International Journal for Parasitology 44 (2014) 339–342

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Succinctus

Modelling parasite aggregation: disentangling statistical and ecological approaches

Laith Yakob [a,*], Ricardo J. Soares Magalhães [a], Darren J. Gray [a], Gabriel Milinovich [a], Nicola Wardrop [b], Rebecca Dunning [c], Jan Barendregt [a], Franziska Bieri [a], Gail M. Williams [a], Archie C.A. Clements [a]

[a] The University of Queensland, School of Population Health, Brisbane, Qld, Australia
[b] University of Southampton, Geography and Environment, Southampton, England, United Kingdom
[c] The University of Queensland, School of Biomedical Sciences, St Lucia, Qld, Australia

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form 25 February 2014
Accepted 26 February 2014
Available online 1 April 2014

Keywords:
Mathematical model
Infection transmission
Intervention
MDA
Soil transmitted helminths
Scenario analysis

ABSTRACT

The overdispersion in macroparasite infection intensity among host populations is commonly simulated using a constant negative binomial aggregation parameter. We describe an alternative to utilising the negative binomial approach and demonstrate important disparities in intervention efficacy projections that can come about from opting for pattern-fitting models that are not process-explicit. We present model output in the context of the epidemiology and control of soil-transmitted helminths due to the significant public health burden imposed by these parasites, but our methods are applicable to other infections with demonstrable aggregation in parasite numbers among hosts.

Some hosts harbour more parasites than others and this heterogeneity is referred to as aggregation in parasite burden. In the context of soil transmitted helminthiasis, disease severity increases proportionally to the parasite burden; distinguishing heavy from light parasite burdens is, therefore, important to policy makers (Bethony et al., 2006). Heavier, chronic parasite burdens have been implicated in a number of adverse health effects including iron deficiency anaemia, low birth weight and inhibited physical and cognitive development in children (Taylor-Robinson et al., 2012). The general rule is that 20% of a population with endemic parasitic disease harbour 80% of the worms: the so-called 20–80 rule (Woolhouse et al., 1997).

Broad recognition of this feature of macroparasitic infection has led to a bifurcation in the way pathogens are modelled. Models of microparasites (viruses, bacteria, etc.) typically simulate infection as a binary variable: an individual is infected or not infected. Incorporation of a measure of the intensity of infection is, however, desirable for models of macroparasite (worm) infections. Unfortunately, the standard method for simulating this aggregation in parasitic load among different hosts is routinely misapplied and this seems to have come about through a conflation of philosophies between statistical and ecological methods for modelling infection (Taylor et al., 1979).

The negative binomial distribution (NBD) has been used to provide statistical fit to biological data for almost a century (Greenwood and Yule, 1920). Bliss and Fisher (1953) used the NBD to fit empirical data from a wide range of different systems, including parasite abundance. These authors compared the NBD with several other statistical models of overdispersion, including the Neyman type A distribution and Fisher's logarithmic distribution, and found the NBD to be 'the most widely adaptable and generally useful' method. Following its inauguration into the biological literature through statistical fitting of data, the NBD was popularised as a method of simulating parasite aggregation among hosts in ecological models developed by Anderson and May (1978). These models have gone on to have considerable influence in the parasitological literature, both theoretical and empirical (Grenfell and Dobson, 1995; Hudson et al., 2002). In the context of macroparasites, where counting the numbers of parasites per host is more feasible, the NBD has become the prevailing method of modelling aggregation.

While there are numerous ways by which aggregation can be incorporated into host infection levels, many of which are already described by Rosà and Pugliese (2002), these often require incorporation of substantially greater biological and/or mathematical

* Corresponding author. Tel.: +61 (0)407 449115.
E-mail address: laith.yakob@uq.edu.au (L. Yakob).

http://dx.doi.org/10.1016/j.ijpara.2014.02.008
0020-7519/

complexity. With greater biological complexity comes less generalisability and projections become increasingly idiosyncratic. With greater mathematical complexity comes a much diminished subset of researchers capable of emulating, or even understanding, the methods. Central to the appeal of the NBD has been its simplicity.

First, we describe the negative binomial aggregation method. Second, we describe how this method that was originally generated to describe observed equilibria has subsequently been misapplied to simulate infection transmission dynamics during and following intervention. Finally, we present what we consider to be the simplest, biologically tractable alternative for modelling parasite aggregation among hosts.

At endemic equilibrium, the relationship between the prevalence ('*P*', proportion infected) and intensity ('*M*', mean parasite burden) of macroparasitic infections can be described by the negative binomial probability distribution (Anderson and May, 1978, 1985). A scaling factor '*k*' determines the extent of heterogeneity whereby a low value denotes extreme heterogeneity (small proportion of the host population harbouring most parasites) and a high '*k*' value generates a more even distribution of parasites among all infected hosts. The mathematical relationship between prevalence and mean parasite burden is described by the following equation:

$$P = 1 - (1 + M/k)^{-k} \tag{1}$$

The NBD model dynamics are described in terms of the mean parasite burden, *M*, which can then be converted into a measure of prevalence, *P*, using Eq. (1). The dynamics are determined from the following ordinary differential equation:

$$\frac{dM}{dt} = \gamma\left(R_0 M^2\left(1 + \left(\frac{1}{k}\right)M\alpha\right)^{-(k+1)} - M\right) \tag{2}$$

R_0 is the basic reproduction number, defined as the average number of secondary infections produced by an index case introduced into a fully susceptible population. This metric represents the threshold for controlling infection (by forcing R_0 below 1) and is calculated by multiplying the transmission coefficient (β) by the average duration of infectiousness ($1/\gamma$):

$$R_0 = \frac{\beta}{\gamma} \tag{3}$$

Soil-transmitted helminth infection control relies on chemotherapeutic drugs targeting the adult parasite stage harboured in the host gut. Mathematically, it has the effect of reducing the duration of infectiousness, thereby reducing the average number of secondary infections. A typical R_0 value for hookworm is 2 and the average duration of infectiousness is approximately 2 years (Anderson and May, 1991). Hence, we can easily calculate the transmission coefficient (which is not directly measureable in most real-world situations).

This highlights a key issue with the NBD method: simulated control reduces the R_0 but the level of parasite aggregation is assumed to remain constant. In other words, the modelled infection dynamics respond equivalently to perturbation regardless of the numerous mechanisms that might actually be responsible for imposing aggregation (Anderson and May, 1991; Grenfell and Dobson, 1995; Isham, 1995; Quinnell et al., 1995; Rosà and Pugliese, 2002). These include genetic heterogeneity in host susceptibility, different exposure-risking behaviours, infection multiplicity and parasite clumping in the environment. However, there is no reason to suspect that these different mechanisms will respond identically (or even similarly) to perturbation. Further, empirical data suggest that the degree of aggregation varies both spatially and temporally (Hudson et al., 1992). In short, use of a static *k* value does not have biological plausibility and it cannot, therefore, be relied upon to inform even the qualitative trajectories of population/infection dynamics produced by ecological models.

In place of the widely used NBD model, we present a simple, but biologically tractable, model in which humans are either susceptible to infection, lightly infected (*L*) or heavily infected (*H*). (This formulation can easily be extended to more categories if required.) The model assumes that the ratio of parasite transmission to host recovery is greater for hosts that are only lightly infected, compared with hosts that are heavily infected, ensuring that the majority of infections are characterised by light parasite burdens. This situation could arise from numerous underlying factors, including increased parasite mortality in higher burdens of nematode infection through elicitation of a more severe host immune response (WHO, 2002).

$$\frac{dL}{dt} = (\beta_L L + \beta_H H)\frac{1 - (L + H)}{2} - \gamma_L L \tag{4}$$

$$\frac{dH}{dt} = (\beta_L L + \beta_H H)\frac{1 - (L + H)}{2} - \gamma_H H \tag{5}$$

In other words $L + H = P$ (prevalence), and:

$$\frac{\beta_L}{\gamma_L} = 4 \times \frac{\beta_H}{\gamma_H} \tag{6}$$

Parameterisation ensures that 75% of the population is infected at equilibrium and that 20% of infections are in the heavily burdened group.

Community-wide drug control of the parasites was simulated to occur in a pulsed fashion every 3 months. This was performed by increasing the parasite mortality rate so that adult parasites died immediately following the day of treatment. Results are presented in terms of parasite prevalence before and during control implementation and in terms of the cumulative days infected for the treated population over the programme's time course of 5 years. Fig. 1 contrasts an infection system governed by our simple model with the classic NBD model incorporating constant *k*. There is a delay in the rate at which prevalence of infection is reduced under the NBD scenario, with the overall result of escalating projected population-level burden of infection over the course of the control programme (cumulative days infected per person are more than three times higher than our simple model's projections). To be clear, this comparison is intended for two purposes: to demonstrate the simplicity of our approach and to highlight the fact that projections of our approach are markedly different from the classically used NBD (with static *k*) which is still being used by modellers to attempt to inform current and future control policy (Anderson et al., 2013). As with any model, the usefulness of our method as a projective tool can only be informed through its adaptation to a specific control setting and its validation with data, which is work in progress.

Here, we have described an intuitive and biologically reasonable alternative to the NBD (static *k*) method for modelling parasite aggregation among hosts. Our model constitutes one of several published alternatives to the constant *k* NBD method. Pacala and Dobson (1988) produced a suite of models of parasite aggregation resulting from density-dependent factors such as parasite mortality and recruitment and parasite-induced host mortality in order to provide indirect evidence for density-dependent effects in natural (uncontrolled) parasite-host systems. Medley et al. (1993) describe a host population that is stratified according to intrinsic host susceptibility levels in order to simulate the effects of control on an aggregated parasite. Quinnell et al. (1995) describe methods which make use of the pattern-fitting NBD but allow *k* to vary with host age (but do not project control effects). Isham (1995), Bottomley et al. (2005) and Fenton et al. (2010) describe individual-based stochastic models of multiparasite species co-infection and