Persistent infections in immunocompromised hosts are rarely sources of new pathogen variants

Supplementary Material

John-Sebastian Eden[1,2], Rebecca H. Chisholm[3,4], Rowena A. Bull[5], Peter A. White[3], Edward C. Holmes[1], Mark M. Tanaka[3,4]

[1] Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney 2006, Australia
[2] Centre for Virus Research, Westmead Millennium Institute, Westmead 2145, Australia
[3] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney 2052, Australia
[4] Evolution & Ecology Research Centre, University of New South Wales, Sydney 2052, Australia
[5] Systems Medicine, Inflammation and Infection Research Centre, School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney 2052, Australia

In this supplementary material we provide

A. Single-outbreak (SIR) model for Norovirus;

B. Extended numerical results of the SIRS model for Norovirus;

C. Taylor series approximation of the population-level relative substitution rate K_c for $c \ll 1$;

D. Extended numerical results of the SIRS model for Norovirus.

E. Extended results of the sensitivity analysis of the SIRS model and SIR model for Norovirus.

A Single-outbreak model for Norovirus

In this model we assume that the outbreak occurs on a time scale much faster than the loss of immunity so that $\theta = \theta_c = 0$. In this case the dynamics are governed by the following system of equations,

$$\begin{aligned}
\frac{dI}{dt} &= \frac{\beta S}{N}\left(I + I_c(1-q)\right) - \gamma I, \\
\frac{dR}{dt} &= \gamma I, \\
\frac{dI_c}{dt} &= \frac{\beta_c S_c}{N}(1-q)\left(I + I_c(1-q)\right) - \gamma_c I_c, \\
\frac{dR_c}{dt} &= \gamma_c I_c.
\end{aligned} \tag{A.1}$$

This model can be rescaled to,

$$\begin{aligned}
\frac{di}{d\tau} &= \frac{\beta(1-c)}{\gamma}(1-i-r)\left(i + i_c\frac{(1-q)c}{1-c}\right) - i, \\
\frac{dr}{d\tau} &= i, \\
\frac{di_c}{d\tau} &= \frac{\gamma_c}{\gamma}\left(\frac{\beta_c c}{\gamma_c}(1-q)^2(1-i_c-r_c)\left(i\frac{1-c}{c(1-q)} + i_c\right) - i_c\right), \\
\frac{dr_c}{d\tau} &= \frac{\gamma_c}{\gamma} i_c.
\end{aligned} \tag{A.2}$$

Here, $I = i(1-c)N$, $R = r(1-c)N$, $S = (1-i-r)(1-c)N$, $I_c = i_c cN$, $R_c = r_c cN$, $S_c = (1-i_c-r_c)cN$, $t = \tau/\gamma$.

B Calculation of R_0 of the SIRS model for Norovirus

Basic reproductive number. Consider the SIRS model for Norovirus

$$\begin{aligned}
\frac{dI}{dt} &= \frac{\beta S}{N}\left(I + I_c(1-q)\right) - \gamma I,\\
\frac{dR}{dt} &= \gamma I - \theta R,\\
\frac{dI_c}{dt} &= \frac{\beta_c S_c}{N}(1-q)\left(I + I_c(1-q)\right) - \gamma_c I_c,\\
\frac{dR_c}{dt} &= \gamma_c I_c - \theta_c R_c.
\end{aligned} \tag{B.1}$$

The next generation method [1] is used to derive the basic reproductive number $\mathcal{R}_0$. This technique is suitable for models where there are more than one class of infectives. Following the method outlined in [1], we define F to be the rate of appearance of new infections in the general population, and F_c to be the rate of appearance of new infections in the immunocompromised population when the population is wholly susceptible so that

$$\begin{aligned}
F &= \beta(1-c)(I + (1-q)I_c)\\
F_c &= \beta_c c(1-q)(I + (1-q)I_c).
\end{aligned}$$

We also define

$$V := V^- - V^+,$$

where V^+ is the rate of transfer of infections into the general population by all other means and V^- is the rate of transfer of infections out of the general population. Similarly, we define

$$V_c := V_c^- - V_c^+,$$

where V_c^+ is the rate of transfer of infections into the immunocompromised population by all other means and V_c^- is the rate of transfer of infections out of the immunocompromised population so that

$$\begin{aligned}
V &= \gamma I\\
V_c &= \gamma_c I_c.
\end{aligned}$$

We form the next generation matrix operator $\mathcal{F}\mathcal{V}^{-1}$ from the matrices of partial derivatives of F, F_c, V and V_c:

$$\begin{aligned}
\mathcal{F} &= \begin{bmatrix} \partial_I F & \partial_{I_c} F \\ \partial_I F_c & \partial_{I_c} F_c \end{bmatrix} = \begin{bmatrix} \beta(1-c) & \beta(1-q)(1-c) \\ c\beta_c(1-q) & c\beta_c(1-q)^2 \end{bmatrix},\\
\mathcal{V} &= \begin{bmatrix} \partial_I V & \partial_{I_c} V \\ \partial_I V_c & \partial_{I_c} V_c \end{bmatrix} = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma_c \end{bmatrix},
\end{aligned}$$

and then calculate $\mathcal{R}_0$ as the largest eigenvalue of the matrix $\mathcal{F}\mathcal{V}^{-1}$. We find that

$$\mathcal{R}_0 = \frac{(1-c)\beta}{\gamma} + \frac{c(1-q)^2\beta_c}{\gamma_c}. \tag{B.2}$$

For $c \ll 1$, $\mathcal{R}_0 \approx \frac{\beta}{\gamma}$.

C Taylor series approximation of the population-level relative substitution rate at endemic equilibrium K_c

Consider the system of equations (B.1). We set each equation in (B.1) to zero to solve for the endemic equilibrium solutions $(\hat{I}, \hat{R}, \hat{I}_c, \hat{R}_c)$ such that $\hat{I} > 0$ and $\hat{I}_c > 0$. We do not reproduce the analytical expressions of these solutions here due to their size. However, they can be easily obtained using Mathematica.

We are interested in the population-level relative substitution rate at endemic equilibrium K_c which depends on these equilibrium solution according to

$$K_c = \left(\frac{k_c}{k}\right)\left(\frac{\hat{I}_c}{\hat{I}}\right). \tag{C.1}$$

The proportion of immunocompromised hosts c is assumed to be small. Therefore we seek a Taylor series expansion of K_c around the point $c = 0$ to approximate the solution for $c \ll 1$. Using Mathematica it can be shown that

$$\begin{aligned} K_c = \left(\frac{k_c}{k}\right)\left(\frac{\hat{I}_c}{\hat{I}}\right) &= \left(\frac{k_c}{k}\right)\left\{\frac{c\beta\beta_c\theta_c(1-q)(\gamma+\theta)}{(\theta_c+\gamma_c)\beta_c(1-q)\theta(\beta-\gamma)+\beta\gamma_c\theta_c(\gamma+\theta)} + O(c^2)\right\} \\ &= c\left(\frac{k_c}{k}\right)\left(\left(1-\frac{\gamma}{\beta}\right)\left(\frac{\frac{\gamma_c}{\theta_c}+1}{\frac{\gamma}{\theta}+1} + \frac{\gamma_c}{\beta_c(1-q)}\right)\right)^{-1} + O(c^2). \end{aligned} \tag{C.2}$$

Therefore, the linear approximation for K_c around $c = 0$ is

$$K_c \approx c\left(\frac{k_c}{k}\right)\left(\left(1-\frac{\gamma}{\beta}\right)\left(\frac{\frac{\gamma_c}{\theta_c}+1}{\frac{\gamma}{\theta}+1} + \frac{\gamma_c}{\beta_c(1-q)}\right)\right)^{-1}. \tag{C.3}$$

This approximation and the full numerical solution of K_c as a function of c are shown in Figure C.1 for one set of parameter values. The linear approximation is clearly close to the full solution for values of the proportion of immunocompromised hosts c up to around 0.03.

From Equation (C.3), it is clear that for small c, K_c increases when either $\beta_c(1-q)/\gamma_c$ or θ_c/γ_c increase, and K_c decreases when either β/γ or θ/γ increase. Hence, the population-level relative substitution rate at endemic equilibrium will increase if the mean number of secondary infections caused by an infective immunocompromised host $(\beta_c(1-q)/\gamma_c)$ increases, the infection duration of infective immunocompromised hosts $(1/\gamma_c)$ increases, the duration of immunity for immunocompromised hosts $(1/\theta_c)$ decreases, the mean number of secondary infections caused by an infective general host (β/γ) decreases, the infection duration of infective general hosts $(1/\gamma)$ decreases, or the duration of immunity for general hosts $(1/\theta)$ increases.

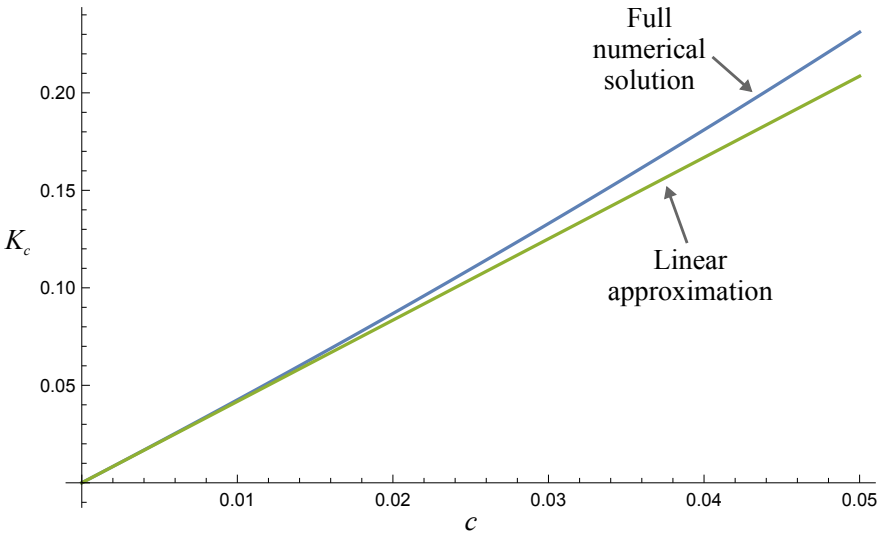

Figure C.1: **The population-level relative substitution rate at endemic equilibrium K_c as a function of the proportion of immunocompromised hosts c.** The blue line indicates the full numerical solution for K_c, and the green line is the linear approximation of K_c for $c \ll 1$. Here, $k_c/k = 1$, $\beta = 1.64/7$, $\beta_c = 1.2\beta$, $q = 0.5$, $N = 1000$, $\gamma = 1/7$, $\gamma_c = 1/31$, $\theta = 1/(5 \times 365)$, $\theta_c = 1/(0.5 \times 365)$.

D Extended numerical results of the SIRS model for Norovirus.

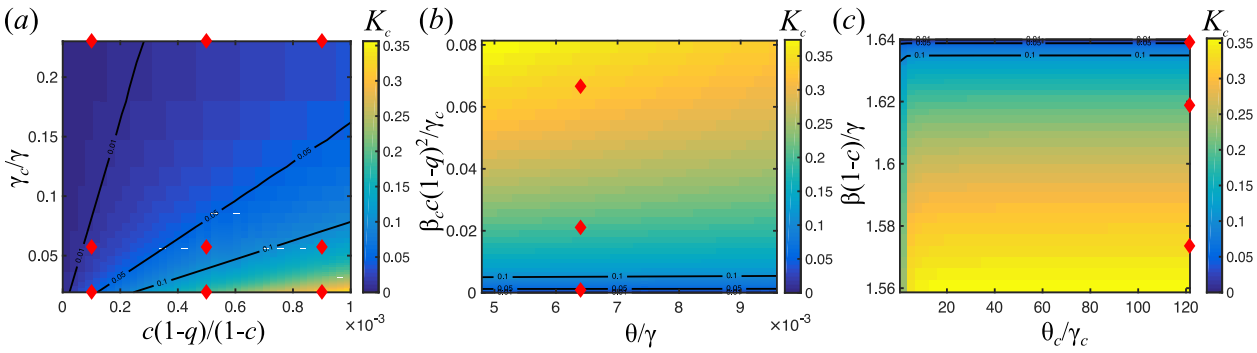

Figure D.1: **Extended numerical results for SIRS model for norovirus.** The relative substitution rate (at the population level) in the immunocompromised population versus the general population K_c when $k_c/k = 5$ and $N = 10^6$. (a) $R_0 = 1.64$, $1/\theta_c = 3$ days, $1/\gamma = 7$ days, $c = 0.001$, $1/\theta = 3$ years, $k = 0.003$, $1/\gamma_c \in [1/12, 1]$ years, $q \in [0, 1]$; (b) $R_0 = 1.64$, $1/\gamma_c = 12$ months, $1/\gamma = 7$ days, $c = 0.001$, $1/\theta_c = 3$ days, $k = 0.003$, $1/\theta \in [2, 4]$ years, $q \in [0, 1]$; (c) $R_0 = 1.64$, $1/\gamma_c = 12$ months, $1/\gamma = 7$ days, $c = 0.001$, $1/\theta = 3$ years, $k = 0.003$, $1/\theta_c \in [3 \text{ days}, 3 \text{ years}]$, $q \in [0, 1]$. Pink diamonds represent the best estimates of the parameters in the parameter space considered. The black lines correspond to the contours $K_c = 0.1, 0.05, 0.01$.

E Extended results of the sensitivity analysis of the SIRS and SIR model for Norovirus.

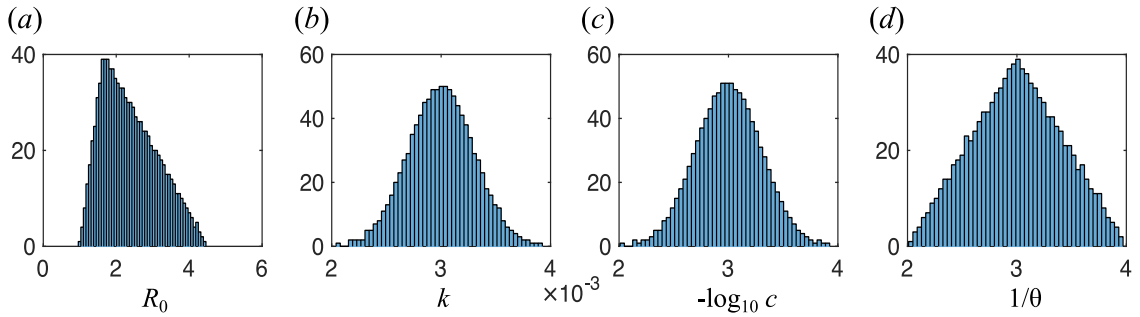

Figure E.1: **Parameter sample distributions used in the sensitivity analysis using the Latin Hypercubic Sampling method.** (a) The basic reproduction number (R_0) was sampled from a triangular distribution with mode 1.64, lower bound 1 and upper bound 4.5; (b) the general population substitution rate (k) was sampled from a normal distribution with mean 0.003 and standard deviation 0.0003; (c) the proportion of the population immunocompromised (c) was sampled from a log normal distribution of base 10 such that $-\log_{10} c \sim \mathcal{N}(3, 0.3)$; (d) the duration of immunity for the general population ($1/\theta$) was sampled from a triangular distribution with mode 3, lower bound 2 and upper bound 4.

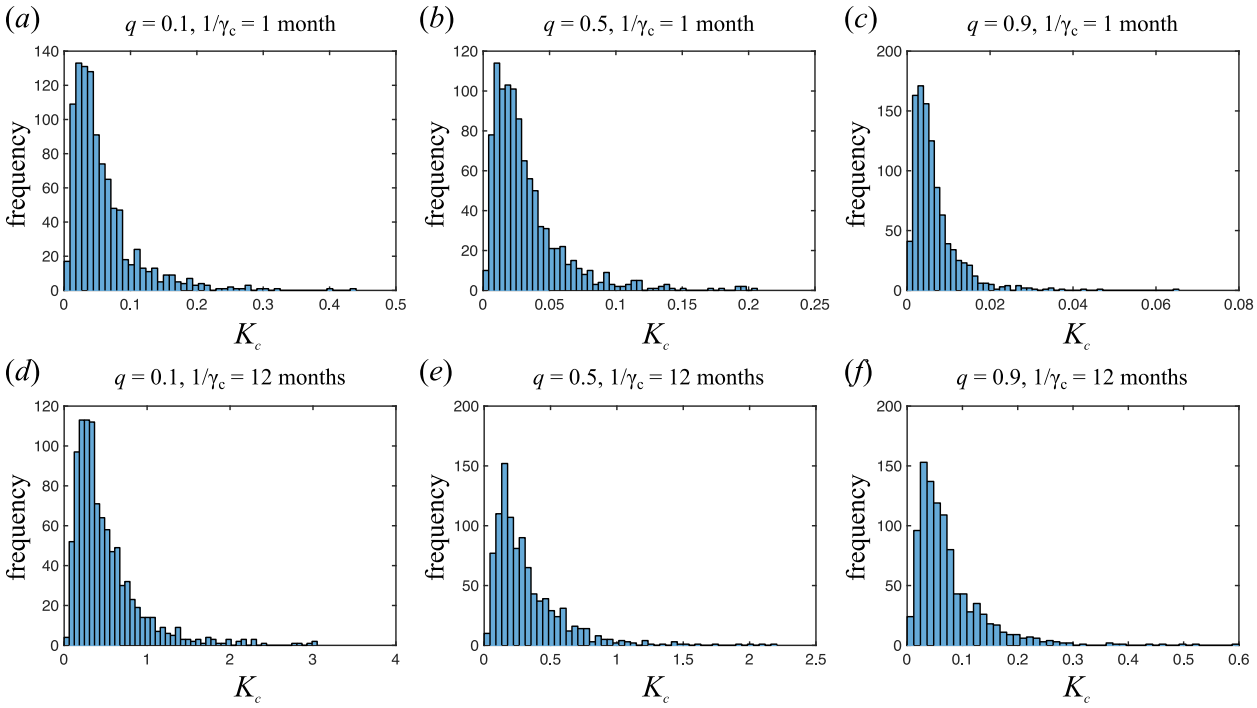

Figure E.2: **Frequency distribution of the relative substitution rate (at the population level) in the immunocompromised population versus the general population** K_c from Latin Hypercubic Sampling of the parameters space. (a)–(c) $1/\gamma_c = 1$ month; (d)–(f) $1/\gamma_c = 12$ months. (a),(d) $q = 0.1$; (b),(e) $q = 0.5$; (c),(f) $q = 0.9$. In all cases $1/\gamma = 7$ days, $N = 10^6$, $1/\theta_c = 3$ days, $k_c/k = 5$.

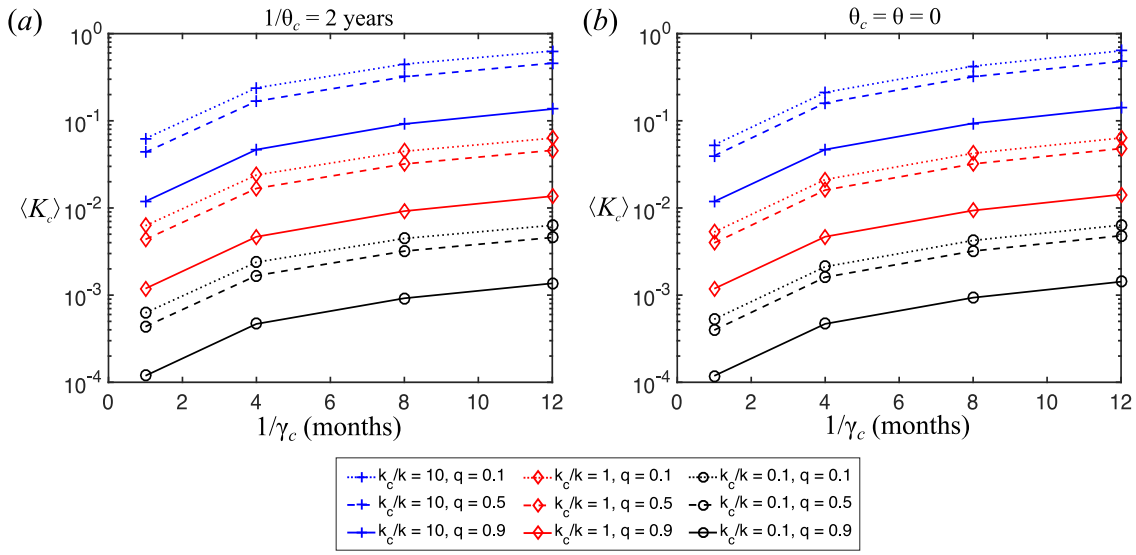

Figure E.3: **The mean relative substitution rate (at the population level) in the immunocompromised population versus the general population** $\langle K_c \rangle$ from Latin Hypercubic Sampling of 1000 points in the parameters space (R_0, k, c, $1/\theta$) as a function of the duration of infection in immunocompromised hosts $1/\gamma_c$, for a range of values of the quarantine parameter q, and the ratio of substitution rates (at the individual host level) k_c/k. (a) SIRS model with $1/\theta_c = 2$ years and $N = 10^6$; (b) SIR model ($\theta_c = \theta = 0$) with $N = 10^4$. Blue lines with square markers indicate $k_c/k = 10$, red lines with diamond markers $k_c/k = 1$, and black lines with circle markers $k_c/k = 0.1$. Dotted lines indicate $q = 0.1$, dashed lines $q = 0.5$ and solid lines $q = 0.9$. Here, $1/\gamma = 7$ days.

Figure E.3 reveals that the SIRS model and SIR model behave similarly when the duration of immunity for the immunocompromised population is large.

References

[1] Heffernan J, Smith R, Wahl L (2005) Perspectives on the basic reproductive ratio. *J R Soc Interface.* 2:281–293.