7 Discussion

This chapter is written primarily in reaction to comments on by various readers on the preceding chapters. While we cannot undertake to implement the various suggestions that have been made as part of the present work, it will be appropriate and useful to consider them in some detail, to suggest, if only partially and at times indirectly, what might be involved in implementing them, and to discuss some of the issues they raise.

Introducing micro-epidemiological parameters

In the present formulation of the model, the calculation of new infections involves the parameters κ and θ that occur in formula (3) of Chapter 3 (the "bare bones" model) and in formulas (21) and (22) of Chapter 3 (the general minimalist model). While these parameters provide a reasonable degree of control over the course of new infections, they are not susceptible to interpretation in terms of such underlying epidemiological parameters as transmission probabilities or percent usage of condoms. Where measurements of these epidemiological parameters are available, this is a significant limitation.

It is not possible to invoke standard epidemiological formulations directly for our discrete time model because the one year periods by which the model advances are far too long for the assumptions involved to hold. At the same time, a discrete formulation involving age or duration becomes impractical if the width of age intervals is not equal to the length of the discrete time periods by which the model advances. Thus to reduce the discrete time interval in our general model, to tenths or hundredths of a year, say, would require similarly reducing the width of age intervals. This would be extremely cumbersome in practice and would in effect represent a movement toward the partial differential equation formulation that we set out to avoid.

We therefore propose a hybrid approach that begins with the formulation of a simple differential equation model involving epidemiologically interpretable parameters. The solution of this equation is then used in our discrete model to generate numbers of new infections over a one year period. Because the model is used only for a one year time period, demographic change (births and deaths) may be neglected, resulting in a very simple formulation. The approach is spelled out in detail in the remainder of this section.

Consider a population of P persons whose size does not change with time consisting at any time t of $S(t)$ susceptible and $I(t)$ infected persons. New infections occur as a result of "associations" of persons in the population. "Association" may be interpreted in various ways, including a single act of sexual intercourse or a partnership that involves many acts of intercourse. The incidence of associations is described by a parameter q that may be thought of as the annual number of associations formed divided by total population. More specifically, the number of meetings that occur over a short time period t to $t+D$ is

$$Dq\,P + o(D) \qquad (1)$$

where $o(D)$ is such that $o(D)/D$ goes to zero with D.

Over short time periods, the chance that an association involves a susceptible and an infected person is

$$2S(t)I(t). \tag{2}$$

Here $S(t)$ is the probability that the first partner to the association is susceptible, $I(t)$ the probability that the second partner is infected, and the combinatorial factor of 2 allows for the possibility that the first partner is infected and the second susceptible. The formula may be rationalized by the hypothesis that the formation of partnerships occurs at random with respect to susceptible-infected status.

Assume finally that the probability that an association of a susceptible and an infected person will result in the infection of the susceptible person is $\mathbf{b}$,

$$\textit{transmission probability} = \mathbf{b} \tag{3}$$

The number of new infections during a short time period t, $t+D$ will then be approximately

$$DqP´2S(t)I(t)´\mathbf{b} \tag{4}$$

This same number of new infections may also be expressed, however, as

$$PI(t+D) - PI(t). \tag{5}$$

Equating these two expressions for the number of new infections during the period t, $t+D$ and canceling out the total population factor P gives

$$\frac{I(t+\Delta)-I(t)}{\Delta} = 2\mathbf{bq}S(t)I(t)\,. \tag{6}$$

Taking the limit as D goes to zero gives

$$I'(t) = kI(t)S(t) \tag{7a}$$

where

$$k = 2\mathbf{bq} \tag{7b}$$

Essentially the same argument yields

$$S'(t) = -kI(t)S(t) \tag{8}$$

so we have the simultaneous equations

$$I'(t) = +kI(t)S(t) \tag{9a}$$

$$S'(t) = -kI(t)S(t) \tag{9b}$$

Dividing the first equation by *I(t)* and the second by *S(t)*, integrating and following the argument that gives the expression for the life table survival function in terms of the force of mortality function gives

$$+\frac{\ln\left[I(t)\right]}{k} = \int_0^t S(u)du \tag{10a}$$

$$-\frac{\ln\left[S(t)\right]}{k} = \int_0^t I(u)du \tag{10b}$$

Because *I(u)* and *S(u)* are proportions summing to one, summing the last two equations gives

$$\frac{1}{k}\ln\left(\frac{I(t)}{S(t)}\right) = \frac{1}{k}\ln\left(\frac{I(t)}{1-I(t)}\right) = t \tag{11}$$

from which it follows that

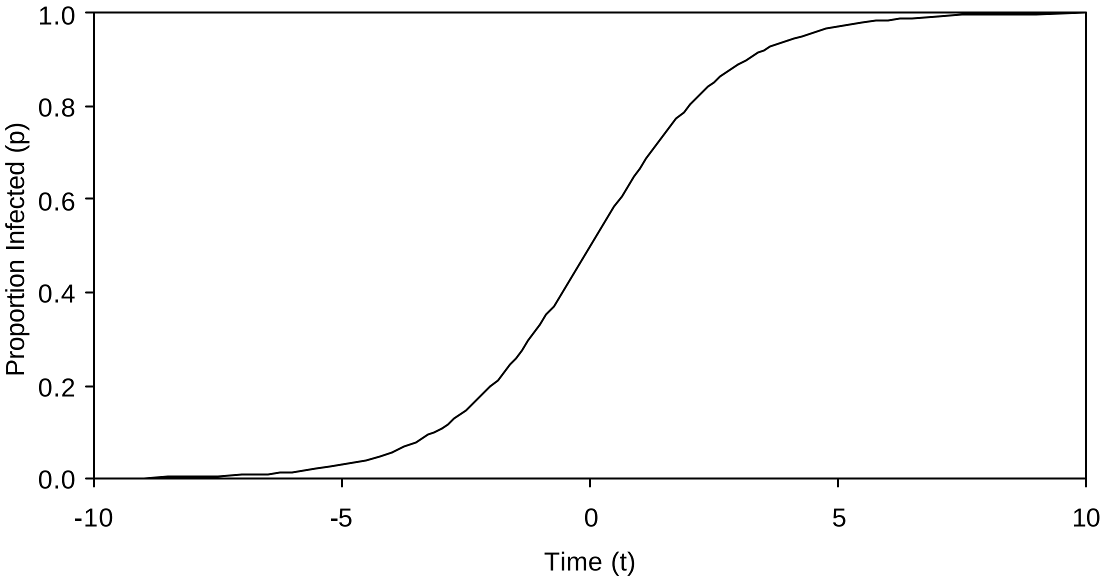
$$I(t) = \frac{\exp(kt)}{1+\exp(kt)} = \frac{1}{1+\exp(-kt)} \tag{12a}$$

$$S(t) = \frac{1}{1+\exp(kt)} \tag{12b}$$

That these do indeed represent a solution to (9a-b) may be verified directly.

The plot of *I(t)* against *t* for *k = 0.7* is shown in Figure 1 below. The model has the property that infection, once introduced, spreads eventually to the entire population. This is a consequence, first, of absence of any class of immune persons, and second, of the absence of population replacement, which might by the continuous introduction of susceptible births keep prevalence below 100 percent.

Figure 1. Proportion infected over time: $I(t) = \dfrac{1}{1+\exp(-kt)}$ with k = 0.7



Now consider the problem of determining the time *t* for which the proportion infected *I(t)* has some given value *p* between 0 and 1. Equating *p* and the right hand side of (12a) and manipulating gives

$$t = \frac{1}{k}\ln\left(\frac{p}{1-p}\right) \tag{13}$$

Substitution of this expression for t in the right hand side of

$$I(t+\Delta) = \frac{1}{1+\exp\{-k(t+\Delta)\}} \tag{14}$$

gives

$$I(t+\Delta) = \frac{1}{1+\left(\dfrac{1-p}{p}\right)\exp(-k\Delta)} \tag{15}$$

The number of new infections during the period t, $t+D$ divided by total population P is thus given by

$$\left(\frac{1}{1+\left(\frac{1-p}{p}\right)\exp(-k\Delta)} \right) - p \tag{16}$$

Tedious but elementary manipulation shows this equal to

$$p(1-p)\left(\frac{1-\exp(-k\Delta)}{p+(1-p)\exp(-k\Delta)} \right), \tag{17}$$

in which the factor $p(1-p)$ provides a point of comparison to the differential equation formulation of formulas (9a-b), with $p = I(t)$ and $1-p = S(t)$.

Consider now two alternative functions expressing the number of new infections per population (new infections divided by the number P of persons in the population) during a one year period when the proportion of infected persons at the beginning of the period is p.

For the first function, apply the differential equation expression in (9a-b), disregarding the finite time period. This gives

$$NewInfections(t) = kp(1-p)\,. \tag{18}$$

For the second function, use the expression (17) incorporating the correction factor for the finite time period with $\Delta=1$,

$$NewInfections(t) = p(1-p)\left(\frac{1-\exp(-k)}{p+(1-p)\exp(-k)} \right). \tag{19}$$

Obviously this may be written

$$NewInfections(t) = p(1-p)\left(\frac{1-\exp(-k)}{p+(1-p)\exp(-k)}\right), \tag{20}$$

so that the second expression differs from the first by the factor

$$\frac{1}{k}\left(\frac{1-\exp(-k)}{p+(1-p)\exp(-k)}\right). \tag{21}$$

For small k, the denominator of the expression in brackets is close to one and the numerator is close to k, so that the expression as a whole is close to one. As expected, then, for small k, there is no difference between the two formulations.

To see the magnitude and pattern of the correction factor (2) when k is not small, consider the following table.

	k= 0.01			**k= 0.1**			**k= 1**		
p	**kp(1-p)**	**Factor**	**Adj p(1-p)**	**kp(1-p)**	**Factor**	**Adj p(1-p)**	**kp(1-p)**	**Factor**	**Adj p(1-p)**
0.10	0.0009	1.0040	0.00090	0.0090	1.0408	0.0094	0.0900	1.4663	0.1320
0.20	0.0016	1.0030	0.00160	0.0160	1.0300	0.0165	0.1600	1.2788	0.2046
0.30	0.0021	1.0020	0.00210	0.0210	1.0195	0.0214	0.2100	1.1338	0.2381
0.40	0.0024	1.0010	0.00240	0.0240	1.0093	0.0242	0.2400	1.0184	0.2444
0.50	0.0025	1.0000	0.00250	0.0250	0.9992	0.0250	0.2500	0.9242	0.2311
0.60	0.0024	0.9990	0.00240	0.0240	0.9893	0.0237	0.2400	0.8460	0.2030
0.70	0.0021	0.9980	0.00210	0.0210	0.9796	0.0206	0.2100	0.7800	0.1638
0.80	0.0016	0.9970	0.00160	0.0160	0.9701	0.0155	0.1600	0.7236	0.1158
0.90	0.0009	0.9960	0.00090	0.0090	0.9608	0.0086	0.0900	0.6748	0.0607

When k is 0.01, the corrections are small enough to be considered negligible. For k of 0.1, they begin to be appreciable, though not large. When k is 1, they become very large indeed. Fits to epidemics with double digit percentage prevalence indicate values of k approaching one, whence the correction for such epidemics is substantial.

A reformulation of the “bare bones” model along the lines suggested here is simple enough. In place of formula (3) of Chapter 2 for the force of new infection, one would use

$$l(t) = i(t) + k(t)\,p(1-p)\times\frac{1}{k(t)}\left(\frac{1-\exp(-k(t))}{p+(1-p)\exp(-k(t))}\right). \tag{22}$$

The *k(t)* terms outside the brackets are logically superfluous but facilitate comparison with the original formulation. In place of the $[k + q(t)]$ of the original formulation we have *k(t)*, which merely saves a symbol. The p of (22) is precisely the proportion infected, written $Z(t)/P(t)$ in formula (3) of Chapter 2. The factor (1-p) reduces the level, which is sensible, since as the proportion of persons infected rises, there are fewer persons left to infect. The factor on the right, finally, corrects for the finite time period.

Simple spreadsheet calculations comparing the original and revised "bare bones" models indicate that they are qualitatively very similar, though there are noticeable differences in the shape of the new infections and prevalence curves. The revised model has the advantage of permitting the essential parameter $k(t)$ to be interpreted as $2bq$, where q describes the frequency of formation of partnerships and b the probability that an infection results from a partnership. This allows changes in the level of $k(t)$ required to mitigate the epidemic to be related to changes in either parameter.

The reformulation of the "bare bones" new infection function may of course be carried over to the general model, which utilizes essentially the same function.

Age structure of new infections

The model of Chapter 3 incorporates sufficient detail to accommodate a far more elaborate consideration of the influence of age structure on new infections that was made in that chapter. Modeling what is known, or suspected, about the importance of age of partners in the spread of the disease is a formidable problem, both conceptually and practically. In this section we consider the modeling of new infections in consideration of age-sex structure.

We begin by imitating the approach of the last section. The infection will be spread through the formation of partnerships among a population of males and females that we will term the *mixing population*. In terms of the model of Chapter 3, the mixing population will be the sexually active population of both sexes. The mixing population will be disaggregated by sex, by whether or not infected, and by age. A parameter q will describe the frequency with which partnerships are formed, with

$$Dq\,P + o(D) \tag{23}$$

giving the number of partnerships formed during some period $t,\ t+D$.

Consider first new infections of females. For a susceptible female age i to become infected during this period by an infected male age j requires that such a female form a partnership with such a male and that this partnership results in transmission of the infection. The transmission probability will simply be taken as a parameter.

To find the number of new infections of susceptible females we consider a time period sufficiently small that we may assume that a single partnership is formed and ask the probability that this partnership results in a new infection of a susceptible female of age i by an infected male of age j. The probability that the partnership involves such a female and such a male may be taken as the product of the proportion of such females in the mixing population times the proportion of such males in the mixing population, which we write simply as

$$\left(\frac{NoSusceptibleFemalesAge(i)}{NoMixingPopulation}\right)\left(\frac{NoInfectedMalesAge(j)}{NoMixingPopulation}\right) \tag{24}$$

For generality we will assume parameters a_i and b_j that give, respectively, the probability that a female age *i* will decline to enter into *any* partnership and the probability that a male age j will decline to enter into *any* partnership. We then let p_{ij}^{f} denote the probability that a female age *i* will enter into a partnership with a male age *j* should the opportunity arise and p_{ji}^{m} the probability that a male age *j* will enter into a partnership with a female age *i* should the opportunity arise. The probability that the partnership in question will result in a new infection of a female, and will then be

$$\left(\frac{SusceptibleFemalesAge(i)}{MixingPopulation}\right)\left(\frac{InfectedMalesAge(j)}{MixingPopulation}\right)\times a_i b_j \times p_{ij}^{f} p_{ji}^{m} \times b_f \tag{25}$$

where for simplicity the probability of transmission of the infection from the male to the female partner is assumed not to depend on the age of either partner. Total new infections of females age *i* are obtained by summing (2) over *j* for the given *i*.

The corresponding probability that a partnership will result in a new infection of a male is

$$\left(\frac{SusceptibleMalesAge(i)}{MixingPopulation}\right)\left(\frac{InfectedFemalesAge(j)}{MixingPopulation}\right)\times b_i a_j \times p_{ij}^{m} p_{ji}^{f} \times b_m , \tag{26}$$

where b_m denotes the probability of transmission from female to male, again assumed not to depend on the age of the partners. Total new infections of males are again obtained by summing (3) over *j* for the given *i*.

The parameters a_i and b_j allow for the possibility that females and/or males may decline partnerships with opposites of any age. Since the model of Chapter 3 distinguishes the sexually inactive from the sexually active, these parameters could be put uniformly to 1 for the sexually active population. For a model that does not distinguish susceptible persons by whether or not they are susceptible to infection, the a_i and b_j parameters could serve this function. They might also be thought of as auxiliary parameters that allow the maximum values in the preference matrices to be 1.

A notable feature of the formulation is the lack of any “dominance” by either sex. The partnerships formed reflect the independent preferences p_{ij}^{f} of females of each age *i* for males of each age *j* and the preferences p_{ij}^{m} of males of each age *i* for females of each age *j*. Males age 60 may accept females age 20 with probability 1, but if females age 20 accept males age 60 with probability 0, no such partnerships will occur.

Another notable feature is that the formulation allows both for preferences for partners of different ages and for variable *strength* of preference for partners in different age group. At one extreme, females and/or males might be completely indifferent to the age of prospective partners, in which case the matrices p_{ij}^{f} and p_{ij}^{m} would consist entirely of 1's. At the other extreme, females and/or males exercise absolute refusal for partners of a particular age group, signified by 0 values in the appropriate cells of the preference matrices.

A further notable feature is that the probability of new infection of a susceptible female age *i* by an infected male age *j* depends not only on the number and age distribution of infected males, but also on the number and age distribution of susceptible females. This is because all susceptible females, not only those age *i* are potential partners of the infected males. Other things being equal, for example, a rise in numbers of susceptible females at ages other than *i* may be expected to decrease the number of new infections of females age *i* because the susceptible females at other ages will tend to draw infected away from susceptible females age *i*. This dependence is evidently lacking in the formulation of Anderson and May (1989: formula 13.62).

The expression for expected numbers of new infections over a short time period is of course heavily under-determined and could at best be used to estimate the product of male and female preferences. The preference probability matrices might be estimated directly by psychometric methods, however, should it be considered worthwhile to do so. The parameters a_i and b_j might also be measured directly, the approach depending on what exactly constitutes being exposed to the risk of infection.

We shall not consider here the formulation and solution of differential equations corresponding to this age-specific model for new infections. The formation of the equations will not be difficult. Whether the method of the preceding section can be directly extended to the resulting simultaneous system is perhaps doubtful, but it may point to an approximate solution that would suffice for some purposes.

Program design

This section addresses relatively technical aspects of the computer implementation of the model that will be of interest (and indeed, be comprehensible) only to readers conversant with programming.

The current implementation of the model, in the statistical computing language **R**, involves a dozen or so **R** programs that all operate on global objects in the current environment. In effect, the entire current environment is utilized as a large, freely structured data object. It was recognized from the outset that this is poor programming practice, but the time pressure for completing the work was intense and proceeding in this way saved significant time and effort in defining suitable data structures for the various programs.

This approach should not be carried forward to further development of the model, however. If further developed in **R**, the first step should be to develop an implementation based on more appropriate data structures.

Discrete versus continuous formulations

Demographers are accustomed to discrete formulations for population projections, though continuous formulations of both the integral equation and partial differential equation formulation are familiar to mathematical demographers. Our argument for pursuing a discrete formulation, noted at the outset, was firstly that it is familiar to demographers and secondly that numerical solution of partial differential equations would in any case involve something very similar to the discrete calculations of familiar (to demographers) “component” projection.

The effort to introduce micro-epidemiological parameters described in the first section of this chapter has shifted our perspective. The events with which demographers deal are relatively infrequent, tending to occur at most once in any given year to any given person. The epidemiological events that lead to new infections occur far more frequently, whence modeling in discrete periods of one year is unsatisfactory.

A direct, differential equation formulation of the “bare bones” model is clearly far simpler than the mathematical contortions involved in deriving the correction factor introduced above, and if age-structured models were not an objective, we would discard the discrete “bare bones” formulation for the considerably simpler differential equation formulation. It is not clear how much is gained by this, given the highly stylized nature of the epidemiological parameters, but something is gained and essentially nothing is lost.

With age-structured models the situation is less clear cut. Simple differential equation models cope with any number of discrete compartments, but to accommodate age, which varies continuous, requires moving to partial differential equations, which are considerably more complicated. If it is possible to remain with one year age intervals, the familiar discrete formulation of demography is substantially simpler and, where analytical solutions are not expected, preferable.

If circumstances require time intervals of less than one year, however, the discrete approach involves complex and tedious, if in principal rather trivial, interpolations to break down age distributions and schedules from single years to smaller units. Given the labour involved, it may be that the partial differential equation route, on which considerable work has been done, might be preferable, primarily because it is more general and generalizable.

Our age-structured model involves not only age, however, but also duration, and a continuous formulation of this model would required doubly partial differential equation models. We are not aware of any such models in the epidemiological literature, which has tended to handle duration dependence with a series of discrete compartments with exponentially distributed departure times. This strikes us as clumsy and inflexible compared to the direct incorporation of duration. Given the difficulties that a partial differential equation model incorporating age and duration would (we suspect) entail, we think that the discrete formulation presented in Chapter 3 may be advantageous.