



**HAL**  
open science

# The effect of intrinsic stochasticity on transmitted HIV drug resistance patterns

Alison J. Marks, Deenan Pillay, Angela R. Mclean

► **To cite this version:**

Alison J. Marks, Deenan Pillay, Angela R. Mclean. The effect of intrinsic stochasticity on transmitted HIV drug resistance patterns. *Journal of Theoretical Biology*, 2009, 262 (1), pp.1-10.1016/j.jtbi.2009.09.017 . hal-00554647

**HAL Id: hal-00554647**

**<https://hal.science/hal-00554647>**

Submitted on 11 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

The effect of intrinsic stochasticity on transmitted HIV drug resistance patterns

Alison J. Marks, Deenan Pillay, Angela R. McLean

PII: S0022-5193(09)00437-8  
DOI: doi:10.1016/j.jtbi.2009.09.017  
Reference: YJTBI5708

To appear in: *Journal of Theoretical Biology*

Received date: 24 October 2008  
Revised date: 29 July 2009  
Accepted date: 10 September 2009

Cite this article as: Alison J. Marks, Deenan Pillay and Angela R. McLean, The effect of intrinsic stochasticity on transmitted HIV drug resistance patterns, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2009.09.017](https://doi.org/10.1016/j.jtbi.2009.09.017)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

The Effect of Intrinsic Stochasticity on Transmitted HIV Drug Resistance Patterns.

Alison J. Marks, Deenan Pillay and Angela R. McLean.

**Author information:**

Dr. Alison J. Marks, (corresponding author),

School of Pharmacy, Bradford University, Richmond Rd., Bradford, BD7 1DP, UK.

E-mail: A.J.Marks@bradford.ac.uk

Tel: +44 (0)1274 234734

Fax: +44 (0)1274 235600

Professor Deenan Pillay,

Department of Infection, University College, London, UK and

Centre for Infections, Health Protection Agency, Colindale, London, UK.

E-mail: d.pillay@ucl.ac.uk

Professor Angela R. McLean,

Institute for Emergent Infections of Humans,

James Martin 21<sup>st</sup> Century School,

Zoology Department,

Oxford University,

South Parks Road, Oxford OX1 3PS, UK.

E-mail: angela.mclean@zoology.oxford.ac.uk

Author Contributions.

AJM, ARM and DP conceived the study. AJM implemented the model. AJM and ARM analysed the data. The paper was principally written by AJM, with contributions and revisions by ARM and DP.

Accepted manuscript

**Abstract**

Estimates of transmitted HIV drug-resistance prevalence vary widely among and within epidemiological surveys. Interpretation of trends from available survey data is therefore difficult. Because the emergence of drug-resistance involves small populations of infected drug-resistant individuals, the role of stochasticity (chance events) is likely to be important. The question addressed here is: how much variability in transmitted HIV drug-resistance prevalence patterns arises due to intrinsic stochasticity alone, i.e., if all starting conditions in the different epidemics surveyed were identical? This 'thought experiment' gives insight into the minimum expected variabilities within and among epidemics. A simple stochastic mathematical model was implemented. Our results show that stochasticity alone can generate a significant degree of variability and that this depends on the size and variation of the pool of new infections when drug treatment is first introduced. The variability in transmitted drug-resistance prevalence *within* an epidemic (i.e. the temporal variability) is large when the annual pool of all new infections is small (fewer than 200, typical of the HIV epidemics in Central European and Scandinavian countries), but diminishes rapidly as that pool grows. Epidemiological surveys involving hundreds of new infections annually are therefore needed to allow meaningful interpretation of temporal trends in transmitted drug-resistance prevalence within individual epidemics. The stochastic variability *among* epidemics shows a similar dependence on the pool of new infections if treatment is introduced after endemic equilibrium is established, but can persist even when there are more than 10,000 new infections annually if drug therapy is introduced earlier. Stochastic models may therefore have an important role to play in interpreting differences in transmitted drug-resistance prevalence trends among epidemiological surveys.

**Keywords**

Viruses, mathematical model, epidemiology.

Accepted manuscript

## 1. Introduction

The use of highly-active anti-retroviral therapy (HAART) for treating HIV infections is increasing. However, resistant strains of HIV-1 arise due to drug failure or poor drug adherence, leading to the potential for drug-resistant HIV strains to be transmitted to susceptible individuals. Transmitted resistance has been documented for each of the three classes of widely-used anti-retroviral drugs: nucleoside reverse-transcriptase inhibitors (NRTIs), non-nucleoside reverse-transcriptase inhibitors (NNRTIs) and protease inhibitors (PIs) (for recent reviews see Tang and Pillay, 2004; Geretti, 2007).

In recent years a number of surveys have been carried out monitoring changes in the proportion of newly infected (i.e. with acute or primary infection) individuals carrying an HIV-1 drug-resistant strain, potentially providing insight into the transmission dynamics of drug-resistant strains. Reported estimates of the transmitted anti-retroviral drug resistance prevalence range from 0.0% (Alexander et al., 1999) to over 25% (Grant et al., 2002; de Mendoza et al., 2005) of recently infected individuals carrying virus with at least one primary resistance mutation. Thus, the variability among epidemics is high. Some of this variability arises due to differing prevailing conditions. For example, treatment regimens in different countries vary, affecting the transmission of drug-resistant virus. The temporal trends within some individual surveys appear to show a decrease in transmitted drug-resistance, (Ammaranond et al., 2003; Bezemer et al., 2004; de Mendoza et al., 2005; Routy et al., 2005; Yerly et al., 2001) whereas others show that the prevalence is increasing (Grant et al., 2002; Little et al., 2002; UK HIV Drug Resistance Database, 2003; UK Group on Transmitted HIV Drug Resistance, 2005). That is, the temporal behaviour of individual epidemics is not consistent among surveys. While it is possible that these

trends are real, arising from temporal changes in clinical conditions or behaviour (Sánchez et al., 2005, 2006), it is important to avoid incorrect interpretations which may lead to ill-informed, detrimental changes in public health policies. For example, observed declines in transmitted drug resistance may lead to decreased support for HIV drug-resistance monitoring or drug adherence counselling (Sánchez et al., 2006).

Two modelling studies (Blower et al., 2000, 2001) have addressed the problem of uncertainty in clinical and biological conditions, but assume that the time evolution of an epidemic is deterministic (see Table 1). The latter assumption is hard to defend when dealing with small populations, when the role of chance events can alter the course of an epidemic. However, although the importance of stochasticity has been highlighted with regard to the within-host evolution of HIV viral strains (Leigh-Brown, 1997; Leigh-Brown and Richman, 1997; Nijhuis et al., 1998; Phillips et al., 2001), its role in influencing the prevalence trends of transmitted drug-resistant viral strains at the population level has not been extensively explored (for a very recent stochastic modelling study with a different emphasis, see Vardavas and Blower, 2007, and for a meeting abstract outlining similar issues to those addressed here, see Brown et al., 2002). Of particular interest here is the role of *intrinsic* stochasticity i.e., stochastic effects or noise arising directly from the inherent (and discrete) nature of interactions between susceptible and infectious individuals. With this in mind, the question that we address in this study is: how much variability in transmitted drug-resistance patterns arises due to intrinsic stochasticity alone, that is, if all starting conditions (population sizes and rate parameters) in the different epidemics surveyed were identical? This 'thought experiment' gives insight into the *minimum* expected variabilities within and among epidemics. The question addressed here is therefore different from that posed in earlier modelling studies (Blower et al., 2000, 2001)



which explore the issue of uncertainty in starting and treatment conditions (see Table 1). These are further sources of uncertainty that would add to the purely stochastic variability among epidemics explored here.

## 2. Methods

The compartment model used in this study is a homosexual transmission model, very similar to that used by Blower et al., (Blower et al., 2000, 2001). It comprises five types of individual: susceptible (S); untreated and infected with wild-type (drug-sensitive) virus ( $I_w$ ); untreated and infected with pure resistant (r) or mixed (w,r) viral strains ( $I_r$ ); treated and infected with either drug-sensitive ( $T_w$ ) or drug-resistant ( $T_r$ ) viral strains. A flow diagram is given in Figure 1 illustrating the compartments and transmission dynamics used in the model. The corresponding differential equations are given in the Appendix. Deterministic simulations were carried out through numerical solution of the ordinary differential equations using a 5<sup>th</sup>-order variable-step Runge-Kutta method (Press et al., 1992). The parameter values were selected to give a basic reproductive ratio  $R_0$  (in the absence of treatment) of 3.33. Estimates of  $R_0$  for HIV are often lower than this (e.g.,  $R_0 = 2.4$ , Bezemer et al., 2008), but we consider our choice to be reasonable for high-risk groups. Indeed, a recent study suggests that in the early stages of the Scandinavian epidemic,  $R_0$  was as high as 15.0 (Amundsen et al., 2004). The choice of a relatively high value for  $R_0$  results in fewer epidemic extinctions and if anything reduces the stochastic effect that we are studying. The model parameters were also selected to be broadly representative of HAART as currently practised (see Table 2 and references therein). The treatment uptake rate was assumed to be 40% p.a., and the rate of (temporarily) ceasing therapy was 10% p.a.. The average time for conversion of an untreated drug-resistant

infection to a drug-sensitive infection ( $1/\gamma$ ) was taken to be 1 year, longer than in previous modelling studies. This represents an average value reflecting the fact that whilst 'growing out' of archived drug-sensitive viral strains (in those with acquired drug-resistance) occurs within a few months, reversion of drug-resistant virus to a drug-sensitive strain by mutation (in those with a transmitted drug-resistant infection) may take years (Brenner et al., 2001).

Deterministic simulations were used for comparison with the stochastic means. A deterministic simulation with a given set of parameters represents the mean behaviour of a collection of epidemics. A repeat deterministic simulation with the same parameters would give an identical result, i.e., there is no error in the deterministic mean. Variability in deterministic simulations arises only from varying the initial parameters (Blower et al., 2000, 2001).

Monte Carlo stochastic simulations were carried out using the Gillespie algorithm (Gillespie, 1976, 1977) and a random number generator (Chandler and Northrop, 1999; Marsaglia and Zaman, 1991) (see Appendix and Table 1) to assess the variability of transmitted drug-resistance prevalence outcomes. During a simulation, for each year we calculated the total number of new infections and the proportion of these that are drug-resistant, giving the prevalence of transmitted (primary) drug resistance in newly infected individuals.

The sensitivity of our results to changes in model parameters was also explored, and it was found that although these affect the predicted mean values of transmitted drug-resistance prevalence, they do not have a major effect on the variability of transmitted drug-resistance prevalence. A dominating driver of variability in this study is the initially small number of drug resistant infections – a reflection of the process of the emergence of drug resistance that is present for any

realistic set of parameter values. However, an important factor in determining the degree of variability in transmitted drug-resistance prevalence is the size and variation of the pool of infected individuals at the time when drug therapy begins. We have used two different treatment introduction times ( $t_i$ ) in our simulations to illustrate this point. However, it is important to note that  $t_i$  by itself cannot be used to predict the variability in transmitted drug-resistance prevalence. It is the effect of  $t_i$  on the mean size and variation of epidemics that is important, and this will depend on the values of the model parameters used.

### 3. Results and Discussion

Stochastic Monte Carlo simulations were used to investigate how intrinsic stochasticity influences variability in the prevalence of transmitted drug-resistant HIV infections. Our results indicate that stochasticity alone can generate a significant degree of variability. We suggest that much of the variability in and among epidemiological surveys can be attributed to intrinsic stochasticity. Our main findings are that the variability in transmitted drug-resistance prevalence *within* an epidemic (i.e. the temporal variability) is large when the annual pool of all new infections is small (fewer than 100-200), but diminishes rapidly as that pool grows. A similar dependence is found for the variability *among* epidemics when treatment is introduced after each epidemic has reached equilibrium. In contrast, if drug treatment is introduced prior to endemic equilibrium, the variability among epidemics can persist even when there are more than 10,000 new infections annually.

#### (a) Epidemiological surveys of transmitted drug-resistance prevalence.

Before describing our simulations, we first illustrate in Figure 2(a) examples of observed epidemiological time series measuring the prevalence of primary (i.e. transmitted) drug resistance mutations in newly-diagnosed HIV-1 infections in countries where HAART is used. The available temporal data sets are very limited, and cases of homosexual and heterosexual HIV transmission, and transmission in injecting drug users are often “lumped” together. In addition, some data sets are collected from single cities, whereas others are heterogeneous, multi-centre studies. We note also that in some of the larger studies (Bennett et al., 2003; UK HIV Drug Resistance Database, 2003; Wensing et al., 2003), the newly diagnosed infections are not necessarily recent infections. This may result in an underestimate of transmitted drug resistance (due to reversion of drug-resistant viral mutations over time) in these studies.

Typical values for the prevalence of primary drug resistance mutations are in the range 10-15%, but there is considerable variability among the surveys, ranging from 0-33%. In some surveys, annual variations in primary HIV drug-resistance prevalence were recorded, but in others the data are “lumped” and the latter are indicated by horizontal lines. In addition, the surveys range from those focussing on new HIV cases arising in a single city (e.g. San Francisco (Grant et al., 2002) , New York (Simon et al., 2002) and Amsterdam (Bezemer et al., 2004)) to national or international multi-centre surveillance schemes such as those carried out in the USA (Bennett et al., 2003; Little et al., 2002), UK (UK HIV Drug Resistance Database, 2003), Switzerland (Yerly et al., 2001) and Europe (Wensing et al., 2003). The temporal variability associated with a single survey is difficult to assess because data are available for only a few years (since the introduction of HAART in 1996) and in addition, the datasets are typically small (tens to hundreds of new HIV cases per

year). In some cases, these comprise comprehensive data collected from small epidemics in individual cities (e.g. San Francisco (Grant et al., 2002)), and in others they consist of samples taken from larger epidemics (e.g. the UK study (UK HIV Drug Resistance Database, 2003)). Despite this, there have been a number of attempts to rationalise and interpret observed trends such as apparent decreases (Ammaranond et al., 2003; de Mendoza et al., 2005; Yerly et al., 2001) or increases (Grant et al., 2002; Little et al., 2002; UK Group on Transmitted HIV Drug Resistance, 2005) in transmitted drug-resistance prevalence.

The available epidemiological time series data shown in Figure 2(a) are included here to illustrate the observed range of variability in transmitted drug-resistance prevalence. It is neither plausible, nor our intention to make a detailed comparison of these epidemiological data with our simple model simulations. Instead, we are trying to illustrate (using a minimal, homosexual HIV transmission model with homogeneous mixing) the simple point that intrinsic stochasticity alone may contribute significantly to the observed variability in transmitted drug-resistance prevalence, and should not be ignored.

#### **(b) Stochastic simulations of transmitted drug-resistance prevalence.**

In our stochastic simulations, a single infected wild-type case ( $I_w$ ) was introduced into a susceptible population of 50,000. Two scenarios were considered, in which the course of an epidemic was followed initially for either 20 or 30 years with no drug treatment, allowing a population of drug-sensitive infected individuals to build up. The choice of two treatment introduction times ( $t_i$ ) was used as a means to generate different average epidemic sizes. A time interval of 20 years until treatment commences ( $t_i = 20$  yr) resulted in small epidemics, with a mean of 396 infections at t

= 20 years, and a mean of 1129 infections (both drug-sensitive and drug-resistant) ten years after treatment commenced. The mean annual number of new infections at  $t=20$  yr was 137. These figures are similar to the HIV statistics of many Central European and Scandinavian countries (EuroHIV, 2007; UNAIDS WHO, 2007).

An interval of 30 years before introduction of drug treatment ( $t_i = 30$  yr) results in larger epidemics, and was chosen to correspond to an equivalent period from HIV emergence in the USA (Gilbert et al., 2007) to first treatment. In this case, there was a mean of 4698 infections at  $t=30$  years, and a mean of 8280 infections (both drug-sensitive and drug-resistant) ten years after treatment begins. The mean annual number of new infections at  $t=30$  years was 1430. The incidence and prevalence of HIV ten years after the start of treatment were 2.5% and 19.4% respectively, typical of HIV epidemics within male homosexual communities in many cities of the USA (CDC MMWR, 2005) and Western Europe (Dougan et al., 2007).

For each scenario, 1000 stochastic simulations were followed, of which 286 ( $t_i = 20$  yr) and 294 ( $t_i = 30$  yr) died out before treatment was initiated. These values are close to the predicted extinction probability of  $1/R_0 = 0.30$ , or 300/1000 simulations. Identical parameters were used in each stochastic simulation (see Table 2), to illustrate the potential variability in transmitted drug-resistance prevalence arising from stochasticity alone. Each stochastic simulation represents a possible realization of the time-course of a single epidemic, whereas a deterministic simulation represents the mean expected behaviour of a collection of epidemics.

Figures 2(b) and 2(c) show the percentage of new infections with drug-resistant virus per annum (p.a.) (i.e., the primary or transmitted drug-resistance prevalence) when  $t_i = 20$  yr and  $t_i = 30$  yr respectively. In the former case, the transmitted drug-resistance prevalence rises to an initial plateau of 15%. However,

after 75 years there is a further rise, leading to an equilibrium value (not shown) of approximately 24%. When  $t_i = 30$  yr, (Figure 2(c)) there is no quasi-equilibrium plateau, and the equilibrium prevalence of 24% is reached 50 years after commencement of drug treatment. The mean of the stochastic simulations follows the corresponding deterministic result quite closely, but the variability in the stochastic prevalence values is high when  $t_i = 20$  yr and moderate when  $t_i = 30$  yr. Thus, the variability in transmitted drug resistance prevalence among epidemics is higher when drug treatment is introduced earlier in an epidemic. The range of simulated values for the prevalence of transmitted drug resistance in Figures 2(b) and 2(c) is similar to that observed in the epidemiological studies (see Figure 2(a)) and indicates that even if all other conditions were equal, a significant amount of variability may be observed due to intrinsic stochasticity alone. When drug treatment is introduced after 20 years, the total number of new infections annually is quite small and the percentage of new infections with drug-resistance shows a great deal of variability (Figure 2(b)). In individual stochastic simulations involving very small numbers (e.g. single figures) the prevalence fluctuates between 0-100% (upper limit not shown in Figures 2(b)). That is, in a given year, all new infections may, by chance, be drug-sensitive or drug-resistant.

We note that most of the epidemiological surveys begin in 1996 or later, coinciding with the introduction of combination therapies (HAART). However, drug treatment in the form of monotherapy had already been in place for some years. This means that by 1996, drug resistance was already present in the infected population, and so the relatively steep initial rise in the transmitted drug-resistance prevalence seen in our simulations (at the onset of drug treatment) are not evident in the epidemiological surveys summarised in Figure 2(a).

**(c) Comparison of simulations with epidemiological surveys.**

As stated in section 3(a) it is not our intention to make an in-depth comparison of our simulations with the epidemiological data illustrated in Figure 2(a). Rather, we are trying to illustrate a simple point, namely that intrinsic stochasticity alone can generate significant variability in transmitted drug-resistance prevalence, and that stochasticity should not be ignored when analysing or modelling epidemiological data relating to transmitted (primary) drug-resistance.

In Figures 3(a) and 3(b) we illustrate nine stochastic simulations selected randomly from Figures 2(b) and 2(c) respectively. We have plotted the simulation data beginning 10 years after introduction of drug treatment, after the initial steep rise in transmitted drug-resistance prevalence. This is to allow a better comparison with the epidemiological surveys, since by 1996 (when data collection and HAART began) drug-resistance was already present in the infected populations as a result of earlier monotherapy.

It is clear that when  $t_i = 20$  yr (Figure 3(a)), there is a high degree of variability in transmitted drug-resistance prevalence both among epidemics and within individual epidemics. When  $t_i = 30$  yr (Figure 3(b)), the variability in transmitted drug-resistance prevalence among epidemics at any point in time, and the temporal variability of individual epidemics, are both reduced. Comparing Figures 3(a) and 3(b) with the nine epidemiological surveys illustrated in Figure 2(a) shows that much of the variability among epidemics recorded in these surveys could be explained as intrinsic stochastic variability arising from initially identical systems. Figures 3(a) and (b) also illustrate the potential for temporal variability within individual epidemics due to stochasticity, and in some cases there are significant annual fluctuations in



prevalence, including both transient increases and decreases. Such effects have been observed in epidemiological surveys, and our simulations predict that they are to be expected on stochastic grounds alone. For example, in a North American study (Grant et al., 2002), NRTI resistance decreased from 30% in 1997 to 5% in 1999, rising to 20% in 2000 then falling again to 15% in 2001.

The epidemiological studies comprise datasets of small numbers (tens to hundreds) of new HIV cases per year, but do not record the number of infected persons in the population, or the sub-set of the susceptible population that is screened. Thus, it is not possible to make a precise comparison with our simulations, in which every simulated new infection is recorded. The annual number of new infections in the stochastic simulations with  $t_i = 20$  yr is of the same magnitude (mean 178 p.a. ten years after treatment commences) as the number of new infections in Central European and Scandinavian countries (EuroHIV, 2007; UNAIDS WHO, 2007); when  $t_i = 30$  yr, the mean annual number of new infections ten years after treatment begins is 987, comparable to the estimated annual number of new HIV cases in countries such as Switzerland and the Netherlands (EuroHIV, 2007; UNAIDS WHO, 2007).

**(d) The annual pool of new infections and transmitted drug-resistance prevalence.**

The transmitted drug-resistance prevalence of interest in this study depends on the annual number of new drug-sensitive infections ( $N_w$ ) and new drug-resistant

infections ( $N_r$ ), and is given by  $\frac{N_r}{(N_w + N_r)} \times 100\%$ . It is therefore useful to examine

the size and variation of the pools of new drug-sensitive and drug-resistant infections for differing treatment introduction times.

Figures 4(a) and 4(b) show individual stochastic simulation results for the number of new infections each year that are drug-sensitive and drug-resistant respectively when drug treatment is introduced after 20 years. Also shown are the mean of the stochastic values and the deterministic outcomes. Figures 4(c) and 4(d) show the corresponding results when treatment is introduced after 30 years. Stochastic simulations that died out before commencement of drug treatment are excluded from the stochastic mean. When  $t_i = 20$  yr, the stochastic mean is less than the deterministic prediction in the range  $t = 60-90$  years (Figures 4(a) and 4(b)) but the values later converge; when  $t_i = 30$  yr, there is also a deviation, with convergence after 80 years. These differences between deterministic and stochastic means occur because many of the stochastic epidemics take a long time to become established. In an individual epidemic, the annual number of new drug-sensitive infections ( $N_w$ ) and the annual number of new drug-resistant infections ( $N_r$ ) are positively correlated. The transmitted drug-resistance prevalence is given by  $\frac{N_r}{(N_w + N_r)} \times 100\%$ , and the stochastic and deterministic means of this quantity show much less deviation (see Figures 2(b) and 2(c)).

Because each simulated epidemic differs due to stochasticity, when treatment is initiated the total number of drug-sensitive individuals  $I_w$  in the population varies, contributing to the variability seen later in time. At the onset of treatment, when  $t_i = 20$  yr the number of new drug-sensitive cases is in the range 0-800 p.a. (Figure 4(a)), with a mean of 137 p.a., whereas when  $t_i = 30$  yr there are up to 4000 new drug-sensitive cases p.a., with a mean of 1430 p.a. (Figure 4(c)) There are no drug-resistant cases initially (see Figures 4(b) and 4(d)), but after some time on drug treatment, cases of acquired drug resistance arise which may subsequently cause new drug-resistant infections by transmission of the drug-resistant viral strain. For example,

when  $t_i = 20$  yr, 10 years after treatment begins there are up to 150 new drug-resistant cases p.a. (with a mean of 23 p.a.) and up to 700 new drug-sensitive cases p.a. (with a mean of 155 p.a.); when  $t_i = 30$  yr, ten years after treatment begins there are up to 300 new drug-resistant cases p.a. (with a mean of 150 p.a.), and up to 1200 new drug-sensitive cases p.a. (mean 780 p.a.). A significant number of the simulated epidemics show large deviations from the mean (and from the deterministic result), illustrating the potential for high variability among epidemics due to stochasticity alone. This variability is more pronounced when  $t_i = 20$ yr, corresponding on average to smaller epidemics.

One potential source of stochastic variability is the timing of first emergence of transmitted drug-resistance. In Figure 5 we illustrate the range of times over which transmitted drug-resistance first appears in the stochastic simulations. In the majority of simulated epidemics transmitted (primary) drug-resistance emerges less than 2 years after drug treatment is first introduced. Thus, the timing of emergence of transmitted drug-resistance does not appear to play a major role in determining the variability in transmitted drug-resistance prevalence.

**(e) Analysis of the stochastic variability in transmitted drug-resistance prevalence within individual epidemics.**

In order to investigate the temporal variability in transmitted drug-resistance prevalence within individual epidemics, for each stochastic simulation (illustrated in Figures 2(b) and 2(c)) a least-squares fit was obtained over a 30 year time period ( $t_i + 10, t_i + 40$ ) yr. A quadratic function was used for the fitting procedure because there was some non-linearity in the simulations. The standard deviation of the residuals (RSD) was calculated and the residual coefficient of variation (RCV) was estimated

as  $(RSD/M) \times 100\%$ , where  $M$  is the mean transmitted drug-resistance prevalence for the individual simulation.

For each stochastic epidemic, the RCV was plotted against the mean annual number of new infections (drug-sensitive and drug-resistant) occurring over the 30 year time interval. This is shown in Figure 6 for  $t_i = 20$  yr,  $t_i = 30$  yr and for treatment introduction after the epidemic has reached equilibrium ( $t_i = 150$  yr). The temporal variability in transmitted drug-resistance prevalence within an epidemic is high when the annual number of new infections is small, but decreases rapidly for larger epidemics. Once the annual number of new infections exceeds about 200, the variability as measured by the RCV is less than 20%. The trends observed are independent of the treatment introduction time  $t_i$ . However, because early introduction of drug treatment results in smaller numbers of new infections, the variability within epidemics on average decreases with  $t_i$  (see Figures 3(a) and 3(b)). If treatment is introduced after an epidemic has reached equilibrium there is much less stochastic variability in the annual number of new infections, a factor that reduces the variability in transmitted drug-resistance prevalence *among* epidemics described later.

The general trends shown in Figure 6 were reproduced for a range of model parameters. However, in cases where the model parameters led to significantly lower/higher transmitted drug-resistance prevalence values, the RCV for a specific number of new infections was correspondingly higher/lower. A better correspondence among different models was obtained by plotting the RCV against the number of new *drug-resistant* infections p.a. rather than the total number of new infections p.a. (See Supplementary Figure S1). This highlights the fact that the dominant driver of variability in transmitted drug-resistance prevalence is the annual number of new

drug-resistant infections, rather than the total annual number of new infections. Overall, our results indicate that the RCV is less than 20% when the annual number of new drug-resistant infections exceeds around 30. For 'typical' HIV epidemics with a transmitted drug-resistance prevalence of 10%-20%, this corresponds to approximately 150-300 new infections p.a.. In summary, the temporal variability in transmitted drug-resistance prevalence within an epidemic depends principally on the annual number of new drug-resistant infections. The timing of introduction of drug treatment is not in itself important, and neither is the value of  $R_0$ . An epidemic with low numbers of new infections will have large variability in transmitted drug resistance prevalence regardless of the cause of those low numbers (e.g., a small susceptible population, a small value of  $R_0$ , or early introduction of drug treatment).

**(f) Analysis of the stochastic variability in transmitted drug-resistance prevalence among epidemics.**

In addition to analysing the temporal variability in transmitted drug-resistance prevalence within an individual epidemic, we can also ask the question: what amount of variability in transmitted drug-resistance prevalence arises *among* initially identical epidemics at any point in time? This thought experiment might aid in understanding some of the observed variability among real epidemics.

To quantify the variability in transmitted drug-resistance prevalence *among* epidemics at any point in time, we calculated the coefficient of variation ( $CV = (\text{standard deviation}/\text{mean}) \times 100\%$ ) as a function of time, for differing times of introduction of drug treatment. This is illustrated in Figure 7 (a). In each case the CV decreases with time, and the variability among epidemics becomes smaller as the time until first introduction of drug treatment ( $t_i$ ) increases. The time of first introduction of

treatment affects the size and variation of the pool of infected individuals at the time when drug therapy begins. At the point of introduction of drug treatment, stochasticity will already have led to a range of epidemics of differing sizes. As can be seen in Figures 4(a) and 4(c), initially identical epidemics can diverge significantly prior to treatment introduction, but tend to converge later, some years after therapy begins. When treatment is introduced after an epidemic has reached equilibrium there is much less variation in the annual number of new infections among epidemics (see, e.g. Figure 6, which shows 'clustering' of the number of new infections for  $t_i = 150$  yr).

The two principal factors affecting the variability in transmitted drug-resistance prevalence among epidemics at a given time are the mean number of new drug-resistant infections p.a., and variations in the dynamics of the drug-sensitive and drug-resistant epidemics (see Section 3(d)). Figure 7(b) shows the variability in transmitted drug-resistance prevalence among epidemics ten years after treatment begins, as a function of the mean number of drug-resistant infections p.a.. Differing mean numbers of drug-resistant infections were obtained by running the model simulations with different susceptible population sizes. Also shown is a typical scatter plot illustrating the temporal variability within individual epidemics (as discussed in Section 3(e)). When treatment is introduced after equilibrium, the trend in variability among epidemics is very similar to the temporal variability within individual epidemics, showing that here the dominant factor is the mean number of new drug-resistant infections. When treatment is introduced relatively early in each epidemic (e.g.  $t_i = 30$  yr) the variability persists at a higher level (CV approximately 20%) as the number of new drug-resistant infections increases. Here, significant divergence in the dynamics of different epidemics gives an additional source of variability.

It would be unrealistic for all epidemics to be exactly the same size when treatment is introduced. However, by comparing our stochastic simulations with a 'thought experiment' in which all epidemics were of equal size at  $t_i$ , we ascertained that in the first 20 years after  $t_i$ , at least 50% of the variability in transmitted drug-resistance prevalence among epidemics arises from the variation in epidemics already existing when drug treatment begins (see Supplementary Figure S2).

The focus so far has been on relatively small epidemics. However, Figure 7(b) indicates how the variability can persist as the number of new infections increases. In order to investigate the variability when the annual number of new infections is very large, we performed simulations with  $t_i = 40$  years, and a susceptible population size  $S=500,000$ , allowing a large population of drug-sensitive infected individuals (mean  $I_w = 67,000$ ) to build up. Ten years after commencement of drug therapy, the mean number of new drug-sensitive infections p.a. was approximately 8000, with 2000 new drug-resistant infections p.a. (compared with 700 new  $I_w$  and 200 new  $I_r$  p.a. when  $S=50,000$  and  $t_i = 40$  yr). Despite a 10-fold increase in the annual number of infections, the coefficients of variation in the first 20 years following introduction of drug treatment are very similar, in the range 15%-20%. This implies that even if the annual pool of new infections is large (around 10,000 p.a., greater than the annual number of newly diagnosed HIV cases in the UK, which was about 9,000 in 2006 (UNAIDS WHO, 2007)), intrinsic stochasticity alone can generate significant variability in transmitted drug-resistance prevalence among epidemics. Further simulations with even larger susceptible populations (up to 5 million with  $t_i = 20, 30$  and 40 yr) showed that the variability among epidemics comprising a mean of up to 30,000 annual infections has a CV of 10% or more for approximately 40 years following introduction of drug treatment.

The variability in transmitted drug-resistance prevalence among epidemics shows some dependence on the model parameters used (see Supplementary Figure S3), but the general conclusion is that if drug treatment is introduced prior to endemic equilibrium, then even when the mean annual number of new infections is of the order of tens of thousands, the variability among epidemics may have a CV of at least 10% for 40 years or more after first introduction of drug treatment. The principal reason for this variability among large epidemics is the fact that at the point of introduction of drug therapy ( $t_i$ ), stochasticity has already given rise to a range of epidemics of differing sizes and dynamics. Examples of this range can be seen in Figures 4 (a) (at  $t = 20$  yr) and 4(c) (at  $t = 30$  yr). In contrast, if treatment is introduced after each epidemic has reached equilibrium, then the stochastic variability in transmitted drug-resistance prevalence among epidemics is much smaller. This is because at equilibrium the epidemic dynamics have converged and the variability now depends principally on the number of new drug-resistant infections p.a. (Figure 7(b)).

Most well-established HIV epidemics (e.g. those in the USA) are assumed to have stabilised very rapidly, prior to introduction of drug treatment (see e.g., Bongaarts et al., 2008). Thus, the stochastic variability among these epidemics at a given time is predicted to be similar to the temporal variability within individual epidemics (see Figure 7(b)). However, a number of epidemics in Eastern European countries did not begin until the late 1990s (Kelly and Amirkhanian, 2003) and have not yet stabilised (Bongaarts et al., 2008). The predicted stochastic variability in transmitted drug-resistance prevalence among these epidemics might therefore be better represented by an 'early treatment introduction' model scenario.



**(g) Limitations of the Study.**

Our model is a simple one, and its limitations include the fact that it is a compartmental model which assumes homogeneous mixing of the susceptible and infected populations. The model structure ignores the role of early (acute) infection, which may affect the detailed transmission dynamics of HIV (Brenner et al., 2007; Hollingsworth et al., 2008; Yerly et al., 2001), and assumes homosexual transmission of a single drug-resistant HIV strain. The stochastic implementation of the model assumes exponentially distributed passage times for all events. Exponential distributions are not always the most realistic for describing infectious periods (Keeling and Grenfell, 1998; Lloyd, 2001) and the detailed dynamics may again be affected.

One of the discrepancies between our model and reality is the fact that the time-scale for reaching epidemic equilibrium before treatment introduction in our model is many decades, whereas real HIV epidemics stabilise much more quickly. One reason for this is that we use a single parameter  $\beta_w$  to describe HIV transmissibility over time. In reality epidemics usually increase rapidly initially (equivalent to a very large  $\beta_w$  or  $R_0$  value) and then, as awareness of the epidemic increases, behavioural changes lead to a slowing down of the spread of infection and a reduction in  $R_0$  (see, e.g., Amundsen et al., 2004; Gran et al., 2008). Another factor contributing to the longer timescale for epidemic stabilisation in our simulations is the fact that we assume exponentially distributed infectious periods, whereas early (acute) infections are known to be much more infectious than the asymptomatic stage (see, e.g., Hollingsworth et al., 2008) and may therefore contribute significantly to an earlier peak in epidemic incidence. However, as mentioned earlier, the drug treatment

introduction times ( $t_i$ ) used in our model were selected principally as a means to generate different average epidemic sizes, and their absolute values are not relevant to the conclusions of our study. Furthermore, our study focuses on the emergence of transmitted drug-resistance and its prevalence (i.e., the *proportion* of all new infections that are drug-resistant). The latter is a relative quantity and its temporal evolution is not significantly affected by the two timescale factors discussed above.

Sensitivity of the model to changes in parameter values was investigated, and there are some differences in results (see Supplementary Figures S1 and S3), but these do not affect the general conclusions of our study. We emphasise that our aim in this study was to use a simple stochastic model to explore the *minimum* expected variabilities in transmitted drug-resistance prevalence within and among epidemics. Uncertainties in starting and treatment conditions (e.g. Blower 2000, 2001), and sampling errors in the epidemiological surveys clearly constitute additional sources of variability. However, the question of how heterogeneous mixing (e.g. contact networks) affects the predicted stochastic variability in transmitted drug-resistance prevalence is worthy of further investigation. A very recent study using a simple susceptible-infectious-susceptible (SIS) epidemic model (i.e., not involving drug treatment) suggests that during the early exponential growth phase heterogeneous mixing leads to slightly lower variability in epidemic sizes, but that at endemic equilibrium the variability is slightly greater than that predicted by a homogeneous mixing model (Dangerfield et al., 2008).

#### **4. Conclusion**

We have shown using a simple model that intrinsic stochasticity plays an important role in generating variability in transmitted drug-resistance prevalence

patterns both among epidemics and within individual epidemics. This is true even when the susceptible population studied is large, because the important factor in governing stochasticity here is the relatively small numbers of new infections (both drug-sensitive and in particular, drug-resistant) occurring each year. For a given set of model parameters the degree of variability depends on the time  $t_i$  at which drug treatment is first introduced, because this affects the size and variation of the pool of new infections. It is important to note that the absolute value of  $t_i$  by itself cannot be used to predict the variability in transmitted drug-resistance prevalence. It is the effect of  $t_i$  on the mean size and variation of the pool of new infections that is important, and this will depend on the values of the model parameters.

A small pool of new infections results in high variability within individual epidemics, and a large variation in the pool of new infections at the time of treatment introduction (see e.g. Figures 4(a) and 4(c)) contributes to the subsequent variability in transmitted drug-resistance prevalence among epidemics. These intrinsic stochastic effects are different from the variability introduced by uncertainty in parameter values (see Table 1). The latter effects cause variability among epidemics additional to that arising from intrinsic stochasticity.

Our results indicate that stochastic temporal variability in transmitted drug-resistance prevalence within individual epidemics is driven by the small annual number of drug-resistant infections, and that when this exceeds around 30, the residual coefficient of variation is less than 20%. For typical HIV epidemics with a transmitted drug-resistance prevalence of 10%-20%, this corresponds to small epidemics with fewer than around 150-300 new infections p.a., such as those occurring in many Central European countries. This suggests that epidemiological surveys involving a minimum of hundreds of new infections annually are required to

allow meaningful interpretation of temporal trends within individual epidemics. The same conclusion applies to the variability in transmitted drug-resistance prevalence *among* epidemics, when comparing epidemics in which treatment was introduced after equilibrium (a situation observed for most well-established HIV epidemics). However, for epidemics where treatment is introduced prior to equilibrium, the uncertainty in interpretation of transmitted drug-resistance prevalence trends among epidemics is much larger and significant variability may persist even when the annual pool of new infections exceeds tens of thousands. This observation may be relevant to HIV epidemics that have emerged relatively recently in some Eastern European countries and have not yet stabilised.

Because none of the epidemiological studies include data on the exact number of infected persons in the population or the sub-set of the total population that is screened, it is not possible to make a precise comparison with our simulations. However, the annual number of new infections in the simulations is of the same magnitude as the pool of new infections from which samples are drawn in epidemiological surveys of cities or small countries, giving credibility to our conclusions. We have shown that intrinsic stochasticity alone causes significant variability in transmitted drug-resistance prevalence patterns. However, there are other sources of variability not addressed here. These include real differences (heterogeneities) among the communities surveyed, temporal changes in behaviour and treatment regimens, and measurement error in initial conditions and parameter values. In addition, whereas our computer simulations record all new HIV cases, the data collected in epidemiological surveys are samples taken from the pool of new HIV infections, and this sampling constitutes a further source of variability. These

simulations of intrinsic stochasticity therefore give an indication of the *minimum* expected inherent variability in drug-resistance prevalence patterns.

In order to understand the role of intrinsic stochasticity in epidemiological surveys it would be useful if these surveys reported an estimate of the size of the underlying pool of infections as well as the sample size.

The role of stochasticity in influencing the transmission dynamics of drug-resistant viral strains in infected populations has been explored previously by Vardavas and Blower, 2007, within the context of the large (around 300,000 infected) HIV epidemic in Botswana. In that paper, the focus is on temporal variability in transmitted drug-resistance prevalence during the first five years following introduction of drug therapy. The paper concludes that stochastic effects are important early on, but does not explore the longer-term dynamics. In contrast, in our study we explore the influence of stochasticity on the longer-term dynamics of transmitted drug resistance both within and among relatively small epidemics and illustrate how this depends on the size and variation of the pool of new infections.

Our message is that significant variability among epidemiological surveys and within individual surveys is expected on stochastic grounds alone, and that caution should be used when interpreting short-term temporal trends. This intrinsic variability is only exposed by the study of stochastic models, which have a key role to play in aiding understanding of drug-resistance transmission. We suggest that when epidemics are relatively small, and/or drug treatment is introduced prior to an epidemic reaching equilibrium, stochastic implementations of deterministic models (e.g. Blower et al., 2000, 2001) should be used to aid interpretation of transmitted drug-resistance prevalence trends in epidemiological surveys. Our study highlights the importance of stochasticity with respect to the transmission of drug-resistant HIV, but

its general conclusions are applicable to the transmission of new drug-resistant strains of any infectious organism.

## Appendix

The differential equations underlying the deterministic model are as follows:

$$\frac{dS}{dt} = B - \mu S - (\lambda_w + \lambda_r)S$$

$$\frac{dI_w}{dt} = -(\mu + \alpha + \pi)I_w + \lambda_w S + \omega T_w + \gamma I_r$$

$$\frac{dI_r}{dt} = -(\mu + \alpha + \pi + \gamma)I_r + \lambda_r S + \omega T_r$$

$$\frac{dT_w}{dt} = -(\mu + \alpha_{tw} + \omega + \phi)T_w + \pi I_w$$

$$\frac{dT_r}{dt} = -(\mu + \alpha_{tr} + \omega)T_r + \pi I_r + \phi T_w$$

where the forces of infection are:

$$\lambda_w = \beta_w (I_w + \rho_w T_w)/N$$

$$\lambda_r = \beta_r (I_r + \rho_r T_r)/N$$

Descriptions of the parameters and their values are given in Table 1. This deterministic model is similar to Blower's (Blower et al., 2000, 2001), but the forces of infection in our model are simpler because patients in the 'drug-resistant' compartments ( $I_r$ ,  $T_r$ ) only transmit drug-resistant virus, whereas in Blower's model they can transmit both drug-resistant and drug-sensitive virus.

The stochastic model was implemented in Fortran using the Gillespie algorithm (Gillespie, 1976, 1977) devised for the study of chemical kinetics. Each compartment contains a discrete number of individuals, in contrast to the continuous differential equation model. The idea is that the probability of a specific process is proportional to

the corresponding rate term in the deterministic model. For example, the process of acquired drug-resistance  $T_w \rightarrow T_r$  has a probability proportional to  $\phi T_w$ . To study the time evolution of the epidemic we need to know:

- At what time will the next process occur?
- Which of the possible processes will it be?

These unknowns can be simulated by using computer-generated random numbers ('Monte Carlo sampling'). A random number generator (Chandler and Northrop, 1999) based on the Marsaglia-Zaman 'subtract-with-borrow' method (Marsaglia and Zaman, 1991) was used. The basic Gillespie algorithm is as follows:

**1.** Calculate the rate term  $a_i$  for each process  $i$  in the model. For example, the process of acquired drug-resistance  $T_w \rightarrow T_r$  has a rate term equal to  $\phi T_w$ , where  $T_w$  is the current number of individuals in compartment  $T_w$ . The probability for each process is given by  $a_i / a_{tot}$  where  $a_{tot}$  is the sum of all the rate terms.

**2.** Select two random numbers  $r_1, r_2$  from a uniform distribution on the interval (0,1).

Calculate the time interval  $\tau$  to the next event, given by  $\tau = -\ln(r_1)/a_{tot}$ . This corresponds to sampling from an exponential distribution of inter-event times, appropriate to a Poisson process. The *single* process that occurs in this time interval is then selected randomly with the appropriate probability, by imagining the interval (0,1) to be partitioned according to the relative size of each rate term, e.g. for a system with two processes, the first occupies the interval  $(0, a_1 / (a_1 + a_2))$  and process 2 occupies the interval  $(a_1 / (a_1 + a_2), 1)$  which has length  $a_2 / (a_1 + a_2)$ . If the random number  $r_2$  lies in the first interval, process 1 is chosen; otherwise process 2 is selected.

**3.** Update the time by  $\tau$ , and update the number of individuals in each compartment according to the process that has occurred. e.g., for drug-failure,  $T_w \rightarrow T_r$ ,  $T_w$  is decreased by one and  $T_r$  is increased by one.

4. Repeat steps 1-3 many times, for a specified length of time.

**Funding**

This work was funded by a Medical Research Council 'Discipline-Hopping' grant.

Accepted manuscript



## References

- Alexander, C. S., Dong, W., Schechter, M. T., O'Shaughnessy, M. V., Strathdee, S. A., Mo, T., Montaner, J. S. G., Harrigan, P. R., 1999. Prevalence of primary HIV drug resistance among seroconverters during an explosive outbreak of HIV infection among injecting drug users. *AIDS* 113, 981-985.
- Ammaranond, P., Cunningham, P., Oelrichs, R., Suzuki, K., Harris, C., Leas, L., Grulich, A., Cooper, D. A., Kelleher, A. D., 2003. No increase in protease resistance and a decrease in reverse transcriptase resistance mutations in primary HIV-1 infection:1992-2001. *AIDS* 17, 264-267.
- Amundsen, E. J., Stigum, H., Rottingen, J.-A., Aalen, O.O, 2004. Definition and estimation of an actual reproduction number describing past infectious disease transmission: application to HIV epidemics among homosexual men in Denmark, Norway and Sweden. *Epidemiol. Infect.* 132, 1139-1149.
- Babiker, A., Darby, S., De Angelis, D., Kwart, D., Porter, K., Beral, V., Darbyshire, J., Day, N., Gill, N., and the Collaborative Group on AIDS Incubation and HIV Survival, 2000. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet* 355, 1131-1137.

Bennett, D. E., Zaidi, I. F., Heneine, W., Woods, T., Garcia-Lerma, J. G., Smith, A.J., McCormick, L., Weinstock, H., 2003. Prevalence of mutations associated with antiretroviral drug resistance among men and women newly diagnosed with HIV in 10 US cities, 1997-2001. *Antivir. Ther.* 8, S133.

Bezemer, D., Jurriaans, S., Prins, M., van der Hoek, L., Prins, J. M., de Wolf, F., Berkhout, B., Coutinho, R., Back, N. K. T., 2004. Declining trend in transmission of drug-resistant HIV-1 in Amsterdam. *AIDS* 18, 1571-1577.

Bezemer, D, de Wolf, F., Boerlijst, M. C., van Sighem, A., Hollingsworth, T. D., Prins, M., Geskus, R. B., Gras, L., Coutinho, R. A., Fraser, C., 2008. A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy. *AIDS* 1071-1077.

Blackham, J., Burns, S., Cameron, S., Cane, P., Chrystie, I., and UK Collaborative Group on HIV Drug Resistance, 2005. Long term probability of detection of HIV-1 drug resistance after starting antiretroviral therapy in routine clinical practice. *AIDS* 19, 487-494.

Blower, S. M., Dowlatabadi, H., 1994. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int. Stat. Rev.* 62, 229-243.

Blower, S. M., Gershengorn, H. B., Grant, R. M., 2000. A tale of two futures: HIV and antiretroviral therapy in San Francisco. *Science* 287, 650-654.

Blower, S. M., Aschenbach, A. N., Gershengorn, H. B., Kahn, J. O., 2001. Predicting the unpredictable: transmission of drug-resistant HIV. *Nat. Med.* 7, 1016-1020.

Bongaarts, J., Buettner T., Heilig, G., Pelletier, F., 2008. Has the HIV Epidemic Peaked? *Population and Development Rev.* 34, 199-224.

Brenner, B.G., Routy, J. P., Petrella, M., Moisi, D.,Oliveira, M., Detorio, M., Spira, B., Essabag, V., Conway, B., Lalonde, R., Sekaly, R. P., Wainberg, M.A., 2001. Persistence and Fitness of Multidrug-Resistant Human Immunodeficiency Virus Type 1 Acquired in Primary Infection. *J. Virol.* 76, 1753-1761.

Brenner, B. G., Roger, M., Routy, J. P., Moisi, D., Ntemgwa, M., Matte, C., Baril, J. G., Thomas, R., Rouleau, D., Bruneau, J., Leblanc, R., Legault, M., Tremblay, C., Charest, H., Wainberg, M.A., 2007. High rates of forward transmission events after acute/early HIV-1 infection. *J. Infect. Dis.* 195, 951-959.

Brown, A. J. L., Frost, S. D., Daar, E., Perrin, L., Richman, D.D., Little, S. J., 2002. An individual-based epidemiological model for the transmission of drug-resistant HIV. *Antivir. Ther.* 7, S187-S187 173 Suppl. 1.

CDC MMWR, 2005. HIV Prevalence, Unrecognised Infection, and HIV Testing Among Men Who Have Sex With Men – Five U.S. Cities, June 2004- April 2005. 54, 597-601.

Chandler, R. E., Northrop, P., 1999. Fortran program randgen.f,  
<http://www.homepages.ucl.ac.uk/~ucakarc/work/software/randgen.f>

Dangerfield, C. E., Ross, J. V., Keeling, M. J., 2008. Integrating stochasticity and network structure into an epidemic model. *J. R. Soc. Interface*, published online 30 October 2008, doi:10.1098/rsif.2008.0410

Deeks, S. G., Wrin, T., Liegler, T., Hoh, R., Hayden, M., Barbour, J. D., Hellmann, N. S., Petropoulos, C. J., McCune, J. M., Hellerstein, M. K., Grant, R. M., 2001. Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *N. Engl. J. Med.* 344, 472-480.

Dougan, S., Evans, B. G., Elford, J., 2007. Sexually transmitted infections in Western Europe among HIV-Positive men who have sex with men. *Sex. Transm. Dis.* 34, 783-790.

EuroHIV, 2007. HIV/AIDS surveillance in Europe: end-year report 2006, No. 75, Saint-Maurice, Institut de Veille Sanitaire. Available at:  
[http://www.eurohiv.org/reports/report\\_75/pdf/report\\_eurohiv\\_75.pdf](http://www.eurohiv.org/reports/report_75/pdf/report_eurohiv_75.pdf)

Geretti, A. M., 2007. Epidemiology of antiretroviral drug resistance in drug-naïve persons. *Curr. Op. Infect. Dis.* 20, 22-32.

Gilbert, M. T. P., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E., Worobey, M., 2007. The Emergence of HIV in the Americas and beyond. *P.N.A.S.* 104, 18566-18570.

Gillespie, D. T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403-434.

Gillespie, D. T., 1977. Exact stochastic simulations of coupled chemical reactions. *J. Phys. Chem.* 81, 2340-2361.

Gran, J. M., Wasmuth, L., Amundsen, E. J., Lindqvist, B. H., Aalen, O. O., 2008. Growth rates in epidemic models: Application to a model for HIV/AIDS progression. *Statistics in Medicine* 27, 4819-4836.

Grant, R. M., Wiley, J. A., Winkelstein, W., 1987. Infectivity of the human-immunodeficiency-virus – estimates from a prospective study of homosexual men. *J. Infect. Dis.* 156, 189-193.

Grant, R. M., Hecht, F. M., Warmerdam, M., Liu, L., Liegler, T., Petropoulos, C. J., Hellmann, N. S., Chesney, M., Busch, M. P., Kahn, J. O., 2002. Time trends in primary HIV-1 drug resistance among recently infected persons. *J.A.M.A.* 288, 181-188.

Hollingsworth, T. D., Anderson, R. M., Fraser, C., 2008. HIV-1 Transmission, by Stage of Infection. *JID* 687-693.

- Keeling, M. J., Grenfell, B.T., 1998. Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Math. Biosci.* 147, 207-226.
- Kelly, J. A., Amirkhanian, Y. A., 2003. The newest epidemic: a review of HIV/AIDS in Central and Eastern Europe. *Int. J. STD and AIDS* 14, 361-371.
- Leigh Brown, A. J., Richman, D. D., 1997. HIV-1: Gambling on the evolution of drug resistance? *Nat. Med.* 3, 268-271.
- Leigh Brown, A. J., 1997. Analysis of HIV-1 *env* gene sequences reveals evidence for a low effective number in the viral population. *P.N.A.S.* 94, 1862-1865.
- Little S. J., Holte S., Routy J. P., Daar E. S., Markowitz M., Collier, A. C., Koup, R. A., Mellors, J. W., Connick, E., Conway, B., Kilby, M., Wang, L., Whitcomb, J. M., Hellmann, N. S., Richman, D. D., 2002. Antiretroviral-drug resistance among patients recently infected with HIV. *N. Engl. J. Med.* 347, 385-394.
- Lloyd, A.L., 2001. Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. *Theor. Pop. Biol.* 60, 59-71.
- Marsaglia, G., Zaman, A., 1991. A new class of random number generators. *Annals of Applied Probability* 1, 462-480.
- de Mendoza, C., Rodriguez, C., Eiros, J. M., Colomina, J., García, F., Leiva, P., Torre-Cisneros, J., Agüero, J., Pedreira, J., Viciano, I., Corral, A., del Romero, J., de

- Lejarazu, R. O., Soriano, V., 2005. Antiretroviral recommendations may influence the rate of transmission of drug-resistant HIV type 1. *Clin. Infect. Dis.* 41, 227-232.
- Monforte, d.'A., Lepri, A. C., Rezza, G., Pezzotti, P., Antinori, A., Phillips, A. N., Angarano, G., Colangeli, V, de Luca, A., Ippolito, G., Caggese, L., Soscia, F., Filice, G., Gritti, F., Narciso, P., Tirelli, U., Moroni, M., 2000. Insights into the reasons for discontinuation of the first highly active antiretroviral therapy (HAART) regimen in a cohort of antiretroviral naïve patients. *AIDS* 14, 499-507.
- Nijhuis, M., Boucher, C. A. B., Schipper, P., Leitner, T., Schuurman, R., Albert, J., 1998. Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. *P.N.A.S.* 95, 14441-14446.
- Pao, D., Andrad, U., Clarke, J., Dean, G., Drake, S., Fisher, M., Green, T., Kumar, S., Murphy, M., Tang, A., Taylor, S., White, D., Underhill, G., Pillay, D., Cane, P., 2004. Long-Term Persistence of Primary Genotypic Resistance after HIV-1 Seroconversion. *J. Acquir. Immune Defic. Syndr.* 37, 1570-1573.
- Phillips, A. N., Youle, M., Johnson, M., Loveday, C., 2001. Use of a stochastic model to develop understanding of the impact of different patterns of antiretroviral drug use on resistance development. *AIDS* 15, 2211-2220.
- Porter, K., Babiker, A. G., Darbyshire, J. H., Pezzotti, P., Bhaskaran, K., Walker, A. S., 2003. Determinants of survival following HIV-1 seroconversion after the introduction of HAART. *Lancet* 362, 1267-1274.

- Press, W. H., Teukolsky, S. A., Vetterling, W.T., Flannery, B. P., 1992. Numerical recipes in Fortran 77 (Volume 1). Cambridge, UK: Cambridge University Press.
- Routy, J. P., Machouf ., Edwardes, M. D., Brenner, B. G., Thomas, R., Trottier, B., Rouleau, D., Tremblay, C. L., Côté, O., Baril, J.-G., Remis, R. S., Sékaly, R. P., Wainberg, M. A., 2004. Factors associated with a decrease in the prevalence of drug resistance in newly HIV-1 infected individuals in Montreal. *AIDS* 18, 2305-2312.
- Sánchez, M. S., Grant, R. M., Porco, T. C., Gross, K. L., Getz, W. M., 2005. A decrease in drug resistance levels of the HIV epidemic can be bad news. *Bull. Math. Biol.* 67, 761-782.
- Sánchez, M. S., Grant, R. M., Porco, T. C., Getz, W. M., 2006. HIV drug-resistant strains as epidemiologic sentinels. *Emerg. Infect. Dis.* 12, 191-197.
- Simon V., Vanderhoeven J., Hurley A., Ramratnam B., Louie M., Dawson, K., Parkin, N., Boden, D., Markowitz, M., 2002. Evolving patterns of HIV-1 resistance to antiretroviral agents in newly infected individuals. *AIDS* 16, 1511-1519.
- Stall, R., Pollack, L., Mills, T. C., Martin, J. N., Osmond, D., Paul, J., Binson, D., Coates, T. J., Catania, J. A., 2001. Use of antiretroviral therapies among HIV-infected men who have sex with men: a household-based sample of 4 major American cities. *Am. J. Publ. Health.* 91, 767-773.



Tang, J. W., Pillay, D., 2004. Transmission of HIV-1 drug-resistance. *J. Clin. Virol.* 30,1-10.

Turner, D., Brenner, B., Routy, J. P., Moisi, D., Rosberger, Z., Roger, M., Wainberg, M. A., 2004. Diminished representation of HIV-1 variants containing select drug resistance-conferring mutations in primary HIV-1 infection. *J. Acquir. Immune Defic. Syndr.* 37, 1627-1631.

UK Group on Transmitted HIV Drug Resistance, 2005. Time trends in primary resistance to HIV drugs in the United Kingdom: multicentre observational study. *B.M.J.* 331,1368-1371.

UK HIV Drug Resistance Database, Oct 23 2003. HIV drug resistance in the United Kingdom. *Commun. Dis. Rep. CDR Wkly.* 13(43).

UNAIDS, WHO, 2007. AIDS epidemic update, Geneva. Available at:

[http://data.unaids.org/pub/EPISlides/2007/2007\\_epiupdate\\_en.pdf](http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf)

Vardavas, R., Blower, S., 2007. The Emergence of HIV Transmitted Resistance in Botswana: 'When Will the WHO Detection Threshold Be Exceeded?' *PLoS ONE* 1:e152.

Wensing A. M. J., van de Vijver D. A. M. C., Asjo B., Balotta C., Canacho., R., de Mendoza C., et al., and the SPREAD Programme, 2003. Prevalence of transmitted drug resistance in Europe is largely influenced by the presence of non-B sequences:

analysis of 1400 patients from 16 countries: the CATCH study. *Antivir. Ther.* 8, S131 (abstract no. 117).

Yerly, S., Vora, S., Rizzardi, P., Chave, J.-P., Vernazza, P. L., Flepp, M., Telenti, A., Battegay, M., Veuthey, A.-L., Bru, J.-P., Rickenbach, M., Hirschel, B., Perrin, L., Swiss HIV Cohort Study, 2001. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. *AIDS* 15, 2287-2292.

Accepted manuscript

**Table 1**

MODELLING UNCERTAINTY IN PARAMETER VALUES.	MODELLING INTRINSIC STOCHASTICITY.
---	---------------------------------------

Rate variables (model parameters) selected from a specified <b>range</b> using Monte Carlo (random number) sampling. e.g. Latin Hypercube Sampling (Blower and Dowlatabadi, 1994).	Starting parameters <b>identical</b> . Model parameters treated as probabilities and events simulated using random numbers. e.g. Gillespie method (Gillespie, 1976, 1977).
Model population is <b>continuous</b> . Model compartments may comprise fractional numbers of individuals.	Model population is <b>discrete</b> (integer-based). Model compartments comprise whole numbers of individuals.
For each set of starting parameters, solve the model differential equations. Each outcome is <b>deterministic</b> and represents the <b>mean</b> behaviour of many epidemics; differences arise due to different starting parameters only.	Use Monte Carlo sampling to simulate the probability of an event (e.g. infection), and the time at which it occurs. Generate random numbers repeatedly to simulate random events during the course of an epidemic. Each <b>stochastic</b> simulation represents an <b>individual</b> epidemic.
Generate an ensemble of deterministic simulations, each with randomly selected starting parameters, and analyse the variability <i>among</i> epidemics. This will depend on the selected parameter range and is independent of the size of the modelled infectious population.	Generate an ensemble of stochastic simulations and analyse the variability <i>within</i> individual epidemics and <i>among</i> epidemics. Modelled variability is inherent and is larger for small infectious populations because the impact of chance events is greater. Ensemble mean is often similar to the corresponding deterministic simulation.

**Table 1.** A Comparison of Two Types of Randomness or Stochasticity and Their Simulation.

Monte Carlo methods, or 'games of chance' involve the generation of random numbers to simulate processes in which there are aspects of uncertainty.

These can be used in different ways to explore the variability in epidemic outcomes due to (a) randomness or uncertainty in biological and clinical input parameters (b) intrinsic stochasticity arising due to chance events.

Accepted manuscript

**Table 2.**

Description of the Model Parameters and their Values. Parameter values are based on those used in previous HIV modelling studies (Blower et al., 2000, 2001; Sánchez et al., 2005, 2006). Detailed source references are given in the Table. All rates have units of  $y^{-1}$ .

Parameter	Value	Description
B	1000	Rate at which susceptible individuals join the population. $B=\mu N$ in absence of disease.
$\mu$	0.02	per capita removal rate (Sánchez et al., 2005).
$\beta_w, \beta_r$	0.4,0.2	Virus transmission rates (sexual contact rate x probability of transmission); w = drug-sensitive strain; r =drug-resistant strain (Grant et al., 1987; Turner et al., 2004).
$\rho_w, \rho_r$	0.02,0.5	Relative transmissibility under treatment ( $0 < \rho_i < 1$ ) (Tang and Pillay, 2004).
$\alpha$	0.1	Disease-induced death rate (Babiker et al., 2000).
$\alpha_{tw}, \alpha_{tr}$	0.05,0.07	Disease-induced death rate under treatment. tw = drug-sensitive strain; tr =drug-resistant strain (Porter et al., 2003).
$\phi$	0.1	Treatment failure rate (leading to acquired drug resistance) (Blackham et al., 2005).
$\omega$	0.1	Rate of ceasing therapy (Monforte et al., 2000).
$\pi$	0.4	Rate of commencing therapy (Stall et al., 2001).
$\gamma$	1.0	Rate of reversion (Brenner et al., 2001; Pao et al., 2004), or ‘growing out’ (Brenner et al., 2001; Deeks et al., 2001) to wild-type virus in untreated individuals.
N	50001 initially.	Total population size = $S + I_w + I_r + T_w + T_r$ $I_w = 1$ ; $S = 50000$ initially.

### Figure Legends.

**Figure 1.** Flow Diagram Illustrating the Different Compartments Described in the Model of Homosexual HIV Transmission Dynamics.

The five compartments are: susceptible (S); untreated and infected with wild-type (drug-sensitive) virus ( $I_w$ ); untreated and infected with pure resistant (r) or mixed (w,r) viral strains ( $I_r$ ); treated and infected with either drug-sensitive ( $T_w$ ) or drug-resistant ( $T_r$ ) viral strains. The model parameters are given in Table 2. The forces of infection are  $\lambda_w = \beta_w (I_w + \rho_w T_w)/N$  and  $\lambda_r = \beta_r (I_r + \rho_r T_r)/N$ .  $1/\phi$  = average time for acquired drug resistance to develop;  $1/\gamma$  = average time for a drug-resistant infection to revert to a drug-sensitive infection in an untreated individual;  $\pi$  = treatment uptake rate;  $\omega$  = rate of ceasing therapy;  $\alpha$  = average disease-induced death rate in absence of therapy;  $\alpha_{tw}, \alpha_{tr}$  = disease-induced death rates for drug-sensitive and drug-resistant treated individuals respectively;  $1/\mu$  = average time for which an individual remains in the sexually active population;  $B$  = rate at which individuals join the susceptible (sexually active) community.

**Figure 2(a).** Epidemiological Time Series Showing the Prevalence of Transmitted Drug-resistant HIV.

Examples of observed epidemiological time series measuring the prevalence of drug resistance mutations in newly-diagnosed HIV-1 infections in countries where HAART is used (taken from Table 1 of Tang and Pillay, 2004). In some surveys, annual variations in primary HIV drug-resistance prevalence were recorded, but in others the data are “lumped” and the latter are indicated by horizontal lines.

References to the original data sources as given in parentheses in the Figure key; total

sample sizes are given below in square brackets: (1) UK HIV Drug Resistance Database, 2003, [n=2025]; (2) Yerly et al., 2001, [n= 197]; (3) Grant et al., 2002, [n= 225]; (4) Simon et al., 2002, [n= 154]; (5) Bennett et al., 2003, [n = 1082]; (6) Wensing et al., 2003, [n= 1369]; (7) Little et al., 2002, [n = 377]; (8) Bezemer et al., 2004, [n = 100]; (9) de Mendoza et al., 2005, [n = 57]. In some of the studies (Bennett et al., 2003; UK HIV Drug Resistance Database, 2003; Wensing et al., 2003), the newly diagnosed infections are not necessarily recent infections.

**Figure 2(b)** Model Simulations Showing the Prevalence of Drug Resistance in Newly Infected Persons.

Drug treatment commenced after 20 years. The deterministic result is shown, together with the individual stochastic simulations and the mean of 714 stochastic simulations. Errorbars indicate  $\pm$  one standard deviation from the mean.

**Figure 2(c)** Model Simulations Showing the Prevalence of Drug Resistance in Newly Infected Persons.

As Figure 2(b), but with drug treatment commencing after 30 years (706 stochastic simulations).

**Figure 3** Stochastic Predictions of Transmitted Drug-resistant HIV Prevalence for Differing Treatment Introduction Times ( $t_i$ ).

Nine stochastic simulations were randomly selected from Figure 2(b) ( $t_i = 20$  yr) and Figure 2(c) ( $t_i = 30$  yr) . Simulations are plotted over the time period 10-18 years after treatment begins, after the initial steep rise in transmitted drug-resistance prevalence.

(a)  $t_i = 20$  yr. The variability is comparable with that observed in epidemiological surveys (Figure 2(a)), despite the fact that each of these stochastic realizations is performed with identical parameter values.

(b)  $t_i = 30$  yr. The temporal variability within individual epidemics is noticeably smaller than in Figure 3(a).

**Figure 4** Stochastic and Deterministic Simulations of New Drug-sensitive and Drug-resistant Infections for Different Treatment Introduction Times ( $t_i$ ).

Individual stochastic simulations are based on results of Figure 2(b) ( $t_i = 20$  yr) and Figure 2(c) ( $t_i = 30$  yr), and illustrate the annual number of new infections. Also shown are the mean of the stochastic values and the deterministic outcomes.

Errorbars indicate  $\pm$  one standard deviation from the mean.

(a)  $t_i = 20$  yr. New drug-sensitive infections

(b)  $t_i = 20$  yr. New drug-resistant infections.

(c)  $t_i = 30$  yr. New drug-sensitive infections.

(d)  $t_i = 30$  yr. New drug-resistant infections.

**Figure 5** Timing of First Emergence of Transmitted Drug-Resistance in Stochastic Simulations, for Different Treatment Introduction Times ( $t_i$ ).

**Figure 6** Stochastic Temporal Variability in Transmitted Drug-Resistance Prevalence Within Individual Epidemics, as a Function of the Mean Annual Number of New Infections.

For each stochastic simulation a least-squares fit (using a quadratic function) was obtained over a 30 year time period ( $t_i + 10$ ,  $t_i + 40$ ) yr and the mean annual number of new infections was calculated over this time interval. The residual coefficient of variation gives an estimate of the temporal variability.



**Figure 7 Stochastic Variability in Transmitted Drug-Resistance Prevalence among Epidemics.**

(a) Comparison of Variability for Different Treatment Introduction Times ( $t_i$ ).

The coefficient of variation ( $CV = (\text{standard deviation}/\text{mean}) \times 100\%$ ), in the transmitted drug-resistance prevalence is shown as a function of time for  $t_i = 15, 20, 30$  and  $40$  yr.

(b) Variability among epidemics 10 years after treatment introduction, as a function of the annual number of New Drug-Resistant Infections.

The variability (CV) is shown for treatment introduction prior to equilibrium ( $t_i = 30$  yr) and after equilibrium ( $t_i = 150$  yr). Also shown is a scatter plot illustrating the typical temporal variability *within* epidemics as a function of the mean annual number of new drug resistant infections.

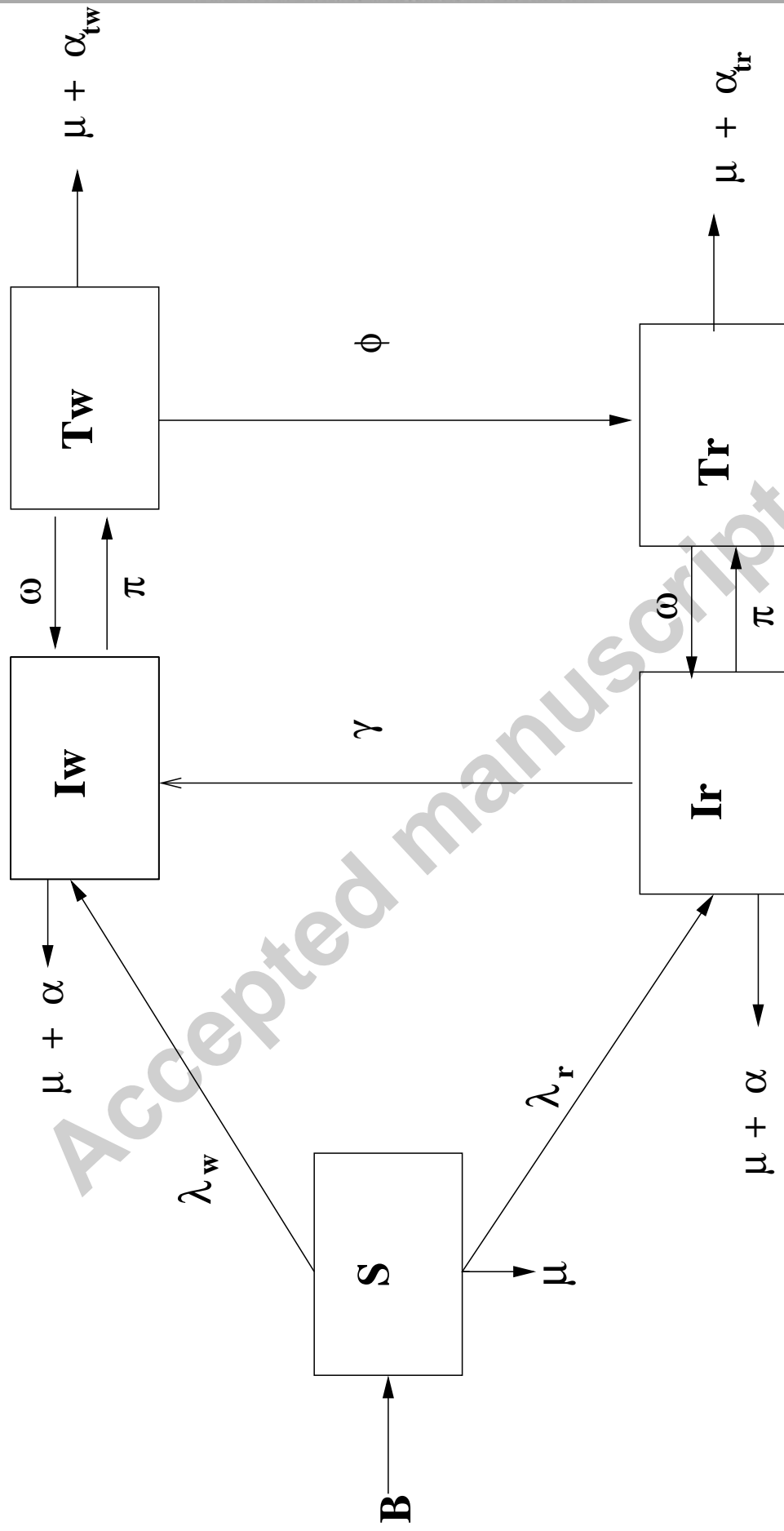
**Supplementary Material.**

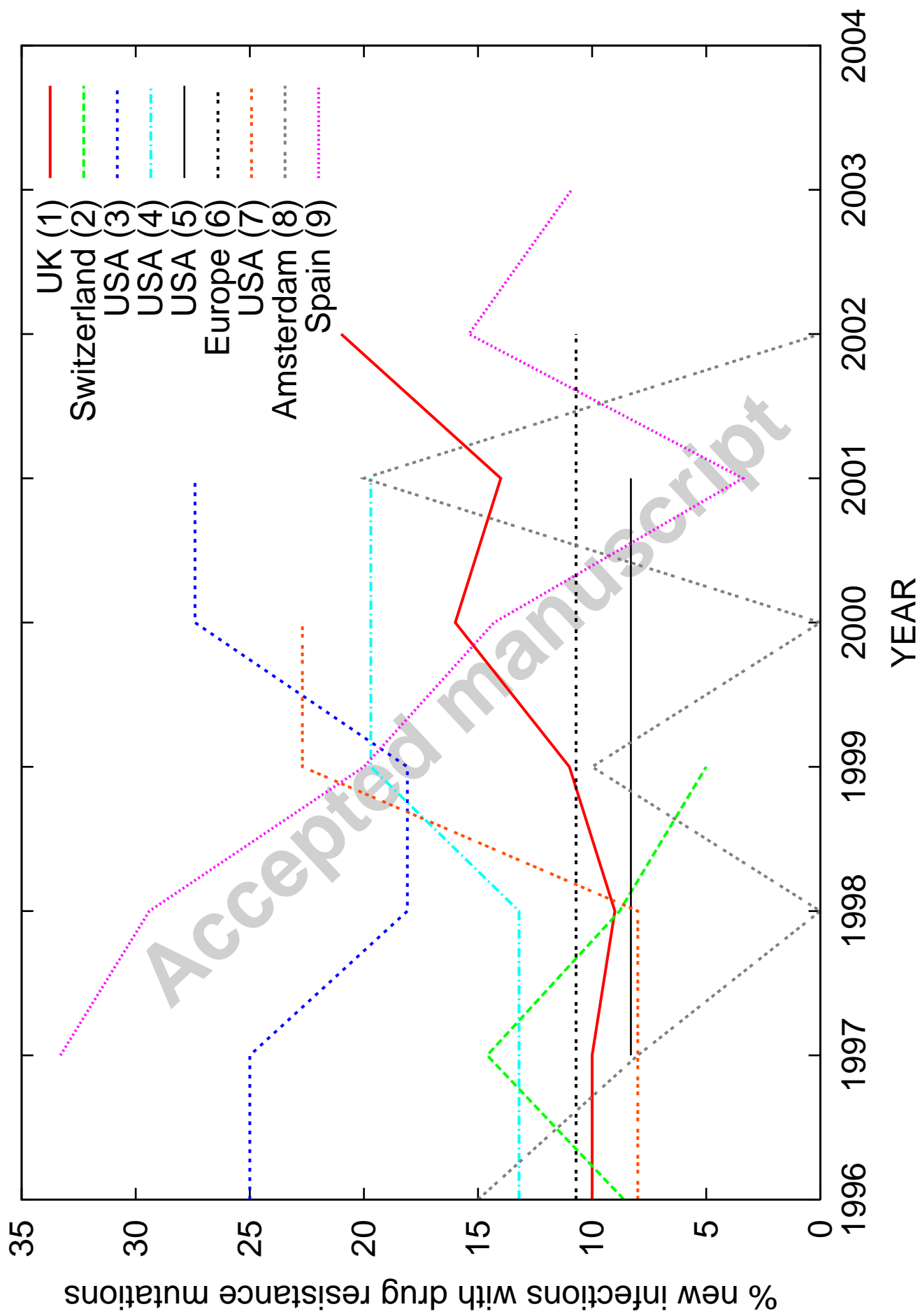
Figure S1. The effect of different model parameters on the stochastic variability in transmitted drug-resistance prevalence within individual epidemics.

Figure S2. Variability among epidemics when they are all of equal size at the time of treatment introduction – a comparison with fully stochastic results.

Figure S3. The effect of different model parameters on the stochastic variability in transmitted drug-resistance prevalence among epidemics.

Accepted manuscript





4. Figure 2a

