ELSEVIER

Biophysical Chemistry 115 (2005) 181–185

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Statistics of infections with diversity in the pathogenicity

Francisco Guinea[a,*], Vincent A.A. Jansen[b], Nico Stollenwerk[c]

[a]Instituto de Ciencia de Materiales de Madrid, CSIC, Cantoblanco, E-28049 Madrid, Spain
[b]School of Biological Sciences, Royal Holloway-University of London, Egham, Surrey TW20 0EX, United Kingdom
[c]Centro de Matematica de Universidade do Porto, Rua do Compo Alegre 687, P-4169-007 Porto, Portugal

Received 30 July 2004; received in revised form 29 October 2004; accepted 10 December 2004
Available online 6 January 2005

Abstract

The statistics of outbreaks in a model for the propagation of meningococcal disease is analyzed, taking into account the possibility that the population is fragmented into weakly connected patches. It is shown that, depending on the size of the sample studied, the ration between the variance and the average of infected cases can vary from unity (Poisson statistics) to ϵ^{-1}, where ϵ is the normalized infection rate.

Keyword: Meningitis contagion; Neisseria meningitidis; Epidemiology; Outbreak

1. Introduction

The meningococcus is a major cause of meningitis and septicaemia. Despite this, infection with the meningococcus is mostly harmless and only rarely leads to disease. Transmission of the disease is almost exclusively through asymptomatic carriers of the disease. A predominant feature of the epidemiology of meningococcal disease is outbreaks of variable scale and duration. The meningococcal population is genetically highly diverse. We have shown, using a mathematical model, that heritable diversity with respect to pathogenic potential can lead to disease outbreaks [1–3].

Meningococcal disease is a notifiable disease in many countries. Therefore, there exist extensive data sets on the incidence of meningococcal disease. The analysis of meningococcal disease data is problematic because the number of asymptomatic carriers at any time, the variable that is probably of most interest, is normally not known because transmission of the pathogen takes place almost exclusively through asymptomatic carriers. Therefore, key epidemiological parameters are difficult to estimate and methods that are standard in epidemiology, such as outbreak reconstruction through contact tracing, cannot easily be applied. For this reason, outbreaks of meningococcal disease are difficult to reconstruct and to detect.

In this paper, we will investigate the statistical structure of an epidemiological model to infer the underlying disease process from data on the number of cases of disease. Such insights have been applied in the analysis of meningococcal disease data [3]. Here, we will investigate the validity of the assumptions made for these inferences and study how the variance in the number of cases of disease depends on the structure of the population.

2. The SIRYX model

We study the SIRYX model, considered in [1–3]. The model is an extension of the SIR model [4]. There are two types of infected individuals, I and Y. The Y's are generated by mutation from the I's at rate $\mu\beta$. For simplicity, we assume that the back mutation rate $Y \rightarrow I$ is nil. The Y population can develop disease at rate $\epsilon\beta$. The parameter ϵ is the pathogenicity: the probability to develop disease upon infection. We define the number of individuals which suffer the disease X. We further simplify the model by assuming

* Corresponding author. Tel.: +34 913349047; fax: +34 913720623.
E-mail address: paco.guinea@icmm.csic.es (F. Guinea).

0301-4622/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.bpc.2004.12.024

that these individuals are removed from the population. The mean field equations are:

$$\frac{dS}{dt} = \alpha R - \beta \frac{S}{N_p}(I+Y)$$

$$\frac{dI}{dt} = \beta(1-\mu)\frac{S}{N_p} I - \gamma I$$

$$\frac{dR}{dt} = \gamma(I+Y) - \alpha R$$

$$\frac{dY}{dt} = \beta(1-\epsilon)\frac{S}{N_p} Y - \gamma Y + \beta\mu\frac{S}{N_p} I$$

$$\frac{dX}{dt} = \beta\epsilon\frac{S}{N_p} Y - \beta\delta X \qquad (1)$$

The only difference with respect to the model studied in [1,3] is the introduction of a the rate δ at which the X individuals are removed from the population (see below). This rate implies that, in the long run, the only stationary situation is the conversion of all individuals into the X type and their eventual disappearance. We will study here quasistationary situations, which arise when δ, $\epsilon \ll 1$.

Following [1], we consider that the system is near its stable point when $\mu=0$ and $Y=0$. The remaining parameters at the fixed point are:

$$\frac{S}{N_p} = \frac{S^*}{N_p} = \frac{\gamma}{\beta}$$

$$\frac{I}{N_p} = \frac{I^*}{N_p} = \frac{\alpha}{\beta}\frac{\beta-\gamma}{\alpha+\gamma} \qquad (2)$$

Assuming that the fixed point values for S and I do not change much for small μ, we can define a simple birth-death model for the variables Y and X. We define $p(Y,t) = \sum_{X=0}^{\infty} p(Y,X,t)$, as the probability of finding the value Y at time t. This function satisfies:

$$\frac{d}{dt}p(Y,t) = [b(Y-1)+c]\,p(Y-1,t) + a(Y+1)p(Y+1,t) - (bY+aY+c)p(Y,t) \qquad (3)$$

where we have defined the death rate, $a=\gamma$, birth rate, $b=\beta(1-\varepsilon)S^*/N_p$ and rate of creation of a new individuals by mutation, $c=\beta\mu(S^*/N_p)I^*$.

We generalize this equation to the case of a system divided into M patches. The main difference is that the birth probability has to be divided into the probability that the contagion is to another individual within the same patch, which we still define as b, and the probability that the contagion leads to a new individual of type Y in another patch, b'. The total infection rate remains $b+(M-1)b'=\beta(1-\epsilon)S^*/N_p$. The generalization of Eq. (3) is:

$$\begin{aligned}\frac{d}{dt}p(\{Y_i\},t) = &\sum_i b(Y_i-1)p\left(Y_i-1,\{Y_j\},t\right) \\ &+ \sum_{j\neq i}\sum_i b'\,Y_j p(Y_i-1,\{Y_i\},t) \\ &+ \sum_i a(Y_i+1)p\left(Y_i+1,\{Y_j\},t\right) \\ &+ \sum_i cp\left(Y_i-1,\{Y_j\},t\right) \\ &- \sum_i bY_i p(\{Y_i\},t) \\ &- \sum_{j\neq i}\sum_i b'\,Y_j p(\{Y_i\},t) \\ &- \sum_i aY_i p(\{Y_i\},t) - cp(\{Y_i\},t)\end{aligned} \qquad (4)$$

Note that in this equation c is the mutation rate within one patch. The total mutation rate is Mc. When $b=b'$, we recover the limit of a well-mixed population, while for $b'=0$ the patches are decoupled. A sketch of the model is shown in Fig. 1.

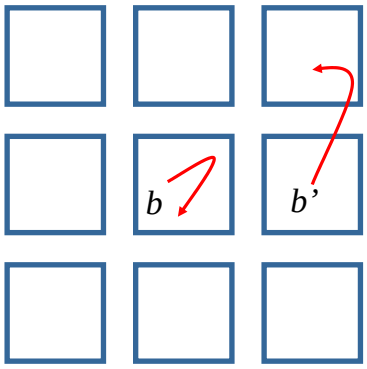

Fig. 1. Sketch of the model used in the text. The probability for contagion within the same patch is b, while the probability of contagion to any neighboring patch is b'.

3. Results

From Eq. (4), we can calculate the ensemble means of different quantities. The details of the calculations are given in Appendix A. The results are:

$$\langle Y_i\rangle = \frac{c}{a-b-(M-1)b'}$$

$$\langle Y_i^2\rangle - \langle Y_i\rangle^2 = \frac{ac[a-b-(M-2)b']}{[a-b-(M-1)b']^2(a-b+b')}$$

$$\langle Y_iY_j\rangle - \langle Y_i\rangle\langle Y_j\rangle = \frac{ab'c}{[a-b-(M-1)b']^2(a-b+b')} \qquad (5)$$

All these quantities vanish when the mutation rate is zero, $c=0$.

The net growth rate is $a\epsilon=a-b-(M-1)b'$. The number of infected cases appear with rate $X_i' = a\epsilon Y_i$.

We now calculate the number of infected individuals, X_i. We study first the case of a single population and a single variable X. The infected individuals are generated from the Y's at rate $a-b-(M-1)b'$. In order to calculate X in a single population, we use as unit of time a^{-1} and assume that the death rate of the X's is δ. We write the mutation rate $c=a$; then, we can write:

$$\begin{aligned}\frac{dP(X,Y,t)}{dt} &= \epsilon[(Y+1)P(Y+1,X-1,t) - YP(X,Y,t)] \\ &+ \delta[(X+1)P(X+1,Y,t) - XP(X,Y,t)] \\ &- \gamma P(X,Y,t)\end{aligned} \tag{6}$$

where we have used as the unit of time a^{-1}, ϵ is now the rate of conversion from Y into X, δ is the death rate of the X's and $\gamma=c/a$ is the mutation rate from I into Y. Using the techniques described in Appendix A, we can write:

$$\begin{aligned}\frac{d\langle X\rangle}{dt} &= \epsilon\langle Y\rangle - \delta\langle X\rangle \\ \frac{d\langle X^2\rangle}{dt} &= 2\epsilon\langle XY\rangle + \langle Y\rangle - 2\delta\langle X^2\rangle + \delta\langle X\rangle \\ \frac{d\langle XY\rangle}{dt} &= -\epsilon\langle XY\rangle + \epsilon\langle Y^2\rangle - \epsilon\langle Y\rangle - \delta\langle XY\rangle + c\langle X\rangle\end{aligned} \tag{7}$$

In a stationary state, the right hand side of these equations is equal to zero and we find:

$$\begin{aligned}\langle X\rangle &= \frac{\epsilon}{\delta}\langle Y\rangle \\ \langle X^2\rangle &= \frac{\langle X\rangle}{2} + \frac{\epsilon}{\delta}\left(\langle XY\rangle + \frac{\langle Y\rangle}{2}\right) \\ \langle XY\rangle &= \frac{\epsilon\langle Y^2\rangle - \epsilon\langle Y\rangle + c\langle X\rangle}{\epsilon+\delta}\end{aligned} \tag{8}$$

We substitute the first and third of these equations into the second, so that:

$$\begin{aligned}\langle X^2\rangle - \langle X\rangle^2 = &\frac{\epsilon}{\epsilon+\delta}\langle Y\rangle + \frac{\epsilon^2}{\delta(\epsilon+\delta)}\left(\langle Y^2\rangle - \langle Y^2\rangle\right) \\ &- \frac{\epsilon^3}{\delta^2(\epsilon+\delta)}\langle Y^2\rangle + \frac{\langle X\rangle}{2} \\ &+ \frac{\epsilon\gamma}{\delta(\epsilon+\gamma)}\langle X\rangle\end{aligned} \tag{9}$$

From Eq. (6), we also obtain $\langle Y\rangle=\gamma/\epsilon$. Inserting this result into Eq. (9), we have:

$$\frac{\langle X^2\rangle - \langle X\rangle^2}{\langle X\rangle} = \frac{1}{2} + \frac{\delta}{\epsilon+\delta} + \frac{\epsilon}{\epsilon+\delta}\frac{\langle Y^2\rangle - \langle Y\rangle^2}{\langle Y\rangle} \tag{10}$$

This equation relates the variance and the average of X. When the mortality rate is very high, $\delta\gg\epsilon$, we have:

$$\left.\frac{\langle X_i^2\rangle - \langle X_i\rangle^2}{\langle X_i\rangle}\right|_{\delta\epsilon} \approx 1 \tag{11}$$

The ratio approaches a constant of order unity and the process seems to have Poisson statistics. This is reasonable, because there is an approximately constant reservoir of Y individuals which can lead to an X individual, which disappears quickly, and the distribution of X cases is not influenced by the fluctuations of Y.

A more interesting regime arises if $\delta\ll\epsilon$ and $\epsilon\ll1$. Then, the r.h.s. in Eq. (10) is dominated by the third term, because $(\langle Y^2\rangle-\langle Y\rangle^2)=\langle Y\rangle\epsilon^{-1}$. We find in this case:

$$\left.\frac{\langle X_i^2\rangle - \langle X_i\rangle^2}{X_i}\right|_{\delta\ll\epsilon\ll1} \approx \frac{\langle Y_i^2\rangle - \langle Y_i\rangle^2}{\langle Y_i\rangle} \tag{12}$$

This result is the basis of the following section. Note that when $\delta=0$ the value of $\langle X_i\rangle$ increases linearly with time.

4. Size effects

Using the results in Appendix A and Eq. (12), we find (for $M\geq2$):

$$\frac{\mathrm{Var}X_i}{\langle X_i\rangle} = \frac{a[a-b-(M-2)b']}{[a-b-(M-1)b'\,](a-b+b')} \tag{13}$$

On the other hand, for the entire system, we obtain:

$$\frac{\mathrm{Var}X}{\langle X\rangle} = \frac{a}{a-b-(M-1)b'} \tag{14}$$

The linear relationship between the variance and the mean is discussed in detail in [3]. For isolated patches, $b'=0$ and $b/a=1-\epsilon$. As expected, the local and global values, Eqs. (13) and (14), coincide, giving a ratio equal to $1/\epsilon$.

In a well-mixed population, we have $b'=b$, the total birth rate is $b_{\mathrm{tot}}=Mb$, and $(Mb)/a=1-\epsilon$. Then, we find:

$$\frac{\mathrm{Var}X_i}{\langle X_i\rangle} = 1 + \frac{1-\epsilon}{M\epsilon} \tag{15}$$

$$\frac{\mathrm{Var}X}{\langle X\rangle} = \frac{1}{\epsilon} \tag{16}$$

For a small subsystem of a well-mixed population ($M\gg1/\epsilon$), we have $\mathrm{Var}X_i/\langle X_i\rangle \approx 1$. This ratio would imply that the process is due to random mutations with Poisson statistics. An analysis of the total variance, however, gives a rather different result. For large (but artificial) subdivisions of the well-mixed system, $M\epsilon\ll1$, and $\mathrm{Var}X_i/\langle X_i\rangle \approx 1=(M\epsilon)$.

It is interesting to analyze the situation in which populations of size N below some size N^* are part of a well-mixed population of size N^*, while larger populations can be considered as isolated, made up of smaller, decoupled populations of size N^*. Then, for populations

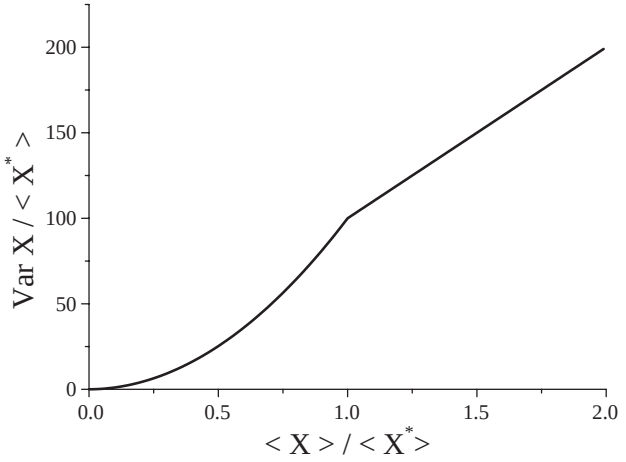

Fig. 2. Dependence of the variance of infected individuals on the average for small populations which are part of a larger, well-mixed population, $\langle X\rangle<\langle X^*\rangle$, or form an isolated population, $\langle X\rangle\geq\langle X^*\rangle$ (see Eq. (17)). The infection rate is ϵ=0.01.

$N\leq N^*$, we can use Eq. (15) with $M=N^*/N=\langle X\rangle/\langle X^*\rangle$ ($\langle X^*\rangle$ is the value of the mean of a population of size N^*), while when $N\geq N^*$ we can use Eq. (16). The variance can be written as:

$$\mathrm{Var}X=\begin{cases}\langle X\rangle+\frac{\langle X\rangle^2(1-\epsilon)}{\langle X^*\rangle\epsilon} & N<N^*\\ \frac{\langle X\rangle}{\epsilon} & N\geq N^*\end{cases} \tag{17}$$

Eq. (17) interpolates between a Poisson like regime for $N\ll N^*$ to a $1/\varepsilon$ ratio between the variance and the mean for $N\geq N^*$. A sketch of the results is shown in Fig. 2. The approximations used here imply that the coupling between different parts of the system are either strongly coupled or totally decoupled. A more realistic approximation where the inter-population contagion rate depends on the patch size is discussed in Appendix B. The calculation gives a smooth interpolation between the Poisson regime and the saturation of the ratio between the variance and the mean for high populations.

Acknowledgements

One of us (FG) is thankful to the Royal Society for a travel grant and to Royal Holloway for hospitality. F.G. also acknowledges financial support from grant MAT2002-0495-C02-01, MCyT (Spain). V.A.A.J. and N.S. acknowledge financial support from The Wellcome Trust Grant 063143.

Appendix A. Calculation of averages

From Eq. (4), one finds the equations:

$$\frac{\mathrm{d}\langle Y_i\rangle}{\mathrm{d}t}=(b-a)\langle Y_i\rangle+b'\sum_{j\neq i}\langle Y_j\rangle+c$$

$$\begin{aligned}\frac{\mathrm{d}\langle Y_i^2\rangle-\langle Y_i\rangle^2}{\mathrm{d}t}&=2(b-a)\left(\langle Y_i^2\rangle-\langle Y_i\rangle^2\right)+(a+b)\langle Y_i\rangle\\&\quad+c+2b'\sum_{j\neq i}\langle Y_j\rangle+b'\sum_i\langle Y_i\rangle\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}\langle Y_iY_j\rangle-\langle Y_i\rangle\langle Y_j\rangle}{\mathrm{d}t}&=2(b-a)\left(\langle Y_iY_j\rangle-\langle Y_i\rangle\langle Y_j\rangle\right)\\&\quad+b'\sum_{i,j\neq k}\langle Y_iY_k\rangle-\langle Y_i\rangle\langle Y_k\rangle\\&\quad+\langle Y_jY_k\rangle-\langle Y_j\rangle\langle Y_k\rangle\\&\quad+b'\sum_{j\neq i}\langle Y_i^2\rangle-\langle Y_i\rangle^2+\langle Y_j^2\rangle\\&\quad-\langle Y_j\rangle^2\end{aligned} \tag{A1}$$

so that:

$$\langle Y_i\rangle=Ke^{-[a-b-(M-1)b']t}+\frac{c}{a-b-(M-1)b'} \tag{A2}$$

where K is a constant determined by the initial conditions. We define:

$$\begin{aligned}\mathcal{C}_{ii}&=\langle Y_i^2\rangle-\langle Y_i\rangle^2\\ \mathcal{C}_{ij}&=\langle Y_iY_j\rangle-\langle Y_i\rangle\langle Y_j\rangle\end{aligned} \tag{A3}$$

These quantities do not depend on the indexes i and j. We can write the two last equations in Eq. (A1) as:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\mathcal{C}_{ii}\\ \mathcal{C}_{ij}\end{pmatrix}&=\begin{pmatrix}2(b-a) & 2(M-1)b'\\ 2b' & 2(b-a)+(M-2)b'\end{pmatrix}\begin{pmatrix}\mathcal{C}_{ii}\\ \mathcal{C}_{ij}\end{pmatrix}\\&\quad+\begin{pmatrix}(a+b)\langle Y_i\rangle+b'\sum_{j\neq i}\langle Y_j\rangle+c\\ 0\end{pmatrix}\end{aligned} \tag{A4}$$

We calculate the value of the limiting value of the quantities $\mathcal{C}_{ii}$ and $\mathcal{C}_{ij}$ at long times. Then, we can make

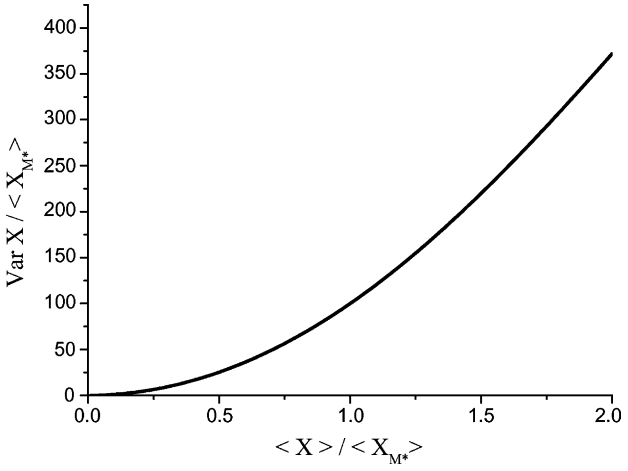

Fig. 3. Dependence of the variance of infected individuals on the average for populations when the inter-population contagion rate is size dependent (see Eq. (B3)). The infection rate is ϵ=0.01.

the l.h.s. in Eq. (A4) equal to zero. In the r.h.s., we can use:

$$\lim_{t\to\infty}\left[(a+b)\langle Y_i\rangle(t)+b'\sum_{j\neq i}\langle Y_j\rangle(t)+c\right] = \frac{2ac}{a-b-(M-1)b'} \tag{A5}$$

This leads to:

$$\lim_{t\to\infty}\begin{pmatrix}C_{ii}\\C_{ij}\end{pmatrix} = -\begin{pmatrix}2(b-a) & 2(M-1)b'\\ 2b' & 2(b-a)+(M-2)b'\end{pmatrix}^{-1} \times \begin{pmatrix}\dfrac{2ac}{a-b-(M-1)b'}\\ 0\end{pmatrix} \tag{A6}$$

This expression leads to the results in Eq. (5).

Appendix B. Size-dependent inter-population contagion rate

We generalize the analysis leading to Eq. (17) to a situation where the contagion rate between patches, $b'(M)$, depends on the size of the patch, M. We keep constant the total contagion rate, $b_{\mathrm{tot}}=b+(M-1)b(M)$. We assume that, for sufficiently large patches $M\ll M^*$, the mutual contagion rate goes to zero. Here, M^* can be used to define a typical size by $N^*=N_{\mathrm{p}}/M^*$, where N_{p} is the size the entire population. For very small sizes, $M\gg M^*$, we have $b'(M)=b$. Hence, the function $f(M/M^*)=b'(M)/b$ should satisfy:

$$f(x)\approx\begin{cases}0 & x\gg 1\\ 1 & x\ll 1\end{cases} \tag{B1}$$

Hence,

$$\frac{b(M)}{a}=\frac{1-\epsilon}{1+(M-1)f(M/M^*)} \tag{B2}$$

and, for a patch of size M, we obtain:

$$\frac{\mathrm{Var}X_M}{\langle X_M\rangle}=\frac{1}{\epsilon}-\frac{(M-1)(1-\epsilon)f(M/M^*)}{\epsilon[Mf(M/M^*)+\epsilon[1-f(M/M^*)]} \tag{B3}$$

An example of this behavior, for $f(x)=1-e^{-1/(2x^2)}$ is shown in Fig. 3.

References

[1] N. Stollenwerk, V.A.A. Jansen, Meningitis, pathogenicity near criticality: the epidemiology of meningococcal disease as a model for accidental pathogens, J. Theor. Biol. 222 (2003) 347–359.
[2] N. Stollenwerk, V.A.A. Jansen, Evolution towards criticality in an epidemiological model for meningococcal disease, Phys. Lett., A 317 (2003) 87–96.
[3] N. Stollenwerk, M.C.J. Maiden, V.A.A. Jansen, Diversity in pathogenicity can cause outbreaks of meningococcal disease, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 10229–10234.
[4] R.M. Anderson, R.M. May, Infectious Diseases of Humans, Oxford U.P., Oxford, 1991.