European Journal of Plant Pathology **110:** 759–761, 2004.

Short communication

Daamen's incidence–severity relationship revisited

G. Hughes[1], N. McRoberts[2] and L.V. Madden[3]
[1]*School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JG, UK (E-mail: ghughes@srv0.bio.ed.ac.uk);* [2]*Crop and Soil Management Research Group, SAC, Auchincruive, Ayr, KA6 5HW, UK;* [3]*Department of Plant Pathology, Ohio State University, Wooster, OH 44691, USA*

Accepted 10 December 2003

Key words: disease incidence, disease severity, binomial distribution, beta distribution, incomplete beta function

Abstract

A relationship between disease incidence and severity first discussed by Daamen (1986) is reviewed. It is shown that Daamen's original analysis is consistent with the use of the incomplete beta function as a statistical analogue of disease severity. Further analysis shows that this use of the incomplete beta function leads to a generalization of Daamen's incidence–severity relationship, permitting a wider range of application than originally envisaged.

In the predecessor of this journal, Daamen (1986) discussed a relationship between disease severity (the proportion of leaf surface visibly diseased) and disease incidence (the proportion of leaves visibly diseased) of powdery mildew (*Blumeria graminis* f. sp. *tritici*) of winter wheat. The derived model was found to be useful for conceptualizing severity–incidence relationships in particular pathosystems where certain simplifying assumptions (mentioned in the following paragraph) were deemed appropriate (Daamen, 1986; Campbell and Madden, 1990; McRoberts et al., 2003). Here we revisit this model with the intention of widening the scope of the original analysis.

First, we offer a derivation that we think follows the spirit, if not the letter, of Daamen's work. Consider a leaf as comprising a number of sites (N_m), each of which may either be occupied by a cluster of pustules or unoccupied. If each leaf has the same total area and all sites are the same size, N_m is constant. It is assumed that the distribution of occupied sites per leaf is binomial, $B(n, p)$ where $n = N_m$ and $p = p_{\text{site}}$, the probability that a site is occupied. The probability that a leaf has no occupied sites is then obtained from the zero term of the distribution as $(1 - p_{\text{site}})^{N_m}$, and the probability that a leaf is diseased (p_{leaf}) is therefore $1-(1-p_{\text{site}})^{N_m}$. If the probabilities p_{leaf} and p_{site} are estimated respectively by disease incidence at the leaf scale (the proportion of leaves visibly diseased, I_{leaf}) and the site scale (the proportion of sites visibly diseased, I_{site}):

$$I_{\text{leaf}} = 1 - (1 - I_{\text{site}})^{N_m} \qquad (1)$$

or

$$I_{\text{site}} = 1 - (1 - I_{\text{leaf}})^{1/N_m} \qquad (2)$$

These are relationships between incidence at two levels in a spatial hierarchy of the type discussed by Hughes et al. (1997). The assumptions about constant leaf and site size imply that the proportion of leaf surface visibly diseased is the same as the proportion of sites visibly diseased. Consider a particular value of I_{site}, denoted $X (0 \leq X \leq 1)$. The corresponding severity, S_X, may be thought of as the area under a standard uniform [0,1] distribution between 0 and X. The corresponding value of I_{leaf} may be calculated from Eq. (1). On this basis, Daamen (1986, Figure 1) calculated graphs of S_X (ordinate) against I_{leaf} (abscissa) for $0 \leq X \leq 1$ and $N_m = 2^i$, $i = 0, 1, \ldots, 6$.

The beta distribution is a continuous probability distribution that, in the form of interest here, is defined for a random variable x, the possible values of which are between 0 and 1 (see Olkin et al., 1980, for details). We denote this distribution Be(a, b) in which a and b, the parameters of the distribution, are positive numbers.

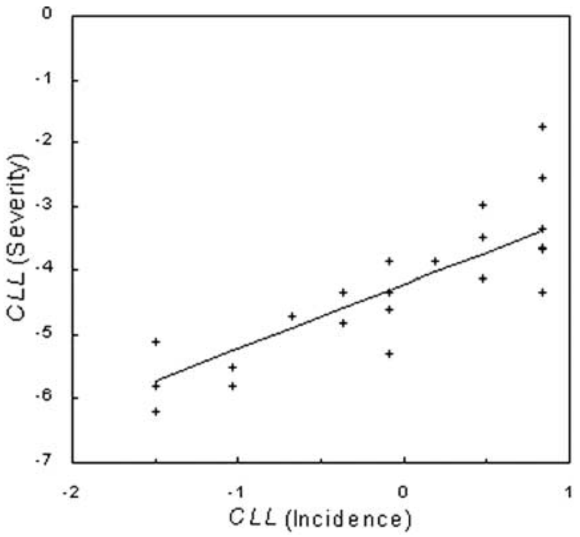

Figure 1. Disease severity and incidence of *B. graminis* f. sp. *tritici* on wheat (cv. Clark) (L.V. Madden and P.E. Lipps, Ohio State University, USA, unpubl.), plotted on complementary log–log (CLL) transformed axes. The fitted line shows $i\mathrm{Be}_X(1, q)$ (severity) against I_{leaf} (incidence at the leaf scale), all values CLL-transformed. The parameter q was estimated ($\hat{q} = 0.015$) from the data via a generalized linear model based on Eq. (6), in which the coefficient of the explanatory variable was set equal to one.

The cumulative distribution function, referred to as the incomplete beta function, is denoted here $i\mathrm{Be}_X(a, b)$ and written as:

$$i\mathrm{Be}_X(a, b) = \frac{\int_0^X x^{a-1}(1-x)^{b-1}\,\mathrm{d}x}{\int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x} \quad (0 \le X \le 1). \tag{3}$$

When a and b are both positive integers, the denominator of the expression on the right hand side (RHS) of Eq. (3) simplifies to $(a-1)!(b-1)!/(a+b-1)!$, and then simplifies further to b^{-1} if $a = 1$. The format of Eq. (3) indicates that $i\mathrm{Be}_X(a, b)$ represents an area expressed as a proportion of the whole. Thus the incomplete beta function may be thought of as a statistical analogue of disease severity, as defined above. This idea motivates the analyses that follow.

As previously, think of X as denoting a particular value of I_{site} with a corresponding value of severity S_X, and a corresponding value of I_{leaf} calculated from Eq. (1). Now note that the standard uniform distribution is a special case of the beta distribution $\mathrm{Be}(a, b)$ with parameters $a = 1$ and $b = 1$ (Olkin et al., 1980), so that when the proportion of leaf surface visibly diseased is the same as the proportion of sites visibly diseased, $S_X = i\mathrm{Be}_X(1, 1)$. The incomplete beta function $i\mathrm{Be}_X(1, 1)$ may be used to calculate a graph of S_X (ordinate) against I_{leaf} (abscissa) for $0 \le X \le 1$ and any given value of N_m, so replicating Daamen's (1986) incidence–severity relationship.

Now think of entire leaves as being diseased or not. The proportion of leaf surface visibly diseased is the same as the proportion of leaves diseased. In this case, the corresponding incidence–severity relationship may be formulated from a beta distribution $\mathrm{Be}(a, b)$ with $a = 1$ and $b = N_m$, using $S_X = i\mathrm{Be}_X(1, N_m)$ as follows. X again denotes a particular value of I_{site} with a corresponding value of severity S_X, and a corresponding value of I_{leaf} calculated from Eq. (1). Recall that N_m is a positive integer, in which case Eq. (3) becomes:

$$i\mathrm{Be}_X(1, N_m) = \int_0^X (1 - I_{\mathrm{site}})^{N_m - 1} N_m \,\mathrm{d}I_{\mathrm{site}} \tag{4a}$$

The solution to Eq. (4a) is:

$$i\mathrm{Be}_X(1, N_m) = 1 - (1 - X)^{N_m} \tag{4b}$$

As X denotes a particular value of I_{site}, the RHS of Eq. (4b) is the same as the RHS of Eq. (1). Therefore, the proportion $i\mathrm{Be}_X(1, N_m)$ (LHS of Eq. (4b)) is the same as the proportion I_{leaf} (LHS of Eq. (1)). When the proportion of leaf surface visibly diseased is the same as the proportion of leaves diseased, the use of the incomplete beta function $i\mathrm{Be}_X(1, N_m)$ to calculate a graph of S_X (ordinate) against I_{leaf} (abscissa) (i.e. retaining the format of Daamen's (1986) Figure 1) for $0 \le X \le 1$ and any given value of N_m results in a straight line relationship between $(0, 0)$ and $(1, 1)$.

In Daamen's (1986) analysis, a leaf is thought of as comprising a number of sites (N_m), each of which may either be occupied or unoccupied. The resulting analysis then effectively treats N_m (actually $q = 1/N_m$) as a parameter to be estimated from data, in which case the relationship between S_X and I_{leaf} is calculated via the assumption that sites may only be occupied (fully) or unoccupied, and the use of either Eqs. (1) or (2) (both involving I_{site}). Now, in the analysis that follows, the previous link between N_m and q is broken and a leaf is thought of as comprising a single site ($N_m = 1$ always) that may be partially occupied. This analysis provides an incidence–severity relationship directly, thus dispensing with the need to invoke Eqs. (1) or (2). As in Daamen's (1986) analysis, q is a parameter to be estimated from data. An incidence–severity relationship may be formulated from a beta distribution $\mathrm{Be}(a, b)$ with $a = 1$ (for simplicity) and $b = q$ $(0 < q \leq 1)$. In this case $S_X = i\mathrm{Be}_X(1, q)$, and the incomplete beta function $i\mathrm{Be}_X(1, q)$ may be used to calculate a graph of S_X (ordinate) against I_{leaf} (abscissa) for $0 \leq X \leq 1$ (X is now just an index variable for I_{leaf} and the corresponding S_X values) (Figure 1).

There is already a wide range of models for the analysis of incidence–severity relationships available to researchers (McRoberts et al., 2003). The analysis outlined here does not add to that range. Instead, it widens the scope of one of those models. Following Daamen (1986, Eq. (9)), we can write:

$$1 - S = (1 - I_{\mathrm{leaf}})^q \qquad (5)$$

Then:

$$\mathrm{CLL}(S) = \ln(q) + \mathrm{CLL}(I_{\mathrm{leaf}}) \qquad (6)$$

in which $\mathrm{CLL}(\bullet)$ denotes the complementary log–log transformation, $\mathrm{CLL}(\bullet) = \ln(-\ln(1 - \bullet))$. Fitting Eq. (6) to data (Figure 1) provides a method of estimating the parameter q of the incomplete beta function $i\mathrm{Be}_X(1, q)$. Previously, the incidence–severity relationship described by Eq. (5) (or Eq. (6)) was calculated using the concept of a leaf comprising sites that may only be occupied or unoccupied, and the use of either Eqs. (1) or (2) (both involving I_{site}). In the analysis presented here, the incidence–severity relationship is calculated directly, without the need to conceptualize sites and hence without invoking Eqs. (1) or (2). The use of the incomplete beta function to represent disease severity has the merit that the statistical description matches the biological variable of interest: both are continuous variables describing an area as a proportion of the whole. Modellers may find this a valuable attribute.

References

Campbell CL and Madden LV (1990) *Introduction to Plant Disease Epidemiology*. John Wiley and Sons, New York

Daamen RA (1986) Measures of disease intensity in powdery mildew (Erysiphe graminis) of winter wheat. 2. Relationships and errors of estimation of pustule number, incidence and severity. Netherlands Journal of Plant Pathology 92: 207–222

Hughes G, McRoberts N, Madden LV and Gottwald TR (1997) Relationships between disease incidence at two levels in a spatial hierarchy. Phytopathology 87: 542–550

McRoberts N, Hughes G and Madden LV (2003) The theoretical basis and practical application of relationships between different disease intensity measurements in plants. Annals of Applied Biology 142: 191–211

Olkin I, Gleser LJ and Derman C (1980) Probability Models and Applications. Macmillan Publishing Co. Inc., New York