standard binomial expression:

$$\text{Prob}\,(C_{t+1} = r) = \frac{S_t!}{r!(S_t - r)!}(1 - q^{C_t})^r(q^{C_t})^{S_t - r} \quad (4)$$

This is the basic stochastic formulation of the Reed-Frost epidemic process. The equation defines the probability of a specified prevalence $(C_{t+1} = r)$ in the subsequent time period, given some set of initial conditions (S_t, C_t and p). By multiplication of probabilities defined by this equation, it is possible to calculate the probability that any specified series of prevalences would occur (e.g., C_{t+1}, C_{t+2}, C_{t+3}, . . .), given some initial conditions. Abbey (12) discussed this formulation, and from it derived a maximum likelihood method for estimating the p value for any given epidemic sequence.

Though the derivation of the stochastic expression 4 is straightforward, and accessible to anyone familiar with the binomial distribution, its implications may not be so immediately clear. This reflects a common difficulty suffered by non-mathematicians when dealing with stochastic equations. Although stochastic epidemics may easily be generated on a computer, in accordance with this equation, the conceptual leap from expression 4 to print-out is sometimes underestimated by biostatisticians who present such material to non-mathematical audiences. It is here that the mechanical analogue to the Reed-Frost model demonstrates its usefulness as a teaching tool.

The mechanical model

The goal is (or was) to design a simple mechanical apparatus which would illustrate the behavior of the stochastic Reed-Frost epidemic. Indeed, the absence of electronic computers when the model was first developed made some such apparatus necessary for the empirical exploration – let alone teaching – of the model. The technique developed at the Johns Hopkins University (Sartwell (8) attributes the idea to Reed) involved the use of colored balls to represent the individuals in the population, and their randomized linear arrangement in a trough to illustrate the outcome of each successive time period

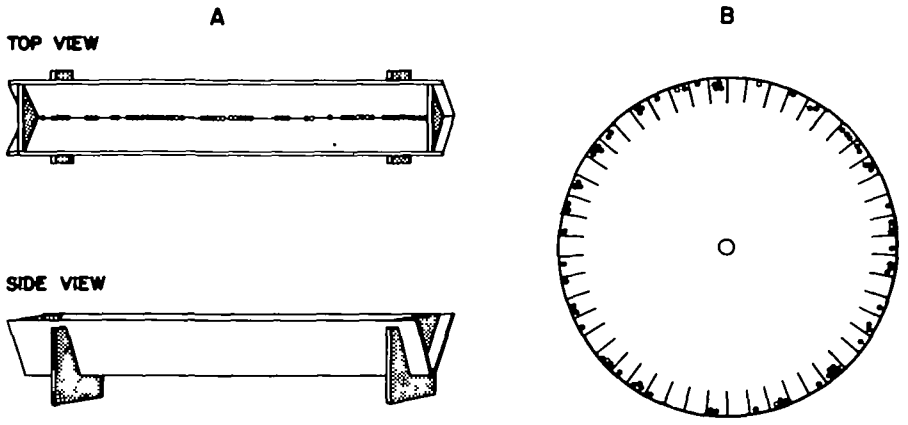

FIGURE 2. Two different types of mechanical Reed-Frost models. 2a is the classic trough, holding a linear arrangement of "susceptible," "case," "immune" and "contact neutralizer" balls. Convenient teaching models may include one fixed and one removable endpiece, to facilitate pouring balls back into a container for randomizing. Sloping the trough makes it easier to achieve a linear arrangement 2b is a roulette-wheel type model, this one containing 50 equal-sized pockets in its circumference. A phonograph turntable may be used to rotate such a wheel. Balls are spun into the wheel independently, and all those which land in the same pocket are said to have "effective contact"

(figure 2A). The simplest convention was described by Elveback and Varma (16) as follows.

Balls of four different colors are required: "susceptibles" (e.g., S = green); "cases" (e.g., C = red); "immune" (e.g., I = blue) and "blocks," or "contact neutralizers" (e.g., N = white). Numbers of balls representing the individuals of each status, during one time period, are placed in a container along with a number of blocks (the determination of their number is discussed below). These are randomized, and poured into a trough *in single file*. It is then considered that all individuals (colored balls) which are *not* separated by a block have contact during that period—thus any "susceptible ball" which is not separated by at least one block from a "case ball" is considered to experience infectious contact. The sequence is considered to be closed-ended, in that no contact occurs between balls at opposite ends of the file. After the result (i.e., the incidence of new cases) has been recorded, the population of balls is then altered accordingly: "case balls" being substituted for susceptibles which experienced infectious contact; and "immune balls" being substituted for cases (see table 1). The entire procedure is then repeated . . . , until the epidemic expires due to absence of cases or exhaustion of susceptibles. The results of three such

TABLE 1

Interpretation of a linear sequence of balls in the standard Reed-Frost trough model. The convention assumes that a case remains infectious for one time period and then becomes permanently immune. Substitutions are made prior to randomization of balls for simulation of the subsequent time period

Random sequence for time period t	Substitutions required for simulation of subsequent time period $(t + 1)$
.	.
.	.
Block	
Susceptible	→ Case
Immune	
Case	→ Immune
Block	
Case	→ Immune
Susceptible	→ Case
Susceptible	→ Case
Block	
Susceptible	
Block	
.	.

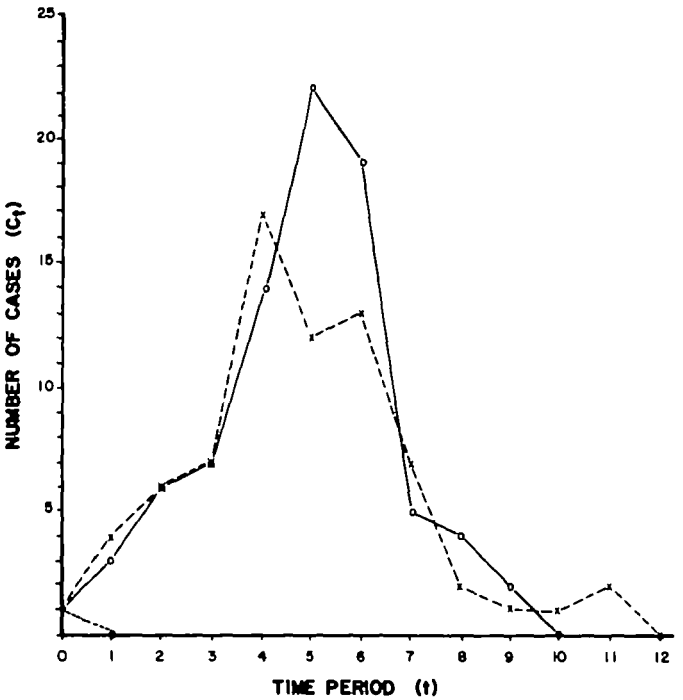

FIGURE 3. Results of three separate epidemic simulations with a Reed-Frost trough model. Each simulation began with 100 initial susceptibles and a single index case. 98 "block" balls were maintained throughout the process. The total numbers of cases to evolve in the three epidemics were 1 (------), 73 (- - -) and 84 (——).

epidemic simulations are presented in figure 3.

Such an exercise provides a vivid demonstration of the role of chance in a simple epidemic process. Of course, the conventions may be changed in numerous ways—for example: new susceptibles may be introduced into the population at any rate; some susceptibles can be converted directly to immune status, to mimic an immunization program; and the duration of infectiousness or immunity may be varied at will.

This is apparently the sort of model originally developed at Johns Hopkins about 1930, and is the one which has been widely used in teaching since that time. It may be noted that any technique for obtaining a random linear ordering of S, C, I and N units may be substituted for the traditional ball and trough apparatus. An obvious alternative would be to use cards in place of the balls, which can then be randomized by shuffling. This may have been the technique used by Horiuchi and Sugiyama (17, 18), who described their technique as "using similar chips and many shufflings." It remains for us to discuss the relationship between such mechanical models and the algebraic formulation 4.

Relationship between the mechanical and mathematical models

The initial problem is to relate the contact probabilities as defined in the mathematical formulation (expressions 1 and 4) to the probabilities of contact between simulated individuals in the mechanical Reed-Frost model.

The probability of contact p was defined as the probability that any two individuals chosen at random from the population would have contact during one time period. This can easily be expressed in terms

of the linear mechanical model, by investigating the probability that two specified colored-ball individuals, X and Y, will not be separated by a block. Assuming that there are n blocks, then, no matter where X should fall, there will be two positions out of a total of $(n + 2)$ for Y to have "effective contact" with X. This is illustrated in figure 4. Therefore, $2/(n + 2)$ defines p, the probability that any two individuals come in contact in a random sequence, and $[1 - 2/(n + 2)] = n/(n + 2)$ gives the probability that two individuals will *not* be in contact in a sequence, or q.

It would be convenient if this simple relationship were sufficient to equate the mechanical and the mathematical forms of the Reed-Frost model. But it is not. A difficulty arises in that the crucial term in the stochastic expression 4 is not the probability that a susceptible individual contacts a single case, but the probability that a susceptible individual will contact *at least one* case during the time period. (This is a crucial difference, as it distinguishes the Reed-Frost model from the earlier mass-action models of Hamer (13) and Soper (14)). In the algebraic formulation this probability could be expressed as $(1 - q^{C_t})$; but it turns out that it cannot be expressed in terms of blanks merely by substituting $n/(n + 2)$ for q. This is so because of a peculiar linear dependence in such a random sequence of balls.

The difficulty does not arise if there is only a single infectious case in the trough "population"; as under this circumstance the probability of contact with one case is equivalent to the probability of contact with at least one case, and equals p, or $2/(n + 2)$. But, let us assume that there were more than a single case, say $C_t = 2$. If we also assume that there were two block balls, then $n = 2$, and $p = 2/(2 + 2) = 0.5$. On the basis of this value of p, the probability that a susceptible contacts at least one of the cases should be, according to expression 1, equal to $1 - (1 - 0.5)^2 = 0.75$. But this does not apply to a linear arrangement of balls, say two blocks, two cases, and one susceptible. A few moments with pencil and paper, exploring the possible permutations and combinations of five such objects, will suffice to convince the reader that the probability the "susceptible ball" contacts at least one "case ball," under the linear arrangement convention, is not 0.75, but 0.7. This is illustrated in table 2.

Rather than depend only upon pencil and paper permuting, a general equation for the contact probabilities in the mechanical model can easily be derived (Appendix A). We find that the probability a

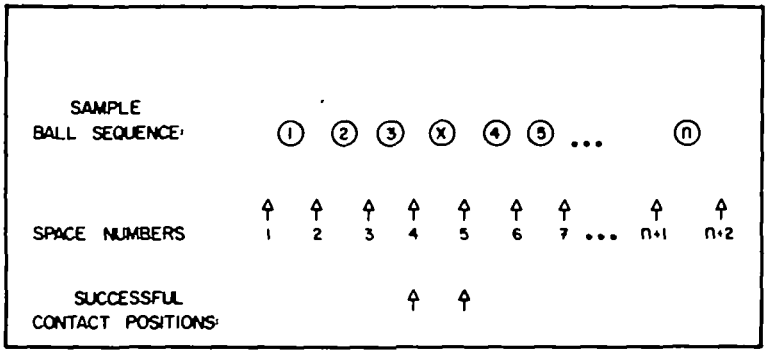

FIGURE 4. Contact probabilities between two colored balls, balls X and Y, in a linear arrangement of n "block" balls: balls $1,2,\cdots n$. Given n "blocks" and ball X, in any linear sequence, there are $(n + 2)$ possible (and equally probable) positions for ball Y. Only two of these (e.g., positions 4 and 5 in the illustrated sequence) involve contact with ball X. The probability of contact between the two colored balls, in the linear model, is therefore $2/(n + 2)$.

TABLE 2

Possible linear permutations of 5 colored balls: two cases (C), two blocks (N), and one susceptible (S). As the two case balls are identical, and the two blocks are identical, there are 5!/(2! × 2!) = 30 possible permutations. These are given below

No	Sequence	No.	Sequence	No	Sequence
1*	NNCCS	11*	NNCSC	21*	NNSCC
2*	NCNCS	12*	NCNSC	22	NSNCC
3	NCCNS	13*	NCSNC	23*	NSCNC
4*	NCCSN	14*	NCSCN	24*	NSCCN
5*	CNNCS	15*	CNNSC	25	SNNCC
6	CNCNS	16	CNSNC	26	SNCNC
7*	CNCSN	17*	CNSCN	27	SNCCN
8	CCNNS	18*	CSNNC	28*	SCNNC
9	CCNSN	19*	CSNCN	29*	SCNCN
10*	CCSNN	20*	CSCNN	30*	SCCNN

* In only 21 of the permutations, marked with an asterisk, does the susceptible contact at least one case. As each of the sequences is equally probable, the probability that the susceptible contacts at least one case, under these conditions, is then 21/30 = 0.7

"susceptible ball" contacts *at least one* "case ball," and hence contracts the infection, is:

Probability a "susceptible ball contacts at least one infected ball"

$$= \frac{C[C + 2n + 1]}{[n + C + 1][n + C]} \qquad (5)$$

where: n = the number of blocks; C = the number of "case balls."

Substitution of n = two blocks and C = two cases into this expression gives 0.7 as the probability that a "susceptible ball" would contact at least one "case ball."

This identifies the problem: given a constant number of blocks n, the linear mechanical model *underestimates* the probability that a susceptible individual contacts at least one of several cases. In so doing, the mechanical model in effect *underestimates* the probability of contact between individuals (i.e., p), if $C > 1$. The generality of this relationship is proven in Appendix A. And the challenge is clear: how does one reconcile the elegant algebraic formulation 4 with the traditional teaching model?

Reconciliation between the mathematical and mechanical models

The imperfect relationship between the algebraic and the mathematical expressions of the Reed-Frost model has been noted by many workers. Though not explicitly discussed in publications, it has been mentioned to the author by several persons familiar with the model. We will discuss here three methods for resolving this discrepancy.

1) *Maintain a constant number of blocks, but alter the assumptions.* The simplest approach is to retain a constant number of blocks throughout the course of an epidemic simulation with the trough model. As noted above, this entails a departure from the simple algebraic formulation, in that the implicit contact probability will not remain constant during the full course of the epidemic, but will decrease with an increase in prevalence (C_t). This may then be explained away, by making the not-unreasonable assumption that some public health measures would be enforced during periods of high prevalence (such as a restriction of movement of individuals), the effect of which would be to lower the basic contact probability within the population.

2) *Adjust the number of blocks at each time period of the simulated epidemic.* The mathematical formulation of the model assumes a constant contact rate "p" throughout the entire course of an epidemic. Insofar as the implied contact probability in the mechanical analogue is dependent upon both C_t and n, it should be possible to adjust the number of blocks (i.e., n) at each time period, so as to ensure a reasonably constant implicit contact rate in the trough model. A method for doing this is presented in the nomogram of figure 5. This chart allows one to select the proper number of blanks (n) necessary to maintain a (nearly) constant implied contact probability, given any number of "case balls" present in that time period. In order

to run an epidemic, one need only decide in advance upon a fixed contact probability, p (nomograms for any desired probability of contact can be easily prepared, using the expression derived in Appendix B). Then, at each period of the epidemic, the number of blocks required is read off the nomogram, based upon the prevalence (C_t) at that time.

This adjustment is not perfect – the step function nature of the nomogram disguises the continuous relationship between n and C_t – but the approximation nevertheless turns out to be very good indeed.

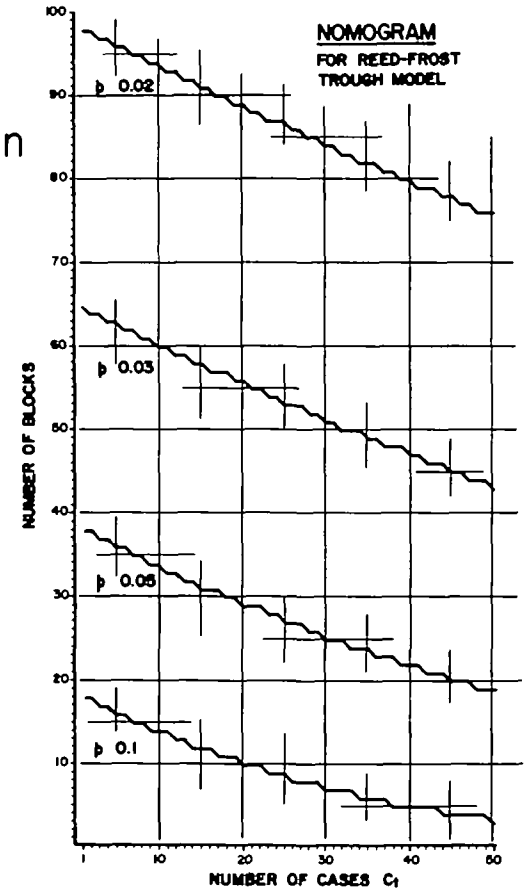

FIGURE 5. Nomogram chart designed for use with the trough analogue of the Reed-Frost model. Given any number of cases (C), and a chosen contact probability of either 0.1, 0.05, 0.03, or 0.02, the proper number of "blocks" (n) may be read off the vertical axis. This adjustment ensures that the implicit probability of contact between "individuals" in the population is as close as possible to the desired value. The derivation of the nomogram values is described in Appendix B.

3) *Development of alternative mechanical models*. The nomogram adjustment discussed above provides a compromise method for approximating the mathematical formulation of the Reed-Frost model with the traditional mechanical trough analogue. An alternative is to design another sort of apparatus which will mimic perfectly the contact probabilities derived in the algebraic model.

An effective solution may be based upon a roulette wheel principle (figure 2B). (I learned of the roulette model from colleagues at Berkeley, where this adaptation was once used in epidemiology teaching.) In this model, colored beads are again used to represent the individuals in the population. They are spun independently into a roulette-type wheel, which contains a number of equal-sized pockets on its circumference. Balls which land together in the same pocket are considered to have effective contact – thus any "susceptible balls" which land in a pocket with at least one "case ball" are considered to experience infectious contact.

In this roulette model, the probability of contact is determined only by the number of pockets on the wheel. The contact rate between individuals, defined as p, is equivalent to $1/H$, where H is the number of pockets. The standard casino roulette wheel is thus appropriate for the simulation of Reed-Frost epidemics with $p = 1/38 \approx 0.0263$. In this apparatus there is no aberration due to increases in prevalence, as long as the pockets are large enough so as never to overflow with balls. If properly designed, such a model should provide a perfect representation of the Reed-Frost assumptions.

Comparison of results using alternative mechanical models

It is of interest to compare the results of Reed-Frost epidemic simulation by each of

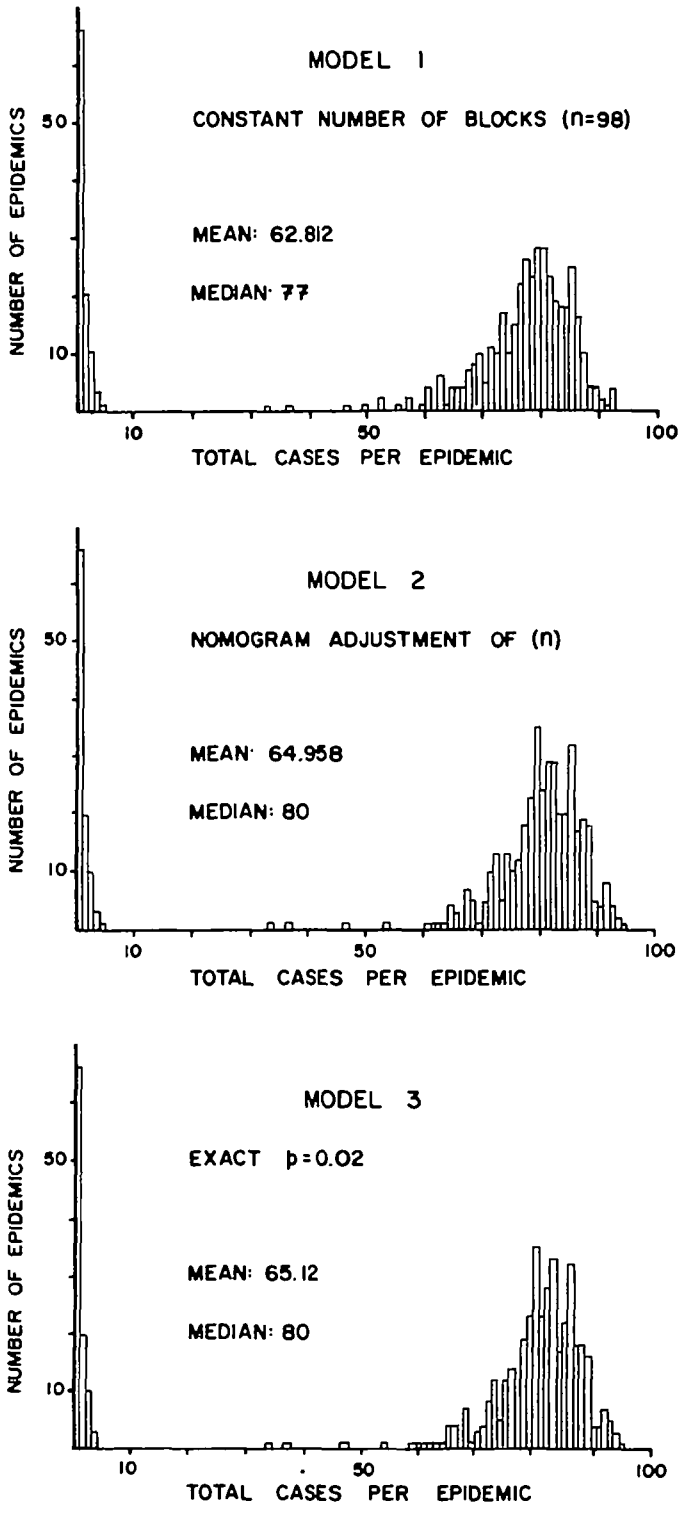

FIGURE 6. Frequency distributions of total epidemic size (total number of cases per epidemic) obtained in 500 simulations according to each of three different mechanical analogues of the Reed-Frost model. Each epidemic began with 100 initial susceptibles ($S_0 = 100$) and a single index case ($C_0 = 1$). The probability of contact between individuals in the population approximated 0.02 in each model

these three mechanical model analogues. Ironically enough, such a comparison is most easily made using electronic computer simulation; as this avoids any biases which might be introduced by the construction and operation of the several mechanical models. Appropriate contact probabilities for each of the three mechanical models can be defined in mathematical terms, and used as the basis for computer simulations.

The results presented here were derived by the simulation of Reed-Frost epidemics with $S_0 = 100$, $C_0 = 1$, and $p = 0.02$. The appropriate probability of infection for a susceptible (i.e., the probability a susceptible contacts at least one case) was calculated at each generation, based upon the number of cases present. For model 1, the probability of infection was adjusted according to expression 5, maintaining a constant n = number of blocks = 98. For model 2, the probability of infection was determined by a two-step process. First the number of blocks (n) was adjusted according to the nomogram (i.e., as the nearest integer solution to expression 14 in Appendix B, with $p = 0.02$). The probability of infection was then calculated by inserting this n into expression 5. For model 3, the probability of infection was obtained on the basis of equation 1, with $q = 0.98$ ($= 1 - p$).

The experience of each susceptible individual was then determined by comparison with a separate random number (uniformly distributed between 0 and 1); and the entire process was then repeated for each time period, until termination of the epidemic. The results are illustrated in figure 6, in terms of the frequency distribution of total epidemic size, based upon 500 separate epidemics generated by each of the three methods.

Not surprisingly, the mean and median epidemic sizes using model 1 are slightly smaller than with either of the other models. This is to be expected, as its conditions entail a lowering of effective contact rate during periods when the prevalence exceeds a single case. On the other hand, it is interesting to note how closely the results of models 2 and 3 concur. This indicates the effectiveness of the nomogram adjustment technique in maintaining the contact rate at a constant level.

Conclusions

The mechanical model developed by Reed and Frost is one of the major landmarks in the history of theoretical epidemiology. Its development coincided with the introduction of stochastic methodology in the late 1920s; and it probably provided the first technique for generating empirical solutions to a probabilistic epidemic model. But its importance for us today is far more than merely historical. Though electronic computers may have replaced such colored-ball models in empirical research on stochastic processes, the dramatic illustration of the play of chance which is afforded by such models makes them as useful today as in the past. In one sense, the model's usefulness today is greater than ever, in that it now may serve as an effective introduction to the large body of stochastic epidemic theory which has accumulated in recent years.

The publication history of the Reed-Frost model is remarkable for its delays. Just as there was a lag of many years before the basic algebraic formulation of the model was first published, so there has been an even longer delay in discussion of the properties of its mechanical analogue. Many of those who have used the model have undoubtedly noted the theoretical problems discussed in this commentary; and a variety of different solutions may well have been developed. Though those who have made such developments may, like Reed and Frost, have considered them as "too slight a contribution" for publication, it is hoped that others will profit from this discussion of the model's properties; and that this may encourage further development and use of this highly instructive and ingenious form of epidemic modelling.

References

1. Bailey NTJ: The Mathematical Theory of Infectious Diseases and Its Applications. London, Charles Griffin and Co, 1975
2. Ross R: Report on the Prevention of Malaria in Mauritius. London, Waterlow, 1908
3. Fine PEM: Ross's *a priori* pathometry—a perspective. Proc R Soc Med 68:547–551, 1975
4. Zinsser H, Wilson EB: Bacterial dissociation and a theory of the rise and decline of epidemic waves. J Prev Med 6:497–514, 1932
5. Muench H: Catalytic Models in Epidemiology. Cambridge, Harvard University Press, 1959
6. Macdonald G: The Epidemiology and Control of Malaria. London, Oxford University Press, 1957
7. Frost WH: Some conceptions of epidemics in general. Am J Epidemiol 103:141–151, 1976
8. Sartwell PE: Memoir on the Reed-Frost epidemic theory. Am J Epidemiol 103:138–140, 1976
9. Wilson EB, Burke MH. The epidemic curve. Proc Natl Acad Sci USA 28:361–367, 1942
10. Elveback LR, Fox JP, Ackerman E, et al: An influenza simulation model for immunization studies. Am J Epidemiol 103:152–165, 1976
11. Fox JP, Elveback L, Scott W, et al: Herd immunity: basic concept and relevance to public health immunization practices. Am J Epidemiol 94:179–189, 1971
12. Abbey H: An examination of the Reed Frost theory of epidemics. Hum Biol 24:201–233, 1952
13. Hamer WH: Epidemic disease in England. Lancet 1:733–739, 1906
14. Soper HE: Interpretation of periodicity in disease-prevalance. J Roy Statist Soc 92:34–73, 1927
15. Costa Maia JDO: Some mathematical developments on the epidemic theory formulated by Reed and Frost. Hum Biol 24:167–200, 1952
16. Elveback L, Varma A: Simulation of mathematical models for public health problems. Public Health Rep 80:1067–1076, 1965
17. Horiuchi K, Sugiyama H: A criticism of the Reed-Frost theory from the standpoint of stochastic epidemiology. Jap J Public Health 2(suppl):355, 1955 (in Japanese)
18. Horiuchi K, Sugiyama H: On the importance of Monte Carlo approach in the research of epidemiology. Osaka City Med J 4:59–62, 1957

Appendix A

Consider a linear sequence of balls in a trough. We assume there are C "case balls" and n "blocks," and calculate the probability that there will be at least one block between the specified susceptible and any "case ball."

Examine the sequence with no case balls, that is, containing only the n blocks and the specified "susceptible ball." By the argument illustrated in figure 4, we see that there are n out of $(n + 2)$ chances for a single "case ball" to miss contact with the susceptible. If we assume that the first "case ball" does in fact fail to contact the susceptible, then, by a similar argument, the probability that a second case *also* fails to contact the susceptible is given by $(n + 1)/(n + 3)$. And the probability that a third case also misses contact with the susceptible would be $(n + 2)/(n + 4)$. This argument is repeated for each of the C "case balls" present. And the probability that *all* of the C "case balls" escape contact with the susceptible is the product of all these probabilities, or

$$\frac{n}{(n+2)} \cdot \frac{(n+1)}{(n+3)} \cdot \frac{(n+2)}{(n+4)} \cdot \ldots \cdot \frac{(n+C-1)}{(n+C+1)}, \quad (6)$$

which reduces to:

$$\frac{n(n+1)}{(n+C+1)(n+C)}. \quad (7)$$

We are more interested in the complement of this probability, or the probability that at least one of the "case balls" does contact the specified susceptible. This is obtained by subtraction:

$$1 - \frac{n(n+1)}{(n+C+1)(n+C)} = \frac{C(C+2n+1)}{(n+C+1)(n+C)}. \quad (8)$$

This is expression 5 in the text. It should be noted that the presence of immunes, or of other susceptibles, has no bearing on this derivation; as their position in the sequence will have no effect on whether or not the *specified* "susceptible ball" contacts at least one "case ball." This is so because of the convention that all colored balls which are not separated by a block have effective contact. Contact occurs "through" susceptibles or immunes, but not through the blocks. For this reason it makes no difference whether "recovered"

case balls are removed from the simulation, or are changed to immune balls – the resultant epidemic curve should not be affected.

Expression 7, which describes the probability that a "susceptible ball" contacts no cases, is analogous to the expression q^{C_i} in the mathematical formulation of the Reed-Frost model. By equating the two expressions, we can derive a general description of the "q" value which is *implicit* in the trough analogue. And this must be the complement of the probability of contact ("p") between individuals which is *implicit* in the trough model convention:

$$(1 - \text{"}p\text{"})^C = \frac{n(n+1)}{(n+C+1)(n+C)} \quad (9)$$

or:

$$\text{"}p\text{"} = 1 - \left[\frac{n(n+1)}{(n+C+1)(n+C)}\right]^{1/C}, \quad (10)$$

where "p" refers to the *implicit* probability of contact between any two individuals in the population, as apparent in the trough model.

According to the basic assumptions of the Reed-Frost model, the value of "p" should remain constant throughout the course of an epidemic. On the other hand, expression 10 suggests that its apparent value in the trough model will be dependent upon the prevalence, i.e., upon the number of "case balls" (C). Indeed, we can demonstrate that if n remains constant as the number of "case balls" in the trough increases, the implicit "p" value will in fact decrease. We compare:

$$\text{"}p\text{"}_C = 1 - \left(\frac{n(n+1)}{(n+C+1)(n+C)}\right)^{1/C}$$

with

$$\text{"}p\text{"}_{C+1} = 1 - \left(\frac{n(n+1)}{(n+C+2)(n+C+1)}\right)^{1/(C+1)}.$$

To show that $\text{"}p\text{"}_C > \text{"}p\text{"}_{C+1}$, we need to prove that:

$$1 - \left(\frac{n(n+1)}{(n+C+1)(n+C)}\right)^{1/C} > 1 - \left(\frac{n(n+1)}{(n+C+2)(n+C+1)}\right)^{1/(C+1)}. \quad (11)$$

By subtracting one from both sides, multiplying through by minus one, and raising both sides to the power C, we have

$$\frac{n(n+1)}{(n+C+1)(n+C)} < \left[\frac{n(n+1)}{(n+C+2)(n+C+1)}\right]^{C/(C+1)}.$$

Since

$$\frac{C}{C+1} = 1 - \frac{1}{C+1},$$

the right hand side can then be factorized, and inverted, and the preceding inequality may be rewritten

$$\frac{n+C+2}{n+C} < \left[\frac{(n+C+2)(n+C+1)}{n(n+1)}\right]^{1/(C+1)},$$

or

$$\left[\frac{n+C+2}{n+C}\right]^{C+1} < \frac{(n+C+2)(n+C+1)}{n(n+1)}. \quad (12)$$

We then prove expression 12 by induction on C.

Expression 12 is clearly true for $C = 1$, by substitution:

$$\left[\frac{n+3}{n+1}\right]^2 < \frac{(n+3)(n+2)}{n(n+1)},$$

as this is equivalent to:

$$n(n+3) < (n+1)(n+2),$$

which is true for all n.

Assuming expression 12, we must show that it remains valid when C is replaced by

$C + 1$, namely

$$\left[\frac{n + C + 3}{n + C + 1}\right]^{C+2} < \frac{(n + C + 3)(n + C + 2)}{n(n + 1)}? \quad (13)$$

Recognizing that

$$\frac{n + C + 3}{n + C + 1} < \frac{n + C + 2}{n + C},$$

or

$$\left[\frac{n + C + 3}{n + C + 1}\right]^{C+1} < \left[\frac{n + C + 2}{n + C}\right]^{C+1},$$

we use expression 12 to write the inequality:

$$\left[\frac{n + C + 3}{n + C + 1}\right]^{C+1} < \frac{(n + C + 2)(n + C + 1)}{n(n + 1)}.$$

Now multiply both sides by $(n + C + 3)/(n + C + 1)$, to get expression 13, which completes the proof.

Appendix B

We wish to adjust the number of blocks in the mechanical model so that the probability a "susceptible ball" contacts at least one "case ball" is commensurate with a constant probability of contact between individuals in the population. We thus wish to maintain the "implicit p" value, as defined in Appendix A, constant throughout a simulation exercise. This may be done by solving equation 9 for n, in terms of p and C. A quadratic is obtained:

$$n^2 + n\left[\frac{2C(1 - \text{"}p\text{"})^C}{(1 - \text{"}p\text{"})^C - 1} + 1\right] + \frac{[C^2 + C](1 - \text{"}p\text{"})^C}{(1 - \text{"}p\text{"})^C - 1} = 0 \quad (14)$$

For any values of $0 < \text{"}p\text{"} < 1$ and $C > 0$, the constant term in this equation must be negative, and thus there can be but a single positive root n. These positive roots are plotted in the nomogram of figure 4. The roots have been rounded off to their nearest integer values, in order to facilitate use of the nomogram with the mechanical Reed-Frost model.

In order to prepare a nomogram for any desired contact probability, one need only substitute that value into expression 14 and solve it repeatedly with successive integer values for C.