Modelling alcohol problems: Total recovery

Caroline Elizabeth Walters[a b *] Brian Straughan[a]

Jeremy Kendal[b]

[a] Department of Mathematics and the Centre for the Coevolution of Biology and Culture, University of Durham, DH1 3LE, U.K.

[b] Department of Anthropology and the Centre for the Coevolution of Biology and Culture, University of Durham, DH1 3LE, U.K.

[*] Corresponding author email: c.e.walters@durham.ac.uk

December 5, 2012

Abstract

Binge drinking in the UK is an increasing problem, resulting in negative health, social and economic effects. Mathematical modelling allows for future predictions to be made and may provide valuable information regarding how to approach solving the problem of binge drinking in the UK. We develop a 3-equation model for alcohol problems, specifically binge drinking, which allows for total recovery. Individuals are split into those that are susceptible to developing an alcohol problem, those with an alcohol problem and those in treatment.

We find that the model has two equilibrium points: one without alcohol problems and one where alcohol problems are endemic in the population. We compare our results with those of an existing model that does not allow for total recovery. We show that without total recovery, the threshold for alcohol problems to become endemic in the population is lowered. The endemic equilibrium solution is also affected, with an increased proportion of the population in the treatment class and a decreased proportion in the susceptible class. Including totally recovery does not determine whether the proportion of individuals with alcohol problems increases or decreases, however it does effect the size of the change. Parameter estimates are made from information regarding binge drinking where we find an increase in the recovery rate decreases the proportion of binge drinkers in the population.

Keywords: Epidemics; ODE model; stability; equilibria; binge drinking

1 Introduction

Smith and Foxcroft [39] report that between 1998 to 2006 there was an overall increase in the proportion of individuals in Great Britain who exceed the recommended alcohol consumption limits, including a doubling of the proportion of women who binge drink. Excessive alcohol consumption can lead to a range of negative health and social effects [16] and it is estimated that alcohol misuse costs the NHS £2.7 billion per year with alcohol related hospital admissions having increased by 100% from 2002/03 to 2009/10 [2]. These figures suggest that there is an increasing trend of alcohol misuse, which is

resulting in costs to health and the economy. Here we devise a predictive mathematical model which may offer an insight into the best strategy for tackling problems with alcohol, in particular binge drinking.

Mathematical models for behaviours such as alcoholism have been developed from epidemiological models for the spread of infectious diseases. One of the first infectious disease models by Kermack and McKendrick [21] considers a constant population where individuals are split into those that are susceptible to catching the disease (S), infected individuals (I) and immune or dead individuals (R). To maintain a constant population, immune individuals and those that have died from the disease enter the removed class, hence models of this form have become known as SIR models. Developments of SIR models and their extensions continue to be employed to describe various scenarios in mathematical epidemics, cf. Murray [32], Wang and Mulone [41], Wang and Ruan [42], Wang and Zhao [43], Boni and Feldman [5], Lou and Ruggeri [24], Buonomo and Lacitignola [6], Capone [9], Keeling and Rohani [20], Li et al. [23], Ma and Li [25], Buonomo and Rionero [7], Buonomo et al. [8], Mulone et al. [31], Rionero [34], Rionero and Vitiello [35].

Another development of such models has been to apply them to situations where it is assumed that social interaction is the key factor in spreading the behaviour. Behaviours which can result in adverse health effects have been represented, such as drinking [36, 3, 26, 28, 37, 30], smoking [38], drug use [44, 29], obesity [18, 15] and eating disorders [14]. Even though the models for each social problem may appear mathematically similar at the onset, there are fundamental differences which must be catered for. For example, a small intake of alcohol may be beneficial to health as shown by the J-shaped curve

of alcohol intake against health problems [27, 22]. For smoking however, the graph of amount smoked against health problems immediately has an increasing gradient, indicating the detrimental effect of smoking on health.

In this paper we develop a three-stage model which represents the effect of social influence on drinking habits, with a particular interest in total recovery. The total population is split into susceptible individuals, individuals with alcohol problems and individuals in treatment. Susceptible individuals are those who do not consume alcohol in a way defined to be problematic. We refer to alcohol problems in general as the model is applicable to a variety of drinking behaviours, for example dependent drinkers who drink every day or binge drinkers who consume many units in one session. The precise definitions of each class must be determined by the nature of the behaviour being modelled, which we demonstrate in section 2.5 using information regarding binge drinking. We consider the recovered class to represent those receiving treatment. Whilst it is possible for individuals to tackle an alcohol problem without professional help, data regarding the number of individuals opting for this approach is unavailable. If such information were to become available then a change to the definition of the class would perhaps be appropriate. The definition of treatment may also vary depending on the nature of the problem and any associated withdrawal effects.

The three subpopulations are similar to the classes defined in the work of Sanchez et al. [36] (also see Benedict [3]) and Mulone and Straughan [30], however neither of these models allow for total recovery. Sanchez et al. [36] found that the basic reproduction number alone is not always the key factor in controlling drinking in the population. Mulone and Straughan [30] extended

their model by splitting binge drinkers into those who admit that they have a problem and those that do not admit. Using data for the north east of England, they conclude that binge drinking is sustainable in the population.

Other models, by Manthey et al. [26], Mubayi et al. [28] and Santonja et al. [37], do not contain a treatment class but instead split the population into three classes depending on the amount of alcohol an individual consumes. Manthey et al. [26] consider a students' 5-year period in a university campus environment, which is deemed too short for recovery to be determined. Mubayi et al. [28] also focused on the drinking habits of students, however they were interested in assessing how a change from low to high risk drinking environments affected the transistion from susceptible to heavy drinker. Santonja et al. [37] do not consider a treatment class despite an individual spending 50 years in the system as the aim of the work is to determine the health and economic costs of risky alcohol consumption. This is determined by the average alcohol intake alone, irrespective of any recovery process.

We have chosen to include a treatment class as we aim to discover the most effective way to reduce the proportion of the population in the alcohol problems class. Such information may be useful to health professionals and policy makers when devising strategies aimed at reducing the proportion of the population suffering from alcohol problems. We also allow for individuals to completely recover from their alcohol problem. The motivation for this stems from recent ideas regarding the nature of recovery.

Best [4] discusses various definitions of recovery and introduces the concept of recovery champions. These champions are individuals who have successfully recovered from misusing alcohol, or other similar problems, and

appear as a role model or an example of success to inspire those currently in treatment. The UK Drug Policy Commission Recovery Consensus Group report [40] does not contain a precise definition of recovery as it is an individual process, i.e. recovery cannot be given a fixed definition which applies to the whole population as it varies depending on the individual. Instead, a set of key principles of recovery are presented. The report concludes that some individuals will always remain in treatment whereas others will feel that they are fully recovered. To accommodate both these options, we allow for individuals to move from the treatment class back to the susceptible population at a given rate.

The model we construct considers a population of N individuals separated into the three subclasses, represented by a system of three ordinary differential equations. Susceptible individuals, denoted by $S(t)$ where t is time, are those without an alcohol problem. We assume that a susceptible individual develops an alcohol problem through interactions with those in the alcohol problems class, $A(t)$. Finally an individual may be in the treatment class, $R(t)$, from which they may relapse and hence return to $A(t)$. Alternatively an individual may remain in treatment for a sufficient length of time so that they totally recover, at which point they return to the susceptible population as they are no longer experiencing difficulties with alcohol.

Using stability analysis we calculate a critical threshold value, R_0, which, once exceeded, determines that alcohol problems will persist in the population. Sensitivity analysis reveals which parameter has the greatest effect on this threshold value and thus may provide valuable insights into the most effective way of tackling alcohol misuse in the population. We then consider

the stability of the endemic equilibrium solution and compare our results with the case where total recovery is not possible. Finally we use numerical simulations to predict the future proportion of binge drinkers in England.

2 The mathematical model

As stated in the introduction, we consider a population of N individuals and split them into three classes, $S(t)$, $A(t)$ and $R(t)$. The probability that a susceptible individual has contact with someone in the alcohol problems class is A/N. Not all such contacts will be sufficient for the susceptible individual to develop an alcohol problem, so we define β to be the rate at which sufficient contacts occur. This gives us the rate at which individuals move from being susceptible to having an alcohol problem as $\beta AS/N$. This sort of transmission term has been employed in modelling drug and alcohol probelms, cf. Sanchez et al. [36], Benedict [3], Manthey et al. [26], Santonja et al. [37]

Individuals may move to the recovery class by entering a treatment programme, which we assume occurs at a constant rate φ. Once in treatment, an individual can either relapse or they can recover. Relapsing back to $A(t)$ is also assumed to happen at a constant rate, ρ, whereas recovery, and hence return to the susceptible class, is assumed to happen at a constant rate γ. We assume that individuals enter and leave the population at the same constant rate μ, where $1/\mu$ represents the average length of time spent in the system.

The dynamics of this SAR system are given by the equations

$$\begin{aligned}
\dot{S} &= \mu N - \frac{\beta AS}{N} + \gamma R - \mu S, \\
\dot{A} &= \frac{\beta AS}{N} + \rho R - (\varphi + \mu)A, \\
\dot{R} &= \varphi A - (\rho + \mu + \gamma)R,
\end{aligned} \tag{2.1}$$

where the total population is given by $N = S + A + R$ with $N > 0$, $S \geq 0$, $A \geq 0$ and $R \geq 0$.

Figure 1 here.

To preserve the direction of flow through the system (see figure 1), we take only positive values for the parameters $\beta, \mu, \varphi, \rho$ and γ. Following the method in Mulone and Straughan [30], we now introduce the variables $s(t) = S(t)/N$, $a(t) = A(t)/N$ and $r(t) = R(t)/N$, which enables us to rewrite system (2.1) as

$$\begin{aligned}
\dot{s} &= \mu - \beta as + \gamma r - \mu s, \\
\dot{a} &= \beta as + \rho r - (\varphi + \mu)a, \\
\dot{r} &= \varphi a - (\rho + \mu + \gamma)r,
\end{aligned} \tag{2.2}$$

where $1 = s + a + r$. As $s = 1 - a - r$, we can reduce system (2.2) to the two equations

$$\begin{aligned}
\dot{a} &= -\beta a^2 - \beta ar + (\beta - \varphi - \mu)a + \rho r, \\
\dot{r} &= \varphi a - (\rho + \mu + \gamma)r.
\end{aligned} \tag{2.3}$$

2.1 Stability analysis

We solve equations (2.3) to find the equilibrium points of the system, which are the problem-free solution $(a, r) = (0, 0)$ and the endemic solution $(a, r) = (\bar{a}, \bar{r})$. We will now analyse the local stability of the problem-free equilibrium solution by considering a linearisation of system (2.3) at $(a, r) = (0, 0)$. The linearisation of equations (2.3) around a general point $(\hat{a}, \hat{r})$ is given by $\dot{\mathbf{a}} = J(\hat{\mathbf{a}})(\mathbf{a} - \hat{\mathbf{a}})$ where J is the Jacobian matrix and $\mathbf{a}$ is the vector

$$\mathbf{a} = \begin{pmatrix} a \\ r \end{pmatrix}. \tag{2.4}$$

The Jacobian matrix at the point $(0, 0)$ is given by

$$J(0,0) = \begin{pmatrix} (\beta - \varphi - \mu) & \rho \\ \varphi & -(\rho + \mu + \gamma) \end{pmatrix},$$

which has eigenvalues

$$\sigma_+ = \frac{-x_1 + \sqrt{{x_1}^2 - 4y_1}}{2} \quad \text{and} \quad \sigma_- = \frac{-x_1 - \sqrt{{x_1}^2 - 4y_1}}{2} \tag{2.5}$$

where

$$x_1 = \varphi + \rho + 2\mu + \gamma - \beta,$$
$$y_1 = -\rho\varphi + (\rho + \mu + \gamma)(\varphi + \mu - \beta).$$

For the problem-free equilibrium point to be locally asymptotically stable we require the real part of both eigenvalues to be negative. This is true provided $x_1 > 0$ and $y_1 > 0$. It is sufficient to consider $y_1 > 0$ only as this condition guarantees $x_1 > 0$ (see appendix A) from which we determine that the inequality

$$\frac{\beta(\rho + \mu + \gamma)}{\mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi} < 1 \tag{2.6}$$

must hold for the equilibrium point to be locally asymptotically stable. If this situation arises then alcohol problems will eventually die out in the population. If inequality (2.6) is reversed then the equilibrium solution is unstable and alcohol problems may persist in the population. We now define the basic reproduction number R_0 to be

$$R_0 := \frac{\beta(\rho+\mu+\gamma)}{\mu(\rho+\mu+\gamma+\varphi)+\gamma\varphi}, \tag{2.7}$$

where $R_0 < 1$ indicates stability and $R_0 > 1$ indicates instability of the problem-free equilibrium solution.

2.2 Endemic equilibrium solution

The second equilibrium solution of system (2.3) is $(a, r) = (\bar{a}, \bar{r})$, where

$$\begin{cases} \bar{a} = \dfrac{\beta(\rho+\mu+\gamma)-\mu(\rho+\mu+\gamma+\varphi)-\gamma\varphi}{\beta(\rho+\mu+\gamma+\varphi)}, \\ \\ \bar{r} = \dfrac{\varphi}{(\rho+\mu+\gamma)} \cdot \dfrac{\beta(\rho+\mu+\gamma)-\mu(\rho+\mu+\gamma+\varphi)-\gamma\varphi}{\beta(\rho+\mu+\gamma+\varphi)}, \end{cases} \tag{2.8}$$

and only exists for $R_0 > 1$. The Jacobian of equations (2.3) at the point $(\bar{a}, \bar{r})$ is

$$J(\bar{a}, \bar{r}) = \begin{pmatrix} -2\beta\bar{a} - \beta\bar{r} + (\beta - \varphi - \mu) & \rho - \beta\bar{a} \\ \varphi & -(\rho+\mu+\gamma) \end{pmatrix} \tag{2.9}$$

and the corresponding eigenvalues are given by

$$\widetilde{\sigma}_+ = \frac{-x_2 + \sqrt{{x_2}^2 - 4y_2}}{2} \quad \text{and} \quad \widetilde{\sigma}_- = \frac{-x_2 - \sqrt{{x_2}^2 - 4y_2}}{2}, \tag{2.10}$$

where

$$x_2 = 2\beta\bar{a} + \beta\bar{r} + \varphi + 2\mu + \rho + \gamma - \beta$$

$$y_2 = \varphi(\beta\bar{a} - \rho) + (\rho+\mu+\gamma)(2\beta\bar{a} + \beta\bar{r} + \varphi + \mu - \beta).$$

For the equilibrium solution $(\bar{a}, \bar{r})$ to be linearly asymptotically stable then $x_2 > 0$ and $y_2 > 0$ must hold. Appendix B.1 gives the calculations which show that $x_2 > 0$ is always true provided $y_2 > 0$, so to find the local stability conditions we need only consider $y_2 > 0$. We can write y_2 in terms of the model parameters only by substituting in the values for $\bar{a}$ and $\bar{r}$, which is shown in appendix B.2. From this we find that the inequality $y_2 > 0$ can be written as

$$\beta(\rho + \mu + \gamma) - \mu(\rho + \mu + \gamma + \varphi) - \gamma\varphi > 0.$$

This can be rearranged to give

$$1 < \frac{\beta(\rho + \mu + \gamma)}{\mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi},$$

which is equivalent to $R_0 > 1$. From this we know that $(\bar{a}, \bar{r})$ is locally asymptotically stable when it exists, hence alcohol problems become endemic provided $R_0 > 1$.

Our analysis reveals that the equilibrium point $(0, 0)$ is locally asymptotically stable for $R_0 < 1$, whereas $(\bar{a}, \bar{r})$ is locally asymptotically stable for $R_0 > 1$. We can see that the value of R_0 determines whether alcohol problems will die out or become endemic in the population and so we consider $R_0 = 1$ to be an invasion threshold value.

2.3 Sensitivity analysis

Having defined $R_0 = 1$ as a threshold for the invasion of alcohol problems, we are now interested in which model parameter has the greatest effect on the R_0 value and hence has the greatest effect in determining whether alcohol problems will persist in the population. To this end, we calculate the

normalised sensitivity index, NSI $= (k/R_0)(\partial R_0/\partial k)$, which indicates how sensitive R_0 is to a change in parameter k, where normalisation allows for a direct comparison between parameters. A negative normalised sensitivity index indicates that an increase in the parameter value results in a decrease in the R_0 value. As we are only interested in the magnitude of the change to the R_0 value, we consider the absolute value. The normalised sensitivity indices for the parameters are

$$\begin{aligned}
\left|\frac{\beta}{R_0}\frac{\partial R_0}{\partial \beta}\right| &= 1, \\
\left|\frac{\mu}{R_0}\frac{\partial R_0}{\partial \mu}\right| &= \frac{\mu(\mu+\gamma)(\rho+\mu+\gamma)+\rho\mu(\rho+\mu+\gamma+\varphi)}{\mu(\mu+\gamma)(\rho+\mu+\gamma)+\rho\mu(\rho+\mu+\gamma+\varphi)+G} &&< 1, \\
\left|\frac{\rho}{R_0}\frac{\partial R_0}{\partial \rho}\right| &= \frac{\rho\varphi(\mu+\gamma)}{\rho\varphi(\mu+\gamma)+\varphi(\mu+\gamma)^2+\mu(\rho+\mu+\gamma)^2} &&< 1, \\
\left|\frac{\varphi}{R_0}\frac{\partial R_0}{\partial \varphi}\right| &= \frac{\varphi(\mu+\gamma)}{\varphi(\mu+\gamma)+\mu(\rho+\mu+\gamma)} &&< 1, \\
\left|\frac{\gamma}{R_0}\frac{\partial R_0}{\partial \gamma}\right| &= \frac{\rho\gamma\varphi}{\rho\gamma\varphi+\gamma\varphi(\mu+\gamma)+\mu(\rho+\mu+\gamma)(\rho+\mu+\gamma+\varphi)} &&< 1,
\end{aligned}$$

where $G = \gamma(\mu+\varphi)(\rho+\mu+\gamma)+\mu\varphi(\rho+\gamma)$.

From the calculations here we can see that R_0 is most sensitive to changes in the value of β, which represents the rate at which social interaction mediates the development of alcohol problems. We can see that equation (2.7) for R_0 has the form $R_0 = c\beta$ where

$$c = \frac{(\rho+\mu+\gamma)}{\mu(\rho+\mu+\gamma+\varphi)+\gamma\varphi}$$

so a factor α change in the β value results in a factor α change in the value of R_0.

2.4 The effect of individuals in treatment returning to the susceptible population

We now compare the model with the situation where movement from $R(t)$ to $S(t)$ is removed, so complete recovery from an alcohol problem is not possible. This is achieved by allowing $\gamma = 0$, resulting in the model proposed by Mulone and Straughan [30]. We focus on how $\gamma = 0$ affects the basic reproduction number and the endemic equilibrium solution. With this comparison we aim to highlight any qualitative differences between the solutions of the two models.

2.4.1 Basic reproduction number

The basic reproduction number is given by equation (2.7). For the case where $\gamma = 0$ we define the basic reproduction number by $\widetilde{R}_0$ where

$$\widetilde{R}_0 = \frac{\beta(\rho + \mu)}{\mu(\rho + \mu + \varphi)}.$$

To study the effect that $\gamma > 0$ has on the basic reproduction, the difference between R_0 and $\widetilde{R}_0$ is calculated,

$$R_0 - \widetilde{R}_0 = \frac{-\beta\rho\varphi\gamma}{\mu\left(\mu + \rho + \varphi\right)\left(\mu^2 + \mu\rho + \mu\gamma + \mu\varphi + \varphi\gamma\right)}. \tag{2.11}$$

As the right-hand side of equation (2.11) is always negative, we conclude that $R_0 < \widetilde{R}_0$ for all possible parameter values. This means that excluding the return to the susceptible class increases the average number of secondary infections which result from a single infected individual entering a wholly susceptible population.

2.4.2 Endemic equilibrium solution

Next we look at the change to the endemic equilibrium solution, $(\bar{s}, \bar{a}, \bar{r})$. The equilibrium value for the susceptible population, $\bar{s}$, is calculated using the expressions for $\bar{a}$ and $\bar{r}$ from system of equations (2.8) in $\bar{s} = 1 - \bar{a} - \bar{r}$, resulting in

$$\bar{s} = \frac{\mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi}{\beta(\rho + \mu + \gamma)}. \tag{2.12}$$

The special case $\gamma = 0$ has the endemic equilibrium solution $(\widetilde{s}, \widetilde{a}, \widetilde{r})$ with

$$\begin{aligned}
\widetilde{s} &= \frac{\mu(\rho + \mu + \varphi)}{\beta(\rho + \mu)}, \\
\widetilde{a} &= \frac{\beta(\rho + \mu) - \mu(\rho + \mu + \varphi)}{\beta(\rho + \mu + \varphi)}, \\
\widetilde{r} &= \frac{\varphi}{(\rho + \mu)} \cdot \frac{\beta(\rho + \mu) - \mu(\rho + \mu + \varphi)}{\beta(\rho + \mu + \varphi)}.
\end{aligned}$$

To study the effect that γ has on each individual class we compare the solution $(\bar{s}, \bar{a}, \bar{r})$ with the $\gamma = 0$ solution $(\widetilde{s}, \widetilde{a}, \widetilde{r})$.

We begin by examining the susceptible population. The equation

$$\bar{s} - \widetilde{s} = \frac{\rho\gamma\varphi}{\beta(\rho + \mu)(\rho + \mu + \gamma)}$$

is always positive, thus $\bar{s} > \widetilde{s}$. Hence, removing the possibility of total recovery reduces the proportion of individuals in the susceptible class.

For the alcohol problems class, the change in the proportion of individuals in the class is not as simple as the previous case. The difference between the two equilibrium values is given by

$$\bar{a} - \widetilde{a} = \frac{\gamma\varphi(\beta - \rho - \mu - \varphi)}{\beta(\rho + \mu + \gamma + \varphi)(\rho + \mu + \varphi)}.$$

The relationship between $\bar{a}$ and $\widetilde{a}$ is determined by the sign of the expression $\beta - \rho - \mu - \varphi$. If $\beta > \rho + \mu + \varphi$ then $\bar{a} > \widetilde{a}$, so preventing a return to

the susceptible class results in a decrease in the proportion of the population with alcohol problems. If $\beta < \rho+\mu+\varphi$ the converse is true so $\bar{a} < \widetilde{a}$. Finally, $\bar{a} = \widetilde{a}$ only when $\beta = \rho + \mu + \varphi$. This expression is independent of γ, so it is possible for the two models to agree on the proportion of alcoholics in the population.

For the recovered class, we find that

$$\bar{r} - \widetilde{r} = \frac{\gamma\varphi F}{\beta(\rho+\mu)(\rho+\mu+\gamma)(\rho+\mu+\varphi)(\rho+\mu+\gamma+\varphi)} \tag{2.13}$$

where

$$F = -\beta(\rho+\mu)(\rho+\mu+\gamma)+\mu(\rho+\mu)(\rho+\mu+\gamma)+\mu\varphi(\mu+\gamma)-\rho\varphi(\rho+\varphi). \tag{2.14}$$

As the denominator of the right-hand side of equation (2.13) is always positive, the sign of $\bar{r}-\widetilde{r}$ depends on the numerator, specifically on the value of F. As we are considering the endemic equilibrium solution, the parameters are constrained by the inequality $R_0 > 1$. We use this information to determine that $F < 0$ must be true and hence $\bar{r} < \widetilde{r}$ (see appendix D). Thus removing the option of returning to the susceptible class results in an increase in the proportion of individuals in the recovered class.

2.4.3 Conclusions

Assuming that β, μ, ρ and φ are fixed, when $\gamma = 0$ the basic reproduction number is increased, i.e. the average number of secondary infections resulting from a single infected being introduced into a wholly susceptible population is increased. For example, we can consider a situation where $\widetilde{R}_0 = 1$, thus $R_0 < 1$, and consider increasing the value of β. This increase instantly results in alcohol problems becoming endemic when $\gamma = 0$. When $\gamma > 0$ however, we

find that the rate at which susceptilbe individuals develop alcohol problems may be increased without resulting in alcohol problems becoming endemic.

Alternatively we could consider R_0 as a strictly decreasing function of γ. In some situations, determined by the other parameter values, an increase in γ changes the stable equilibrium from the endemic to the alcohol problems-free solution. Increased γ values indicate that more individuals recover from an alcohol problem, perhaps achievable by improvements to treatment services which discourage individuals from relapsing. An example situation is shown in figure 2. When $\gamma = 0.550$, $R_0 = 1$ and the stable equilibrium moves from the endemic solution to the alcohol problems-free solution. This example highlights the importance in understanding the key variables that affect drinking behaviour. By comparing the model without total recovery (equating to considering $\gamma = 0$) with the model where $\gamma > 0.550$, we see that the value of γ affects the qualitative nature of the system and hence offers different predictions.

Figure 2 here.

When $\gamma = 0$, the endemic equilibrium solution has a decreased proportion of susceptible individuals and an increased proportion of those in treatment. This result is intuitive as $\gamma = 0$ prevents individuals moving from the treatment class back to the susceptible population. The relationship between γ and the alcohol problems class is not so obvious as whether there is an increase or a decrease does not depend on the value of γ, however the magnitude of the effect does. Recall that if $\beta < \rho + \mu + \varphi$ then $\gamma = 0$ increases the proportion of individuals in the alcohol class but if $\beta > \rho + \mu + \varphi$ then

the proportion is decreased.

Assuming ρ, μ and φ are fixed, then the susceptible population is increased when $\gamma > 0$. As γ relates to totally recovered individuals, $\gamma > 0$ also results in a reduction in the treatment class size and hence fewer individuals available to relapse. If β is small then we have an increased number of individuals in the susceptible class, each with only a small chance of developing alcohol problems. The only way to repopulate $A(t)$ is by individuals who relapse or by susceptible individuals developing an alcohol problem. By combining small β with $\gamma > 0$, the number of individuals available to relapse is small and the chance that susceptible individuals will develop alcohol problems is low. From this we conclude that $\gamma > 0$ decreases the proportion of individuals with alcohol problems when β is small. Conversely, if β is large and $\gamma > 0$ then we have an increased number of susceptible individuals, each with a large chance of developing alcohol problems, so we intuitively expect an increase in the proportion of those with alcohol problems in the population.

2.5 A model for binge drinking in England

We now estimate the parameter values based on recent information regarding binge drinking in England. Social influence, for example through social norms and peer pressure, is often considered to play a key role in binge drinking [13, 12]. As our model represents the effect of social influence on drinking behaviour, it is appropriate to apply it to the situation of binge drinking. We shall consider those who binge drink to form the alcohol problems class.

The term binge drinking has not been strictly defined, however according

to Deacon et al. [10] a binge drinker is usually regarded as someone who regularly consumes at least twice the guideline daily units of alcohol during the heaviest drinking day of the week. The UK guidelines state that binge drinking is consuming 8+ units for men and 6+ units for women in a single session. It is possible that an individual may occasionally binge drink in accordance with these guidelines, however this behaviour may be very rare and hence not indicative of a drinking problem. The data available in Deacon et al. [10] considers one week only so may include information on infrequent binge drinkers, however it can be used as an upper bound when wishing to determine the proportion of regular binge drinkers in the population.

According to Jones et al. [19], the government aims to reduce the harm caused by 18-24 year old binge drinkers so we shall restrict our population to this age group. In Britain binge drinking is most prevalent among young adults, however it is not restricted to this age group with those that binge drink in their early 20s being more likely to do so in their 40s than those that do not binge drink [17]. With this in mind we argue that tackling the current problem of binge drinking will not only reduce antisocial behaviour and alcohol related accidents now, but may also contribute to reducing the number of individuals with alcohol-related illnesses and alcohol dependence in the future. The 18-24 year old age group spend a total of 7 years so we take $\mu = 1/7 = 0.143$. Information for the number of binge drinkers in treatment could not be obtained so we shall assume that it is the same proportion as for dependent drinkers, which is 6% of the drinking population according to the 2011 National Institute for Health & Clinical Excellence (NICE) report [33]. We therefore consider the maximum annual probability of entering treatment

to be 0.06, which we can convert to a rate using

$$\text{rate} = \frac{-\ln(1-p)}{t} \tag{2.15}$$

where p is the probability of an event over the time period t. Using this equation we calculate a maximum value of $\varphi = 0.0619$.

El Sheikh and Bashir [11] report that 35% of alcoholics in treatment relapse within the first 2 weeks and 58% within the first 3 months. After 4 years, 90% are expected to have relapsed [1]. We use the data for the greatest time interval, which is 4 years, as we believe this will give the most accurate information and use equation (2.15) to estimate $\rho = 0.576$. Best [4] indicates that an individual experiencing a 4/5 year period without consuming alcohol can be considered as recovered. If 90% of individuals relapse after 4 years in treatment then we know that 10% remain in treatment so we can use this in equation (2.15) to calculate the rate of recovery. We obtain a value of 0.0263 which we then divide across a 4-year time period to give a maximum value of $\gamma = 0.00659$. An estimate for β is difficult to determine so we will consider the minimum β value which ensures alcohol problems become endemic in the population, defined as $\beta_{\min}$. We find that this minimum value is $\beta_{\min} = 0.156$, calculated using the parameter values stated above and the equation $R_0 = 1$.

Deacon et al. [10] give the 2005 percentages for adults that binge drink as 19.3% for males and 8.1% for females. Assuming an even sex-ratio, this averages to 13.7% of the adult population so we take an initial value of $a(0) = 0.137$. As we assume that 6% of binge drinkers are in treatment we take $r(0) = 0.00874$.

According to Smith and Foxcroft [39], there has been an increase in the number of people drinking over the guideline weekly amounts from 1988-

2006. We find that the value $\beta_{\min} = 0.156$ results in a decrease in the binge drinking population from our $a(0)$ value so this is not an appropriate lower bound. Instead the lowest value we consider in $\beta = 0.2$ as this results in a continuation of the trend. When a susceptible individual meets a binge drinker, the likelihood that they also become a binge drinker is proportional to the value of β. As β increases a susceptible individual is more likely to become a binge drinker so as we increase the value of β we expect an increase in the proportion of binge drinkers in the population.

Figure 3 here.

We now take starting values $(a(0), r(0)) = (0.137, 0.00874)$ and parameter values $\mu = 0.143$, $\varphi = 0.0619$, $\rho = 0.576$, $\gamma = 0.00659$ and let β take the values 0.2, 0.4, 0.6 and 0.8. Figure 3 shows how the fractions in each of the classes change over time for the different β values. The graphs plateau at the equilibrium solution values. It can be seen from figures 3a and 3b that the greatest increase in the proportion of binge drinkers in the population occurs when β changes from a value of 0.2 to 0.4. This 0.2 increase in β results in a change from 20% of the population binge drinking to 56%. Subsequent increases in β do not have such a great effect on the proportion of binge drinkers in the population. This highlights that for large β values any inaccuracy in the estimate for β will not greatly affect the results. If β is small however, then any inaccuracies could greatly alter the predicted outcome. Figure 4 shows this relationship.

Figure 4 here.

Figure 5 shows the phase portrait in the a, r-plane of the endemic equilibrium solution for the model where $\gamma > 0$ and for the case where $\gamma = 0$. The parameter values have been taken as above, along with $\beta = 0.4$. We know that $\mu+\rho+\varphi = 0.7809$ is greater than the value $\beta = 0.4$, so from our analysis is section 2.4 we expect taking $\gamma = 0$ to increase both the equilibrium value for $a(t)$ and $r(t)$. Figure 5 shows that this is indeed the case.

Figure 5 here.

3 Discussion

We have constructed a model for alcohol problems in a population which allows for individuals to totally recover and return to the susceptible population. The threshold $R_0 = 1$ was found, where $R_0 < 1$ indicates that alcohol problems will die out and $R_0 > 1$ determines that alcohol problems become endemic in the population. We found that the R_0 value was most sensitive to changes in the parameter β, which affects the rate at which susceptible individuals develop an alcohol problem. Decreasing β results in a decrease in the value for R_0. This indicates that efforts to reduce alcohol problems in the population should focus on preventing susceptible individuals from developing an alcohol problem.

We compared this model with the special case $\gamma = 0$, presented in Mulone and Straughan [30], which prevents individuals from returning to the susceptible class. We found that taking $\gamma = 0$ increased the value of the basic reproduction number and lead to an increase in the proportion of recov-

ered individuals and a decrease in the proportion of susceptible individuals. Whether the proportion of individuals with alcohol problems increases or decreases is not determined by including totally recovery, however the size of the change is affected by the value of γ. Thus our analysis reveals that the effect of γ on $\bar{a}$ is not straightforward. If the situation were such that β were large then totally recovery would have an adverse effect on reducing alcohol misuse in the population.

Estimates were made for the parameters using data for binge drinking in England. We were particularly interested in the effect of social influence on binge drinking and so we considered β values in the interval $[0.2, 0.8]$. Simulations using these values revealed that any inaccuracies in the β value could have a great effect on the proportion of binge drinkers in the population if β was small. For larger β values any inaccuracies did not have such a great effect.

One assumption of the model is that fully recovered individuals who have returned to the susceptible class have the same likelihood of developing an alcohol problem for the second time as someone who has had no prior problem with alcohol. A development for the future would be to include a fourth class of fully recovered individuals, as in the four equation smoking model by Sharomi and Gumel [38], rather than assuming that they return to the susceptible class. This would distinguish fully recovered individuals from susceptible individuals without prior alcohol problems. Alternatively the champion effect discussed by Best [4] may be better represented by assuming that those in treatment recover because of interactions with a recovery champion in the susceptible class, which would introduce more terms into

the equations.

The mathematics presented here offers a valuable insight into understanding patterns in drinking behaviour which are affected by social influence. Future work will continue to explore mathematical applications to the study of human behaviours.

Acknowledgements

This work was supported in part by a grant from the Leverhulme Trust, "Tipping points: mathematics, metaphors and meanings", and in part by a DTA award from EPSRC, a RCUK academic research fellowship to JK and a Durham University Social Sciences and Health Interdisciplinary Scholarship to CW.

We would like to thank Ms Jill Smith and Dr David Chappel (NEPHO) for helpful conversations.

References

[1] Alcohol relapse and craving: 90 percent have at least one relapse after treatment. http://alcoholism.about.com/cs/alerts/l/blnaa06.htm.

[2] Alcohol Concern. Making alcohol a health priority: Opportunities to reduce alcohol harms and rising costs. http://www.alcoholconcern.org.uk/publications/policy-reports/making-alcohol-a-health-priority, 2011.

[3] B. Benedict. Modeling alcoholism as a contagious disease: How "infected" drinking buddies spread problem drinking. *SIAM News*, 40(3), 2007.

[4] D. Best. Scottish drugs recovery consortium: Digesting the evidence. http://www.sdrconsortium.org/assets/files/DigestingTheEvidenceResearch.pdf, 2010.

[5] M. F. Boni and M. W. Feldman. Evolution of antibiotic resistance by human and bacterial niche construction. *Evolution*, 59:477–491, 2005.

[6] B. Buonomo and D. Lacitignola. On the dynamics of an SEIR epidemic model with a convex incidence rate. *Ricerche di Matematica*, 57:261–281, 2008.

[7] B. Buonomo and S. Rionero. On the lyapunov stability for SIRS epidemic models with general nonlinear incidence rate. *Applied Mathematics and Computation*, 217:4010–4016, 2010.

[8] B. Buonomo, D. Lacitignola, and S. Rionero. Effect of prey growth and predator cannibalism rate on the stability of a structured population model. *Nonlinear Analysis: Real World Applications*, 11:1170–1181, 2010.

[9] F. Capone. On the dynamics of predator-prey models with the beddington–de angelis functional response, under robin boundary conditions. *Ricerche di Matematica*, 57:137–157, 2008.

[10] L. Deacon, S. Hughes, K. Tocque, and M. A. Bellis. Indications of public health in the English regions. 8: Alcohol, 2007.

[11] S. E. G. El Sheikh and T. Z. Bashir. High-risk relapse situations and self-efficacy: Comparison between alcoholics and heroin addicts. *Addictive Behaviors*, 29(4):753–758, 2004.

[12] D. P. French and R. Cooke. Using the theory of planned behaviour to understand binge drinking: The importance of beliefs for developing interventions. *British Journal of Health Psychology*, 17:1–17, 2012.

[13] J. S. Gill. Reported levels of alcohol consumption and binge drinking within the UK undergraduate student population over the last 25 years. *Alcohol and Alcoholism*, 37:109–120, 2002.

[14] B. Gonzalez, E. Huerta-Sanchez, A. Ortiz-Nieves, T. Vazquez-Alvarez, and C. Kribs-Zaleta. Am I too fat? Bulimia as an epidemic. *Journal of Mathematical Psychology*, 47(5-6):515–526, 2003.

[15] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis. Infectious

disease modeling of social contagion in networks. *PLoS Computational Biology*, 6, 11 2010. doi: 10.1371/journal.pcbi.1000968.

[16] House of Commons Science and Technology Committee. Alcohol guidelines: Eleventh report of session 2010–12, 2012.

[17] Institute of Alcohol Studies. IAS factsheet: Binge drinking – Nature, prevalence and causes. `http://www.ias.org.uk/resources/factsheets/binge_drinking.pdf`, 2010.

[18] L. Jodar, F. J. Santonja, and G. Gonzalez-Parra. Modeling dynamics of infant obesity in the region of Valencia, Spain. *Computers & Mathematics with Applications*, 56(3):679–689, 2008.

[19] L. Jones, M. A. Bellis, D. Dedman, H. Sumnall, and K. Tocque. Alcohol-attributable fractions for England: Alcohol-attributable mortality and hospital admissions. `http://www.nwph.net/nwpho/publications/alcoholattributablefractions.pdf`, 2008.

[20] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in humans and Animals*. Princeton University Press, New Jersey, USA, 2008.

[21] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115: 700–721, 1927.

[22] R. A. Kloner and S. H. Rezkalla. To drink or not to drink? that is the question. *Circulation*, 116:1306–1317, 2007.

[23] J. Li, Z. Ma, and F. Zhang. Stability analysis for an epidemic model

with stage structure. *Nonlinear Analysis: Real World Applications*, 9: 1672 – 1679, 2008.

[24] J. Lou and T. Ruggeri. A time delay model about aids-related cancer: equilibria, cycles and chaotic behavior. *Ricerche di Matematica*, 56: 195–208, 2007.

[25] Z. Ma and J. Li. *Dynamical modeling and analysis of epidemics*. World Scientific, Singapore, 2009.

[26] J. L. Manthey, A. Aidoob, and K.Y.Ward. Campus drinking: An epidemiological model. *Journal of Biological Dynamics*, 2:346–356, 2008.

[27] M. Marmot and E. Brunner. Alcohol and cardiovascular disease: the status of the u shaped curve. *BMJ*, 303:565–568, 1991.

[28] A. Mubayi, P. E. Greenwood, C. Castillo-Chavez, P. Gruenewald, and D. M. Gorman. Impact of relative residence times in highly distinct environments on the distribution of heavy drinkers. *Socio-Economic Planning Sciences*, 44:45–56, 2010.

[29] G. Mulone and B. Straughan. A note on heroin epidemics. *Mathematical Biosciences*, 218:138–141, 2009.

[30] G. Mulone and B. Straughan. Modelling binge drinking. *International Journal of Biomathematics*, 2011. doi: 10.1142/S1793524511001453.

[31] G. Mulone, S. Rionero, and W. Wang. The effect of density-dependent dispersal on the stability of populations. *Nonlinear Analysis*, 74: 4831–4846, 2011.

[32] J. D. Murray. *Mathmatical Biology.* Springer-Verlag, Berlin, Germany, 1993.

[33] National Institute for Health & Clinical Excellence (NICE) report. Alcohol use disorders: Diagnosis, assessment and management of harmful drinking and alcohol dependence, CG115. `http://www.nice.org.uk/guidance/CG115`, 2011.

[34] S. Rionero. On the nonlinear stability of nonautonomous binary systems. *Nonlinear Analysis*, 75:2338–2348, 2012.

[35] S. Rionero and M. Vitiello. Long-time behavior of the solutions of murray–thomas model for interacting chemicals. *Mathematics and Computers in Simulation*, 82:1597–1614, 2012.

[36] F. Sanchez, X. Wang, C. Castillo-Chavez, D. M. Gorman, and P. J. Gruenewald. Drinking as an epidemic - A simple mathemathical model with recovery and relapse. In K. Witkiewitz and G. A. Marlatt, editors, *Therapist's guide to evidence-based relapse prevention.* Academic Press, New York, 2007.

[37] F.-J. Santonja, E. Sanchez, M. Rubio, and J.-L. Morera. Alcohol consumption in Spain and its economic cost: A mathematical modeling approach. *Mathematical and Computer Modelling*, 52:999–1003, 2010.

[38] O. Sharomi and A. B. Gumel. Curtailing smoking dynamics: A mathematical modeling approach. *Applied Mathematics and Computation*, 195:475–499, 2008.

[39] L. Smith and D. Foxcroft. Drinking in the UK: An exploration of trends. http://www.jrf.org.uk/sites/files/jrf/UK-alcohol-trends-FULL.pdf, 2009.

[40] The UK Drug Policy Commission Recovery Consensus Group report. A vision of recovery. http://www.ukdpc.org.uk/Recovery_Consensus_Statement.shtml, July 2008.

[41] W. Wang and G. Mulone. Threshold of disease transmission in a patch environment. *Journal of Mathematical Analysis and Applications*, 285: 321–335, 2003.

[42] W. Wang and S. Ruan. Simulating the SARS outbreak in Beijing with limited data. *Journal of Theoretical Biology*, 227:369–379, 2004.

[43] W. Wang and X.-Q. Zhao. An epidemic model in a patchy environment. *Mathematical Biosciences*, 190:97–112, 2004.

[44] E. White and C. Comiskey. Heroin epidemics, treatment and ODE modelling. *Mathematical Biosciences*, 208:312–324, 2007.

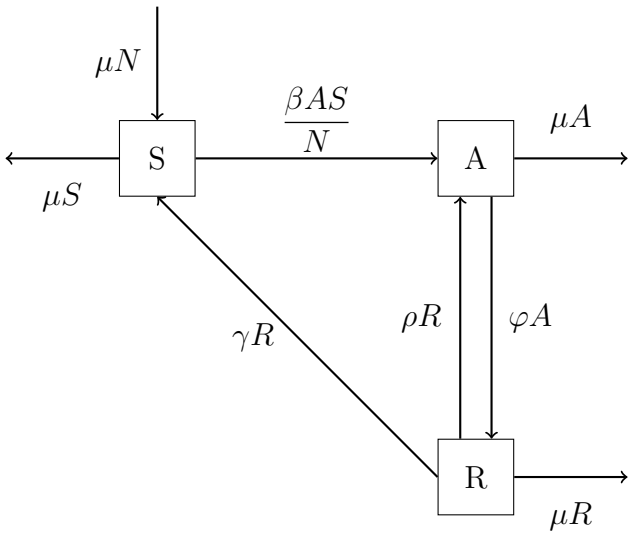

Figure 1: Flow diagram showing the movement between the three subpopulations $S(t)$, $A(t)$ and $R(t)$.

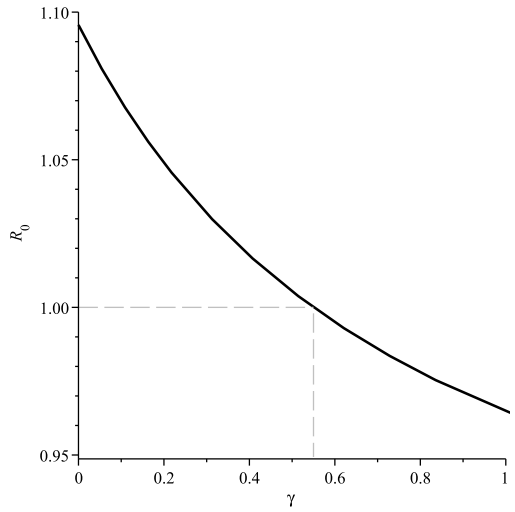

Figure 2: A change in the stability of the two equilibrium solutions occurs when $R_0 = 1$ at $\gamma = 0.550$. The fixed parameter values are $\mu = 0.25$, $\beta = 0.3$, $\rho = 0.8$ and $\phi = 0.1$.

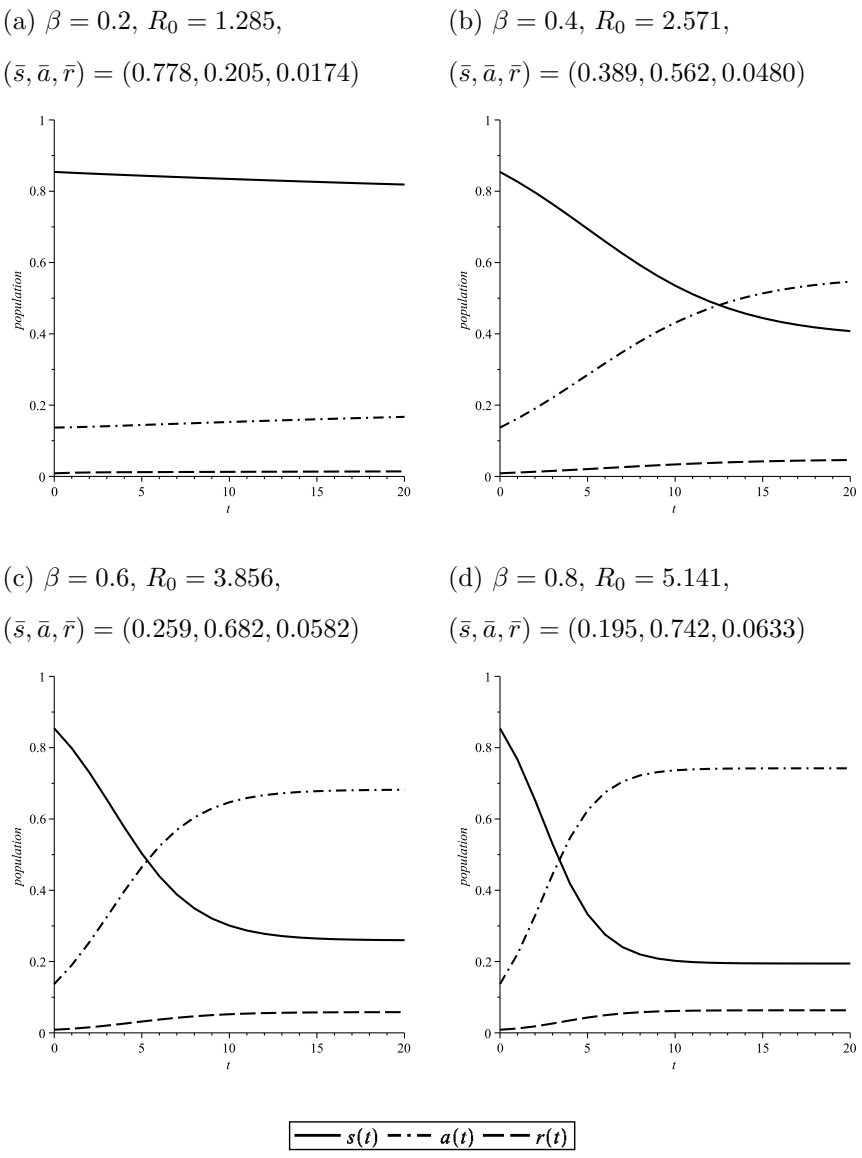

Figure 3: Simulations showing how the value of the parameter β affects the endemic equilibrium solution. The other parameters have values $\mu = 0.143$, $\varphi = 0.0619$, $\rho = 0.576$ and $\gamma = 0.00659$.

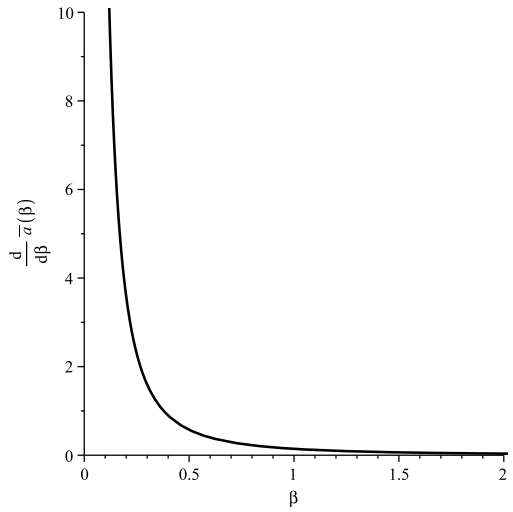

Figure 4: Graph showing the rate of change of $\bar{a}$ with respect to β, where we consider $\bar{a}$ to be a function of β only. As β increases, the rate of change in $\bar{a}$ tends towards 0.

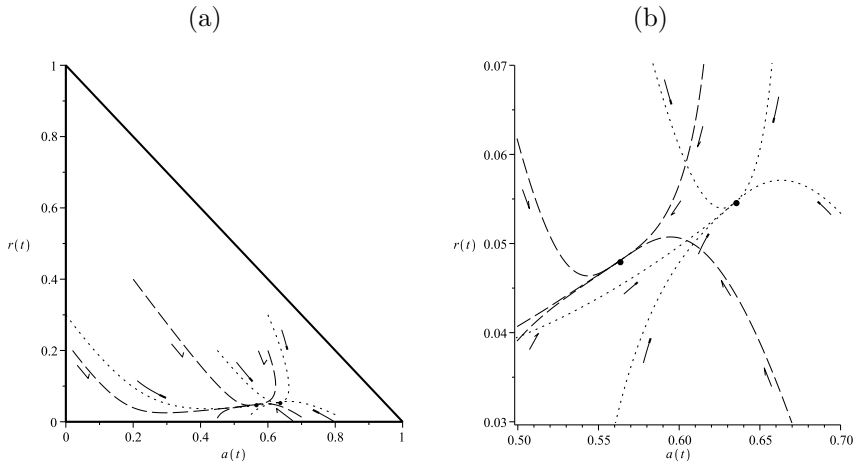

Figure 5: Phase portraits showing the endemic equilibrium point for $\gamma = 0.00659$ and for $\gamma = 0$, represented by the dashed and dotted lines, respectively. The other parameters take the values $\mu = 0.143$, $\beta = 0.4$, $\varphi = 0.0619$ and $\rho = 0.576$. The triangle with vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$ in figure 5a is the boundary of the positive invariant region $D = \{(a, r) \in \mathbb{R}^2 : a \geq 0, r \geq 0, a + r \leq 1\}$, where all solutions lie. A proof of the positive invariance of D is included in appendix C. Figure 5b shows the behaviour of the system close to the equilibrium solutions.

A Proof of $y_1 > 0$ implies $x_1 > 0$

We need to show that $x_1 > 0$ always holds provided $y_1 > 0$, where

$$\begin{aligned} x_1 &= \varphi + \rho + 2\mu + \gamma - \beta, \\ y_1 &= -\rho\varphi + (\rho + \mu + \gamma)(\varphi + \mu - \beta). \end{aligned}$$

We first consider the inequality $y_1 > 0$ which can be written in terms of the parameters as

$$\begin{aligned} & 0 < -\rho\varphi + (\rho + \mu + \gamma)(\mu + \varphi - \beta) \\ \Leftrightarrow \quad & 0 < -\beta(\rho + \mu + \gamma) + \mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi \\ \Leftrightarrow \quad & \beta(\rho + \mu + \gamma) < \mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi. \end{aligned} \tag{A.1}$$

We now consider the necessary condition for $x_1 > 0$ by rewriting this inequality in terms of the parameters,

$$\begin{aligned} & 0 < \varphi + \rho + 2\mu + \gamma - \beta \\ \Leftrightarrow \quad & \beta < \varphi + \rho + 2\mu + \gamma. \end{aligned} \tag{A.2}$$

We now multiply inequality (A.2) by $(\rho + \mu + \gamma)$ so that it may be directly compared with (A.1), which results in

$$\begin{aligned} \beta(\rho + \mu + \gamma) < \quad & (\varphi + \rho + 2\mu + \gamma)(\rho + \mu + \gamma) \\ \Leftrightarrow \quad \beta(\rho + \mu + \gamma) < \quad & \mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi + \mu^2 + \rho\varphi \\ & + (\rho + \gamma)(2\mu + \rho + \gamma). \end{aligned} \tag{A.3}$$

By comparison, we see that inequality (A.1) imposes a stronger comdition on $\beta(\rho + \mu + \gamma)$ than inequality (A.3). From this we conclude that $y_1 > 0$ implies $x_1 > 0$.

B Endemic equilibrium solution calculations

B.1 Proof of $y_2 > 0$ implies $x_2 > 0$

We need to show that $x_2 > 0$ always holds provided $y_2 > 0$, where

$$x_2 = 2\beta\bar{a} + \beta\bar{r} + \varphi + 2\mu + \rho + \gamma - \beta,$$
$$y_2 = \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)(2\beta\bar{a} + \beta\bar{r} + \varphi + \mu - \beta).$$

We can write y_2 in terms of x_2 as

$$\begin{aligned} y_2 &= \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)([2\beta\bar{a} + \beta\bar{r} + \varphi + 2\mu + \rho + \gamma - \beta] - [\rho + \mu + \gamma]) \\ &= \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)(x_2 - [\rho + \mu + \gamma]), \end{aligned} \tag{B.1}$$

and, from (2.8), the equation for $\bar{a}$ in terms of y_2 as

$$\bar{a} = \frac{y_2}{\beta(\rho + \mu + \gamma + \varphi)}. \tag{B.2}$$

Substituting equation (B.2) into equation (B.1) gives

$$y_2 = \frac{\varphi y_2}{\rho + \mu + \gamma + \varphi} - \rho\varphi + (\rho + \mu + \gamma)(x_2 - [\rho + \mu + \gamma]),$$

from which we find the equation for x_2,

$$x_2 = \frac{y_2}{\rho + \mu + \gamma + \varphi} + \frac{\rho\varphi}{\rho + \mu + \gamma} + \rho + \mu + \gamma. \tag{B.3}$$

From equation (B.3) we see that $x_2 > 0$ is always true if $y_2 > 0$.

B.2 Simplification of y_2

We have

$$y_2 = \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)(2\beta\bar{a} + \beta\bar{r} + \varphi + \mu - \beta)$$

and we want to write this equation in terms of the model parameters only. We begin by substituting for $\bar{r}$ using

$$\bar{r} = \frac{\varphi}{\rho + \mu + \gamma}\bar{a},$$

which follows from equations (2.8). This gives

$$\begin{aligned} y_2 &= \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)(2\beta\bar{a} + \frac{\beta\varphi}{\rho + \mu + \gamma}\bar{a} + \varphi + \mu - \beta) \\ &= \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)(\varphi + \mu - \beta) + \beta\bar{a}(\rho + \mu + \gamma)(2 + \frac{\varphi}{\rho + \mu + \gamma}) \\ &= \varphi(\beta\bar{a} - \rho) + (\rho + \mu + \gamma)(\varphi + \mu - \beta) + \beta\bar{a}(2\rho + 2\mu + 2\gamma + \varphi) \\ &= 2\beta\bar{a}(\rho + \mu + \gamma + \varphi) - \beta(\rho + \mu + \gamma) + \mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi. \end{aligned}$$

Using equation (2.8) for $\bar{a}$, we write y_2 in terms of the parameters only as

$$\begin{aligned} y_2 &= 2\beta(\rho + \mu + \gamma) - 2\mu(\rho + \mu + \gamma + \varphi) - 2\gamma\varphi - \beta(\rho + \mu + \gamma) + \mu(\rho + \mu + \gamma + \varphi) + \gamma\varphi \\ &= \beta(\rho + \mu + \gamma) - \mu(\rho + \mu + \gamma + \varphi) - \gamma\varphi. \end{aligned}$$

C Positive invariant region

We show that provided we always take our initial conditions to lie in D, the solution will always be in D. We do this by considering the direction field at the boundary, ∂D, which is the triangle in the ar-plane with vertices $(0, 0), (1, 0)$ and $(0, 1)$. We want to show that the direction field at ∂D always enters D. This ensures that any trajectory starting in D remains in D.

The boundary will be considered as the union of six sets: each of the three vertices, and each of the three edges minus the vertices. Firstly we shall look at the direction field across the line $r = 0$ for $a \in (0, 1)$. To determine the

direction field along this boundary line we consider equations (2.3) along $r = 0$. This gives

$$\begin{aligned} \dot{a} &= -\beta a^2 + (\beta - \varphi - \mu)a, \\ \dot{r} &= \varphi a. \end{aligned} \tag{C.1}$$

As $a > 0$ along the boundary, equation (C.1) determines that $\dot{r} > 0$ along the boundary line $r = 0$. This is sufficient for us to determine that the direction field arrows at the boundary line always point into D. Similarly we now evaluate equations (2.3) along the bounday line $a = 0$ with $r \in (0, 1)$ to obtain

$$\begin{aligned} \dot{a} &= \rho r, \\ \dot{r} &= -(\rho + \mu + \gamma)r. \end{aligned} \tag{C.2}$$

As $r > 0$ we can conclude from equation (C.2) that $\dot{a} > 0$, hence all direction field arrows along this boundary line point into the region D.

The final boundary line is $a + r = 1$ for $(a, r) \in (0, 1)^2$. Along this line we can write equations (2.3) in terms of one variable by using $r = 1 - a$ which gives the equations

$$\begin{aligned} \dot{a} &= -(\rho + \mu + \varphi)a + \rho, \\ \dot{r} &= (\rho + \mu + \gamma + \varphi)a - (\rho + \mu + \gamma). \end{aligned}$$

To determine the direction that arrows cross the boundary line $a + r = 1$ we use the vector dot product. The vector dot product for two vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos\theta \tag{C.3}$$

where θ is the angle between the two vectors. We consider the vector $\begin{pmatrix}1\\1\end{pmatrix}$, which is orthogonal to the boundary line, and dot this with the vector $\begin{pmatrix}\dot{a}\\\dot{r}\end{pmatrix}$. We find that

$$\begin{pmatrix}\dot{a}\\\dot{r}\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix} = -(\rho+\mu+\varphi)a+\rho+(\rho+\mu+\gamma+\varphi)a-(\rho+\mu+\gamma)$$

$$= -\mu-(1-a)\gamma. \tag{C.4}$$

As $(1-a)>0$, the right hand side of equation (C.4) is negative. By applying the vector dot product formula (C.3) we conclude that $\cos\theta<0$ so $\theta\in(\pi/2,3\pi/2)$. For these values of θ, the direction field always crosses the boundary line $a+r=1$ in a direction which enters the region D.

We now consider the direction field at each of the vertices of the boundary triangle. No trajectories can pass through the point $(0,0)$ as $(\dot{a},\dot{r})=(0,0)$. At the point $(1,0)$,

$$\dot{a} = -(\varphi+\mu),$$

$$\dot{r} = \varphi.$$

The equation of the line passing through $(0,1)$ for which $(\dot{a},\dot{r})$ is the direction vector is $r=-\varphi(1-a)/(\varphi+\mu)$ which has a gradient less than that of the boundary line $a+r=1$, hence trajectories passing through the boundary point $(1,0)$ will always enter D. At the point $(0,1)$,

$$\dot{a} = \rho,$$

$$\dot{r} = -(\rho+\mu+\gamma).$$

The equation of the line passing through $(0,1)$ corresponding to the direction vector $(\dot{a},\dot{r})$ is $r=-(\rho+\mu+\gamma)a/\rho+1$. As the magnitude of the gradient of

this line is greater than that of the boundary line $a + r = 1$ we can conclude that all trajectories passing through the point $(0, 1)$ will always enter the feasible region.

D Endemic equilibrium solultion comparison

We show that if the inequality $R_0 > 1$ is satified then the inequality $F < 0$ must also be true. We begin by considering the inequality $F < 0$, which gives

$$0 > -\beta(\rho+\mu)(\rho+\mu+\gamma) + \mu(\rho+\mu)(\rho+\mu+\gamma) + \mu\varphi(\mu+\gamma) - \rho\varphi(\rho+\varphi).$$

This rearranges to

$$\beta(\rho+\mu)(\rho+\mu+\gamma) > \mu(\rho+\mu)(\rho+\mu+\gamma) + \mu\varphi(\mu+\gamma) - \rho\varphi(\rho+\varphi). \quad \text{(D.1)}$$

We now look at the constraints on the parameter values which come from $R_0 > 1$. This can be written as

$$\beta(\rho+\mu+\gamma) > \mu(\rho+\mu+\gamma+\varphi) + \gamma\varphi. \quad \text{(D.2)}$$

By multiplying both sides of inequality (D.2) by $(\rho + \mu)$ we get

$$\beta(\rho+\mu)(\rho+\mu+\gamma) > \mu(\rho+\mu)(\rho+\mu+\gamma) + \mu\varphi(\mu+\gamma) + \rho\varphi(\mu+\gamma), \quad \text{(D.3)}$$

which has the same left hand side as inequality (D.1). We now compare inequalites (D.1) and (D.3) and find that inequality (D.3) imposes the greatest lower bound on the expression $\beta(\rho + \mu)(\rho + \mu + \gamma)$. From this we conclude that if the parameter values satify $R_0 > 1$ then they will satify $F < 0$.