READER'S OPINION

Vectorial capacity, basic reproduction number, force of infection and all that: formal notation to complete and adjust their classical concepts and equations

Eduardo Massad[1,2]/[+], Francisco Antonio Bezerra Coutinho[1]

[1]LIM 01-Hospital de Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brasil
[2]London School of Hygiene and Tropical Medicine, London University, London, UK

A dimensional analysis of the classical equations related to the dynamics of vector-borne infections is presented. It is provided a formal notation to complete the expressions for the Ross' Threshold Theorem, the Macdonald's basic reproduction "rate" and sporozoite "rate", Garret-Jones' vectorial capacity and Dietz-Molineaux-Thomas' force of infection. The analysis was intended to provide a formal notation that complete the classical equations proposed by these authors.

Key words: vector-borne infections - basic reproduction rate - sporozoite rate - vectorial capacity - force of infection - dimensional analysis

It is now accepted that infectious diseases will contribute a proportionately smaller burden of disease over the next two decades as non-communicable diseases emerge as public health problems. Infectious diseases, however, will continue to contribute proportionately more in the poorest quintile of the population (Molineaux 2003).

Currently the burden of vector-born infections poses a heavy toll on the major fraction of the world. Dengue represents a threat to the integrity of almost half the world's population and malaria still kills a child every 35 sec (Healthy Environments for Children Alliance) (available from: who.int/heca/infomaterials/vector-borne.pdf), not to mention several hundred thousand new yearly cases of leishmaniasis, African trypanosomiasis, among other infections transmitted by the bites of insects. In Asia, Japanese encephalitis puts three billion people at risk every year. Malaria and dengue are also major public health problems in south-east Asia. While 8% of the world population lives in the eastern Mediterranean Region, almost 11% of the global burden of vector-borne diseases is found there. Malaria is among the most prevalent causes of death and illness in Africa, where 90% of the globally reported cases of malaria take place. An estimated one million deaths from malaria occur each year, most of them in children under five years old. In Latin America and the Caribbean, dengue haemorrhagic fever has infested all major cities. Nearly 95 million people in the western Pacific are at risk of contracting malaria. To make things worse, the burden of many of these diseases is borne largely by developing countries (Hill et al. 2005).

Since the seminal work by Ronald Ross, mathematical models have provided a great deal of theoretical support for understanding the complex dynamics of those infections, in addition to the important role those models have played in designing and assessing control strategies. Key concepts like the basic reproduction number, vectorial capacity and the force of infection derived from the theoretical works on vector-borne infections are currently central to the quantification of transmission, as well as to the proposal of public health measures to control them. Some of the original mathematical expressions of these concepts suffer from minor (although important) formal inadequacies. These dimensionality inadequacies are related to incomplete notation.

In this note we revisit some of the central tenets of vector-borne infections theory from the perspective of dimensional analysis. As mentioned above, we point to some small (but important) formal inadequacies in some of the theoretical developments of some important quantifiers of risk and transmission of vector-borne infections.

Dimensional analysis - At the heart of dimensional analysis is the concept of similarity, due to Newton (1686). In physical terms, similarity refers to some equivalence between two things or phenomena that are actually different (Sonin 2012).

Essentially, the arguments of transcendental functions such as exponential, trigonometric and logarithmic functions, or of inhomogeneous polynomials, must be dimensionless quantities (Gibbings 2011). Therefore, an expression like log(30*km*) is meaningless because the argument has an explicit dimension (*km*).

More generally, dimensional analysis is, or should be, a crucial step in any mathematical approach to the real world, but surprisingly dimensional methods are so little understood and applied in the biological sciences (Stahl 1961).

Ross' Threshold Theorem - Ronald Ross, the father of the theory behind the dynamics of vector-borne infections, in a set of publications (Ross 1911), proposed what he called "pathometry", a term he later defined as "the quantitative study of disease" (Ross 1923, Bailey 1982). From a set of simple assumptions, Ross derived a system of equations for the incidence of malaria, beginning with two difference equations (Fine 1975). Let a_t be the number of uninfected and z_t the number of infected at time (t). Then

$$a_{t+1} = (1-h)va_t + HVz_t$$
$$z_{t+1} = hva_t + (1-H)Vz_t$$
(1)

Where, according to Fine (1975), "h is the infection rate H is the recovery rate, v, V refer to births, deaths, immigrations and emigrations of affected and unaffected individuals (respectively)". From the dimensional point of view, it is difficult to understand system (1). For example, h and H are rates and, therefore, have dimension T^{-1}. Therefore expressions like $(1-h)$ and $(1-H)$ are, as written, dimensionally incorrect. Below we show how to circumvent these difficulties.

The so-called Threshold Theorem refers to the critical density of mosquitoes, a, below which the disease would disappear from the human population. Ross proposed that this could be calculated by

$$a = \frac{r}{b^2} \frac{1}{i} \frac{1}{s}$$
(2)

where r is the recovery rate, b is the proportion of uninfected mosquitoes which feed on man, i is the proportion of infected individuals who are infectious and s is the proportion of mosquitoes which survive through the extrinsic incubation period. Note that according to this equation a has dimension T^{-1} which is more like a rate than a mosquito density.

Macdonald's basic reproduction "rate" and sporozoite "rate" - The year of 1952 is a landmark in the history of mathematical epidemiology. It was when George Macdonald proposed his expression for what he defined as the "basic reproduction rate" (R_0) of malaria. In his words, it is defined as the number of secondary cases produced by a single infected individual (index case) along his/her infectious period in an entirely susceptible population (Macdonald 1952a, Dietz 1993, Massad et al. 1994, 2010, Heesterbeek 2002, Lopez et al. 2002, Burattini et al. 2008). His derivation of R_0 is as follows. Let us begin by assuming that the index case is a human host. The question to be answered is how many human secondary infections this index case produces in his/her entire infectiousness period.

Let N_m be the number of female mosquitoes. Let a be the average daily biting rate female anophelines inflict in the human population. The number of bites in the human population per units of time is, therefore, $N_m a$. Let N_h be the number of humans and r be the rate of recovery from parasitemia in the human cases. Therefore, the index case produces $(N_m a/N_h r)c$ infected mosquitoes,

where c is the probability that a mosquito gets the infection after biting an infective human. Those $(N_m a/N_h r)$ c_n infected mosquitoes, in turn, produce $a(N_m a/N_h r)$ $c[1/-1n(p)]bp^n$ new human cases in the first generation, where $1/-1n(p)$ is the average life expectance of mosquitoes, b is the probability that a human gets the infection after being bitten by an infective mosquito, p is the probability that the mosquitoes survive one day and so p^n is the fraction of the infected mosquito population that survives through the extrinsic incubation period n days of the parasite. Note that, once infective a mosquito is assumed to remain so for life. Therefore, the expression for R_0 is (Macdonald 1952a):

$$R_0 = a \frac{N_m a}{N_h r} c \frac{1}{-1n(p)} bp^n$$
(3)

or calling $N_m/N_h = m$, we get the original expression:

$$R_0 = \frac{ma^2 bp^n}{-r1n(p)}$$
(4)

Similarly, if we begin with an infective mosquito as an index case and compute the number of infected mosquitoes this index case produces in the first generation we get the same expression. Note that in his original formulation Macdonald did not consider the parameter c.

In equation (4), it is possible to identify the dimensional problem in the term $1/-1n(p)$. If, on the one hand, the parameter p has a dimension (per day), then $1n(p)$ is ill defined. On the other hand, if p is dimensionless, then R_0 ends up with dimension T^{-1}, being, therefore, like a rate. We shall see below how to overcome these difficulties.

In the same year, Macdonald (1952b) proposed another useful quantity for the epidemiology of malaria, namely the "sporozoite rate" (S), which he defined as the proportion of mosquitoes with sporozoites in the salivary glands (therefore it cannot be a "rate"). His definition of S was the lifespan of the mosquitoes with sporozoites in their salivary glands and hence it can be deduced as follows. Let y be the proportion of human hosts infected (a prevalence). If the mosquitoes bite an average a bites per day then the average number of infective bites (to the mosquito) per day is ay and the probability of biting at least once a day is $(1-e^{-ay})$ (here the problem begins: if a is a rate it cannot be the argument of the exponential). Macdonald continues his argument by assuming that the probability of no infective bites is e^{-ay}. Now p was defined as the probability of mosquitoes survival for one day hence the probability of survival for one day in a non-infected state is pe^{-ay}. Therefore, the mosquitoes' life expectancy in a non-infected state is

$$\int_0^\infty (pe^{-ay})^t \, dt = \frac{1}{ay - 1n(p)}$$
(5)

but we are interested only in the life expectancy of the mosquitoes after the extrinsic incubation period n days of the parasite, that is:

$$\frac{p^n}{ay - \ln(p)} \tag{6}$$

So, the sporozoite "rate" can be expressed as

$$S = \frac{\left[\dfrac{p^n}{-\ln(p)} - \dfrac{p^n}{ay - \ln(p)}\right]}{\dfrac{1}{-\ln(p)}} \tag{7}$$

which simplifies to

$$S = \frac{p^n}{ay - \ln(p)} \tag{8}$$

Again, if the parameter *p* has a dimension (per day), then $\ln(p)$ is ill defined; if *p* is dimensionless, then the denominator of equation (8) is meaningless.

Garret-Jones' vectorial capacity - In 1964, Garret-Jones (1964a, b) (Garret-Jones & Shidrawi 1969, Dye 1990, 1992) proposed a new parameter to estimate the risk of malaria introduction which he called the "vectorial capacity" denoted *C*. This parameter is a measure which is essentially independent of the prevalence of *Plasmodium* infection. Strictly defined, the vectorial capacity is the daily rate at which future inoculations arise from a currently infective case, provided that all females mosquitoes biting that case become infected (Dye 1986). According to Garret-Jones, *C* could be obtained by the product of the number of bites per person per day [man-biting rate (*ma*)], the number of bites per day by each female mosquito [man-biting habit (*a*)] and the expectation of a mosquitoe's infective life [$p^n/-\ln(p)$], such that:

$$C = \frac{ma^2 p^n}{-\ln(p)} \tag{9}$$

Here the problem is even more serious. If *p* has a dimension (per day), then $\ln(p)$ is ill defined, but if *p* is dimensionless, then expression (9) ends up with dimension T^{-2}! Again, this can be easily corrected as we show below.

Dietz-Molineaux-Thomas' force of infection - In what can be considered the most important work on modelling malaria Dietz et al. (1974, 1980) extended the concept of vectorial capacity for a multi strains mosquitoes' populations such that if there are *J* vectors populations, with time-dependent densities [$m_j(t)$], man-biting habits (a_j) and daily survival probabilities, then the total vectorial capacity is given by

$$C(t) = \sum_j^J \frac{m_j(t)a^2 p_j^n}{-\ln(p_j)} \tag{10}$$

They then defined a time-dependent inoculation rate *h(t)* (the force of infection) as

$$h(t) = g[1-\exp(-C(t-n)y(t-n)] \tag{11}$$

which, for stable situations reduces to h = g[1-exp(-Cy)], where *g* is equivalent to *b* in Macdonald's expressions (the authors call *g* the hosts' susceptibility but we prefer to call it vector competence). At first sight the latter expression is dimensionally incorrect because *C* has dimension T^{-1}. However, it is possible to understand the author's intention and in the next section we propose a notational completion to make things clearer.

Completing the notation to solve some dimensional problems - Let us begin by trying to complete equations (1). For this we define a time interval *T*. Let a_{t+T} be the number of uninfected individuals at time *t+T*. This is equal to the number of uninfected individuals at time *t*, vTa_t that were not infected in the interval *T*, (1-hT) plus the recovered $HTVTz_t$. The reader can see that by making *T* = 1 one recovers equation (1), but should not forget that *p T* = 1 is one unit time (e.g. 1 day). The same applies for the infected equation.

Now, the way to make equation (2) dimensionally correct is to multiply the recovery rate *r* by a time unit *T*, which is made equal to 1.

Now, the dimensional inadequacies of the equations related to the basic reproduction number (R_0), the sporozoite rate (*S*) and the vectorial capacity (*C*) can be entirely solved when we use the modern notation. To the best of our knowledge the introduction of the terms *μ* for the mortality rate of the mosquitoes and $e^{-\mu\tau}$ for the probability of mosquitoes survival trough the extrinsic incubation period *τ*, is due to Aron and May (1982). In their paper, R_0 appears dimensionally correct:

$$R_0 = ma^2 be^{-\mu\tau}/r\mu \tag{12}$$

From equation (12) it is possible to write equation (4) by identify $p = e^{-\mu}$ and therefore $\ln(p) = \mu$. This, however, is dimensionally incorrect. However, it is important to note that $p = e^{\mu T}$, where $_T = 1$, such that = $-\ln(p)/T$. Therefore, equations (4), (6) and (9) should have their term $-\ln(p)$ replaced by , which makes they dimensionally correct.

The assumption behind this deduction is an exponential distribution for the mosquitoes' survival curve such that if we suppose that a mosquito female is infected as soon as she emerges as adult, it is possible to calculate what is her life expectancy in the infective condition, that is, after the extrinsic incubation period is elapsed:

$$\int_\tau^\infty e^{-\mu t}\, dt = \frac{e^{-\mu\tau}}{\mu} \tag{13}$$

which is the term that substitutes $p^n/-\ln(p)$ in the original Macdonald notation.

Now, equation (11) for the time-dependent force of infection has a hidden time interval in the sense that what is computed is the number of new infections that occurred in a give interval Δ*t*, such that in that specific interval a certain number of potentially infective bites is *C*Δ*t*, which makes equation (11) dimensionally correct. So, equation (11) should read

$$h(t) = g[1-\exp(-C(t-n)y(t-n)\Delta t] \tag{14}$$

which reduces to (11) if $\Delta t = 1$. We should not, however, forget that $\Delta t = 1$ is one unit time (e.g. 1 day).

This note was intended exclusively to provide complete notational formality to the classical equations in the field of vector-borne infections. It would be preposterous to propose a "correction" to these equations since we are pretty sure that all the authors mentioned in this note knew exactly what they were doing. It is just that they took for granted that the hidden time dimension (T=1) would be readily noticed by the readers of their paper. However, perhaps due to the fact that we are both physicists, we decided that we could contribute to readers interested in understanding in dept the theory behind the rich dynamics of vector-borne infections. Therefore, we hope that the analysis presented above can have a didactical role for the new generations of researchers in this fascinating area.

REFERENCES

Aron JL, May RM 1982. The population dynamics of malaria. In: RM Anderson (ed.), *Population dynamics of infectious diseases*, Chapman and Hall, London/New York, p. 139-179.

Bailey NTJ 1982. *The biomathematics of malaria*, Charles Griffin & Company, London/Hugh Wycombe, 210 pp.

Burattini MN, Chen M, Chow A, Coutinho FAB, Goh KT, Lopez LF, Ma S, Massad E 2008. Modelling the control strategies against dengue in Singapore. *Epidemiol Infect 136*: 309-319.

Dietz K 1993. The estimation of the basic reproduction number for infectious diseases. *Stat Meth Med Res 2*: 23-41.

Dietz K, Molineaux L, Thomas A 1974. A malaria model tested in the African savannah. *Bull WHO 50*: 347-357.

Dietz K, Molineaux L, Thomas A 1980. A mathematical model of trasnmission. In L Molineux, G Gramiccia (eds.), *The Garki Project*, World Health Organization, Geneva, p. 261-288.

Dye 1986. Vectorial capacity: must we measure all its components? *Parasitol Today 2*: 203-209.

Dye C 1990. Epidemiological significance of vector-parasite interactions. *Parasitology 101*: 409-415.

Dye C 1992. The analysis of parasite transmission by bloodsucking insects. *Ann Rev Entomol 37*: 1-19.

Fine PEM 1975. Ross's *a priori* pathometry - a perspective. *Proc Roy Soc Med 68*: 547-551.

Garret-Jones C 1964a. The Human Blood Index of malaria vectors in relation to epidemiological assessment. *Bull WHO 30*: 241-261.

Garret-Jones C 1964b. Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. *Nature 204*: 1173-1175.

Garret-Jones C, Shidrawi GR 1969. Malaria vectorial capacity of a population of *Anopheles gambiae*. *Bull WHO 40*: 531-545.

Gibbings JC 2011. *Dimensional analysis*, Springer, New York, 297 pp.

Heesterbeek JAP 2002. A brief history of R_0 and a recipe for its calculation. *Acta Biotheor 50*: 189-2002.

Hill CA, Kafatos FC, Stansfield SK, Collins FH 2005. Arthropod-borne diseases: vector control in the genomic era. *Nat Rev Microbiol 3*: 262-268.

Lopez LF, Coutinho FAB, Burattini MN, Massad E 2002. Threshold conditions for infection persistence in complex host-vectors interactions. *Comptes Rendus Biologies 325*: 1073-1084.

Macdonald G 1952a. The analysis of equilibrium in malaria. *Trop Dis Bull 49*: 813-828.

Macdonald G 1952b. The analysis of the sporozoite rate. *Trop Dis Bull 49*: 569-585.

Massad E, Coutinho FAB, Burattini MN, Amaku M 2010. Estimation of R_0 from the initial phase of an outbreak of a vector-born infection. *Trop Med Int Health 15*: 120-126.

Massad E, Coutinho FAB, Yang HM, De Carvalho HB, Mesquita F, Burattini MN 1994. The basic reproduction ratio of HIV among intravenous-drug-users. *Math Biosc 123*: 227-247.

Molyneaux DH 2003. Common themes in changing vector-borne disease scenarios. *Trans R Soc Trop Med Hyg 97*: 129-132.

Newton I 1686. *Philosophiae naturalis principia mathematica* (latin edition), Watchmaker Publishing (2011), Edinburgh, 538 pp.

Ross R 1911. *The Prevention of malaria*, 2nd ed., John Murray, London, 772 pp.

Ross R 1923. *Memoirs: whit full account of the great malaria problem and its solution*, John Murray, London, 547 pp.

Sonin AA 2012. [Assessed 2 May 2012]. The physical basis of dimensional analysis, 2nd ed. Available from: web.mit.edu/2.25/www/pdf/DA_unified.pdf.

Stahl WR 1961. Dimensional analysis in mathematical biology. *Bull Math Biophys 23*: 355-376.