Biology and mathematics: the need for balance – An illustration from tsetse biology

John Hargrove - Senior Research Consultant at SACEMA

At a 2012 conference I fired a salvo at the abuse of the approach to mathematical epidemiology where disease systems are represented as a series of boxes, with flow between the boxes indicating the rate at which individuals move between various states (Figure 1). There is nothing inherently wrong in such a representation: compartmental modelling has a long and illustrious history and has led to important insights in a wide variety of scientific fields.

The offensiveness arises when the analysis of some allegedly "new" system actually involves nothing new mathematically, and where the models are derived and published without reference either to the realities of the disease concerned or to data.

The authors then have no idea even whether the models make any sense, which they often do not, let alone whether they can be used to describe existing data. As a natural consequence there is no attempt at parsimony, no attempt to see which boxes and arrows are strictly necessary. Hence or otherwise the existence of previous models of a given disease does not act as a deterrent to the production of further similar models. Quite the contrary: the temptation – too seldom resisted – is simply to add another box, another arrow to Figure 1, stir vigorously and, "Hey Presto", another model and another publication.

Said publications often involve pages of lengthy equations designed to frighten young students and
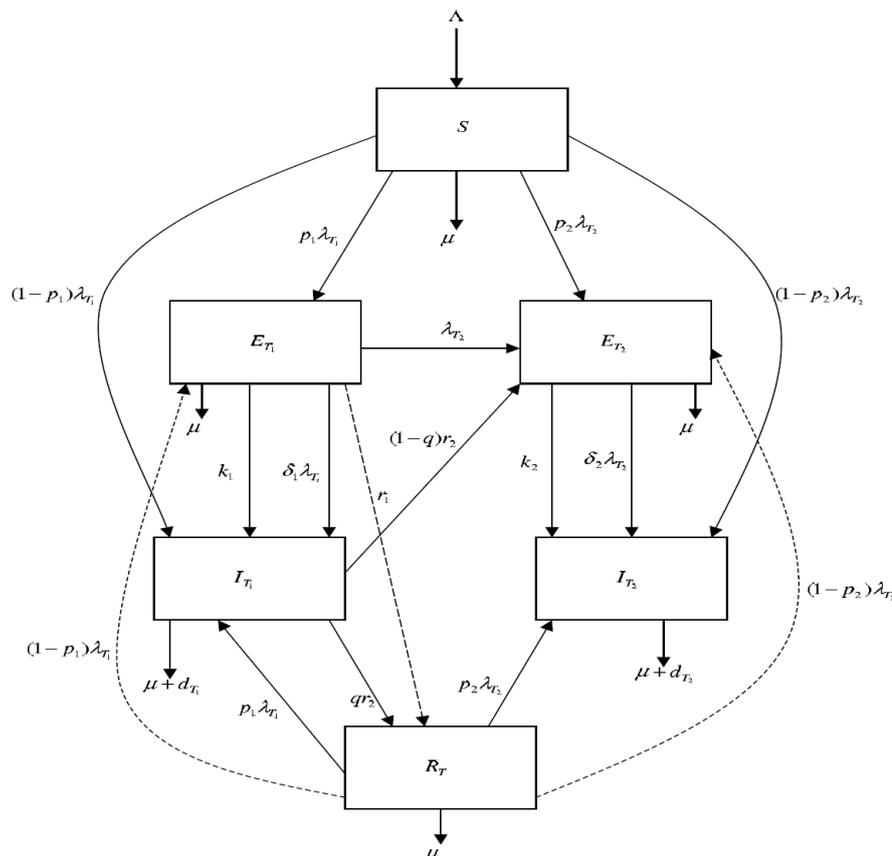

Figure 1. A "Pick a Box" model of an imaginary disease. The boxes represent various states in a disease system and the arrows possible movements between various states at rates indicated by adjacent symbols.

intimidate elderly clinicians. The latter group in particular will, like the characters in Hans Christian Anderson's story, not wish to appear as fools by entering into mathematical discussion and will not, therefore, dare to point out that the Emperor, or the model in our case, has no clothes.

As an illustration of the problems that can arise I considered the problem of estimating mortality from age distributions of populations of wild animals – using as my example the tsetse flies (*Glossina spp*) with which animals I have had a love affair spanning the past four decades (1). The mathematics involved in making such estimates is not difficult but illustrates, nonetheless, the dangers of accepting at face value the results of a mathematical analysis, and of ignoring the detailed reality of the biology.

Minimum requirements for the estimation of mortality from a cross-sectional survey of an animal population are that one can estimate the ages of the sampled animals, that all ages are sampled with equal probability, and that the underlying population has a stable age distribution. It is then a simple matter to derive a maximum likelihood estimate of the mortality from the rate at which numbers fall off with age. However, when one applies the mathematical solution to field data for tsetse flies, such as those we have derived over many years in the Zambezi Valley of Zimbabwe, the resulting estimates suggest that mortality takes a minimum value at the hottest times of the year. But these are the times when populations are at greatest stress and numbers crash – at those times, in fact, when mortality must actually be highest.

Careful consideration suggests that, at the hottest times of the year, there is an increase in mortality among immature, and very young adult flies which is sufficient to completely destabilise the age distribution and lead to manifestly absurd mortality estimates. This provides a good example of an apparently sound piece of mathematics which, when applied to real data, provides obviously erroneous results. The mathematical analysis does, however, have the saving grace that it was applied to data, so that the problems could be identified. Far more serious objections arise when, as happens too frequently, mathematical models make no attempt to address the real world in such a way that they can be tested. One must, of course, acknowledge the imperfection of all models: but, unless the model both accounts for the known biology of the problem under investigation, and is also challenged with data, the existence and nature of any imperfections will likely not be detected.

John Hargrove, Senior Research Consultant at SACEMA. Areas of interest: tsetse and trypanosomiasis biology and control. *jhargrove@sun.ac.za*

Reference:
1. Hargrove J. An example from the world of tsetse flies. Mathematical Biosciences and Engineering. 2013;10(3):691 – 704.