An introduction to mathematical models in sexually transmitted disease epidemiology

G P Garnett

Mathematical models serve a number of roles in understanding sexually transmitted infection epidemiology and control. This article seeks to provide the non-mathematician with a description of their construction and use and presents illustrative examples from sexually transmitted infection epidemiology.

The epidemiology of infectious diseases has moved beyond identifying aetiological agents and risk factors to a more detailed understanding of the mechanisms controlling the distribution of infections and disease in populations. Mathematical models provide an explicit framework within which to develop and communicate an understanding of infectious disease transmission dynamics. Because we can identify the contacts necessary for the spread of sexually transmitted infections (STIs) they provide an interesting subject for mathematical models and substantial progress has been made in model development. However, the language of mathematics can be intimidating for those unused to it. This can often lead to interesting work being ignored, or, more significantly, the uncritical acceptance of the results of models. Often mathematical modellers are responsible, either by describing complex and irrelevant detail or, alternatively, simply presenting a “black box” without ever explaining the assumptions that are critical to the results derived. In the presentation of mathematical models a balance is required so that sufficient and comprehensible descriptions can move the subject forward under the critical scrutiny of the scientific community.

This paper aims to provide an overview of the terms and methods used in mathematical models and a brief illustration of the use of models in sexually transmitted disease (STD) epidemiology.

THE FUNCTIONS OF MATHEMATICAL MODELS

The process of describing a system like the spread of an infectious disease forces one to recognise the assumptions made, the data available to estimate parameter values, and allows for qualitative or quantitative predictions that can be tested by comparison with experimental or observed patterns. Thus, the central role of creating and analysing mathematical models is to develop our understanding of a system. Once the transmission dynamics of an infectious disease are appropriately described by a model it is possible to evaluate the potential impact of proposed interventions. Models should assist in the identification of successful interventions, their necessary scale, and the role or ability of new technologies to deliver public health benefits. Often mathematical models are called upon to forecast the future. This has been particularly true for HIV where many models have been used with varying success. It is important to understand the limitations placed on forecasts by the poor quality of available data, uncertainty about parameter values, non-linearities in the system, and chance events.

In sexually transmitted disease epidemiology mathematical models can describe the position of individuals within the network of sexual partnerships via which infections spread allowing a more complete identification of risks for acquiring and transmitting infection. The population patterns of STI incidence can be simulated based upon descriptions of patterns of sexual behaviour and pathogen biology and compared with observed patterns to test our understanding. Subsequently, the consequences of health policy, such as poor access to care delaying STD treatment, or the use of screening for asymptomatic cases, can be calculated.

TERMINOLOGY

The methods used in modelling reflect both aims and stage of development. Models can be categorised by a number of key criteria.

Compartmental versus distributional

There are different ways of representing the modelled variables. For example, a group of people with an infection could all be grouped together as a “compartment” of infecteds, \( Y \), or we could explore the distribution of severity of disease, \( s \), in the population, where \( y(s) \) is the number with disease of severity \( s \) and the total population of infecteds \( Y \) is the sum of those in all the different levels of severity, \( s \). A compartmental model of gonorrhoea might divide the population into those susceptible, symptomatically and asymptptomatically infected, whereas a distributional model could describe gradations of symptoms, perhaps related to the size of the gonococcal colony or the level of immune response.

Discrete versus continuous

Change in the model population can take place either as a smooth continuous process or in discrete steps. In the former differential calculus developed to explore changes in one variable with a diminishingly small change in another variable is used, whereas in discrete models, difference equations, which reflect the change over the
whole time step are used. The sexual interactions allowing
the spread of STIs occurs continually, so most models of STIs
are “continuous.” However, for ease of calculation some might
use discrete time steps. For example, in demography because
of the availability of fertility and mortality rates in 5 year age
groups the United Nations uses 5 year time steps to project
population growth. To include the impact of STDs on fertility
and mortality in such models requires their impact to be
apportioned over 5 year periods.

**Deterministic versus stochastic**

In a deterministic model events are not subject to chance and
two realisations of a model using the same parameters and
exact starting conditions will give exactly the same results.
However, results can diverge in the case of deterministic chaos
because it is impossible to exactly specify starting conditions
and the value of variables. There is debate over the role of
chaos in measles where small differences in the fraction
of susceptible individuals. Many health economic models make linear
assumptions—that is, treating one more individual reduces
the number of cases by one. However, there are knock on
effects which depend upon the epidemiological context. For
example, in figure 1A the impact of a prophylactic vaccine in a
homogeneous population (that is, everyone has the same pattern
of risk behaviour) is illustrated. If we assume absolute
protection, increasing coverage has a linear impact on
prevalence, because it simply removes individuals from the at
risk population, whereas the impact of a vaccine that reduces
susceptibility increases as it moves towards eliminating infec-

**Analytical versus numerical solution**

Often for simple models exact mathematical solutions are
possible which provide powerful insights into the relation
between parameter values and results. Such solutions are
often more precise and elegant. The conceptual understanding
of the impact of a vaccine shown in figure 1A comes from
analytical solutions. However, as models increase in complexity
it is often necessary to resort to numerical solution. Here
specific parameters are entered and the results calculated. The
inclusion of sex, age, and sexual activity in the model used to
illustrate the impact of a potential HIV vaccine (fig 1B) neces-
sitates numerical solution. Standard methods are available
which reduce the errors in deriving results. Sensitivity
analysis allows the impact of varying parameters to be

---

**Figure 1** The relation between vaccine efficacy and the prevalence of infection. (A) In a homogeneous population where individuals are fully susceptible, vaccinated, or infected, the vaccine provides two types of protection: “degree”—where for all vaccine recipients efficacy is the fraction of challenges from which they are protected and “take”—where efficacy is the fraction of individuals protected from all challenges. (B) The predicted impact in a model stratified according to sex, age, and sexual activity of an HIV vaccine, with a mean duration of protection of 10 years, introduced after a decade of spread in a generalised HIV epidemic (protection is lost at a constant rate generating an exponentially distributed duration of protection). There are assumed to be two levels of efficacy (50% and 95%) and two types of protection—degree and take.

---

www.sextransinf.com
MODELS AND POLICY
A key function of models is to predict the consequences of changes such as those caused by interventions. The models provide a tool to translate the changes in patterns of behaviour or biology into an impact on infection and disease. Through exposure to economic models policy makers are predisposed towards their use in decision making and models have the advantage that they can be rapidly deployed when there is insufficient time for field and laboratory studies. Models have significant roles if they generate counterintuitive results (for example, the potential increase in early syphilis that could result from treating latent syphilis\(^{1,2}\) or include the knock on consequences of an intervention not captured in standard health economic analyses (for example, the potential benefits of an HIV vaccine that could reduce the infectiousness of breakthrough infections\(^{3}\)). It is often not necessary to include in full detail the complexity of the system to generate valid results, only sufficient complexity is required. For example, in exploring the elimination of syphilis, stochastic models that represent individual infections are required. In deterministic large population approximations an impossibly tiny fraction of infectious individuals could persist and subsequently reintroduce infection. Validation of models is important when appropriate data are available to test the models. However, convincing explanation of the causes of results and factors that have the potential to invalidate them often have to suffice. The following examples illustrate the use of mathematical models in sexually transmitted disease epidemiology.

Models as a framework for data analysis: a global HIV model
Mathematical models cannot replace surveillance data, but as data accrue they can provide a framework to analyse and communicate results. Recent work by a UNAIDS reference group has generated a simple and flexible model of the HIV epidemic which provides a framework for the analysis of seroprevalence data from around the globe.\(^{24}\) Using maximum likelihood methods the parameters for this model can be estimated from the prevalence data available (fig 3). The initial rate of growth of the epidemic depends largely on a “transmission coefficient”, \(r\), where the epidemic peaks depends on the initial fraction at risk, \(I_0\), and where the epidemic stabilises depends upon the strength of changes in recruitment to the at risk population in response to AIDS deaths (determined by the value of \(\phi\)). These three values, along with the start time of the epidemic, are estimated from local seroprevalence data, whereas the period between HIV and death, and the birth and death rates are estimated separately and entered. Examples of the use of the model in Uganda and Benin are illustrated (fig 3). The reliability of the resulting HIV curves depends upon the availability and validity of prevalence data. Because such a model is meant to be universally applicable and based on the observed outcome of the epidemic it sheds light upon the local causes of HIV spread and does not include the detail required to provide insight into interventions.

Understanding interventions: mass antibiotic administration
Within mathematical models we can explore the relation between the outcome of an intervention and its impact. In developing such models data from a wide range of sources can be combined to define the epidemiological context. In the planning stage of an intervention modelling enables us to evaluate the intervention’s potential and set meaningful targets. Additionally, models assist the interpretation of intervention trials, explaining observed results and how they might translate to a different epidemiological context. For example, the expected impact of mass antibiotic administration on the prevalence of trichomoniasis and gonorrhoea is

---

**Figure 2** The endemic prevalence a sexually transmitted infection in a heterosexual population stratified according to rates of sexual partner change. The SI model is described by a susceptible-infectious-susceptible compartmental model. The endemic (steady state) prevalence is plotted for different biological parameter values with an average rate of partner change of 2 for different patterns of mixing defined on a scale from assortative (like with like) to random according to rates of sexual partner change.
compared with syndromic management and illustrated in figure 4. The coverage and interval between rounds represented is that estimated in the Rakai trial and the results of the deterministic model are compared with those from the trial. The rebound of prevalence following a single round of antibiotic administration is illustrated. This is expected to be faster for gonorrhoea with its high transmission probability than for trichomoniasis. The model is in better agreement for trichomoniasis than gonorrhoea. In the latter case the deterministic model predicts a second rapid rebound after the second round of treatment. However, the low prevalence of infection following treatment will increase the influence of chance delays in the growth of the epidemic. The slower bounce back for trichomoniasis explains the observed reduction in prevalence in the second round of screening. Because of the previous low level of symptom recognition syndromic management is expected to have very little impact on the prevalence of infection. The observed impact indicates either a background trend of decreased STD prevalence, perhaps associated with AIDS mortality and HIV education.

The construction of a simple mathematical model

The model described here is a representation of an infection that causes no acquired immunity. The flow diagram is a means of illustrating the model population, which is compartmentalised into two categories or "state variables"—susceptible or infected.

The critical assumptions made about our infection above are that

- the basic reproductive number
- the proportion of the population susceptible:

\[ R_0 \]

\[ \frac{dX}{dt} = \text{rate of change in } X \text{ with respect to time.} \]

People can be in one of these two states, which are represented by boxes. The movement of people in the population is represented by arrows. Such a flow diagram is simply turned into a set of ordinary differential equations. An ordinary differential equation represents the instantaneous rate of change in a state variable with respect to another variable, in this case time. We have an equation for each of the two state variables, \( \frac{dX}{dt} \), for the rate of change in susceptible numbers with respect to time, and \( \frac{dY}{dt} \), for those infectious. The flows shown as arrows are calculated using the terms on the right hand side of the equations: a flow out of a box is taken away from a state variable and is a negative term, whereas a flow into a box is added to the state variable, and is a positive term. In the illustration the flows are numbered to show how the terms in the two equations correspond to the arrows in the flow diagram. To derive the terms we need to make assumptions about the flows—for example, initial entry into the susceptible population is by births, termed B in our equation. This could be a number that is consistently recruited each year regardless of the population size. Alternatively, it could be a function of the population size, N, and the birth rate for the average individual in the population, r, such that B = rN. To maintain a constant population size we assume the birth rate equals the death rate, B = µN. The critical assumptions made about our infection above are that an average susceptible X makes c contacts per unit time; that the fraction of these contacts that are infectious is simply the proportion of the total population infected, \( Y/N \) and that there is a chance \( \beta \) of transmission on each contact between an infected and a susceptible individual. Thus, the instantaneous total incidence of infection is \( X \beta(Y/N) \) and the instantaneous incidence per susceptible (often described as the force of infection, \( \lambda \)) is \( c\beta(Y/N) \). The recovery rate in the above equations is assumed to occur at a constant per infected rate \( \nu \), irrespective of how long someone has been infected. This assumes that a cohort of infecteds will decay exponentially with an average duration of infection given by \( 1/(\nu + \mu) \).

This simple model can be solved analytically, with the equilibrium values being those where the two rates of change are zero. One equilibrium, the disease free state, occurs when there are no infections. The other equilibrium, the endemic steady state, occurs when each new infection causes one more new infection. The value of the basic reproductive number \( R_0 \), which is given by \( R_0 = c\beta/(\nu + \mu) \), determines which equilibrium is stable. When it is less than one that the disease free state is stable. Alternatively the endemic steady state is stable when it is above one. It should be noted that the basic reproductive number holds for the disease free boundary, as prevalence increases exposures are wasted on those already infected. The reproductive number at time \( t \), depends on the basic reproductive number and the proportion of the population susceptible: \( R_t = R_0 (Y/N) \). At the steady state its value equals one, and because the proportion infected is \( 1 - (X/N) \) we can re-arrange the equation to derive the endemic prevalence \( Y/N = 1 - (1/R_t) \).
messages, or that exposure to the trial in the control population had an influence on other epidemiological parameters—for example, recognition of symptoms or condom use. In retrospect, more rapid follow up and administration of antibiotics would have enhanced gonorrhoea control within this trial, but may not have been practical. Model results suggest that the main role of presumptive therapy should be more frequent use in those with a high risk of acquiring and transmitting infection.

The impact of new technology: low efficacy HIV vaccines

The above conclusions are the more believable because they agree with our intuition (admittedly with the great advantage of hindsight). However, results may not always agree with intuition. For example, we are very used to the high efficacy of vaccines against simple viral infections—smallpox, measles, mumps, rubella, etc—which rely on rapid replication and further transmission to new susceptibles. It might be that the sexually transmitted infections prove a greater challenge with vaccines of lower efficacy. The intuition born of experience with earlier vaccines might lead us to reject such products, but model results demonstrate that a low efficacy HIV vaccine could have a substantial epidemiological impact. However, this should be interpreted with caution. Further analysis of the same model illustrates that a short duration of protection could be a more serious failing (fig 1B). It is important to note that in this model an individual with many sexual partners maintains a high risk over a long period, increasing the significance of the duration of protection from a vaccine. A more detailed analysis of patterns of risk, delivery schedules, and vaccine duration is required to fully understand the potential of particular hypothetical HIV vaccine properties.

GOALS FOR FUTURE MODEL DEVELOPMENT

Mathematical models of sexually transmitted infections have become more common and more sophisticated, but there is more to be done to demonstrate their worth. Theoretically, understanding the behaviour of infections within the dynamic network of sexual partnerships is a great challenge, particularly since the quality of data on network structures will always be limited and biased. An improved understanding of the pattern of infectiousness within and between infections is another important area where data are sparse and models can have an important role. In the field of interventions the influence of epidemiological context needs to be explored in more detail, but also the connection between the input into
interventions and the changing parameter values in mathematical models needs to be better described. Modelling should be an iterative process, with a dialogue between theoreticians, experimentalists, field workers, and policymakers. Models can have an important advocacy role, particularly when they have an attractive user interface, and modelling tools are likely to become increasingly available as user friendly software is developed. Hand in hand with this there is a responsibility not to mislead, which involves the education of the end user about what can and cannot be delivered by mathematical models and the confidence that we can place in our theoretical understanding.

ACKNOWLEDGEMENTS
Thanks to the Royal Society, the Medical Research Council, and the Wellcome Trust for funding and Dr John Williams and Dr Nick Grassy for help with examples.

REFERENCES